PEAK-OVER-THRESHOLD FLOOD FREQUENCY ANALYSIS OF STREAMFLOW SERIES FOR INSULAR NEWFOUNDLAND

KENNETH GORDON TAYLOR

0-612-62434-X

PEAK-OVER-THRESHOLD FLOOD FREQUENCY ANALYSIS

OF

STREAMFLOW SERIES FOR INSULAR NEWFOUNDLAND

By

© Kenneth Gordon Taylor

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements for the

Degree of Master of Engineering

FACULTY OF ENGINEERING AND APPLIED SCIENCE

MEMORIAL UNIVERSITY OF NEWFOUNDLAND

APRIL 2001

ST. JOHN'S NEWFOUNDLAND CANADA

Abstract

In this thesis, regional models for the prediction of flood quantiles for streams on the island of Newfoundland are developed using historical streamflow data which has been subject to peak-over-threshold analysis. The Peak-Over-Threshold method of flood frequency analysis allows extraction of more relevant data from a historical flow series than would be available using the conventional annual maximum flow method. As a result, the peak-over-threshold method is of particular interest in regions where data on streamflows is limited. This is the case in Newfoundland.

Streamflow series from 63 rivers on the island of Newfoundland are considered. This data is modelled using a Poisson arrival process and the Exponential and Pareto magnitude distributions. Results from single-station peak-over-threshold analysis are compared to those obtained from the annual maxima series modelled using the 3-Parameter Lognormal and Generalized Extreme Value distributions. The island is divided into hydrologically homogeneous regions. Hydrologically homogeneous regions are defined as geographic areas in which flood flows are identically distributed except for scale. Regional index flood estimators are developed using the data generated from the peak-over-threshold approach.

For the quantile estimates generated for the 63 data series analysed, there is no statistically significant difference between the central position of the results of the 3-Parameter Lognormal,

Generalized Extreme Value, Poisson-Exponential, and Poisson-Pareto models. Model error for the single station analysis is tested using a bootstrap approach. For the standard error of quantile estimates generated by resampling, the Poisson-Exponential Distribution model exhibited comparable standard error for lower quantiles and lower standard error for higher quantiles. Because of this, the Poisson-Exponential model was determined to be the most robust for a variety of quantiles. Although the Poisson-Pareto distribution is more flexible, it appears to be inferior to the Poisson-Exponential model in this case.

Regional models were developed using an index flood approach. The index flood was taken as the two-year return period flood, Q(2), and regional estimators for index floods for each region were developed by non-linear regression on physical basin descriptors. Regional models developed using nonlinear regression exhibited better fit to the underlying data than did the models produced using the traditional log-linear method. The nonlinear models exhibited lower bias, and also less estimation error. The ratios of Q(T)/Q(2) were easily calculated, and allowed estimation of flood quantiles for stations in the regions with a reasonably good fit to the expected values. For most regions RMSE was less than 10% of the mean of the expected values. The estimated values from application of the index flood technique tended to overestimate the quantile slightly and results were somewhat positively skewed from expected values. This will tend to produce more conservative (higher) estimates of flood quantiles.

In the Southwest Region the equation which performed best (generated estimates with the lowest error) relied on three basin descriptors. The number of gauge records available in this region was only six. The coefficients developed for this equation are also somewhat suspect as they suggest a significant scaling of the result. In this region, the use of the whole island equation may provide a more reliable result and is recommended.

Quantile estimates generated using the index flood method produced the poorest results in the Northwest Region. However, results were still reasonable and at lower quantiles, the RMSE was less than 10% of the mean expected value. When the estimators derived for the whole island were applied to this region they produced slightly better results. Thus with the exception of the Northwest Region, the use of regional index floods produced improved quantile estimates when compared to the estimates produced by equations developed for the whole island.

Acknowledgements

I am grateful for the guidance, advice, and patience of Dr. Leonard Lye in the preparation of this thesis.

I am also grateful for the financial support and accommodation in work schedules provided by my employers during the course of my research. These employers have included the Department of Environment and Labour, The College of the North Atlantic, The SGE Group, and Fracflow Consultants.

Thanks go to Ed Byrne for his help in developing computer programs and data structuring.

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

This chapter starts with a discussion of hydrologic modelling, how models are developed, and the types of models commonly applied to predict peak flows. Following this general discussion, the objectives and methodology of this thesis are explained, and the outline of the thesis is given. At this time, it should be noted that the work of this thesis is concerned only with the island portion of the province of Newfoundland and any reference to Newfoundland in this thesis is intended to indicate only the island portion.

1.1 Background

One of the most difficult problems in hydrology is the prediction of frequency of occurrence of future streamflow magnitudes, or flood quantiles. A flood quantile is a flood event of known or estimated probability of recurrence. That is, for 100 years of data the 75^{th} quantile is the flood event not exceeded 75% of the time. When studying this problem, the engineer or hydrologist wants to develop a model by which he can predict the probability that a future event of a given magnitude will occur within some time period of interest. For example, an event with a magnitude which occurs on average once every hundred years has a probability of occurrence of 1/100 in any given year. Such an event is referred to as a hundred-year event or as an event with a 100-year return period.

The hydrologist and engineer must understand these events in terms of their probability of occurrence over the life of a structure. The accurate prediction of flood quantiles is difficult, but this information is critical in the design of bridges, culverts, dams, flood protection works, and other works which are impacted by the flow of a stream. Failure to design these structures with sufficient capacity can result in failures which are costly and result in loss of human life. Alternatively, structures which are designed with excessive capacity are unacceptably costly to construct. Hence, the ability to provide accurate probabilistic models of flood events has significance from both an economic and environmental standpoint (Bobée and Rasmussen, 1995).

Hydrologic models allow the hydrologist or engineer to reduce complex physical systems to components and to make predictions of hydrologic behaviour in a deterministic or probabilistic sense (Haan, Johnson and Brakensiek, 1982). However, all models are incomplete approximations of real behaviour, and the output information from a model is seldom an exact representation of the actual response of the real system to the same inputs. Additionally, models are generally designed to predict only limited components of system output response. Thus a model designed to predict flood quantiles may predict the magnitude of a flood corresponding to a particular probability of occurrence but may say nothing about the duration of that flood or the shape of the flood hydrograph. Generally, as the amount of explanatory information integrated into the model and the amount of information contained in the model output increases, the complexity of the model and the uncertainty associated with the model output increases as well. The task of the modeller is to model the actual system as closely as possible while keeping the model as simple as possible.

The simplest type of flood model occurs where individuals living adjacent to a stream witness a flood event and subsequently adjust their construction practices to accommodate this high flow condition. Over a long period, information on the behaviour of a stream is passed along through the group and an understanding of the stream's behaviour, a model, becomes cultural information. The individuals involved do not require any understanding of the underlying processes related to the high flow or any knowledge of probability concepts to apply their model.

Models which require intimate knowledge of the behaviour of a particular stream over an extended period of time are limited in their application to the stream upon which the knowledge is based. To extrapolate the behaviour of unobserved streams from knowledge about observed streams, mathematical models are used. Information about streams with known behaviour is utilized to develop an idea of the behaviour of a stream which has not been observed. Where components of mathematical models are considered to be free from random variation, the model is defined as deterministic (Haan, Johnson and Brakensiek, 1982). The Associate Committee on Hydrology (1989) describes the *flood envelope chart* as an example of a deterministic approach. An example of this type of chart from the work of Neill (1986), is included as Figure 1.1. High flood discharges may be plotted against drainage area to show a relationship. This relationship may be expressed as an equation:

$$Q=C*A^{B} \tag{1}$$

Where Q is a high discharge of unknown return period, *A* being the area of the drainage basin, and *C* and *B* being coefficients determined by the modeller. The selection of drainage area as a primary predictor of flood flows is a logical one since the amount of water available for streamflow is directly related to the amount which is collected over the drainage basin area. This approach is based on a collection of observed maximum flows for a number of rivers in a region and no probability of occurrence is implied. The assumption of similar hydrologic behaviour among streams in close proximity is implicit in this model. The concept of regional hydrologic homogeneity will be discussed in some detail later in this thesis.

In statistical modelling, the modeller uses known information about the event of interest and the underlying explanatory phenomena to develop models which allow inferences about future events. The mathematical model provides a simplified explanation of how the explanatory variables influence the variable of interest. The quality of the model is determined to a large degree by the modeller's understanding of the relationship between the variables, and by the amount and quality of relevant explanatory information which it uses to produce its outcome.

Some models use an underlying phenomenon, such as rainfall, to obtain model inputs with known probability of occurrence. The model then relates these inputs to streamflow in terms of basin characteristics. The rational formula is an example of this type of model.

$$Q = kCIA \tag{2}$$

Where Q is a discharge of known return period, A being the area of the drainage basin, I being a rainfall event of known intensity, duration and frequency of occurrence, C being a coefficient related to surface characteristics of the drainage basin, and k being a conversion factor to allow use of metric units. The rational formula is an attempt to model the output characteristics of a stream (streamflow) based on the physical relationship between the system input (rainfall) and the drainage system it must pass through. In this type of model, the inputs are estimated using a statistical model of rainfall, basin characteristics are estimated from maps or field data, equations are derived relating rainfall inputs to streamflow, and these equations are calibrated to the streamflow conditions for known inputs.

In models like the rational formula, the quality of the input data has a significant influence on the reliability of the outcome. For the rational formula to work well, long rainfall records are required containing not just daily rainfall amounts but information about rainfall intensity. The rainfall data must come from a source in close proximity to the stream which is being studied. In addition, the model implies that the probability distribution of basin output is the same as that of input, which may not necessarily be the case. The rational formula works best for small uncomplicated drainage basins where rainfall input produces output response quickly and there are few attenuating features. In large complex basins, the input signal takes much longer to propagate through the system and is moderated by a number of processes. The output of large systems may not reflect the shape of the input signal. Thus, for large basins there may be problems with application of models like the rational formula.

Statistical methods have long been applied to historical streamflow data to estimate the frequency of occurrence of future streamflow magnitudes. If a long streamflow record exists for the stream under consideration, an appropriate probability model may be fitted to this long data record to yield good estimates of flood quantiles and results from the model may be calibrated against known data points within the record. For example, if a probability model is fitted to a data series with 100 years of data, the calculated magnitude of an event with probability of 0.04, may be compared to the fourth highest recorded flow in the 100 years.

The most common methods use series of Annual Maximum Flows (AMF) from gauged streams. In this approach, only the maximum flow in any year is considered relevant. Other information about flow magnitudes is discarded. Probability distributions are fitted to the annual series to produce estimates of flood quantiles, Q_T, for gauged streams.

The 3-Parameter Lognormal Distribution and the Generalized Extreme Value Distribution are two probability models which have been applied for prediction of flood quantiles from AMF data series. Other distributions are also available, including the Log-Pearson Type III, and the Wakeby distribution. However, Beersing (1990) found that the Log-Pearson Type III and Wakeby distributions exhibited poorer fit for Newfoundland data than the 3-Parameter Lognormal and Generalized Extreme Value. Only the 3-Parameter Lognormal and Generalized Extreme Value distributions have been considered for modelling of AMF series in this thesis.

Where the record of historical streamflows is short, fitting probability distributions to AMF series can be problematic. Obtaining a satisfactory fit may be difficult, and once a fit is obtained the outcomes may be unstable and highly dependent on individual values in the data. Some researchers have found that the limited availability of data reduces the utility of sophisticated probabilistic models and that simpler models perform just as well for these limited data sets (Bobée and Rasmussen, 1995). One way to combat this problem is to extract more data from the historical records available. Where the amount of historical data which is available for the construction of models is very limited, the peak-over-threshold method of analysis offers certain advantages. The primary advantage of the peak-over-threshold method, compared to the conventional annual maximum flow approach, is that it allows the incorporation of more explanatory information in model formulation. The inclusion of more explanatory information should improve the quality of model outputs.

The Peak-Over-Threshold (POT) method is a statistical approach which allows extraction of more data from a streamflow record than would be available using the AMF approach (NERC, 1975). The POT method is also known as the Partial-Duration-Series (PDS) method. In the POT approach, all independent flow peaks above a set threshold are considered relevant. The POT method can be particularly useful when the period of record is short because POT series can be selected to contain a larger number of peaks than the AMF series (ACH, 1989). The threshold may be adjusted to increase or decrease the amount of information considered. The larger amounts of data generated using the POT method should permit better fitting of probability distributions. This additional information constitutes the added value in using this approach rather than the more conventional AMF method. However, results must be evaluated against known stream behaviour, and it is incorrect to assume that the use of a larger number of peaks will necessarily produce a more efficient model (NERC, 1975).

The modelling of POT data is generally done by combining a Poisson recurrence process with another distribution for magnitude. The Exponential Distribution and Pareto Distribution are popular choices and their use is well supported in a number of studies. The Exponential distribution has the advantage that it is simple and requires the estimation of only one parameter. The Pareto distribution is more flexible but requires estimation of two parameters. The use of both of these distributions is examined in this thesis.

As has been discussed in the preceding paragraphs, where the designer has access to long or short streamflow records, some understanding of the distribution of flood peaks for the stream may be reached. However, in many cases there is no data available for the location of interest. In these cases the designer must use regional models to estimate flood quantiles. A regional model is a model of drainage basin output (streamflow) which relates the output to basin descriptors and which has as an underlying assumption, the concept that basins with a hydrologically homogeneous region will behave in a similar manner. These models use flood frequency information from hydrologically similar streams to predict flood quantiles for the stream of interest. In cases where there is some streamflow information but it is limited, quantile estimates from regional models may be better than those obtained from distributions fitted to the data for the location. Such equations allow estimation of the flood quantile, $Q_{T,}$ at a specific site based on regional equations. These equations may be developed for any region with similar hydrologic conditions throughout. In general, regional estimators are useful for improving flood quantile estimation at sites where little hydrologic information exists, and are essential for estimating flood quantiles at sites where no hydrologic data are available (Ashkar, 1994). Regionalisation is probably one of the most promising ways to improve flood quantile estimates (Bobée and Rasmussen, 1995).

The hydrologist or engineer must exercise care in using either deterministic or statistical models. Model calculations generally require the assumption of homogeneity of response between the watershed under study and the watershed used to construct the model. Models are generally devised using data from a restricted study region and individuals using these models must be sure that the assumptions and conditions of the model apply to the stream which they are studying. For example, the United States Department of Agriculture developed the SCS Curve Number Method (SCS, 1972) to simulate rainfall-runoff relationships for small agricultural basins. This method is unsuitable for areas with frozen ground and runoff from melting snow, and is of limited use in simulating rainfall-runoff events in the cold Canadian climate (ACH, 1989).

Both deterministic and statistical models may produce results which deviate significantly from actual streamflow behaviour. When interpreting model results, it is important that the designer exercise judgement and use local historical knowledge of the stream's behaviour to evaluate model outputs.

1.2 Research Objectives

A number of methods are used in the prediction of peak flows for Newfoundland. These include the Rational Method, SCS Curve Number Method, channel capacity methods, and local historical knowledge. Recent advances include the work of Beersing (1990) *Regional Flood Frequency Analysis for the Island of Newfoundland*, and the work of Susan Richter (1994) in her thesis *Relationships of Flow and Basin Variables on the Island of Newfoundland, Canada, with a Regional Application*.

The purpose of this work is to investigate the use of peak-over-threshold analysis to construct improved regional models for prediction of flood quantiles for insular Newfoundland. Caissie and El-Jabi (1991a) indicated that the POT method could be applied as successfully to Newfoundland flow series as it could to flow series for any other province. They also indicated that the POT method was found to work well in the eastern regions of Canada. They considered fifteen (15) gauge records for the island portion of Newfoundland, and the island was treated as one homogeneous region.

In this thesis, single station quantile estimators will be constructed by fitting probability distributions to streamflow data extracted using the peak-over-threshold method. Sixty-three (63) stations are used in this analysis. Using these single station quantile estimates, the island will be divided into regions and regional models will be constructed relating basin descriptors to flood quantiles. The index flood is related to quantile estimates obtained using POT analysis, and regional quantile estimators are produced.

This thesis incorporates more streamflow records than previous studies, extracts more data from each series by using the POT method, and generates regional quantile estimates using non-linear regression. This should produce better estimates of flood quantiles than those currently available.

1.3 Research Methodology

This thesis applies the peak-over-threshold method to generate flood quantiles for streamflow records for the island portion of Newfoundland. Regional quantile estimator models are constructed for the island. An extensive literature review is part of this research and the results of this review are contained in the first few chapters of this document. The last two chapters contain the experimental results and conclusions based on the literature review and the results. The general methodology applied in this research is explained below:

1. Incorporate the maximum number of suitable flow records into the data set.
2. Perform AMF and POT analysis of selected flow records.
3. Fit probability models to extracted data using the three parameter log-normal (3LN) and generalized extreme value (GEV) distributions for annual maximum flood (AMF) series, and the Poisson-Exponential (PED) and Poisson-Pareto (PPD) distributions for the POT series.
4. Compare the output of AMF and POT models for prediction of flood quantiles for stations with historical records.
5. Divide the island into regions and test regions for hydrologic homogeneity.
6. Develop regional equations to estimate flood quantiles from basin parameters.

1.4 Organization of this Document

This thesis starts with an introduction to the concepts of flood frequency analysis and the reasons for the application of the POT method to data series for the Island of Newfoundland. In Chapter 2, the study area is described and the known hydrologic characteristics discussed. In Chapter 3, the methods for flood frequency analysis of single data sets using annual maximum floods, AMF, and POT approaches are discussed and the quantile estimators derived for a number of probability distributions. In Chapter 4, the rationale for regionalisation and the methods for defining regions are discussed. In Chapter 5, drainage basin descriptors and the methods used to develop regional models are explained. In Chapter 6, the results of experimental analysis are presented. And finally, in Chapter 7 some conclusions are made regarding the expected and obtained results.

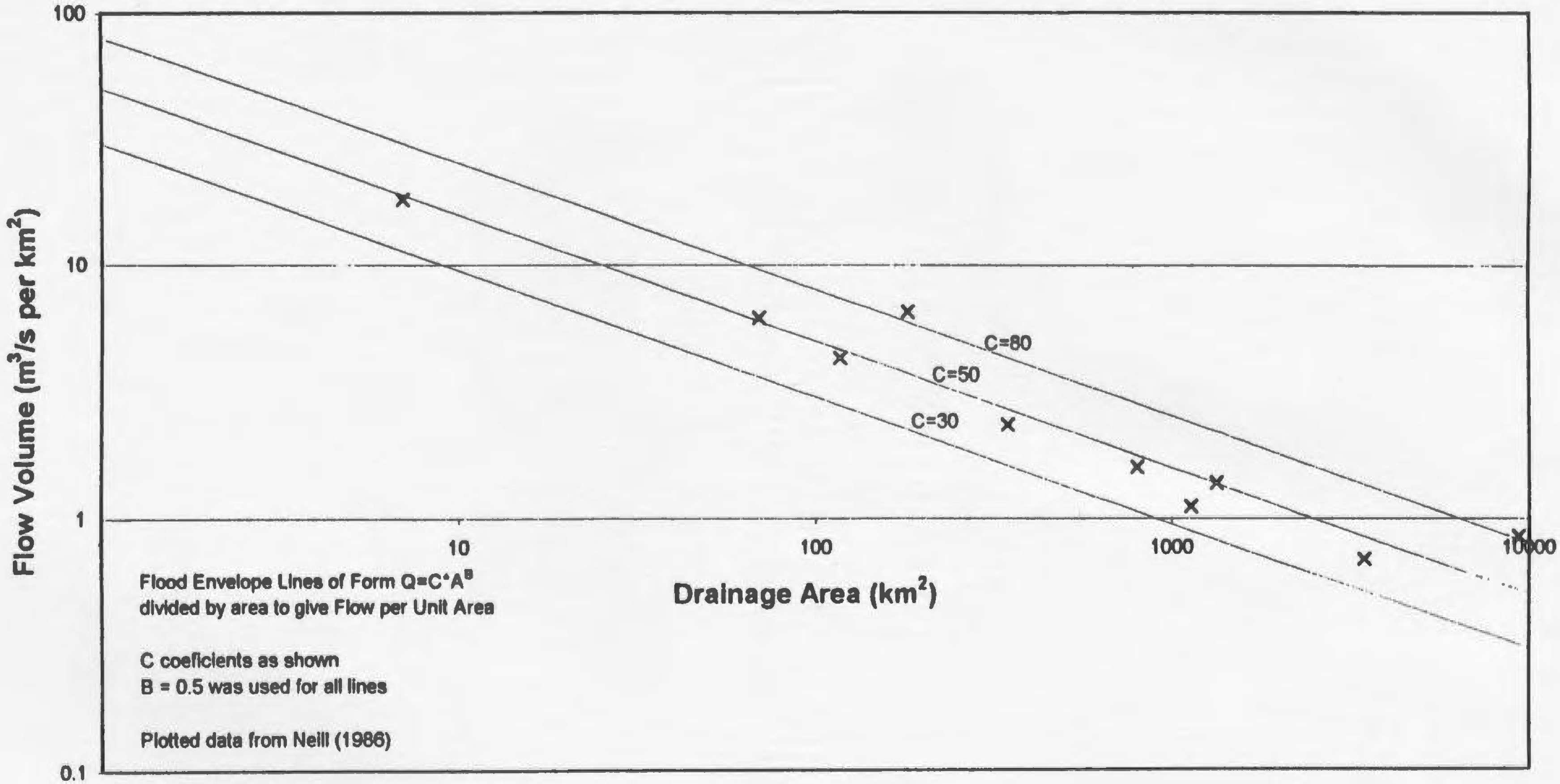

Figure 1.1 Flood Envelope Chart with data from Neill (1986)

2. DESCRIPTION OF STUDY AREA

In this chapter the location, topography, climate and hydrology of Insular Newfoundland are discussed. In addition, in the final section of this chapter, the availability of streamflow data for the island is presented.

2.1 Location of Study Area

The Province of Newfoundland and Labrador, the easternmost province of Canada, consists of an island portion, Newfoundland, and a continental portion, Labrador, as shown in Figure 2.1. The island portion has an area of 111 390 square kilometres (DOE, 1992). The island is subject to continental weather from Canada as well as the Eastern seaboard of the United States. The waters of the Gulf of St. Lawrence and North Atlantic surround the island and moderate continental effects while the Labrador Current and Gulf Stream both act to influence island climates. Because of these influences, the streamflow records for Newfoundland do not always exhibit the same behaviour as records at similar latitudes in Canada.

2.2 Topography and Land Use

Cassie and El-Jabi (1991a) treated Newfoundland was treated as a single homogeneous hydrologic region. However, the island of Newfoundland has a diverse geographic makeup. The *Water Resources Atlas of Newfoundland* (DOE, 1992), states that, while most of the terrain is hilly and rugged, the nature of the landforms, surficial geology, and ground cover vary greatly and from east to west. A map of the island is shown in Figure 2.2.

The west coast is dominated by the Long Range Mountains, a part of a chain which stretches as far south as New England. In Newfoundland these mountains extend from the southwestern tip of the island to the end of the Northern Peninsula. The terrain ranges from 200 to 600 metres in elevation with some higher peaks (DOE, 1992). The mountains and long coastal inlets have profound localized impacts on the hydrology of this area. Much of this area is sparsely populated but timber harvesting activity is prevalent throughout. The Southwestern corner of the island is exposed to incoming storms and moist ocean air. Strong orographic influences may dominate the local hydrology. This area is sparsely populated, and much of the terrain is barren.

Terrain in the central region ranges in elevation from 200 to 300 metres (DOE, 1992). This area is also sparsely populated and timber harvesting is prevalent. The Avalon zone is connected to the main body of the Island by a narrow isthmus. This region has lower more undulating topography with isolated peaks to 300 metres (DOE, 1992). This area is the most densely settled area of the province and contains the provincial capital.

2.3 Climate

Newfoundland is subject to varying weather patterns influenced by latitude, general atmospheric circulation, continental weather, and ocean currents. The normal seasonal conditions of Canada are prevalent, but there may be variations because of the strong influence of the surrounding ocean. A mild winter and cool summer are typical (DOE, 1992).

Temperature varies across the island with five degrees Celsius the average for the Avalon and Burin Peninsula regions and one degree Celsius average for the Northern Peninsula (DOE, 1992). Mean annual precipitation varies from 779 millimetres to 1644 millimetres across the island (DOE, 1992). Richter (1994) described the climate as cool, moist and maritime, characterized by unsettled weather with few extremes of temperature or precipitation.

The island is positioned in the belt of westerly trade winds (Richter, 1994). Prevailing winds flow from west to east bringing air and weather patterns from Eastern North America. Storms tend to cross the island in a generally southwest to northwest direction (Richter, 1994). In summer the prevailing westerly flow delivers warm air from the continent, and in winter cold continental air is delivered to the region. The continental influence on local air temperatures is moderated by surrounding water and tends to decrease as one moves to the east. Variations in the position of the jet stream may produce winter conditions with incoming cold air from eastern Canada, or warm air from the eastern seaboard of the United States. Some parts of Newfoundland frequently experience midwinter warming which may persist for days.

Ocean circulation also has a major impact on Newfoundland weather. Along the Northern and Northeastern coasts the Labrador Current delivers cold water throughout the year. Along the South coast there is a strong impact from the warm Gulf Stream and many inlets remain ice free. The cold Labrador Current and the warm Gulf Stream converge at the southeast tip of the island and produce variability in atmospheric conditions. Fog is common in this region.

2.4 General Hydrology

Newfoundland streamflow records typically follow the normal patterns for continental North America. There are usually a spring peak and a fall peak with the spring peak being the most significant. However, as a result of the climate variability mentioned earlier, there are Newfoundland streams which do not fit the continental hydrology pattern or are subject to more variability than normal. The hydrographs in Figure 2.3 illustrate the variety of hydrologic patterns which are present in Newfoundland.

The streamflow records presented in Figure 2.3 are produced from the average daily flow over the period of record for each gauging station, smoothed by a seven day moving average. This average daily flow approach was adopted because it is more representative of general behaviour of the stream than one year of record. Torrent River, Figure 2.3a, on the Northern Peninsula, has a large spring runoff with much lower peaks later in the year. This river is most likely exhibiting a significant melt-out in the spring which produces the peak streamflow for this basin. However, some additional high flows occur as a result of storms later in the year. Northeast Pond River, Figure 2.3b, in the southeastern part of the island typically has its highest peak flow in the spring but also has significant events in the fall. In this area of the island, the occurrence of peak flows is less associated with snowmelt and more associated with storm events and rain-on-snow events. The Humber River, Figure 2.3c, shows a significant spring peak around April/May and then much lower peaks in the fall. In this basin, snowmelt produces significant runoff which generates high spring flows, but this river has a large basin which tends to attenuate the influence of storm events. Gander River, Figure 2.3d, shows a high spring peak, most likely associated with snowmelt, and some fall peaks which are associated with a fall storm events.

Surface water is more important than groundwater in Newfoundland (Richter, 1994). The island geology with a few exceptions is characterised by bedrock with a thin veneer of glacial till (Richter, 1994). Infiltration effects and aquifer storage do not have the significant impacts on flood flows which they have in regions with deep soil cover. This would lead to an expectation of quick runoff and basins which were highly responsive to variations in rainfall input. However, in many basins, the presence of numerous water bodies and swamps flattens flood hydrographs (ACH, 1989).

Causes of flooding on the island of Newfoundland include rainfall alone, rainfall plus melting snow, ice jamming, and tidal events (ACH, 1989). Severe flooding which occurred in 1983 involved rainfall, melting snow, and ice jamming (ACH, 1989).

2.5 Seasonal Effects

Seasonal variations may be a source of problems in flood series analysis (Ashkar, 1994). The peaks associated with spring and fall may be different enough in mean and variance to comprise two different populations. Most annual flood series in Canada contain floods of two types which, on occasion, comprise two populations (ACH, 1989). It may not be feasible to assume that the daily flows of May have the same distribution as those of December (Taesombut and Yevjevich, 1978).

In the Avalon and Burin Peninsula areas of the island most peak flows are the result of rainfall combined with melting snow (Beersing, 1990). However, peak flows have occurred in every month of the year and are not strongly grouped into one season. In the Central area most of the peak flows occur in April and May and are primarily caused by melting snow (Beersing, 1990). In the Northwest area melting snow is also the most prevalent cause of peak flows but peaks occur from

April to June (Beersing, 1990). In the Southwest area most peak flows occur most often between October and December as a result of rainfall events (Beersing, 1990).

A treatment which divides the flow record into seasons complicates the preparation of frequency analysis considerably (ACH, 1989). There is little reason to perform this division unless treatment as a single population produces a peculiar problem (ACH, 1989). In addition, for long data series, peak size tends to override seasonal effects (NERC, 1975). Ashkar (1994) considered only spring peaks. However, this type of data censoring is what one is trying to avoid by using the POT method.

In Newfoundland, peak flows occur in the periods April through June and November through February with little distinction as to timing between rainfall only and rain with melting snow events (ACH, 1989). For a study of flood quantiles the timing of the flood within the hydrologic year is of less interest than its magnitude. Because seasonal effects are poorly defined for Newfoundland, and because the modelling of seasonal effects increases model complexity significantly, seasonal variations were not modelled in this thesis.

2.6 Availability of Data

Data on streamflow is limited for much of Newfoundland. An area of 111 390 square kilometres has 93 active hydrometric stations. Most streamflow record are short and long records are biased toward larger watersheds.

Including both active and discontinued locations, data are available for one hundred and eleven numbered hydrometric stations at various locations throughout the province. Records vary in length from one year to about seventy years. Physiographic data for gauged basins are available from the Department of Environment and Labour, Government of Newfoundland.

Climate records are available from the Atmospheric Environment Service, Environment Canada. However, the climate network is sparse and most stations are coastal and at low elevation, making the data of limited use for hydrologic analysis (Richter, 1994).

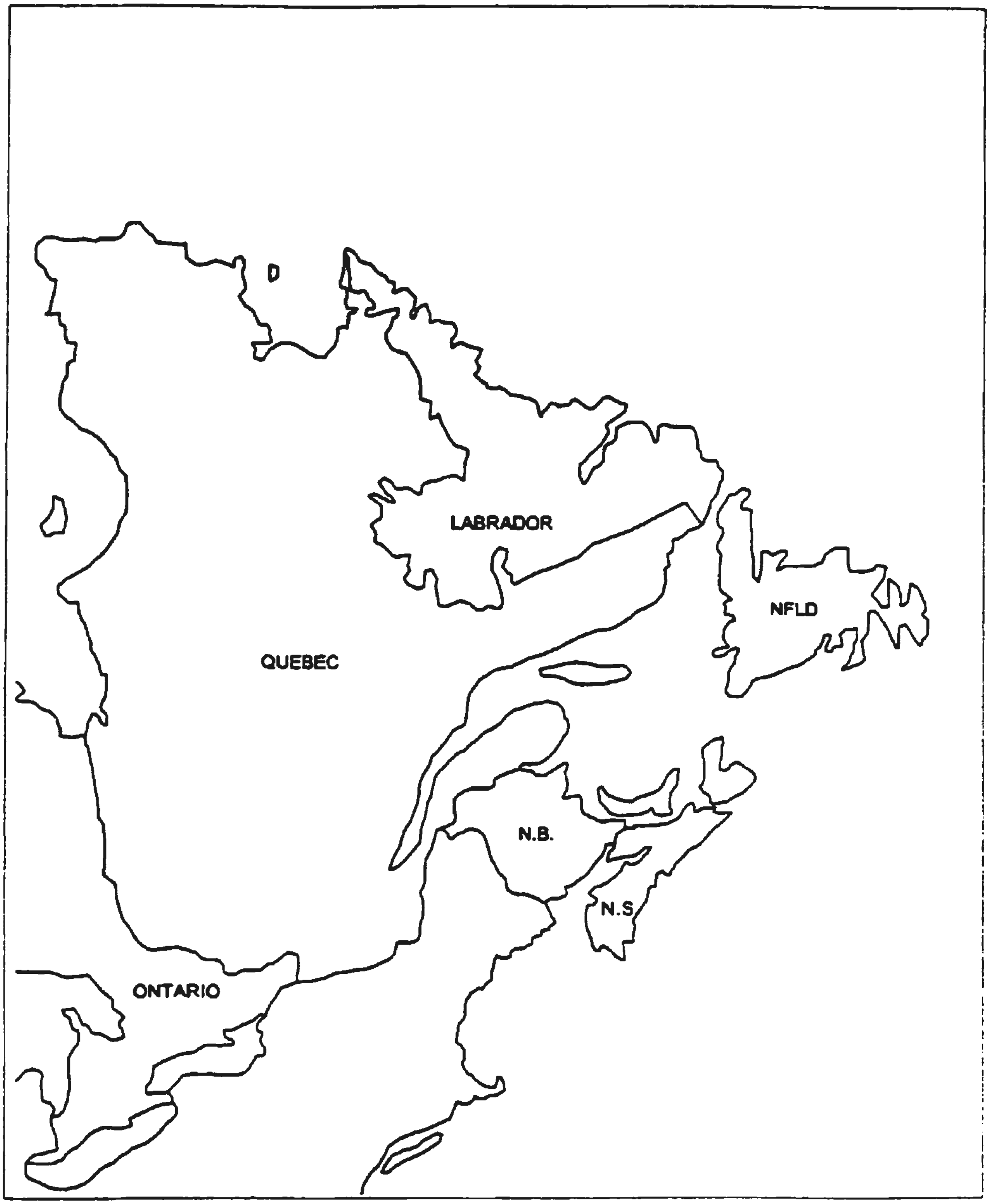

Figure 2.1 Map of Eastern Canada

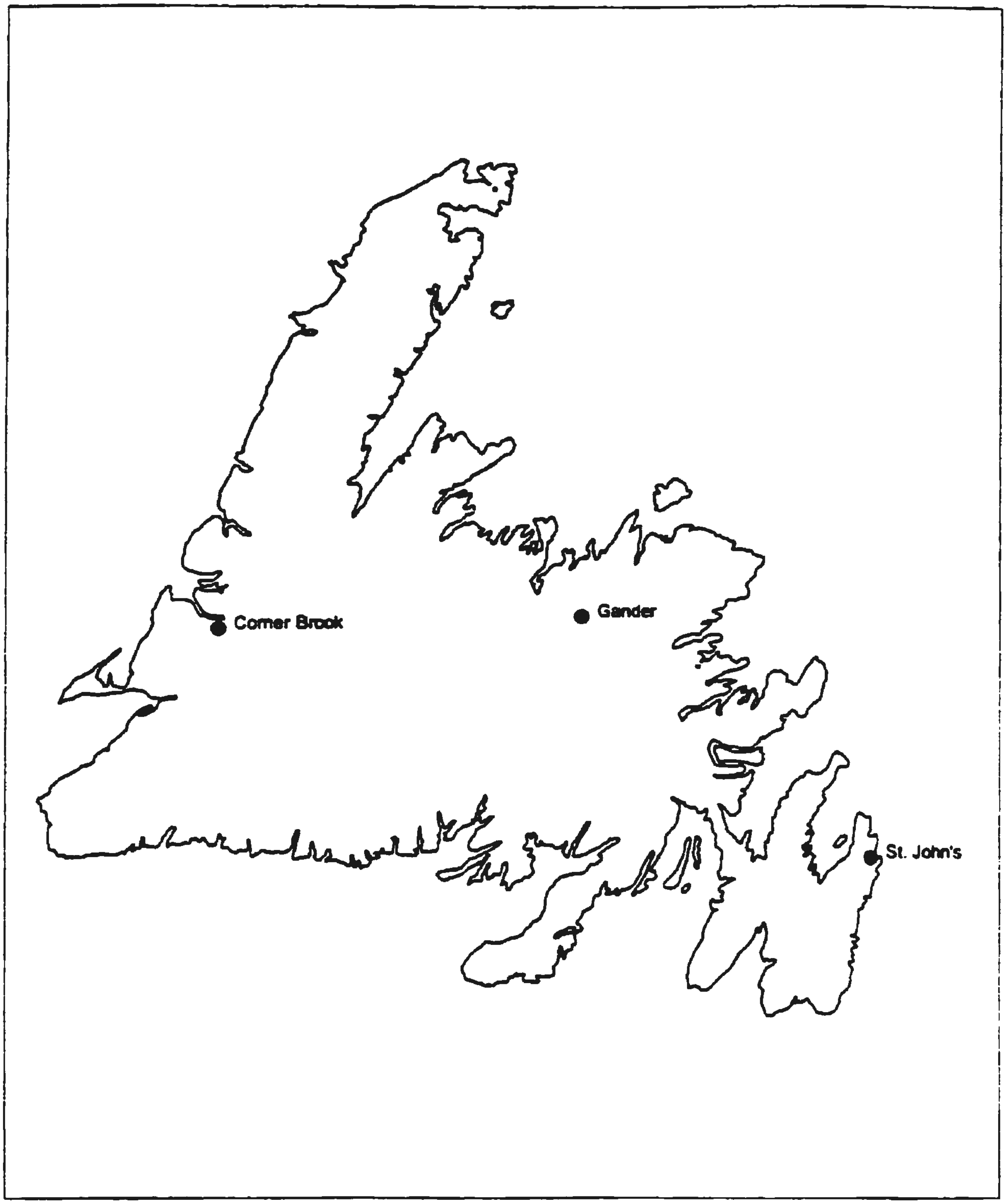

Figure 2.2 Map of Newfoundland

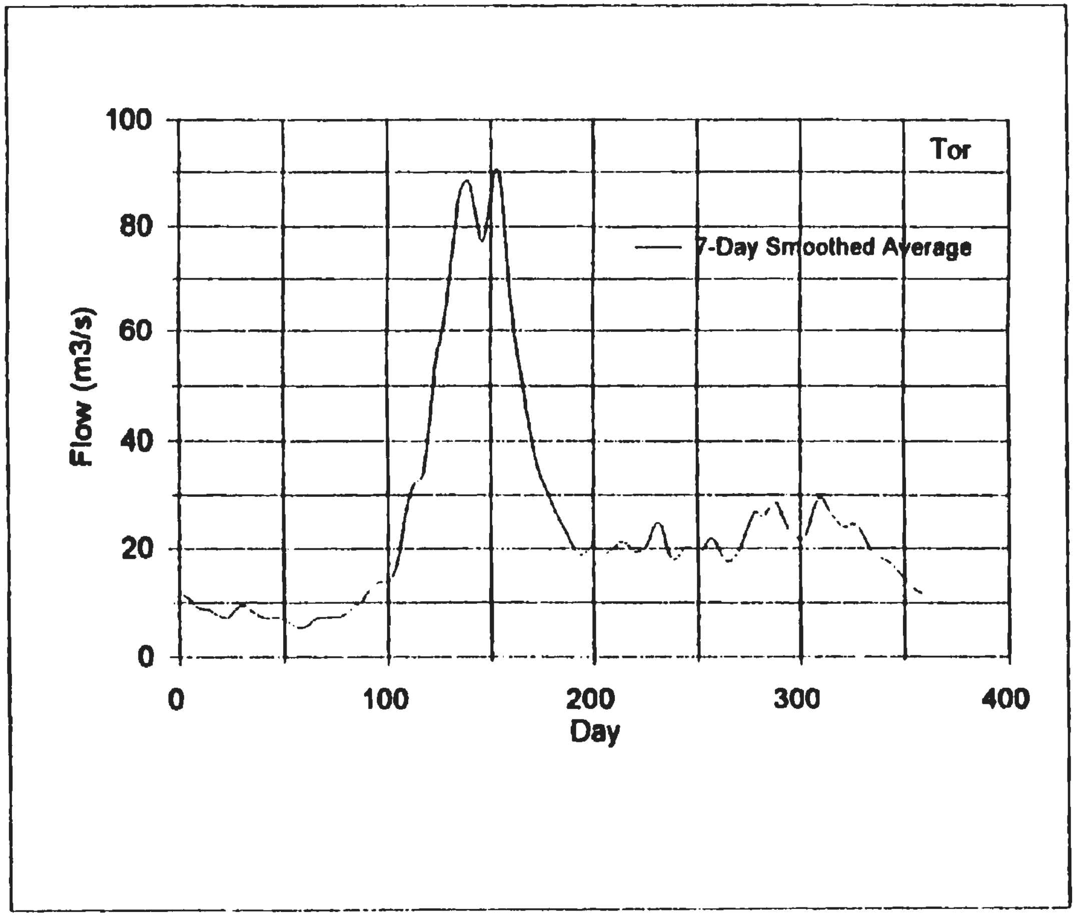

Figure 2.3a **Torrent River at Bristols Pool - 02YC001**

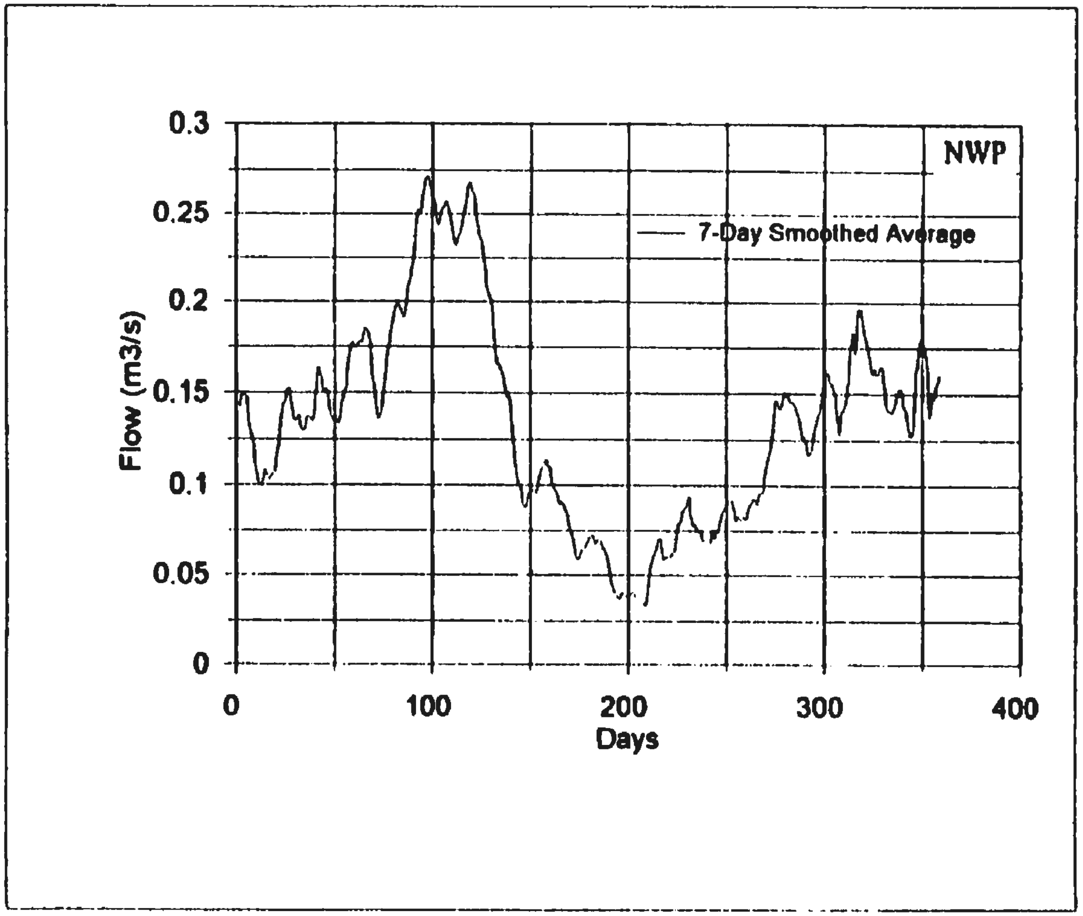

Figure 2.3b Northeast Pond River - 02ZM006

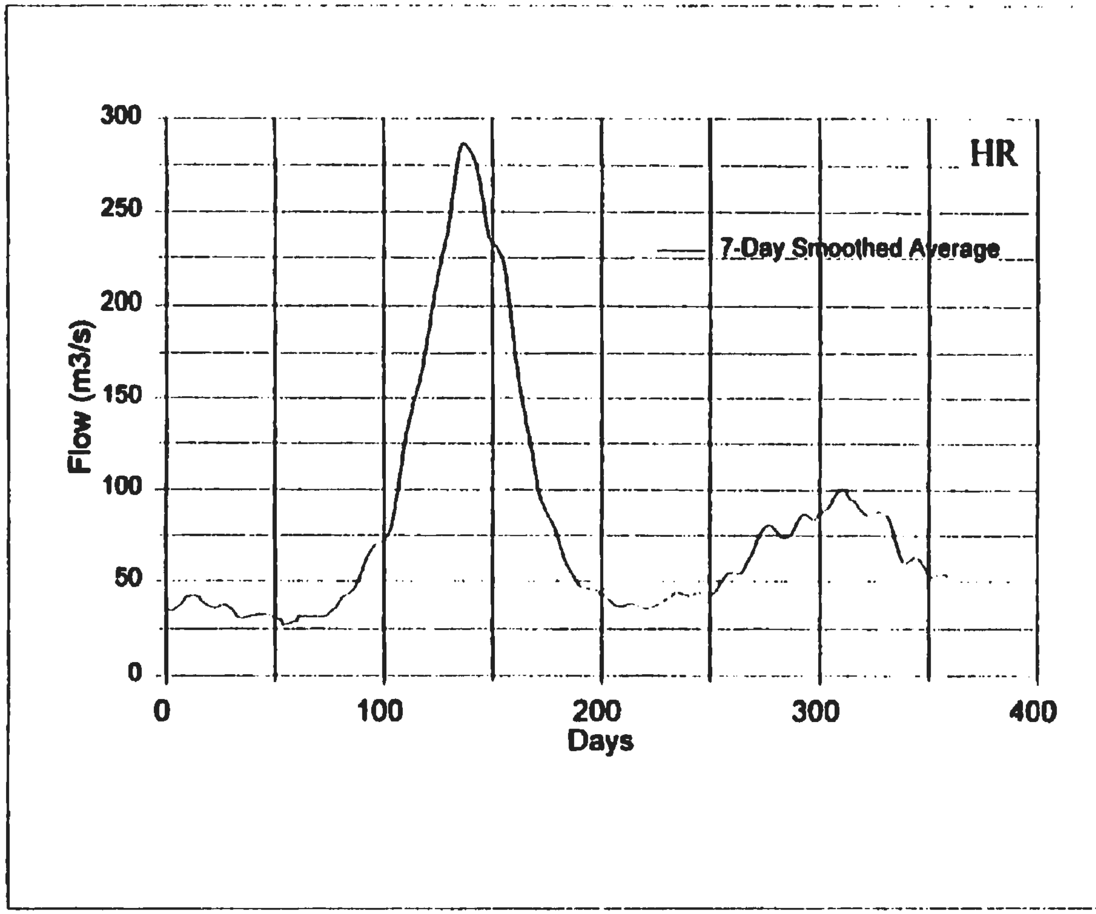

Figure 2.3c Humber River at Reidville - 02YL001

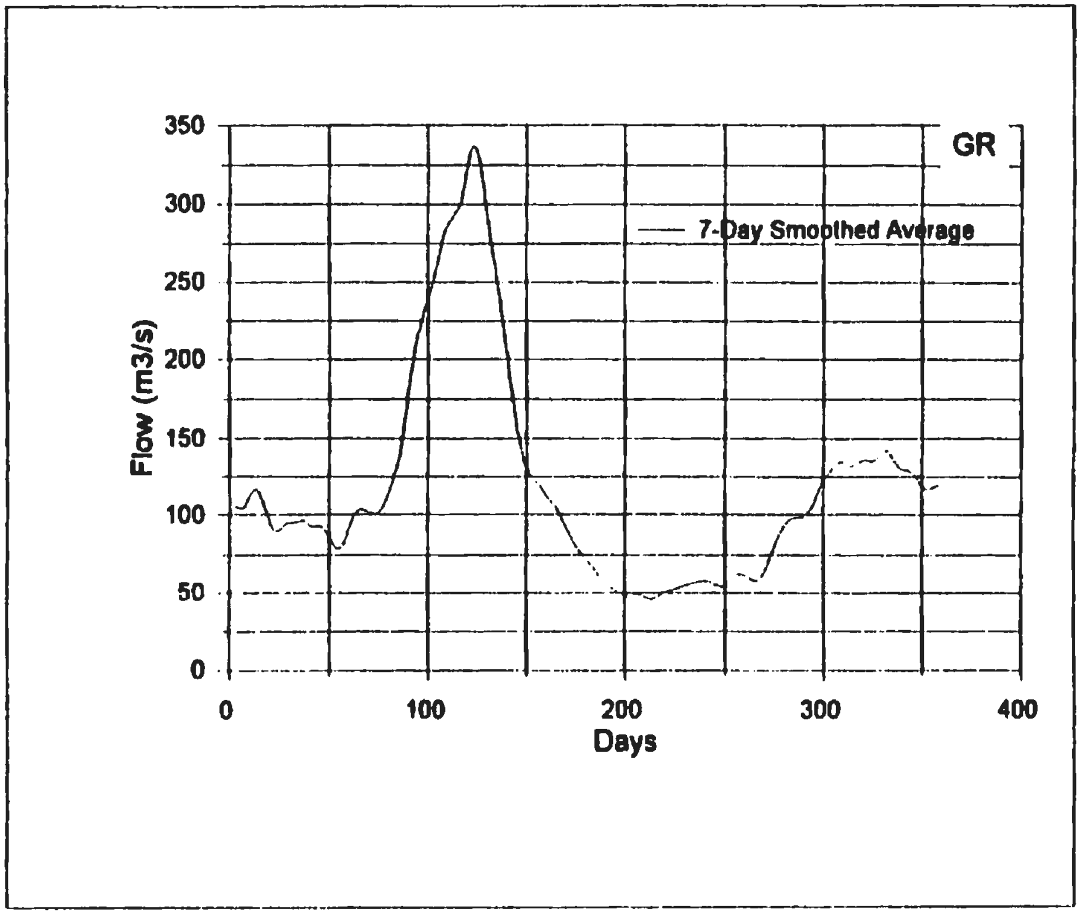

Figure 2.3d Gander River at Big Chute - 02YQ001

3.0 SINGLE STATION ANALYSIS

This chapter discusses the approaches to fitting frequency distributions to data sets for individual gauged streams. Analysis using the Annual Maximum Flow Series and Peak-Over-Threshold Series are compared and then each method is discussed in some detail. Probability distributions associated with each approach are discussed. Finally, the quantile estimators for each method are explained.

3.1 Peak-Over-Threshold versus Annual Maxima

When applying any statistical method, it is preferable that the maximum amount of raw data is incorporated into the analysis. By including more data, a statistical model can be made to fit nature more closely and to model the system under study through a wide range of conditions and states. However, the researcher is not always interested in the total behaviour of the system. In most cases, the results desired relate to the centre of the data and the upper and lower extremes. Models developed to predict probability of occurrence of streamflow maximums will generally not be improved by inputting data which relates to the low flow characteristics. This additional data does nothing to improve the behaviour of the model and increases the computational load. There is always a trade off between inclusiveness and utility.

As an example, a record of daily flows contains a large amount of data. Five years of data contains approximately 1826 data points. As one can see from Figure 3.1, there is a lot of information in the data set, however it is difficult to make this information meaningful in terms of peak flow events.

The Annual Maximum Flow, AMF, series excludes everything except the maximum flow for a given year. Any flow events within the year with magnitude less than the annual maximum are discarded. This data may be used to construct models which estimate probability of occurrence of future flood magnitudes. The disadvantage of this method is that multiple events in any year may be higher than the maximum in another year, but these events are discarded if lower than the annual maximum. The advantage is that it is simple to extract the annual maxima, and as one can see from Figure 3.2, the amount of data which must be manipulated is greatly reduced.

The Peak-Over-Threshold, POT, series is generated using a different approach than the annual maximum flow series. In the POT approach, all events which exceed a specific threshold, q_o, are counted as flood events and are included in the extracted data series. As shown in Figure 3.3, this can produce a greater number of events than the AMF series while keeping the amount of data manageable. In addition, by proper selection of q_o, the modeller may include more events which are more representative of peak flow conditions than would be available using the AMF approach. As mentioned above, the AMF method may discard significant events which may be included in analysis using a POT approach. The use of more data, and the use of data which more directly reflect peak flow conditions, should result in improved fit of flood quantile estimators. However, the idea that more data are necessarily better is not always true. In some cases the AMF approach produces smaller estimate variability than the POT approach. (NERC, 1975)

Where there is a scarcity of streamflow data, the researcher is faced with a problem. How does one estimate the probability of future events with only a limited knowledge of what has gone before? The solution to this problem, in some cases, may lie in a more intensive examination of the data which does exist. It is possible that additional information has been suppressed by the application of methods like AMF, which excludes all but the maximum event in any given year. An alternative approach like the POT method, which extracts more information from the available data, may allow a researcher to better fit a probability distribution to the streamflow series. Thus, peak-over-threshold based estimation procedures may be useful in estimating floods when there is a limited amount of data (Ashkar, 1994).

The main strength of POT models in comparison to AMF is that, by appropriate selection of the threshold they allow a better inclusion of events which are to be considered floods (Ashkar, 1994). Taesombut and Yevjevich (1978) suggested that some of the problems of short streamflow records could be overcome by the consideration of all the flood peaks above a carefully set threshold. The estimates generated from POT series should be subject to lesser uncertainty than those generated from AMF data if the threshold is selected properly (Taesombut and Yevjevich, 1978).

POT thresholds are usually selected to include a greater number of events than would be produced by the AMF method. Generally the fit of POT models is better than AMF for low quantiles, and is known to deteriorate at higher quantiles where the threshold is selected too low (Wang, 1991). Care should be exercised in the use of the POT method to derive quantile estimates for events with high return periods (NERC, 1975). POT outcomes may depart significantly from those developed using AMF series.

Where the threshold is selected to produce an average of one flood per year, equivalent to the AMF method, the POT model and AMF model have similar efficiency for high quantile estimation (Wang, 1991). In addition, for long records the estimate produced from POT and AMF series will tend to converge.

3.2 Annual Maxima Models

In the AMF approach, a probability distribution is fitted to the series of annual maximum flows. This allows one to predict the probability of occurrence of a given flood magnitude. The series of annual maximum flow events is generally assumed to be independent and stationary (Bobée and

Rasmussen, 1995). The independence of individual flood events in the series makes sense, since most of these events are separated by substantial time periods. A number of tests are also available to test the data series for serial correlation or trend. If a serial correlation or trend is detected in the data, the fundamental assumptions of the probability model are violated and measures must be taken to model the data differently.

A number of probability distributions have been proposed as models for flood frequency. Some distributions are better at modelling the behaviour of the data within the range of the data set, and some distributions are better at modelling the estimated values outside the known data (Bobée and Rasmussen, 1995). In this thesis, the focus is on prediction of flood quantiles, most of which are outside the known data. A number of models have been discussed in literature and some have been specifically developed for the purpose of predicting frequency of occurrence of flood flows. Most notable among the models used for flood frequency analysis are the Three Parameter Log-normal (3LN), Generalized Extreme Value (GEV), Gumbel, Wakeby, and Log-Pearson. All of these models have relative advantages and disadvantages.

In *Regional Flood Frequency Analysis for the Island of Newfoundland* (Beersing, 1990), the annual maxima series was modelled using the best fitting of the GEV and 3LN. Of the thirty-nine stations considered, the GEV model had the best fit for eighteen stations, and the 3LN best fit the other twenty-one (Beersing, 1990). The results for both models were very close, within five percent, and using the criteria of that study, either model would have made an adequate fit (Beersing, 1990). In this thesis, results of the 3LN and GEV fit to the annual maximum flow series are used for comparison to the results of peak-over-threshold modelling. The application of these distributions is explained in the following two sections.

3.2.1 Lognormal Distribution

The two parameter Log-normal, 2LN, and three parameter Log-normal, 3LN, models are adaptations which allow the use of the Normal, or Gaussian, distribution to predict flood quantiles. The Three Parameter Log-normal distribution has been used extensively throughout Canada and the United States (ACH, 1989). The model is well understood and works reasonably well for many flood series.

Normal distribution curves may be completely described by two parameters, their mean and variance. However, the familiar bell shaped curve of the Normal distribution has a range along the x-axis described by $-\infty < x < \infty$, while most hydrologic phenomena have a lower bound of zero (R. L. Bras, 1990). To overcome the inconsistency between the data and the distribution, the data may be transformed into logarithmic space. In the 2LN distribution, a new transformed variable is developed, $y = ln(x)$, and for the 2LN model the two parameters are the mean and variance of the transformed variable. Although this transformation resolves the issue of the lower bound, a better fit may generally be obtained by the introduction of a third parameter which modifies the position of the data prior to transforming it into logarithmic space. In the 3LN distribution, the new transformed variable is $y = \ln(x-\xi)$, and for the 3LN model the three parameters are the lower bound ξ, and the mean and variance of the transformed variable.

For positively skewed data the parameter ξ, is a lower bound which may be estimated from the x-data using a formula given in Maidment (1992):

$$\xi = \frac{x_{max} x_{min} - x_{median}^2}{x_{max} + x_{min} - 2x_{median}} \tag{3}$$

This process works well for positively skewed data. However, the 3LN distribution does not work well for negatively skewed data. When the data are negatively skewed the formula standard deviation (σ_x), and skew (g_1) of the x-values. The second step is to solve the foll for the transformed y-values changes to $y = \ln(\xi - x)$ and ξ becomes a positive upper bound. A more general method for derivation of the lower or upper bound using method of moments estimators is elaborated by Pilon and Harvey (1994), and is preferable to the estimator given by Maidment (1992) when data may be negatively skewed. The first step is to obtain the mean (μ_x), owing equation for c:

$$c^3+3c-g_1=0 \tag{4}$$

In most applications the x-values are positively skewed. Following Pilon and Harvey (1994), the lower bound may be estimated for positively skewed x-values using Equation 5.

$$\xi=\mu_x-\frac{\sigma_x}{c} \tag{5}$$

Where x-values are negatively skewed, the upper bound may be estimated using Equation 6.

$$\xi=\mu_x+\frac{\sigma_x}{c} \tag{6}$$

The transformed variable $y = ln(x-\xi)$ or $y = ln(\xi-x)$, has a range $-\infty < x < \infty$, consistent with the Normal distribution, which has a probability density distribution which is effectively described by Equation 7:

$$f(y)=\frac{1}{\sigma_y\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{y-\mu_y}{\sigma_y})^2) \tag{7}$$

Within the transformed space one can find μ_y, the sample mean and σ_y, the sample standard deviation of y. Although a simple equation for the cumulative distribution function is not available, it is easy to estimate the probability of non-exceedence of any y- value by use of the standard deviate $z = (y-\mu_y)/\sigma_y$, and standard tables for the calculation of cumulative probability for the Normal distribution.

3.2.2 Generalized Extreme Value Distribution

The Generalized Extreme Value, GEV, distribution has also been applied with success in most regions of Canada. The GEV type distributions are divided into three classes corresponding to the shape parameter, k. (Pilon and Harvey, 1994).. If k<0 the distribution is a Frechet's Type II, EV2, if k=0 it is a Gumbel Type I, EV1, and if K>0 it is a Weibull Type III, EV3 (Martins and Stedinger, 2000). The k-value is generally in the range -0.6<k<0.6 (Pilon and Harvey, 1994). The cumulative probability distribution function is effectively described by Equation 8:

$$P(x) = e^{-[1-\frac{k}{\alpha}(x-\xi)]^{1/k}} \tag{8}$$

This can be seen to be equivalent to the equation describing the Poisson-Pareto distribution used with the POT series in later sections.

The derivation of the parameters of the GEV is somewhat more complex than for the 3LN, and distribution parameters are typically estimated using probability weighted moments. The distribution is described by three parameters: ξ, is a bound or location parameter, α is a scale parameter, and k the shape parameter. In a paper on the Pareto distribution, Rosbjerg et.al. (1992), indicated that method of moments estimators were as efficient as others for parameters of the GEV and Pareto distributions. Using method of moments, the scale parameter α, and the shape parameter k, may be estimated as shown in Equations 9 and 10 respectively.

$$\alpha=\frac{1}{2}\mu(\frac{\mu^2}{\sigma^2}+1) \tag{9}$$

$$k=\frac{1}{2}(\frac{\mu^2}{\sigma^2}-1) \tag{10}$$

If k=0, the distribution is defined as a two parameter EV1, or Gumbel Distribution. If k is less than zero then the distribution is an EV2 and the lower bound ξ, is defined by equation 12. If k is greater than zero then the distribution is defined as an EV3 and the upper bound ξ, is defined by Equation 11.

$$\xi = \mu + \alpha / k \tag{11}$$

Alternately, the parameters may be estimated using L-moments or maximum likelihood estimators. In the L-moment approach the value of k is estimated first (Martins and Stedinger, 2000).

$$k = 7.859z + 2.9554z^2 \tag{12}$$

and

$$z = 2/(\tau_3 + 3) - \ln 2/\ln 3 \tag{13}$$

where

$$\tau_3 = \lambda_3 / \lambda_2 \tag{14}$$

and λ_1, λ_2, and λ_3 are the L-moment estimators. A number of methods are available for estimating the λ_i values and are discussed in depth in Pilon and Harvey (1994) and Martins and Stedinger (2000). Once k has been calculated, the other parameters are calculated easily.

$$\alpha = \lambda_2 k / \{(1 - 2^{-k)}\Gamma(1 + k)\} \tag{15}$$

$$\xi = \lambda_1 + \alpha\{\Gamma(1 + k) - 1\}/k \tag{16}$$

GEV parameter estimates using method of moments or L-moments have both been found satisfactory by a number of researchers (Martins and Stedinger, 2000). Maximum likelihood estimators, MLE's, have also been used to estimate GEV parameters but have performed poorly for small samples (Martins and Stedinger, 2000).

3.3 Peak-over-threshold Approach

3.3.1 Setting the Threshold

The first critical decision in the design of a POT model is the selection of a threshold value. Some researchers select the threshold, q_o, based on physical constraints which determine whether or not an event is relevant. Other researchers have indicated that the threshold should be selected to produce a preselected recurrence rate for flood events. Still others have suggested selecting the threshold to produce a POT series which has characteristics of the distribution used to model it.

High thresholds are those which produce average peak recurrence rate of less than one event per year, extracting less peaks than would be contained in the AMF series. For high thresholds, the quality of low quantile estimates will tend to deteriorate, but the quality of high quantile estimates may improve slightly (Wang, 1991). The improvement in high quantile estimates is limited, and where short data records are being studied it is difficult to justify using less than one peak per year. In *Hydrology of Floods in Canada* (ACH, 1989), and in the work of Taesombut and Yevjevich (1978), a minimum of 1.65 peaks per year is recommended for POT series. High threshold series are not considered in this thesis.

Low thresholds produce large mean annual recurrence rates. The difficulty with low thresholds is that independence of peak events may be compromised. In addition, additional peaks introduced by lowering the threshold correspond to events with a high probability of occurrence. These events contain less information related to flood events which have a relatively low probability of occurrence. Thus, the calculation load is increased with no increase in model performance.

Some researchers argue that q_o should be selected based on real physical conditions of the stream (Caissie & El-Jabi, 1991b). These physical constraints may include bank-full conditions, hydraulic capacity of the stream, percentage of mean flow or other parameters. This approach produces a series of events which can be identified as floods, but the POT series produced may not be amenable to statistical treatment.

Other researchers adopt an approach where an average annual rate of exceedance, λ, is preset and q_o is adjusted to produce this value of λ. In general, the base is selected low enough that at least one event in each year is included. Taesombut and Yevjevich (1978) found that where $\lambda \geq 1.65$, the results of models constructed from POT series had less variance than those from AMF series.

Cassie and El-Jabi (1991b) suggested using the mean to variance ratio of flood recurrence to set q_o. Assuming a Poisson arrival process for flood recurrence, $\lambda/\sigma^2 = 1$, where σ^2 is the variance in recurrence rate. Because it is generated from the data, this threshold setting method has the advantage of being somewhat more robust and less arbitrary than the preceding two. However, in using this approach significant numbers of iterations may be required to obtain a threshold which satisfies the $\lambda=\sigma^2$ criteria, and the $\lambda=\sigma^2$ criteria may be satisfied at thresholds which produce very high or very low recurrence rates. Some judgement may be required on the part of the researcher to determine if the threshold selected using this method will produce the type of data series desired.

3.3.2 Selecting Independent Peaks

One major concern of users of peak-over-threshold analysis, is that the sequence of events extracted might be dependent since some peaks may occur on the recession limb of a prior event (Taesombut and Yevjevich, 1978). However, for the proper application of most statistical models, each event must be separate and distinct. A variety of methods have been proposed to ensure the independence of events.

Ashkar (1994) set two criteria for independent flood peaks:

(1) Two consecutive peaks must be separated by at least seven days;

(2) The flow between two consecutive peaks must drop below a specified fraction (50%) of the lesser of the two peaks.

Taesombut and Yevjevich (1978) suggested the Water Resources Council guideline:

(1) Five day separation plus the natural logarithm of the drainage area in square miles;

(2) The flow between two consecutive peaks must drop below 75% of the lower of the two peaks.

For the purposes of this thesis, two criteria were used to exclude dependant peaks:

(1) a minimum seven-day separation;

(2) at least one intervening daily maximum flow below 50% of the lesser of the two peaks.

As stated earlier, groundwater effects on flood flows are limited in Newfoundland because the soil layer is typically thin till over bedrock (Richter, 1994). Thus, the recession limb of flood events is fairly short, and where flows have dropped below 50% of the lower of the two peaks, there is reasonable security in assuming that the effects of the prior event are insignificant in the development of the second. If the threshold is taken adequately high and the criteria for independence applied as given above, the assumption that individual peaks are independent events should be a reasonable one.

3.3.3 Modelling Recurrence Distribution

The second major decision in applying the POT methodology is the distribution selected for modelling recurrence of flood events. This has generally been done with a Poisson recurrence model, but a variety of tenable distributions have been proposed for recurrence (Taesombut and Yevjevich, 1978). Common recurrence distributions are shown in Table 3.1.

Flood peaks may be defined as successes in a series of randomly spaced Bernoulli trials, each representing the occurrence of a peak (Taesombut and Yevjevich, 1978). Where the events are independent, this implies a Poisson arrival process (Taesombut and Yevjevich, 1978). Given a series of length N years, and an average exceedence rate of λ, the total number of expected peaks M is defined as $M=N\lambda$ (NERC, 1975).

For a Poisson process, λ defines the value of the mean and variance of the distribution. Generally this follows the formula of Equation 17:

$$p(x)=\frac{\lambda^{x}\, e^{-\lambda}}{x!} \tag{17}$$

where $x = 0, 1, 2, \ldots$

Which, considering the probability of exceedence any number of times P(1,2, ... n), in a period T, gives:

$$P(1,2,\ldots,\infty) = 1 - e^{-\lambda} \tag{18}$$

This can then be manipulated simply to produce a probability of non-exceedence: the probability that no flow will exceed a given threshold (NERC 1975):

$$P(0) = 1 - P(1,2,\ldots,\infty) = e^{-\lambda} \tag{19}$$

3.3.4 Modelling Magnitude Distribution

We see above that one can produce a probability that flow does or does not exceed q_o, but so far we do not know anything about the magnitude of these exceedence events. The size of the peaks above q_o may be modelled using a continuous distribution such as the exponential (Taesombut and Yevjevich 1978). A variety of tenable distributions for magnitude have been proposed and are shown in Table 3.2.

Taesombut and Yevjevich (1978) found that the exponential distribution had the best fit for magnitude of exceedences. The exponential is the most frequently used distribution for modelling exceedences and, since only one parameter is estimated, may lead to a more precise prediction of flood quantiles than a more complex model (Rosbjerg et.al., 1992).

A probability of non-exceedence for any given flood magnitude which has Poisson recurrence was described by Ekanayake and Cruise (1994):

$$P(x)=e^{-\lambda(1-F(x))} \tag{20}$$

where $x = q - q_o$ and $F(x)$ is the distribution of the magnitude of flood exceedences. If $F(x)$ follows the standard form of the exponential distribution, then

$$F(x)=1-e^{-x/\beta} \tag{21}$$

where β equals μ, the mean of the x values. This may be substituted into Equation 20,

$$P(x)=e^{-\lambda e^{-x/\beta}} \tag{22}$$

which yields the probability of non-exceedence or cumulative distribution function for an event of magnitude x. This model, which looks at a peak-over-threshold series as having a Poisson arrival process and exponential magnitude distribution, may be referred to as the PED model. The results of the PED model follow the same shape as the Gumbel Distribution used to model AMF series (NERC, 1975).

Any distribution for magnitude may be satisfactorily substituted for F(x) if it satisfies the data. Ashkar (1994) described the Pareto distribution as:

$$F(x)=1-(1-\frac{kx}{\alpha})^{\frac{1}{k}} \tag{23}$$

Where α and k are the scale parameter and shape parameter respectively.

Rosbjerg et.al. (1992), expressed these parameters using the method of moments:

$$\alpha=\frac{1}{2}\mu(\frac{\mu^2}{\sigma^2}+1) \tag{24}$$

$$k=\frac{1}{2}(\frac{\mu^2}{\sigma^2}-1) \tag{25}$$

Where μ is the sample mean and σ^2 the sample variance of the magnitude of peak-over-threshold events. These method of moments estimators are simple to use, and were found to be as efficient as estimation by probability weighted moments (Rosbjerg et.al., 1992).

The Pareto distribution equation may be substituted into Equation 20 to give:

$$P(x)=e^{-\lambda(1-\frac{kx}{\alpha})^{\frac{1}{k}}} \tag{26}$$

The advantage of the Pareto distribution is its flexibility and ease of use. The distribution parameters are easily obtained and should produce more consistently reliable results than less flexible single parameter models. This model, with a Poisson arrival process and a Pareto magnitude distribution, may be referred to as the PPD model, and follows the GEV Distribution as used to model flood quantiles for the AMF series.

3.4 Quantile Estimators

By manipulating the form of the cumulative distribution function for the flood frequency distributions, equations may be developed to produce flood quantile estimates. Where the data extracted as peaks over threshold is assumed to have a Poisson recurrence distribution and an Exponential magnitude distribution, the estimate of the flood with probability of exceedence P=1/T, is given by Equation 27:

$$Q(T)=q_o+\beta\ln\lambda+\beta\ln T \tag{27}$$

Where the data extracted as peaks over threshold is assumed to have a Poisson recurrence distribution and a Pareto magnitude distribution, the estimate of the flood with probability of exceedence P=1/T, is given by Equation 28:

$$Q(T)=q_o+\frac{a}{k}[1-(\frac{1}{\lambda T})^k] \tag{28}$$

Where data is extracted as a series of annual maxima, and is assumed to have a 3LN distribution, the quantile estimator of the flood with probability of exceedence P=1/T, is given by Equation 29 (Maidment, 1992):

$$x_T=\xi+e^{(\mu_y+\sigma_y z_T)} \tag{29}$$

where y is the transformed variable, μ_Y is the mean of y, σ_Y is its standard deviation, and ξ is a lower bound parameter described earlier. The constant z_T is the normal score corresponding to the probability of non-exceedence for a given return period "T." These z-scores may be obtained from standard tables.

Where data is extracted as a series of annual maxima, and is assumed to have a GEV distribution, the quantile estimator of the flood with probability of exceedence P=1/T, is given by Equation 30 (Maidment 1992):

$$X_T = \xi + \frac{\alpha}{k}[1-(-\ln(1-1/T))^k)] \tag{30}$$

As mentioned previously, the GEV and 3LN distributions were found to have similar efficiency in fitting the AMF series for Newfoundland (Beersing 1990). The GEV model and 3LN model will be used in this thesis for comparison to the PED model and PPD model.

Table 3.1 Recurrence Distributions for Peaks over Threshold Model (after Taesombut and Yevjevich, 1978).

Distribution	**Parameters**	**Comments**
Poisson	λ	most popular approach
Mixed Poisson	λ_1, λ_2	accounts for seasonal variation
Hyper Poisson	λ, θ	
Negative Binomial	r	
Mixed Geometric	$\theta_1, \theta_2, \gamma, \alpha$	
Non-parametric	$a_1, a_2, a_3 \ldots$	based on data

Table 3.2 Magnitude Distributions for Peaks over Threshold Model (after Taesombut and Yevjevich, 1978).

Distribution	**Parameters**	**Comments**
Exponential	β	Simplest
Gamma	β, γ	
Pearson Type III	X_o, β, γ	
Weibull	a, b	
Mixed Exponential	β_1, β_2	
Pareto	k, α	Flexible, includes exponential as a special case
Normal	μ, σ	
Non-parametric	$a_1, a_2, a_3 \ldots$	Based on data

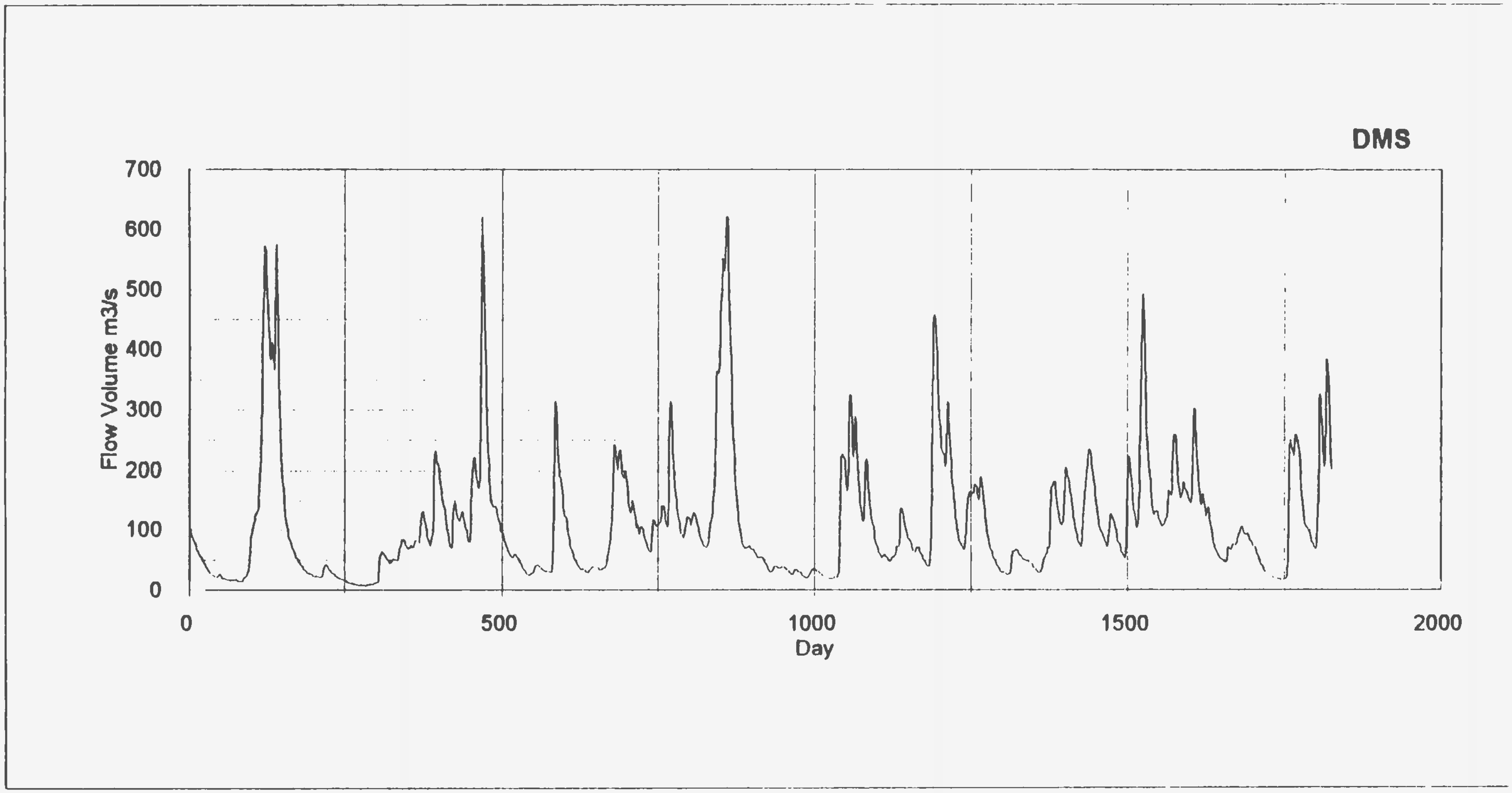

Figure 3.1 Daily Maxima Series

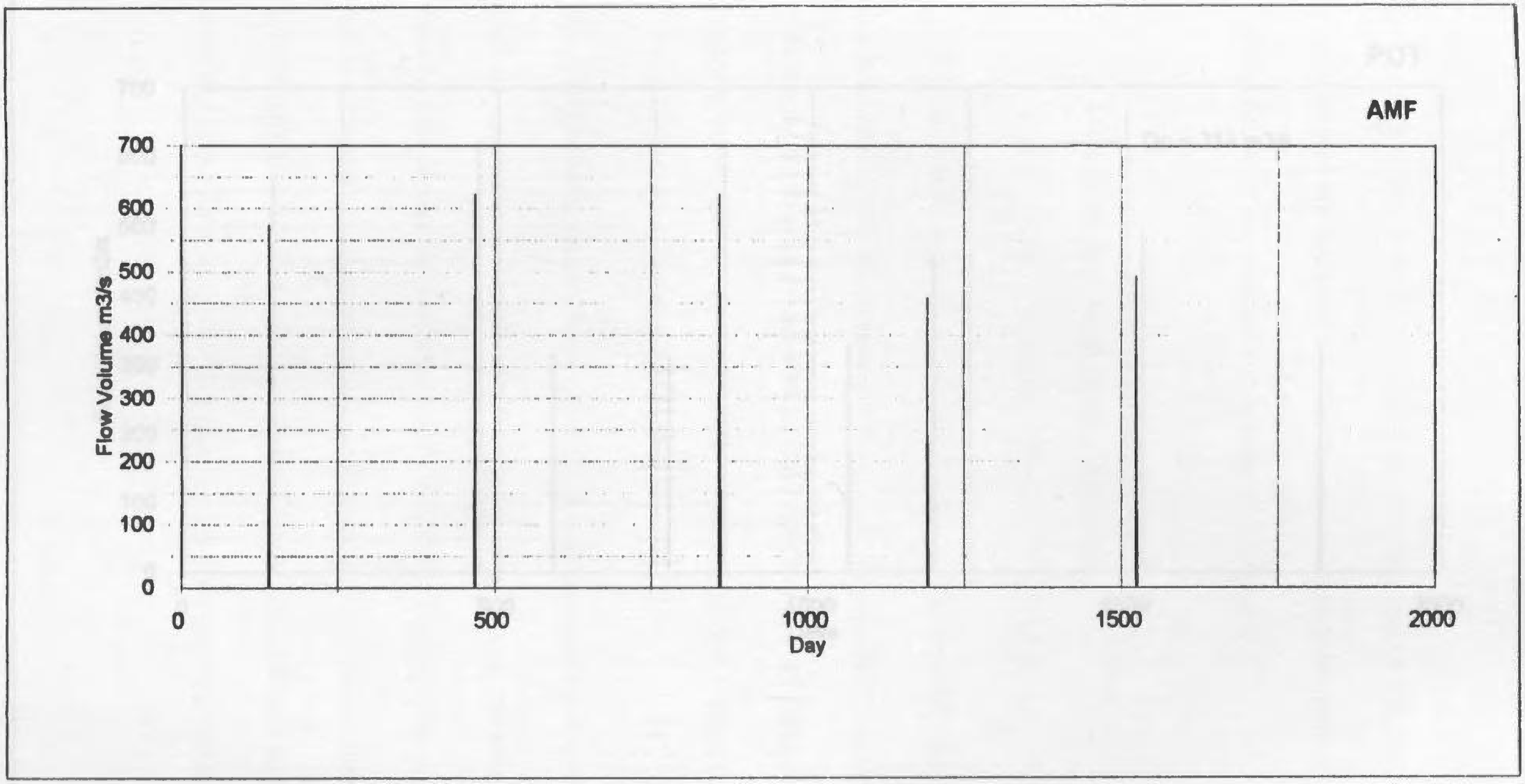

Figure 3.2 Annual Maxima Series

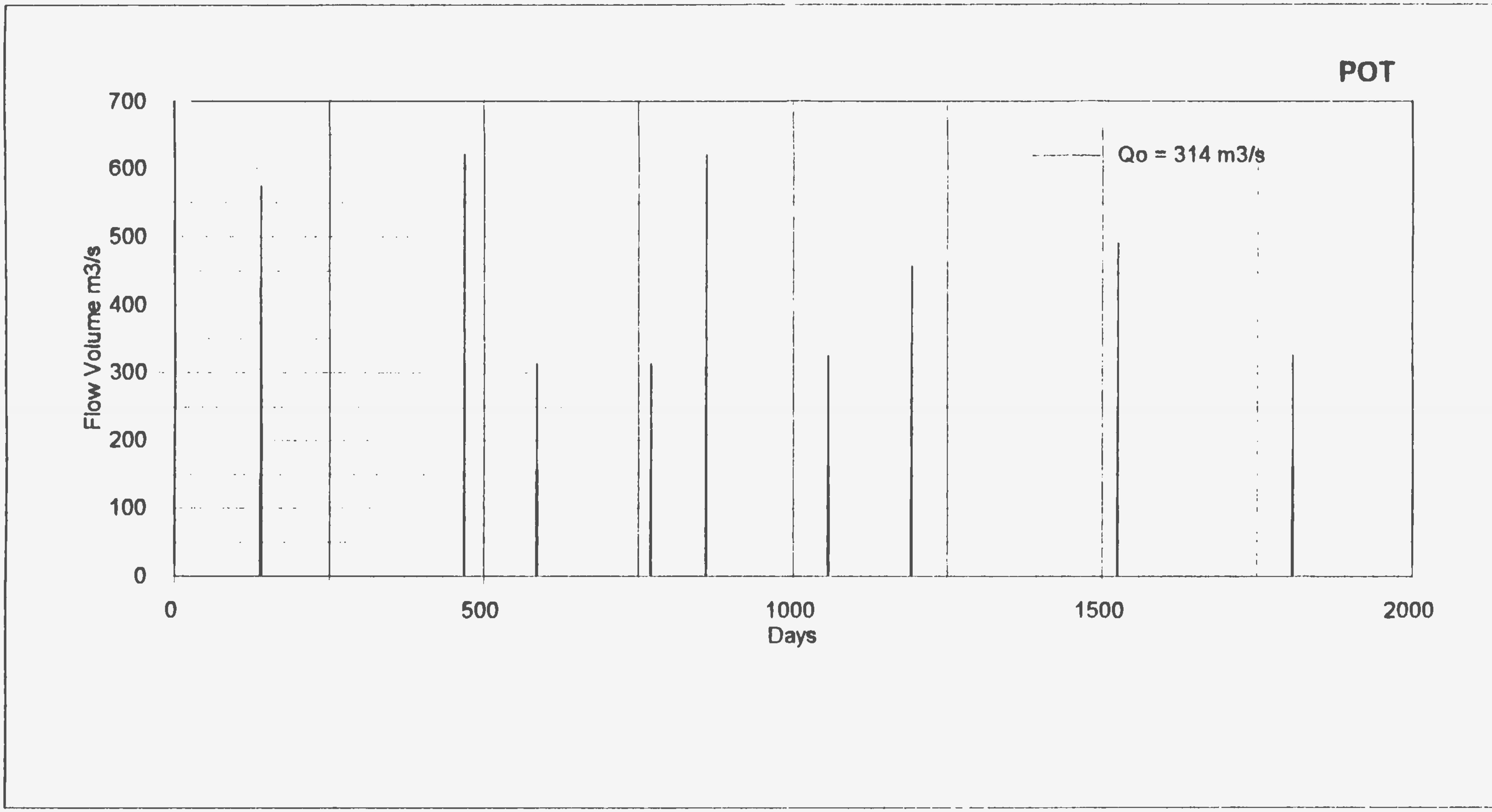

Figure 3.3 Peak-Over-Threshold Series

4.0 REGIONALISATION

In this chapter the method of regionalisation is discussed. Reasons for using the regional approach are given, and methods for determining regional groupings are considered. Previous regional delineations for Newfoundland are also discussed.

4.1 Reasons for Regional Analysis

In simplest terms, regional analysis assumes that one stream in a region will have hydrologic behaviour similar to other streams in that region. Regional flood frequency analysis of streamflow data involves grouping streams with similar hydrologic properties into regions and developing regional equations which estimate flood quantiles from basin descriptors.

The effective estimation of flood quantiles for a gauged stream may require single station analysis, regional analysis, or a combination of both. Where long streamflow records exist, the flood quantiles predicted by single station analysis may be excellent. In fact these estimates may be superior to regional estimates (ACH 1989). However, where streamflow records are short, the errors in single station quantile estimates are correspondingly large. There are problems with identifying the distribution which best fits the data and with estimating the parameters for the distribution (Bobée and Rasmussen, 1995). In these cases, the quality of quantile estimates can be improved by the application of regional equations (ACH 1989).

To estimate flood frequency for ungauged basins a regional approach must be used (Caissie & El-Jabi, 1991a). Obviously, since no gauge data exists for the study stream, inferences based on the behaviour of adjacent gauged streams are necessary to make predictions about the behaviour of the stream under study. This is true of both statistical and deterministic models. For ungauged basins, any model which uses data from neighbouring basins is making an assumption of similar response between the study basin and its neighbours. The use of popular models like the rational method or SCS method assumes some level of homogeneity between the study basin and the basins used to calibrate those models.

Regional analysis is generally recognized as a powerful means to improve flood quantile estimates (Bobée and Rasmussen, 1995). There has been some resistence to the broad application of regional analysis. However, in Newfoundland, the local regulatory agency has encouraged local practitioners to adopt the RFFA of Beersing (1990). While this has met with widespread acceptance, the reality is that many practitioners apply this method without concern for the statistical nature of the approach or for the parameter boundaries discussed in the research. The method is often applied in a deterministic manner.

Of primary interest in this thesis, is the usefulness of regionalisation for the island of Newfoundland. If quantile estimates produced from four regional equations are not significantly superior to estimates based on a single region, then there is no benefit in regionalisation.

4.2 Region Delineation

The delineation of regions is a complex procedure. The usual approach is to group basins into areas with similar geographic, hydrologic, and climatic characteristics. Most research has relied on physical properties of basins to determine regional boundaries (Richter, 1994). Typical physical characteristics include location, elevation, topography, ground cover, and exposure to prevailing winds. However, the use of geographically contiguous regions has been criticized as being arbitrary (Bobée and Rasmussen, 1995). In any study of historical flow records, the statistical properties of these records must be given substantial weight when grouping the stations into hydrologically similar regions. A methodology for delineating similar regions should be based on both physical properties of the basins, and statistical analysis of basin response (Ashkar, 1994).

In practice, most regions are defined geographically, using a combination of physical characteristics and gauge record information. Regional boundaries may be defined loosely using physical parameters, then gauge statistics may be tested to determine if a basin should be a member of a region, or of some adjacent region. The purpose of these tests is to detect stations having flow records which are not homogeneous with the general pattern for a region. When nonconforming stations are detected, the boundaries may be adjusted so that the stations are reassigned to a region with similar hydrologic response. Once regions have been delineated and stations tested for homogeneity, regional quantile estimations can be developed.

Within any homogeneous region, gauged stations should produce data which is consistent with other stations within the region. A variety of hydrologic parameters are commonly used to test for homogeneity including mean annual runoff per unit area, mean peak flow per unit area, and coefficient of variation or coefficient of standard error.

A popular statistic for testing regional homogeneity is the ratio of the ten-year flood quantile to the mean annual flood (Beersing, 1990). A variation of this is the ratio of the ten-year flood quantile to the two-year quantile. First the quantile ratio is calculated for each gauge in the region, then the summary statistics of mean and standard deviation of Q(10)/Q(2) are calculated. Assuming that the data are normally distributed, the stations are tested against the supposition that all stations within a homogeneous region should produce results within some confidence interval set by the researcher; 95% is commonly used.

Other popular test statistics include the coefficient of standard deviation, CS, and coefficient of variation, CV. The coefficient of standard deviation for any flood quantile may be calculated as the ratio of its standard deviation to that quantile's mean value. The use of this ratio allows comparison of standard deviation across basins of differing size across a region. The coefficient of variation is similarly calculated. However, testing for homogeneity with the coefficient of variation has been found to be a weak test which accepts homogeneity too often (Richter, 1994).

Some researchers have looked at methods of grouping basins in a data space which is not geographical (Richter, 1994). In some cases, basins in the same geographic area may exhibit very different streamflow behaviour. The set of all gauges in a study area may be grouped into regions according to a test parameter applied to gauge data. Some parameters used to derive station clusters include mean flow per unit area, quantile variation, skew and kurtosis (Richter, 1994).

The Region Of Influence, ROI, approach dispenses completely with geographic groupings (Bobée and Rasmussen, 1995). Each study site is treated as the centre of gravity of a multidimensional space in which vectors correspond to a variety of statistical or descriptive characteristics. These descriptive characteristics are weighted with respect to their influence on the central site (Bobée and Rasmussen, 1995). Distance in the multidimensional space is measured in terms of difference between characteristics of the central site and the regional sites, rather than physical distance.

Another alternative is cluster analysis. In this approach, characteristics are selected which are thought to relate the response at the study site to the behaviour at the gauge sites. Starting from the study site and working in the same type of multidimensional space used in the ROI approach, gauge sites which are most similar are clustered to the study site until the difference in characteristics reaches a cutoff point.

In addition, some researchers have sought to group sites into regions based on the nature of the statistical description which best fits their flood frequency data (Bobée and Rasmussen, 1995). A number of distribution characteristics including coefficient of variation and skew have been used as the basis for regional delineation. The use of various L-moments has gained some popularity among proponents of this method of regional grouping (Bobée and Rasmussen, 1995).

It must be understood that regional groupings based on statistical properties of basin response do not necessarily translate into geographical groupings. One additional problem with this approach is that statistical data is required to assign any stream under study to a non-geographic region and this data is unavailable for ungauged streams.

4.3 Hydrologic Regionalisation in Newfoundland

Caissie and El-Jabi (1991a), analysed records from fifteen (15) hydrometric stations, and treated Newfoundland as one homogeneous region. However, Newfoundland has varied landforms and climate influences. There may be some benefit to dividing the island into regions.

The Atlantic Development Board (1969) divided Newfoundland into four hydrologic regions: Avalon & Burin Peninsula, South & East Coast, West Coast and Great Northern Peninsula, and Northeast Coast.

In the DOE (1984) study *Regional Flood Frequency Analysis for the Island of Newfoundland*, the island was divided into two regions: North and South. This division was based on the causative factors behind peak flow events (DOE, 1984).

Beersing (1990) divided the island into four regions. This division was based on mean annual flow per unit area and time of occurrence of peak flows. The regions delineated also make sense from an examination of the topography and geography of the island. The Eastern Region comprises the Avalon and Burin Peninsulas. This area has generally low relief, and is subject to mixed weather produced by the confluence of the Gulf Stream and Labrador Current. The Central Region includes the central landmass of the province, and includes both coastal and non-coastal areas. This region's interior is less subject to oceanic effects and experiences greater extremes of cold and heat than coastal areas. The Northwest Region is defined by the Humber Valley and Northern Peninsula. This area is characterized by the large watershed of the Humber River, and Long Range Mountains and a coastal plain along the Northern Peninsula. The Southwest Region includes the southwest tip of the island. This area also has strong relief and may be subject to strong orographic influences. In general, the Southwest is the first area affected by incoming storms as they move from the waters of the Gulf of St. Lawrence onto the land.

Beersing (1990) used thirty nine (39) gauge records and divided the island into four homogeneous regions. Peak flow series were extracted using the AMF approach and flood quantiles were estimated using either the Three Parameter Lognormal Distribution or the Generalized Extreme Value Distribution. Regional quantile estimates were generated by regression on log-transformed data.

Richter (1994) investigated a variety of methods for delineating homogeneous hydrologic regions. This included analysis in non-geographic data space. Richter (1994) found that mathematically rigorous methods for region delineation did not significantly improve model outcomes when compared to the regions of Beersing (1990).

In this thesis, the regions delineated by Beersing (1990) were adopted as the initial regional divisions and were then tested for hydrologic homogeneity. The use of one region for the entire island was also evaluated. There are a number of methods available to determine the grouping of hydrologic stations into regions. A brief description of some of these methods has been provided in section 4.2 of this thesis. These approaches have been described somewhat exhaustively by Richter (1994). These methods were not applied in this work and as such, any further discussion of the methods would be beyond the scope of this work. Details of the results of regional homogeneity testing are presented in Chapter 6 of this thesis. Briefly, the stations within each region are tested to see if they meet the criteria that the Q10/Q2 ratio for a station is within the 95% confidence interval for Q10/Q2 ratios described for the region's population.

This test may be compared to the popular test of Dalrymple (1960). Fill and Stedinger (1995) provided a critical appraisal of the Dalrymple test of regional homogeneity. Following Dalrymple (1960), they describe a hydrologically homogeneous region as one where flood flows when scaled by their mean Q(T)/μ are identically distributed. This implies that for any homogeneous region, all ratio Q(T)/μ should fall within some confidence interval which can be defined for that region. The test suggested by Dalrymple (1960) assumes a Gumbel probability distribution and thus the mean flood flow is qual to Q(2.33). Essentially, for any station a return period value T is calculated based on the fit of Q(10) to the distribution curve plotted for the region. This calculated T-value is compared to the Lower and Upper limits of the 95% confidence interval for T for record of length N.

In this thesis, the Q10/Q2 ratio is analogous to the Q(T)/μ ratio discussed by Fill and Sedinger (1995) and which forms the basis of the Dalrymple (1960) test. The testing of a station for acceptance within the confidence interval for this ratio is a valid test statistic which should produce results similar to the Dalrymple (1960) test and analogous approaches.

There are a number of papers of specific relevance in developing models for Newfoundland. The work of Caissie and El Jabi (1991a, 1991b) provided useful information on the development of truncation levels and regions for the island. However, they used a very small data set of only fifteen stations for the island and treated it as one region. In addition, the formula which they developed for truncation level did not perform well for the data analysed here. The work of Beersing (1990) was important in the selection of hydrologically homogeneous regions and in the analysis of AMF series. One criticism of the work of Beersing (1990), is that he extended a number of flow record artificially and thus may have reduced variability in some of his data sets. However, the hydrologic regions developed by Beersing have provided results as good as more rigorously defined regions (Richter, 1994). The work of Richter (1994) provides much valuable information on the hydrology of Newfoundland including regionalization and regional modelling of flows. Richter (1994) points out that the deficiencies in hydrologic input (rainfall) data for the island seriously impact on the development of accurate flow models. Indeed, regionalization is described as one method of overcoming this problem by grouping stations into regions with similar hydrologic input characteristics.

5.0 REGIONAL MODELLING

5.1 Parameters of Regional Models

In regional flood frequency analysis, the equations which relate flood magnitude to probability of occurrence are represented as functions of physical descriptors. Some of the possible basin descriptors are categorized and listed in categories in Table 5.1.

Any number of these parameters, $X_1 \ldots X_n$, may be included in a regional model. A properly constructed model will incorporate only those parameters which add significant information to the outcomes. To be useful for regional peak flow models these physical parameters must have some basic properties:

1. They can be readily extracted from the information available for the basin
2. They must contain relevant information about the streamflow of the basin
3. It must be possible to express them as a numerical value

The objective is to create equations which will allow the user to compute flood quantiles for both gauged and ungauged streams within homogeneous regions. Many researchers have noted that finding the proper basin characteristics to include in a regional model is more important than fitting the best model to those characteristics (Richter, 1994).

Some characteristics, such as basin geology, cannot be expressed satisfactorily as a numerical index (Richter, 1994). While an understanding of these characteristics and their interactions may give a researcher a much better understanding of the processes occurring within a drainage basin, they are of limited value when developing modelling equations.

Some parameters are commonly used in most models. Drainage area is included in almost all models, firstly because it seems logical to include it, and secondly because it is usually strongly correlated to streamflow magnitude.

Richter (1994) states that Riggs (1973) listed three physical descriptors (drainage area, the basin slope, percent lakes and swamps) and one climate descriptor (mean annual precipitation) as explaining most variability in basin response.

Caissie and El-Jabi (1991a) used drainage area, areas of lakes and swamps, area of forest, and drainage density as explanatory variables for estimating flood quantiles for Newfoundland streams.

Richter (1994) found that drainage area, area controlled by lakes and swamps, fraction of barren area, and distance of the basin from defined lines were the most important explanatory variables for estimating mean annual flow for ungauged Newfoundland streams.

In the DOE (1984) study five parameters were selected: drainage area, mean annual runoff, percent area controlled by lakes and swamps, shape factor, and latitude. One general model for the island and two regional models for the north and south regions were developed using combinations of these variables. The parameters of these models are listed in Table 5.2. The Mean Annual Runoff, MAR, occurs in all of the equations and is the most important variable after drainage area. However, Lye and Moore (1991) identified MAR as a problematic variable, because it had a very high influence on model output, it was difficult to estimate accurately, and it was derived using a parameter, DA, already included in the model. Beersing (1990) also felt that the use of MAR in these equations was problematic because equation results were very sensitive to MAR and the descriptor was difficult to obtain accurately for ungauged streams.

Richter (1994) discussed the use of Effective Precipitation, EffP, expressed as an average annual runoff depth over a basin, which is equivalent to MAR. This derived variable may be used as a proxy for precipitation input. There is an understandable desire to include precipitation input as an explanatory variable in a study of flow series. EffP and its analog, MAR, have been identified as very significant predictors of peak flow magnitudes. Where there is no base precipitation data or data is very limited, a proxy variable may be introduced to represent this data. However, precipitation is a result of atmospheric processes, not basin processes. Where inferences are made about precipitation series from flow series data, special care must be taken to allow for the damping and amplifying effects which basin processes may generate.

Beersing (1991) selected different parameters for each of the four regions which he used. The parameters selected are listed in Table 5.3. In Newfoundland the influence of lakes and swamps can be quite significant in determining the flow regime of a stream. Both the area of lakes and swamps and the area controlled by these lakes and swamps are important. To provide a descriptor which represents both the area of lakes and swamps and their influence area, Beersing (1990) used a Lakes and Swamps Factor, LSF:

$$LSF = 1 + FACLS - \frac{FLSAR}{1 + FACLS} \tag{31}$$

Where FLSAR is the fraction of the drainage basin occupied by lakes and swamps, and FACLS is the area controlled by lakes and swamps.

Some techniques are available to select model variables prior to regression analysis. By selecting explanatory variables properly the amount of analysis can be reduced and problems such as cross-correlation can be avoided.

A simple analysis is done by generating multiple plots of basin variables against other basin variables. This technique was employed by Richter (1994), who provided an extensive set of plots. As expected, the magnitudes of the mean flood and average daily flow were strongly correlated to Drainage Area (DA). Slope (SLP) was positively correlated to flood magnitude. Richter (1994) indicated that Shape (SHP) also appeared to be significant.

Care must be used in interpreting these types of plots. The influence of some factors, notably drainage area, is dominant and may mask the influence of other factors (Richter, 1994). In general, the relationships of various descriptors tend to confirm the relationships put forward in other research. Flow magnitude is strongly correlated to drainage area, while the basin slope, the fraction of the area controlled by lakes and swamps and other basin characteristics have varying amounts of influence on flood magnitude.

5.2 Developing Models by Regression

Regional flood frequency models are commonly constructed by the techniques of linear and nonlinear regression. Software is readily available to perform both linear and non-linear regression. In general terms, all regression approaches construct a relationship between explanatory variables and outcomes and seek to minimize error. Error is defined as the difference between model outcomes and expected values.

By simple linear regression and multiple linear regression, flood quantiles for gauged basins may be related to physical descriptors of those basins. These regional quantile estimators must produce results which are consistent with the results of single station estimates within the region. Linear regression models represent results as having a straight-line relationship with their explanatory variables. The goal is to find an equation for a line that minimizes the sum of squared errors.

Equations from multiple linear regression on untransformed data take the form given in Equation 32. This form is not very popular for the study of hydrologic phenomena. Although it may produce usable results, this model form does not relate the physical parameters to each other in any meaningful way.

$$Q(T) = a_0 + a_1 x_1 + a_2 x_2 \dots \tag{32}$$

To make explanatory variable and outcomes more amenable to linear regression, a variety of transforms are used. With the data in the transformed space, models are constructed using linear regression, then transformed back into the real domain. One popular approach is the power transform, where all data is transformed by taking the logarithm. Inside the transformed log-space, the regression equations for estimating regional flood quantiles take the form given in Equation 33:

$$\ln(Q(T)) = \ln(a_0) + a_1 \ln(x_1) + a_2 \ln(x_2) + \dots \tag{33}$$

When the reverse transform is performed, the equation parameters are reorganized into a nonlinear form:

$$Q(T)=a_0 x_1^{a_1} x_2^{a_2} x_3^{a_3} \ldots \tag{34}$$

Where Q(T) is the expected flow for some return period T, a_i is a coefficient derived from multiple regression in log-space, and x_i is some physical parameter of the drainage basin. The derived values, $a_1 \ldots a_n$, are only valid for the return period for which they were calculated.

The nonlinear relationship of Equation 34 is derived using linear regression. In this approach a nonlinear relationship is transformed such that it can be handled by linear means. Once the linear regression is performed, the equation may be transformed back to its original nonlinear form. The transformation of data in this manner distorts the model error. Errors and bias which are generated in the transformed space must also be un-transformed for analysis of how well the equations fit the data.

Nonlinear regression resolves the problem of transformation generated bias. This method, like linear regression, attempts to minimize the sum of the squared error, where error is measured as the distance of the data from the model curve. Because the equations being manipulated in the regression are not linear, more computing effort is required than for linear regression techniques.

Nonlinear regression requires that you initially define the expected relationship between the result and explanatory variables. Because of this, nonlinear regression requires a deeper initial understanding of the interaction of results and explanatory variables. Generally, the approximated model equation is of the form given in Equation 34. The model is then fitted to the data using the estimated parameters, and by repeated adjustment of model parameters error is minimized. The output values finally arrived at may depend to some extent on the parameter values set initially. To compensate for this it is important that the initial values make sense on a physical basis. Variables which are initially assigned strong positive relationships must have a strong physical explanation for this relationship. This understanding of how the variable relate to the outcome is important, because relationships developed using this method will not produce an equation which can be plotted and confirmed by visual examination.

5.3 Regional Estimators

Given that flood quantiles have been modelled by analysis of gauged basins, and that adequate physiographic information is available for these gauged basins, there are two approaches which may be taken in the development of regional quantile estimators:

1. **Regression on Quantiles**: For each region and each return period, T, develop equations which correlate recurrence probability and flood quantile magnitude based on hydrologic and physiographic data.

2. **Index Flood**: For each region, develop an index flood equation based on hydrologic and physiographic data. Develop a rating curve which correlates flood quantile magnitude to the index flood.

The disadvantage of the first approach, regression on quantiles, is that a large number of equations must be developed. Each equation can only be applied for its specific return period and its specific region. If a practitioner needs quantile estimates for return periods other than those given, he must interpolate. The advantage of this approach is that variation in basin response for different size events is well modelled.

In the second approach, the index flood equation for a region is developed based on the relationship of an index flood to basin physiographic characteristics. Flood quantiles are described by their relationship to this index flood (Caissie & El-Jabi, 1991a). For an index flood to perform well for a region, the ratio of flood quantiles, Q(T), to the index flood must be consistent throughout the region.

In regional models based on the series of annual maxima, the index flood is often taken as the mean annual flood. Richter (1994) refers to this value as Qavgfld, and indicates that it is frequently used as an index flood in regional flood frequency analysis.

Determination of the index flood is complicated by the use of POT series. The mean annual flood is not equal to the average value of all the POT peaks. In addition, the average recurrence rate, λ, may not be constant from station to station. For an individual station, where the model is PED the average annual flood can be estimated from the model parameters by using Equation 35 (NERC, 1975). This approach produces results very similar to those produced using the series of annual maxima.

$$\mu = q_o + \beta \ln\lambda + 0.5772\beta \tag{35}$$

An alternative to using the mean annual flood or similar average flow, is to use a low return period flood quantile as the index flood. The use of the estimate of the two-year return flood, Q(2), is an example of this approach. For the PED series, the estimate of Q(2) should be as good as the estimate produced by Equation 35, since this equation is of the same form as the estimator for Q(2).

The disadvantage of the index flood approach is that errors in estimating the index flood equation will be carried through into quantile estimates. Richter (1994) indicates that errors in estimates of the index flood are a large source of error in estimates of flood quantiles. Variations in basin response to events of differing sizes may be poorly modelled. The main advantage is that calculations are very much simplified. Caissie and El-Jabi (1991a) felt that the regression on quantiles approach was superior to the index flood method in most regions. For Newfoundland, however, the results of index flood and regression on quantiles were similar (Caissie and El-Jabi, 1991a)

The index flood method is a powerful technique. For any basin, only one estimate of the index flood is required. Quantile estimates may then be obtained by simple mathematical or graphical relationship to the index flood. In addition, where errors in extraction of relevant physiographic parameters affect the reliability of the index flood, the same errors will similarly affect individual quantile estimators. The index flood approach is the method of regional quantile estimation which is investigated in this thesis.

Table 5.1 Parameters for Regional Models.

Climate:	Mean Annual Precipitation (MAP)
	Effective Precipitation (EffP)
	Annual Dry Days/Wet Days
Streamflow	Mean Annual Runoff (MAR)
	Mean Annual Flow (MAF)
	Mean Annual Flood (MAFL)
Basin Physiography	Drainage Area (DA)
	Land Slope (SLP)
	Perimeter (P)
	Shape Coefficient (SHP)
	Mean Elevation of Basin (MELE)
	Basin Length-Width Ratio
	Latitude of basin Centroid (LAT)
	Longitude of basin Centroid (LONG)
	Channel Length (L)
	Channel Slope (S)
	Channel Shape
	Stream Order
	Drainage Density (DRD)
	Area of lakes and swamps (ALS)
	Influence area of lakes and swamps (ACLS)

Table 5.1 Parameters for Regional Models (continued).

Surface Conditions	Ground Cover Type
	Area of Forest (AF)
	Area of Pasture (AP)
	Area of Barren (AB)
Soil Type	Rock
	Soil Classification
	Soil Permeability
	Soil Depth
Moisture Conditions	Moisture condition of Ground Cover
	Moisture condition of Soil

Table 5.2 Explanatory Variables from DOE 1984.

Region	Explanatory Variables
Entire Island	drainage area, mean annual runoff, percent area controlled by lakes and swamps, and shape factor
North	drainage area, mean annual runoff, latitude
South	drainage area, mean annual runoff, percent area controlled by lakes and swamps, and shape factor

Table 5.3 Explanatory Variables from Beersing 1990.

Region	Explanatory Variables
Avalon	drainage area, lakes and swamps factor, drainage density
Central	drainage area, drainage density
Northwest	drainage area, lakes and swamps factor, drainage density, slope of main channel
Southwest	drainage area, lakes and swamps factor

6.0 RESULTS AND DISCUSSION

In this chapter, the selection of streamflow series, the testing of single station models, the testing of regional homogeneity, and the development and testing of single sation and regional models is discussed.

6.1 Selection of Data Series for Analysis

The streamflow series used in this thesis include data from federal and provincial gauging stations, available as HYDAT CD-ROM Version 1.05.8, compiled by Environment Canada.

Four criteria were applied when selecting data sets from the one hundred eleven records available for active and discontinued hydrometric stations for the island portion of Newfoundland:

1. Each station must have at least 10 years of data
2. Any structural control of flows upstream must be insignificant
3. Records must be reasonably complete (no missing years)
4. Urbanized streams are excluded

Applying the above criteria to the one hundred eleven records available, sixty-three data series were found to be suitable for analysis. Seventeen omitted stations were regulated, and five omitted streams had diversions. A further sixteen stations were omitted because of short records, and six stations were omitted because they were in urban areas. Three stations were also omitted because they provided information which could be obtained from longer records at other locations in their watershed. One station was omitted because of missing data.

Data series for the single station analysis were tested for trend and independence using the standard measures of these properties as contained in CFA 3.1 (Pilon and Harvey, 1994). A number of series were found to have some problems.

Trend was detected in the AMF series for station 02ZF001 at 5% significance. More detailed graphical analysis of this data showed trend to be weakly defined. Regression of values on position explained only a small portion variability. In addition, the POT series data did not exhibit any significant trend. This series was retained for analysis in its entirety.

Trend was detected in the AMF series for station 02YK002. This was attributed to a diversion which was installed on this stream. Only 23 years of data following the diversion were retained.

Trend was detected in the AMF series for station 02ZH001. This basin was subject to a fire in the 1960s and this is the probable cause of this apparent trend. Regression of values on position explained only a small portion of variability (r-square = 8.3%). The POT series showed no evidence of trend. This series was retained in its entirety.

A detailed analysis of trend, independence, randomness and outliers for Newfoundland streamflow records is presented in the work of Rollings (1999).

The sixty-three stations selected for analysis included the thirty-nine (39) stations used by Beersing (1990) in *Regional Flood Frequency Analysis for the Island of Newfoundland*, and the fifteen (15) stations used by Caissie and El-Jabi (1991a) in their analysis of Newfoundland streamflows. A complete listing of the hydrometric stations used in this analysis is included in Table 6.1.

6.2 POT Data Extraction and Computer Program

As part of this research, a computer program was developed to set a threshold and extract peaks. The program set an initial threshold, extracted all values above that threshold, applied peak independence criteria, discarded values which failed independence criteria, and calculated mean and variance of recurrence of extracted peaks. The mean and variance of recurrence ware compared and evaluated against the Poisson distribution criteria. If the recurrence statistics were not within acceptable tolerances (usually < 0.1 difference), the program reset the threshold and repeated the procedure until a satisfactory threshold was found.

Caissie and El-Jabi (1991b) produced an equation for estimation of q_o for Newfoundland streamflow records based on mean annual flood levels:

$$q_o = 0.587 \times MAFL - 2.514 \tag{36}$$

In initial tests of the extraction program, the estimate of Equation 28 was used to get a starting value for the threshold. However, on many occasions this estimator predicted a threshold which produced low recurrence rates, and the mean and variance of recurrence failed to converge. Because of this, the estimator used to obtain an initial threshold was modified to produce a lower initial estimate.

While the Poisson recurrence distribution criteria were used to set thresholds for peak extraction, there were some occasions where, the mean and variance of recurrence converged only at very high recurrence rates (eight to ten peaks per year). This recurrence level increases the calculation load significantly in later analysis, and may compromise the independence of peak-over-threshold events. For these reasons, where the Poisson criteria produced high recurrence rates, the threshold was set higher and peak-over-threshold series extracted with between three and five peaks per year.

All extracted series were tested to see that their recurrence pattern fit that expected for a Poisson arrival process. This was done using the Kolmogorov-Smirnov test. All extracted series passed the Kolmogorov-Smirnov test, and thus were determined to be reasonably well fitted by a Poisson Distribution.

6.3 Comparison of Results of Single Station Analysis

Flood quantiles were modelled for series of annual maxima using the Three Parameter Log-Normal Distribution (3LN), and the Generalized Extreme Value Distribution (GEV). For the Peak-over-threshold data series, flood quantiles were modelled using the Poisson-Exponential Distribution (PED) and the Poisson-Pareto Distribution (PPD).

The 3LN and GEV models have been used to model series of annual maxima for Newfoundland in the past. These methods were used by Beersing in the Regional Flood Frequency Analysis (Beersing, 1990). In general, he found that both approaches produced acceptable results for flood series in Newfoundland. However, the 3LN method is best suited to positively skewed data. Some of the annual maxima series for Newfoundland exhibit negative skew. Where a series of annual maxima exhibits negative skew, the 3LN method is not well suited to describing the distribution of the data and derivation of the distribution parameters is more difficult than with positively skewed data. Because of these difficulties in fitting the model, there were six annual maxima series for which the 3LN model was not fitted as part of this research. The GEV model was fitted to the sixty-three annual maxima series.

The Poisson-Exponential Distribution (PED) and Poisson-Pareto Distribution(PPD) models were fitted to the sixty-three peak-over-threshold data series for Newfoundland. The Poisson component of these distributions is derived during the extraction of the peaks over threshold data, and the Poisson parameter λ, is equal to the recurrence rate for the peaks. The Exponential Distribution is the simplest magnitude distribution to derive, as it only has one parameter β. However, this reduces the flexibility of this distribution. The Pareto Distribution is more complex, requiring the derivation of α and k, the shape and scale parameters. Although the additional parameters of the Pareto model increases model complexity and add some model error, the increased flexibility of the Pareto distribution should allow it to fit the data more closely.

The first comparison of the output of the four flood quantile models under consideration was a comparison of central position for the model outputs. The extracted AMF and POT data sets were modelled using 3LN and GEV for the AMF, and PED and PPD for the POT. Quantile estimates were generated for 2, 5,10, 25, 50,100, 500, and 1000 year return periods. This was done for all 63 sets of station data (57 for 3LN model). These results were then compared using ANOVA. Using this method, the central positions of quantile estimators for the different flood quantiles can be compared across distributions. Examination of the data presented in Table 6.2 shows that, for the four distributions considered, the means of model outputs were similar for all of the models considered. Examining the mean values for each quantile estimator, and considering the upper and lower limits of the 95% t-confidence interval for the mean, all of the models have outputs which, for each quantile level, are not statistically significantly different.

The box-plots in Figure 6.3 provide graphical confirmation of the above conclusion. For each group of quantile estimates, the position of the means and medians for the four models are both similar. For each group of quantile estimates the data sets are similarly positively skewed (mean greater than median). Some differences in the model results are apparent in Figure 6.3. For all of the quantile estimates, the 3LN distribution has a somewhat larger inter-quartile range (IQR) indicated by a larger box, and this effect becomes more pronounced at the higher quantiles. For quantile estimates of 25 years return or greater, the PED distribution exhibits a smaller IQR than the other distributions. For the two highest quantiles, the PPD data exhibits larger IQR than the PED data ad the PPD data has high outliers.

Based on the ANOVA analysis of the quantile estimates and the examination of the boxplots, it would appear that all of the models produce similar results, and that the PED has slightly less variability at higher quantiles.

The second comparison of the outputs of the four flood quantile models under consideration was a comparison of the robustness of the models, or sensitivity of the model to variations in the underlying data set. The better model not only fits the data closely but is resistant to variations in the underlying data. To test this quality a resampling approach was used.

For each set of AMF and POT data, the model parameters and quantile estimates were generated for the underlying data set. The underlying data sets were then resampled with replacement and new model parameters and quantile estimates calculated based on the resampled data. Thus a set of new model outcomes was produced from data sets which contained only the data available from the original but with variation from the original. For any quantile, calculation of the standard error (standard deviation) of the produced quantile estimates gives a measure of the sensitivity of the model to variations in the data. Comparison of the standard error of results from different models allows a comparison of the relative robustness of the models.

For the 3LN model, resamplng sometimes produced data sets for which the method used to derive the model parameters failed. For some of the original 63 data sets this failure on resampled data occurred in a large proportion (>25%) of the resampling events. Where this occurred, the standard error was not calculated for the resampled data. On this basis, in addition to the 6 series omitted because the underlying data set could not be fitted, an additional 20 series were omitted from analysis of standard error of 3LN quantile estimates.

The 3LN distribution is commonly used for single station analysis and has met with good success in the island of Newfoundland (Beersing, 1990). However, the distribution does not work well for data with negative skew. In this work a number of short series and series with skew close to zero were analysed. During resampling it is easy for skew to be shifted slightly thus causing a 3LN model intended for positively skewed data to fail. However, the comparison of central position and error for the distributions analysed should remain valid. The fact that no statistically significant difference was found in central position of the distributions tends to confirm this.

Similarly to the analysis of the central position, the standard error was analysed using ANOVA. Examination of the data presented in Table 6.3 indicates that for lower quantile estimates the standard error of the model outcomes is similar for all the models. Using the mean standard error and 95% t-confidence interval, to compare the model outcomes for the 2, 5, 10 and 25 year quantile groups, there is no statistically significant difference between the standard error of the model outcomes within each quantile group. For the 50 year quantile estimates, the standard error of PED outcomes is the lowest of the four, and the 3LN is the highest. In fact, while the mean of the PED standard error is within the 95% t-confidence interval of the 3LN, the mean standard error of the 3LN is higher than the upper limit of the confidence interval for the PED standard error. At the 100 year quantile level the standard error of the 3LN and PED are significantly statistically different, and while the GEV and PPD outcomes are higher than the upper limit of the 95% t-interval for the PED, the PED is barely within the lower limits of the confidence interval for the GEV and PPD outcomes.

At the 500 and 1000 year quantiles, the standard error of the PED outcomes is significantly statistically different than that of the 3LN, GEV, and PPD. Based on this analysis, it would appear that the four models exhibit similar standard error for low quantiles, with the PED model exhibiting better performance at higher quantiles.

Examination of the box plots of Figure 6.4 tends to confirm the results of the ANOVA analysis. For the lower quantiles, the standard error is similar for all models. At higher quantiles, starting at about Q(50), the size of the IQR, indicated by the height of the box, begins to be noticeably smaller for the PED outcomes. Indeed, for the higher quantiles, the position of the PED standard error median is lower, and the box is significantly smaller. In addition, there are fewer outliers for the PED data and the outliers are closer to the expected range. This tends to indicate that the PED model has comparable performance at lower quantiles, and better performance at higher quantiles.

Overall, for both AMF and POT series, the PED model had the lowest standard error in model outcomes for resampled data. The quantile estimates from PED models were consistent with those of the other methods over the range of return periods under consideration. This seems to indicate that the PED model produces a reasonably good fit to the data and is more resistant to changes in the data. Thus, among the models tested, the PED model is determined to provide the best estimates.

6.4 Results of Regional Homogeneity Testing

As discussed in Section 4.3 of this thesis, the division of Newfoundland into four hydrologically homogeneous regions as defined by Beersing (1990) was adopted for this research. This approach was examined by Richter (1994), who found that more complex methods of delineating regions did not improve the performance of regional models. As a check on the validity of these regions, homogeneity testing was done on the island as a single region and on the four regions delineated by Beersing (1990). The stations within the regions were tested for homogeneity using the ratio of the ten-year and two-year flood quantiles, Q(10)/Q(2). These quantiles were selected as reliable indicators because all stations had at least ten years of data. The ratio Q(10)/Q(2) was calculated for all stations in a region, and the mean and standard deviation of the ratio was computed. All stations were then tested to be within the 95% and 99% t-confidence interval about the mean. The stations were also tested using the nonparametric outlier criteria of the boxplot (LL = QL - 1.5IQR, UL = QU + 1.5IQR).

Testing the whole island as one region, two stations failed for the 95% t-confidence interval and one failed for the 99% t-confidence limits. Station 02YD001 failed for the 95% t-confidence interval but passed for 99%. This station also failed the non-parametric outlier criteria. Station 02ZM009 failed at both the 95% and 99% t-confidence levels and was well below the lower limit. Station 02ZM009 also failed the non-parametric outlier criteria.

For the Avalon Region one station, 02ZM009, failed for the 95% t-confidence interval but passed for the 99% interval. Station 02ZM009 also failed the non-parametric outlier criteria. This station is located at the southeastern corner of the Avalon Peninsula and is highly exposed to the oceanic weather effects which occur in this region.

For the Central Region all stations passed for the 95% and 99% t-confidence intervals and for the non-parametric outlier criteria. For the Northwest Region all stations passed for the 95% and 99% t-confidence intervals and for the non-parametric outlier criteria. Station 02YD001, which was marked as an outlier for the whole island region, was not an outlier in the northwest region. For the

Southwest Region all stations passed at both 95% and 99% t-confidence intervals and for the non-parametric outlier criteria.

6.5 Results of Regional Modelling

6.5.1 Model Generation by Linear and Nonlinear Regression

As discussed in Section 5.2 of this thesis, regional models typically follow the nonlinear form given in Equation 34, repeated here:

$$Q(T)=a_0 x_1^{a_1} x_2^{a_2} x_3^{a_3} \ldots \tag{34}$$

Traditionally, nonlinear models for flood quantiles, as shown above, have been derived by transforming the data into log-space, performing linear regression, and then transforming the equations back into normal space and applying them to the data. This method introduces bias into the equations as a result of the transformation. Development of regional models by direct nonlinear regression should produce superior results to the traditional log-linear method. The bias inherent to the logarithmic transformation is not generated, and the fitting of the model coefficients is performed in the real data space.

In this thesis, the traditional log-linear method of model development was used to generate regional models for the two-year return flood quantile. Direct nonlinear regression was also used to generate regional models for the two-year return flood quantile. The regional model outcomes from linear and nonlinear regression are compared to each other on the basis of their goodness of fit to the expected flood quantile values.

Variables considered in the development of regional equations were limited to physical descriptors of basin characteristics. This data was available from the Newfoundland Department of Environment and Labour. Variables related to basin position were eliminated because regionalisation effectively addresses position. Variables related to mean annual runoff and other analogs for precipitation were eliminated as well. Variables related to soils, infiltration rates, and soil permeability were eliminated because information on these basin properties was not readily available.

Explanatory variables were then included and excluded following an iterative process. The drainage area was selected as the first explanatory variable for all regions. Following this, slope, fractional area of lakes and swamps, lakes and swamps factor, drainage density, and shape were considered. Factors such as fractional area of barrens and forest were also considered but were not found to improve the performance of estimates. The order of variable testing and the combinations of variables tested was determined by the author in an organized sequence. The performance of variables was judged based on the r-square, mean error, and root mean square error of the regional estimate developed.

Model parameters which were considered as possible explanatory variables included the drainage area, the basin slope, the fraction of the basin controlled by lakes and swamps, the lakes and swamps factor, the drainage density, the and shape. Drainage area is typically the most significant component of regional models because the system inputs (rainfall, fog, or melting snow) are distributed over the basin at some depth and the input volume is the product of the drainage area and the input depth. Drainage area was found to be the most significant parameter for the regional models developed here.

Two parameters were considered for addressing the influence of lakes and swamps: Fractional Area Controlled by Lakes and Swamps (FACLS), and Lakes and Swamps Factor (LSF). The FACLS is calculated simply as the ratio of the area of the basin hydrologically controlled by lakes and swamps to the entire area of the basin. The calculation of the LSF, as explained in Section 5.1 of this thesis, is done using the FACLS and the fractional area of lakes and swamps (FLSAR), and is slightly more complicated. The influence of lakes and swamps in a basin is typically to mitigate the height of flood peaks, and the use of the FACLS is intended to allow the model to include this attenuating effect. The LSF was adopted by Beersing (1990) to include the effect of the open water surfaces of lakes and swamps which reduce infiltration in the drainage basin. In this thesis, only one of the FACLS or LSF was included in any regional model - the one which produced the best fit.

The basin slope (SLP) was considered potentially significant because steeper basins tend to concentrate water more rapidly, and thus will tend to respond to shorter duration and higher intensity precipitation inputs. Drainage density (DRD) is computed as the ratio of the length of all the streams in a watershed to the area of the watershed, and gives a measure of how well drained the basin is. The implication is that an increase in drainage density will produce an increase inflow. The shape parameter (SHP) is a measure of how elongated a basin is, with a more elongated basin having a higher shape factor. Shape is calculated using a simple formula (Beersing 1990):

$$SHAPE = 0.28 \times Perimeter \div \sqrt{DrainageArea} \tag{37}$$

A number of parameters which are popular for the development of regional models were not employed. No parameter for precipitation was included in the analysis. This information was excluded because the climate network for Newfoundland is sparse and availability of accurate precipitation is limited. The problems with the use of precipitation data or its analog, mean annual runoff, have been discussed at some length by Lye and Moore (1991), and Beersing (1990). The use of Latitude and Longitude or Northing and Easting parameters was not considered. Where a regional model is applied, the region is assumed to be hydrologically homogeneous so position within the region should not influence the model outputs.

6.5.2 Comparison of Linear and Nonlinear Models

A number of measures of model fit are available to compare model outcomes for the regional models. The three measures selected to compare the model outcomes are the adjusted R-square value, the mean error (ME), and root mean square error (RMSE). The adjusted R-square value indicates how much of the variability of the dependant data is explained by the model. Error was calculated as the difference of the predicted value less the expected value of Q(2), and the mean error (ME) was calculated as the simple average of the error. This approach gives an indication of the location of the central position of the model outputs compared to the expected value, and allows one to get an indication of the bias of the model. The root mean square error is calculated at the root of the average of the error squared. The RMSE is a measure of the average size of the deviation between the predicted and actual values of the dependant variable.

In general, the models generated by nonlinear regression produced higher R-square values and lower RMSE for the same model parameters. Mean error, ME, was consistently smaller and positively skewed for the nonlinear models. This indicates that the models derived using nonlinear regression had less bias, and their bias was to slightly overestimate the flood quantile. Considering the error properties of the models, the nonlinear derived models were generally better than those derived using the log-linear method.

Results for log-linear and nonlinear regression on the whole island, are presented in tables 6.4 and 6.5, respectively. For the whole island, the best fit was obtained using the drainage area, slope, lakes and swamps factor, and drainage density. Results for the Avalon region are presented in Tables 6.6 and 6.7 respectively, and show that the best fit for the Avalon region is obtained by nonlinear regression using DA, LSF, and DRD. It should be noted that the model gave good results when just DA and LSF were used, and the improvement in the fit by the addition of DRD was slight. For the Central Region, nonlinear regression using DA and FACLS produced the best model. The addition of slope to the equation produced a slightly higher R-square value and a slightly lower ME, but increased RMSE. For the Northwest Region, nonlinear regression on DA, SLP, and DRD produced the best result with the highest R-quare value, and ME and RMSE which were very close to the lowest for the model results. For this regional equation the addition of LSF did improve the RMSE slightly, but the R-squared and ME values were made worse. For the Southwest Region, nonlinear regression on DA, SLP, and SHP gave the best estimate, with a much higher R-square value, and ME and RMSE than any other combination of parameters tested.

In general, the nonlinear regression models outperformed the log-linear regression models. For the same parameters, the nonlinear models exhibited higher R-squared values, and lower RMSE. The mean error, ME, which was a measure of bias, was much better for the nonlinear models than for the log-linear models.

The parameters FACLS and LSF, contribute similar information to the model, and for most models the addition of either of these parameters produced similar results. Since the FACLS is simpler to derive, it is probably the best choice for representing the effect of lakes and swamps in the models.

6.6 Index Floods

The index flood method for estimating flood quantiles for regions is an approach with a long and successful history. This was the method of developing regional quantile estimators for the whole island and the four regions considered in this thesis.

The index flood selected was the 2-year quantile estimate, Q(2). Other popular choices for the index flood include the mean daily maximum flow and the mean annual maximum. The process of generating the index flood curves for each region is a simple one. First the flood quantiles for various return periods are calculated for each station. In this case the PED quantile estimates were generated for the 2, 5, 10, 25, 50, 100, 500, and 1000 year return periods. The Q(T)/Q(2) ratio for each station and each quantile was then calculated. Then for each region, the mean ratio of

Q(T)/Q(2) ratio was calculated for each quantile. This mean Q(T)/Q(2) ratio allows the estimation of Q(T) for an ungauged site once the index flood Q(2) is known. Estimates of Q(2) for ungauged stations may be generated using the formulas developed for each region in section 6.5 of this thesis. Once the index flood for any site is known, estimates of quantiles may be calculated by the following formula:

$$Q(T) = \frac{Q(T)_R}{Q(2)_R} \times Q(2)_s \tag{38}$$

Where $Q(T)_R/Q(2)_R$ is the known ratio of the flood quantile to the index flood for the region.

An analysis of the errors associated with quantile prediction using the index floods and ratios derived in this thesis is presented in Table 6.16. The mean error, ME, is typically quite small compared to the mean estimate and is also somewhat positively skewed, indicating that the estimates tend to be somewhat higher than the expected values. For most regions the RMSE is quite small at low quantiles and remains at less than 10% of the mean expected value even at the highest quantile estimates.

For the Northwest region, however, the performance of the estimators is not as good as for the other regions, and RMSE is >10% for quantile estimates above the 25 year return period. The Q(T)/Q(2) ratios for the whole island were applied to generate estimates for the 19 stations in the

Northwest region These estimators did produce somewhat lower RMSE for the samples, but they also behaved poorly at the higher quantiles and had RMSE > 10% of the mean expected value for quantiles of Q(50) and higher. These estimates were also somewhat negatively biased and tended to underestimate the expected flow.

The mean and median ratio values for estimation of quantiles from the index flood Q(2), are given in Table 6.15. In this thesis, the mean ratios were used to generate estimates for flood quantiles at each gauging site. Figure 6.6(a-e) allows graphic interpretation to determine flood quantile ratios for return periods other than those used to generate the curve.

Table 6.1 Hydrometric Series for the Entire Island.

No.	Station No.	Station Name	Record Years
1	02YA001	St. Genevieve River	28
2	02YA002	Bartlett's River	12
3	02YC001	Torrent River at Bristols Pool	39
4	02YD001	Beaver Brook	20
5	02YD002	Northeast Brook near Roddickton	18
6	02YE001	Greavett Brook	14
7	02YF001	Cat Arm River	15
8	02YG001	Main River at Paradise Pool	12
9	02YH001	Bottom Creek near Rocky Harb.	13
10	02YJ001	Harrys River	30
11	02YJ003	Pinchgut Brook	11
12	02YK002	Lewaseechjeech Brook at L.Grand Lake	23
13	02YK004	Hinds Brk. near Grand Lake	24
14	02YK005	Sheffield Brook near TCH	26
15	02YK007	Glide Brook	13
16	02YK008	Boot Brook	13
17	02YL001	Upper Humber R. near Reidville	70
18	02YL004	South Brook at Pasadena	15
19	02YL005	Rattler Brook near Mcivers	13
20	02YM003	South West Brook near Baie Verte	18
21	02YN002	Lloyds R. below King George IV Lake	17
22	02YO006	Peters River near Botwood	17
23	02YO007	Leech Brook	13
24	02YO008	Great Rattling Brk. Above tote Rv.	14
25	02YO010	Junction Brook near Badger	12
26	02YP001	Shoal Arm Brook	15
27	02YQ001	Gander R. at Big Chute	49
28	02YQ004	NW Gander River near Gander Lake	15
29	02YQ005	Salmon River near Glenwood	11
30	02YR001	Middle Brook Near Gambo	39
31	02YR002	Ragged Harbour River	20
32	02YR003	Indian Bay Brook near NW Arm	17
33	02YS001	Terra Nova Riv at Eight Mile Bridge	34

Table 6.1 Hydrometric Series for the Entire Island (continued).

34	02YS003	Southwest Brook at Terra Nova Park	31
35	02ZA001	Little Barachois Brook neat St.Georges	19
36	02ZA002	Highlands River at TCH	16
37	02ZA003	Little Codroy R. Near Doyles	15
38	02ZB001	Isle Aux Morts River	36
39	02ZC002	Grandy Brook	16
40	02ZE001	Salmon River at Long Pond	22
41	02ZF001	Bay du Nord River	48
42	02ZG001	Garnish River	40
43	02ZG002	Tides Brook	20
44	02ZG003	Salmonier River near Lamaline	18
45	02ZG004	Rattle Brook near Boat Harbour	17
46	02ZH001	Pipers Hole Riv. At Mothers Brk.	46
47	02ZH002	Come By Chance River	30
48	02ZJ001	Southern Bay River near Sthrn Bay	22
49	02ZJ002	Salmon Cove River near Champneys	15
50	02ZJ003	Shoal Harbour River	12
51	02ZK001	Rocky River near Colinette	50
52	02ZK002	Northeast River near Placentia	16
53	02ZK003	Little Barachois Riv. Near Placentia	15
54	02ZK004	Little Salmonier Riv. Near North Harbour	15
55	02ZK005	Trout Brook	11
56	02ZL003	Spout Cove Brook	18
57	02ZL004	Shearstown Brook at Shearstown	15
58	02ZL005	Big Brook at Lead Cove	13
59	02ZM006	Northeast Pond River at NE Pond	45
60	02ZM009	Seal Cove Brook near Cappahayden	19
61	02ZM016	South Riv near Holyrood	15
62	02ZN001	Northwest Brook at NW Pond	30
63	02ZN002	St Shotts Riv	13

Table 6.1a Hydrometric Series for the Avalon Region.

No.	Station No.	Station Name	Record Years
1	02ZG001	Garnish River	40
2	02ZG002	Tides Brook	20
3	02ZG003	Salmonier River near Lamaline	18
4	02ZG004	Rattle Brook near Boat Harbour	17
5	02ZH001	Pipers Hole Riv. At Mothers Brk.	46
6	02ZH002	Come By Chance River	30
7	02ZK001	Rocky River near Colinette	50
8	02ZK002	Northeast River near Placentia	16
9	02ZK003	Little Barachois Riv. Near Placentia	15
10	02ZK004	Little Salmonier Riv. Near North Harbour	15
11	02ZK005	Trout Brook	11
12	02ZL003	Spout Cove Brook	18
13	02ZL004	Shearstown Brook at Shearstown	15
14	02ZL005	Big Brook at Lead Cove	13
15	02ZM006	Northeast Pond River at NE Pond	45
16	02ZM009	Seal Cove Brook near Cappahayden	19
17	02ZM016	South Riv. near Holyrood	15
18	02ZN001	Northwest Brook at NW Pond	30
19	02ZN002	St. Shotts Riv.	13

Table 6.1b Hydrometric Series for the Central Region.

No.	Station No.	Station Name	Record Years
1	02YN002	Lloyds R. below King George IV Lake	17
2	02YO006	Peters River near Botwood	17
3	02YO007	Leech Brook	13
4	02YO008	Great Rattling Brk. Above tote Rv.	14
5	02YO010	Junction Brook near Badger	12
6	02YP001	Shoal Arm Brook	15
7	02YQ001	Gander R. at Big Chute	49
8	02YQ004	NW Gander River near Gander Lake	15
9	02YQ005	Salmon River near Glenwood	11
10	02YR001	Middle Brook Near Gambo	39
11	02YR002	Ragged Harbour River	20
12	02YR003	Indian Bay Brook near NW Arm	17
13	02YS001	Terra Nova Riv at Eight Mile Bridge	34
14	02YS003	Southwest Brook at Terra Nova Park	31
15	02ZE001	Salmon River at Long Pond	22
16	02ZF001	Bay du Nord River	48
17	02ZJ001	Southern Bay River near Sthrn Bay	22
18	02ZJ002	Salmon Cove River near Champneys	15
19	02ZJ003	Shoal Harbour River	12

Table 6.1c Hydrometric Series for the Northwest Region.

No.	Station No.	Station Name	Record Years
1	02YA001	St. Genevieve River	28
2	02YA002	Bartlett's River	12
3	02YC001	Torrent River at Bristols Pool	39
4	02YD001	Beaver Brook	20
5	02YD002	Northeast Brook near Roddickton	18
6	02YE001	Greavett Brook	14
7	02YF001	Cat Arm River	15
8	02YG001	Main River at Paradise Pool	12
9	02YH001	Bottom Creek near Rocky Harb.	13
10	02YJ003	Pinchgut Brook	11
11	02YK002	Lewaseechjeech Brook at L.Grand Lake	23
12	02YK004	Hinds Brk. near Grand Lake	24
13	02YK005	Sheffield Brook near TCH	26
14	02YK007	Glide Brook	13
15	02YK008	Boot Brook	13
16	02YL001	Upper Humber R. near Reidville	70
17	02YL004	South Brook at Pasadena	15
18	02YL005	Rattler Brook near Mcivers	13
19	02YM003	South West Brook near Baie Verte	18

Table 6.1d Hydrometric Series for the Southwestern Region.

No.	Station No.	Station Name	Record Years
1	02YJ001	Harrys River	30
2	02ZA001	Little Barachois Brook neat St.Georges	19
3	02ZA002	Highlands River at TCH	16
4	02ZA003	Little Codroy R. Near Doyles	15
5	02ZB001	Isle Aux Morts River	36
6	02ZC002	Grandy Brook	16

Table 6.2 Mean and Upper and lower 95% t-confidence limit for quantile values derived using four distributions.

Quantile	Distribution							
	LN3		GEV		PExp		PPar	
	Mean		Mean		Mean		Mean	
	LL	UL	LL	UL	LL	UL	LL	UL
2	96.9		90.8		92.9		92.6	
	59.8	134.0	58.4	123.1	61.8	124.0	61.7	123.5
5	121.0		119.7		118.3		118.3	
	75.6	166.4	77.9	161.5	79.0	157.6	78.4	158.2
10	140.9		138.4		137.5		138.3	
	89.0	192.8	90.8	186.0	92.0	183.0	90.8	185.7
25	167.8		162.1		162.8		165.9	
	107.5	228.1	107.4	216.7	109.2	216.5	107.2	224.5
50	186.0		179.9		182.0		188.0	
	120.2	251.9	120.2	239.6	122.2	241.9	119.6	256.3
100	206.5		198.2		201.2		211.4	
	134.5	278.5	133.6	262.8	135.1	267.3	132.0	290.8
500	256.7		243.9		245.8		272.6	
	169.8	343.5	167.7	320.1	165.3	326.3	160.6	384.7
1000	280.0		265.9		265.0		302.8	
	186.2	373.8	184.4	347.4	178.2	351.7	172.8	432.8

Table 6.3 Mean and Upper and Lower 95% confidence limit for standard error of quantile values derived using four distributions

Quantile	Distribution							
	LN3		GEV		PExp		PPar	
	Mean		Mean		Mean		Mean	
	LL	UL	LL	UL	LL	UL	LL	UL
2	5.44		7.66		5.96		5.927	
	3.583	7.291	4.44	10.87	3.988	7.932	4.048	7.807
5	7.10		9.33		8.84		8.46	
	4.78	9.42	6.06	12.61	6.00	11.69	5.80	11.11
10	9.45		10.79		11.03		10.95	
	6.46	12.45	7.56	14.02	7.53	14.53	7.55	14.35
25	15.24		14.58		13.91		15.47	
	10.58	19.91	10.38	18.79	9.54	18.28	10.72	20.22
50	21.27		19.57		16.08		20.12	
	14.93	27.60	12.67	25.46	11.05	21.11	13.97	26.26
100	30.13		26.41		18.28		26.06	
	21.34	38.92	18.08	34.73	12.59	23.97	18.06	34.05
500	63.13		50.01		23.35		45.68	
	44.31	81.96	32.96	67.07	16.12	30.57	30.70	60.65
1000	85.4		64.3		25.53		57.14	
	58.3	112.5	41.8	86.8	17.64	33.42	37.46	76.83

Table 6.4 Whole Island Results of Log-Linear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	ao	a1	a2	a3						
Q(2)	0.8396	0.8						86.1	-5.55	55.50
Q(2)	0.4762	0.947	0.393					89.0	-5.172	42.90
Q(2)	0.447	0.940	0.348	-0.320				89.9	-5.191	35.65
Q(2)	0.913	0.932	0.349		-1.10			90.2	-4.56	32.159
Q(2)	0.6643	0.992	0.344		-0.952	0.428		91.8	-4.56	32.667
Q(2)	0.8025	0.885	0.349		-1.19		0.755	90.7	-6.75	36.67
Q(2)	0.575	0.944	0.343		-1.04	0.438	0.797	92.4	-6.62	37.36

Table 6.5 Whole Island Fits of Nonlinear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP	%		
	ao	a1	a2	a3	a4	a5	a6			
Q(2)	2.122	0.670						80.36	0.834	50.20
Q(2)	0.932	0.855	0.368					86.43	0.673	41.16
Q(2)	0.637	0.889	0.377	-0.601				92.76	0.242	29.72
Q(2)	1.645	0.883	0.4119		-1.321			93.06	2.37	29.33
Q(2)	1.597	0.896	0.355		-1.386	0.237		93.61	0.526	27.66
Q(2)	1.571	0.887	0.421		-1.408		0.078	92.91	-0.043	29.15

Table 6.6 Avalon Region Fits of Log-Linear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	ao	al	a2	a3						
Q(2)	0.706	0.883						93.7	1.69	16.61
Q(2)	0.745	0.867	-0.039					93.3	1.54	15.85
Q(2)	0.608	0.907				0.600		95.1	0.0035	9.52
Q(2)	0.773	0.897			-0.326	0.581		94.9	0.175	9.26
Q(2)	0.604	0.905				0.063	0.033	94.7	0.091	9.40
Q(2)	0.601	0.906		-0.085		0.588		94.8	0.066	9.25

Table 6.7 Avalon Region Fits of Nonlinear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	a0	al	a2	a3	a4	a5	a6			
Q(2)	1.522	0.728						94.8	0.541	9.796
Q(2)	1.685	0.679	-0.21					95.5	-0.172	8.895
Q(2)	0.992	0.812				0.382		95.7	0.184	8.64
Q(2)	1.75	0.723					-0.178	94.6	0.815	9.71
Q(2)	1.295	0.748		-0.265				95.1	0.315	9.25
Q(2)	2.978	0.694			-0.966			95.8	0.710	8.56
Q(2)	1.889	0.773			-0.869	0.361		96.5	0.313	7.61

Table 6.8 Central Region Fits of Log-Linear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	ao	al	a2	a3						
Q(2)	0.486	0.843						92.0	-7.77	59.81
Q(2)	0.259	1.03	0.402					92.8	–1.05	59.13
Q(2)	0.461	0.831		-0.714				93.2	-4.58	28.10
Q(2)	1.019	0.828			-1.19			92.8	-6.02	31.53
Q(2)	0.385	0.917				0.518		93.8	-8.80	59.59
Q(2)	0.206	1.10	0.402			0.518		94.7	-5.08	59.43
Q(2)	0.3012	0.960	0.280	-0.573				93.9	-3.90	29.79

Table 6.9 Central Region Fits of Nonlinear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	a0	al	a2	a3	a4	a5	a6			
Q(2)	0.950	0.762						86.13	3.25	58.37
Q(2)	0.798	0.853	0.287					87.15	4.70	54.32
Q(2)	0.465	0.829		-0.809				96.99	-1.28	26.38
Q(2)	1.089	0.844			-1.495			96.79	-2.23	27.16
Q(2)	0.712	0.827				0.294		85.84	3.47	57.31
Q(2)	1.313	0.815					-1.131	87.58	2.04	53.62

Table 6.10 Northwest Region Fits of Log-Linear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	ao	a1	a2	a3						
Q(2)	0.5862	0.857						85.5	-13.35	49.14
Q(2)	0.339	0.984	0.481					90.9	-8.54	37.30
Q(2)	0.2698	1.01	0.462	-0.256				91.9	-8.78	34.34
Q(2)	0.5075	1.027	0.47		-1.15			92.5	-5.04	31.38
Q(2)	0.313	1.01	0.479			0.133		90.6	-4.90	30.77
Q(2)	0.475	1.04	0.465		-1.11	0.089		92.1	-3.90	30.23
Q(2)	0.3396	0.983	0.451				0.019	90.3	-8.20	37.14

Table 6.11 Northwest Region Fits of Nonlinear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	a0	a1	a2	a3	a4	a5	a6	(%)		
Q(2)	0.469	0.918						89.6	1.46	39.16
Q(2)	0.234	1.057	0.498					93.37	-0.517	30.37
Q(2)	0.272	1.026	0.47	-0.266				93.28	1.166	29.58
Q(2)	0.408	1.024	0.476		-0.612			93.16	0.413	29.89
Q(2)	0.326	1.015	0.443			0.225		93.98	0.496	27.98
Q(2)	0.304	1.019	0.445		0.088	0.234		93.55	0.626	27.96
Q(2)	0.104	1.133	0.545				0.597	93.43	-1.763	29.26

Table 6.12 Southwest Region Fits of Log-Linear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	ao	a1	a2	a3						
Q(2)	3.03	0.685						48.1	-5.65	56.98
Q(2)	0.0000203	2.89	2.62					78.3	-6.64	34.82
Q(2)	0.000111	2.71	2.26		1.55			80.9	1.91	18.46
Q(2)	0.000001122	3.66	3.36				1.96	98.2	-7.51	12.64

Table 6.13 Southwest Region Fits of Nonlinear Regression.

Quantile	Parameters							R2	ME	RMSE
	Coef	DA	SLP	FACLS	LSF	DRD	SHP			
	ao	a1	a2	a3						
Q(2)	15.83	0.40						26.9	0.45	49.75
Q(2)	0.0000353	2.80	2.51					54.8	0.32	34.04
Q(2)	0.000309	2.33	1.82	-0.648				80.4	-0.48	18.52
Q(2)	0.00199	2.22	1.69		-2.12			86.2	0.30	15.61
Q(2)	0.000000192	4.03	3.72				-2.23	96.1	-0.18	7.57

Table 6.14 Regional Equations for the 2-Year Return Period Flood Quantile.

Region	Parameter						
	Coef	DA	SLP	FACLS	LSF	DRD	SHP
Island	1.645	0.883	0.4119		-1.321	0.237	
Eastern	1.889	0.773			-0.869	0.361	
Central	0.461	0.831		-0.714			
Northwest	0.3259	1.015	0.4431			0.225	
Southwest	0.000000192	4.026	3.722				-2.228

Table 6.15 Index Flood Ratios.

Region	Means of Ratios Q(T)/Q(2)						
	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Island	1.279	1.490	1.769	1.981	2.192	2.682	2.893
Avalon	1.264	1.463	1.726	1.926	2.125	2.588	2.787
Central	1.271	1.476	1.747	1.951	2.156	2.632	2.837
Northwest	1.307	1.539	1.845	2.078	2.310	2.849	3.081
Southwest	1.267	1.468	1.736	1.937	2.139	2.608	2.810
Region	Medians of Ratios Q(T)/Q(2)						
	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Island	1.281	1.495	1.776	1.989	2.203	2.697	2.910
Avalon	1.265	1.466	1.732	1.933	2.134	2.602	2.803
Central	1.269	1.472	1.740	1.944	2.147	2.618	2.821
Northwest	1.298	1.523	1.821	2.047	2.273	2.797	3.022
Southwest	1.268	1.471	1.739	1.942	2.145	2.616	2.819

Table 6.16 Errors in Quantile Estimates Generated using the Index Flood Ratios.

Whole Island							
Quantile	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Mean Error	0.579	1.017	1.595	2.033	2.471	3.487	3.925
RMSE	4.282	7.523	11.803	15.042	18.282	25.802	29.042
Mean Value	118.299	137.486	162.848	182.034	201.220	245.769	264.955
% Mean Error	0.489	0.739	0.980	1.117	1.228	1.419	1.481
% RMSE	3.620	5.472	7.248	8.263	9.085	10.499	10.961
Avalon Region							
Quantile	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Mean Error	0.183	0.322	0.505	0.644	0.782	1.104	1.243
RMSE	0.838	1.471	2.309	2.943	3.576	5.048	5.681
Mean Value	50.296	58.122	68.467	76.292	84.118	102.289	110.114
Mean % Error	0.364	0.554	0.738	0.844	0.930	1.080	1.129
% RMSE	1.666	2.532	3.373	3.857	4.252	4.935	5.160
Central Region							
Quantile	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Mean Error	0.086	0.152	0.238	0.303	0.368	0.520	0.585
RMSE	3.708	6.512	10.220	13.024	15.830	22.341	25.146
Mean Value	162.673	188.849	223.451	249.627	275.803	336.580	362.756
Mean % Error	0.053	0.080	0.107	0.122	0.133	0.154	0.161
% RMSE	2.279	3.448	4.574	5.217	5.740	6.638	6.932

Table 6.16 Error in Quantile Estimates Generated using the Index Flood Ratios (continued).

Northwest Region							
Quantile	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Mean Error	1.007	1.767	2.777	3.538	4.299	6.069	6.832
RMSE	7.713	13.550	21.262	27.097	32.932	46.481	52.318
Mean Value	123.519	144.395	171.986	192.859	213.733	262.200	283.073
Mean % Error	0.815	1.224	1.614	1.835	2.012	2.315	2.414
% RMSE	6.245	9.384	12.363	14.050	15.408	17.727	18.482
Southwest Region							
Quantile	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)
Mean Error	0.953	1.674	2.626	3.347	4.068	5.741	6.462
RMSE	2.364	4.153	6.516	8.305	10.093	14.245	16.034
Mean Value	176.598	204.281	240.878	268.562	296.246	360.526	388.210
Mean % Error	0.540	0.819	1.090	1.246	1.373	1.592	1.664
% RMSE	1.339	2.033	2.705	3.092	3.407	3.951	4.130

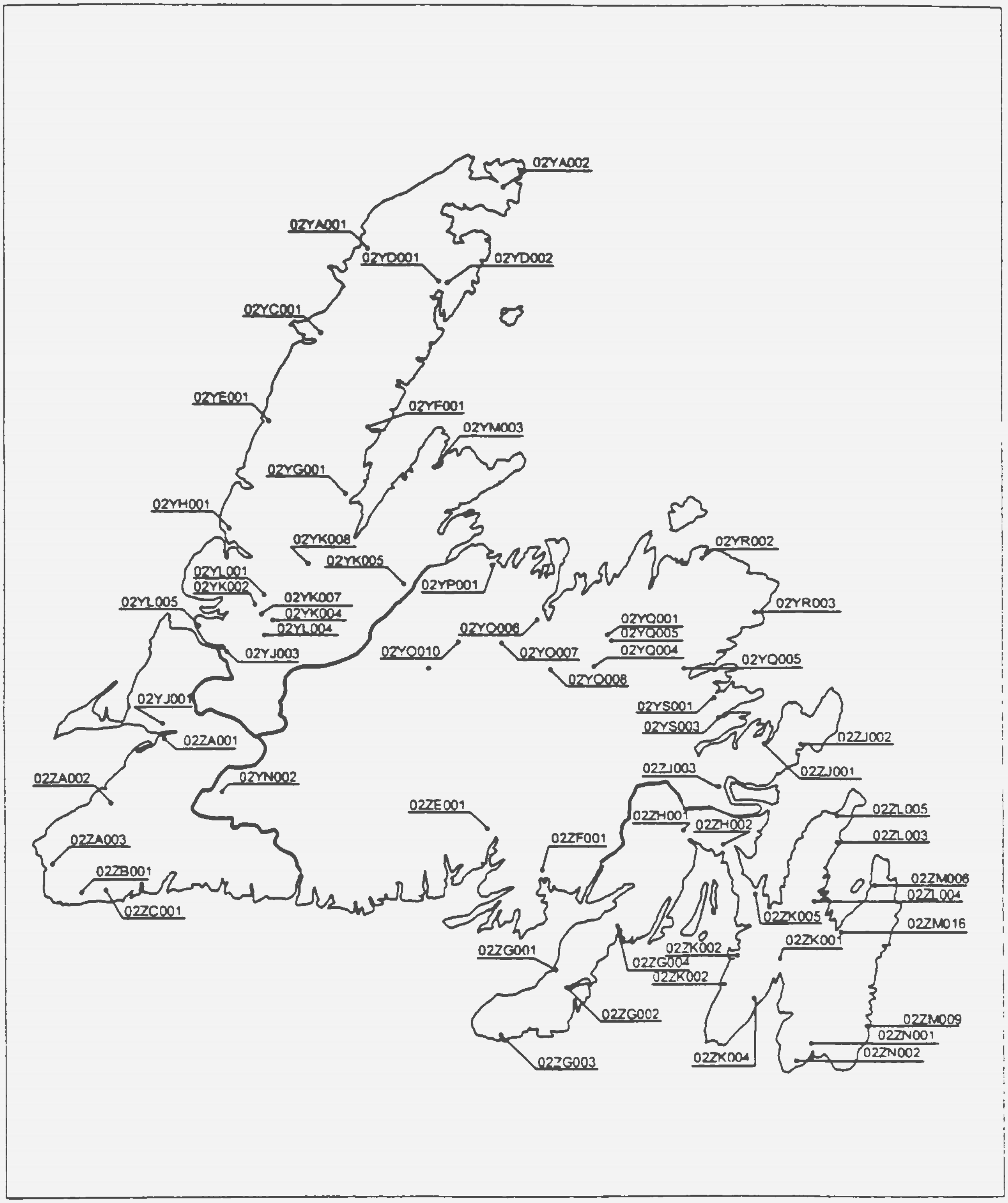

Figure 6.1 Map of Newfoundland Showing Stations

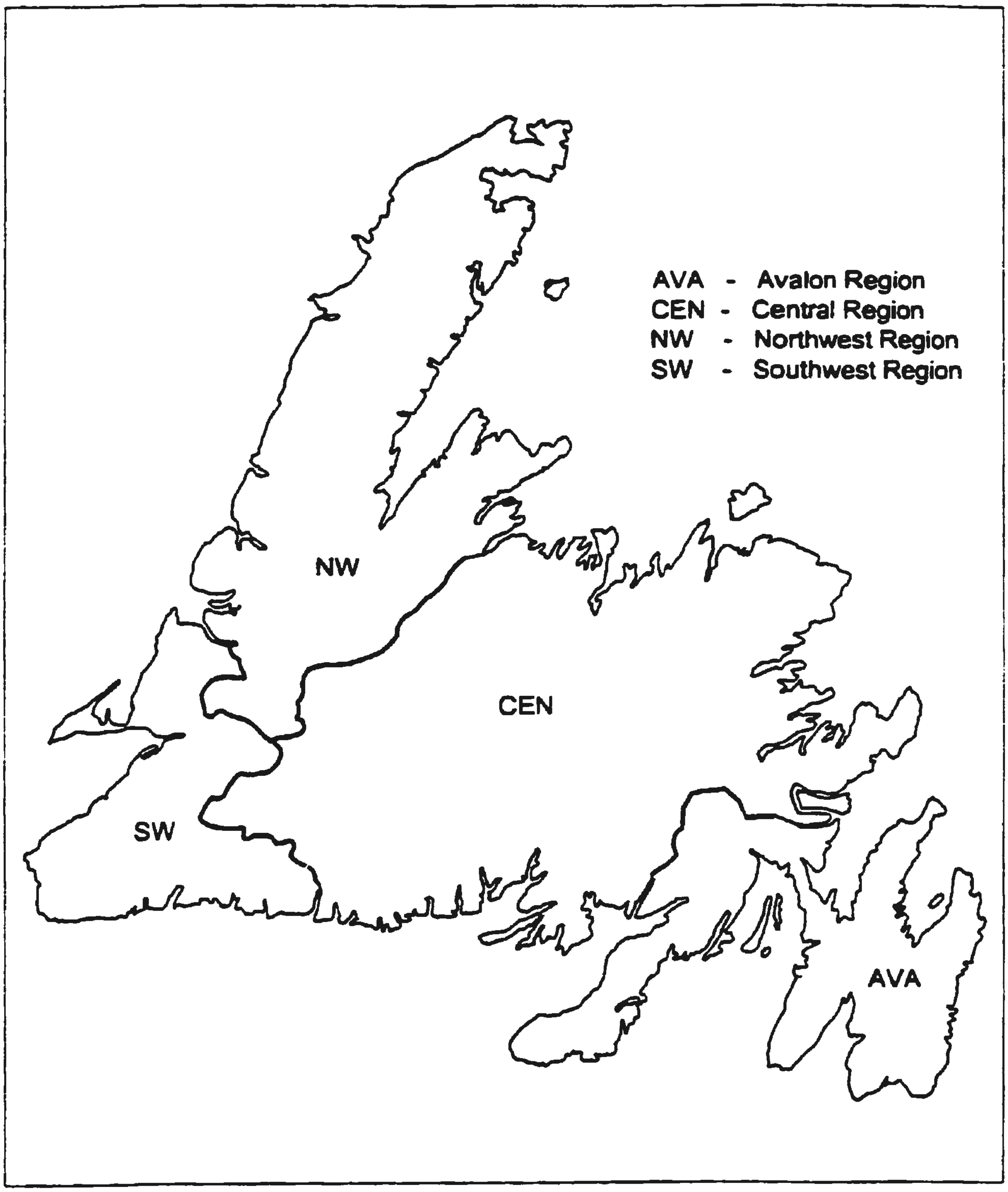

Figure 6.2 Map of Newfoundland Showing Regions

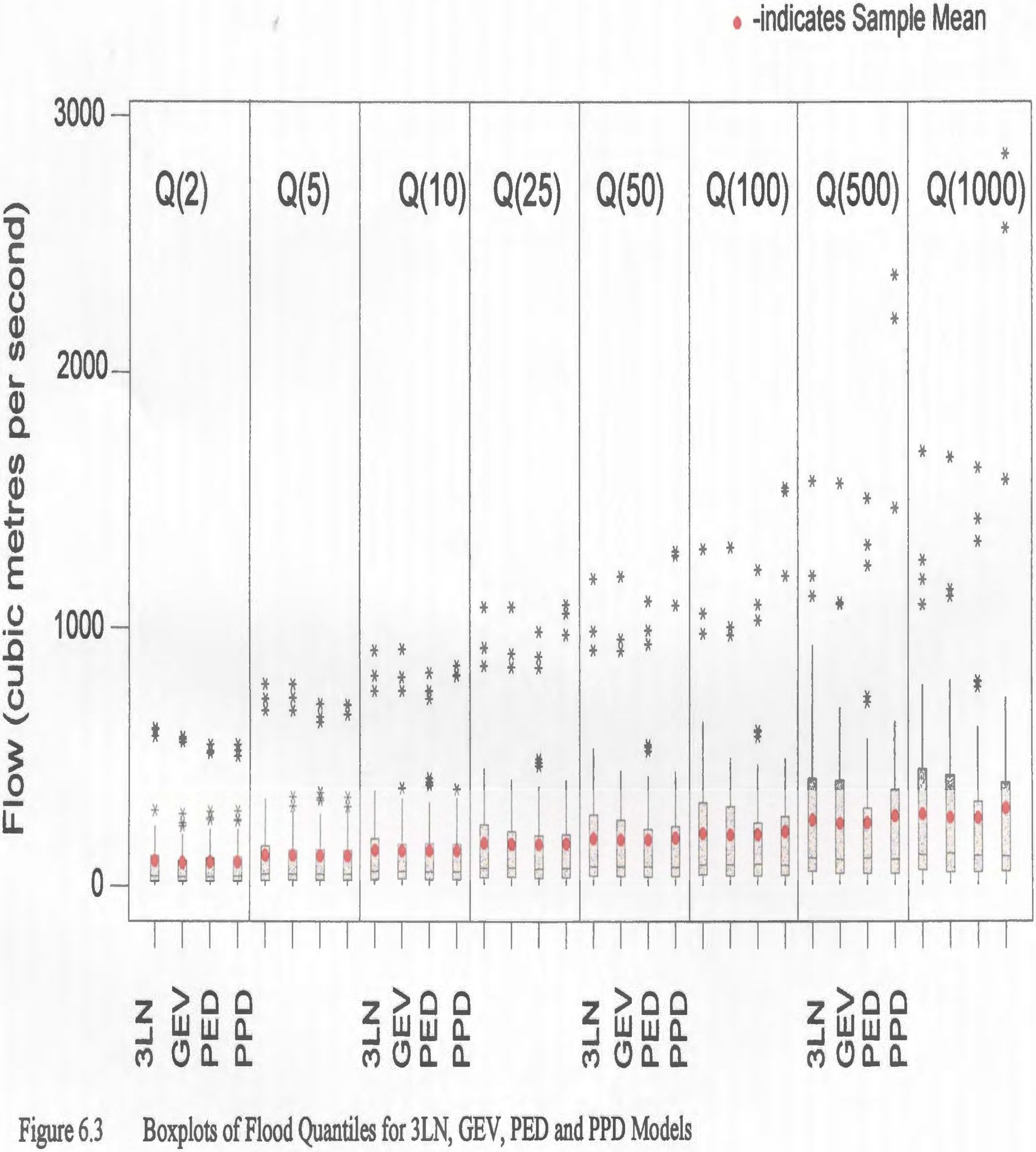

Figure 6.3 Boxplots of Flood Quantiles for 3LN, GEV, PED and PPD Models

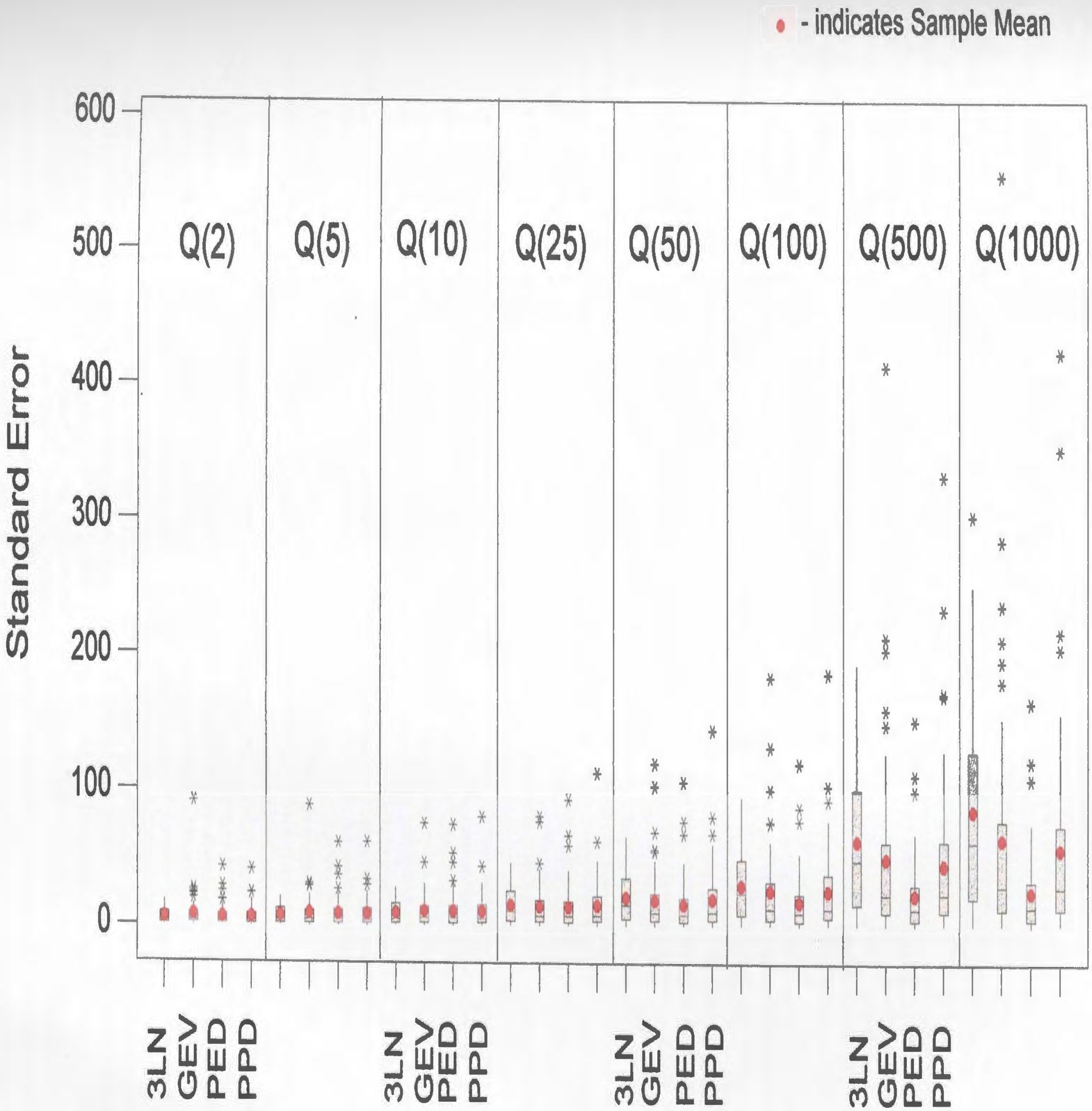

Figure 6.4 Boxplots of Standard Error of Flood Quantiles for 3LN, GEV, PED and PPD Models

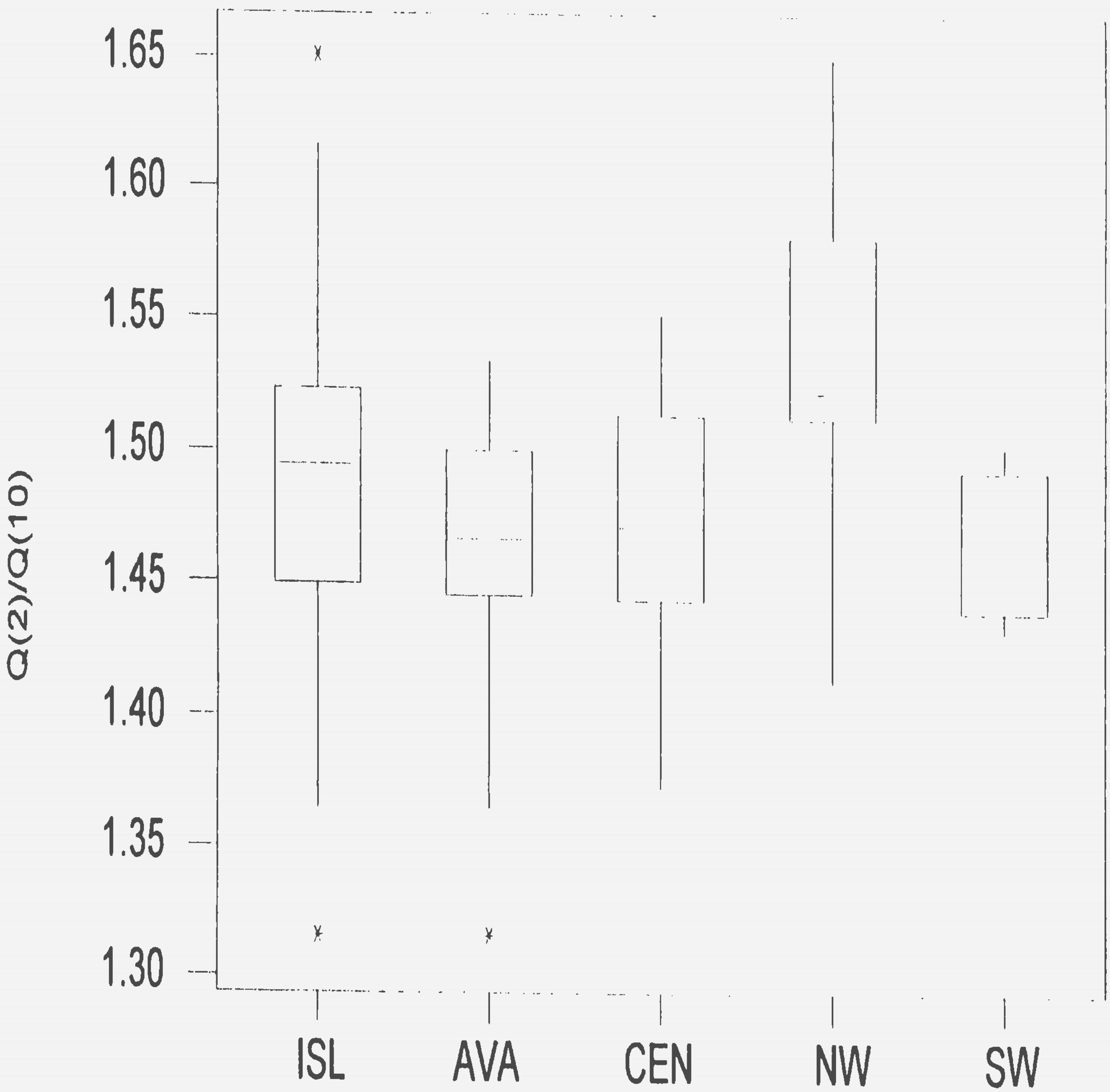

Figure 6.5 Boxplots of Q(2)/Q(10) for the Island, Avalon, Central, Northwest, and Southwest Regions

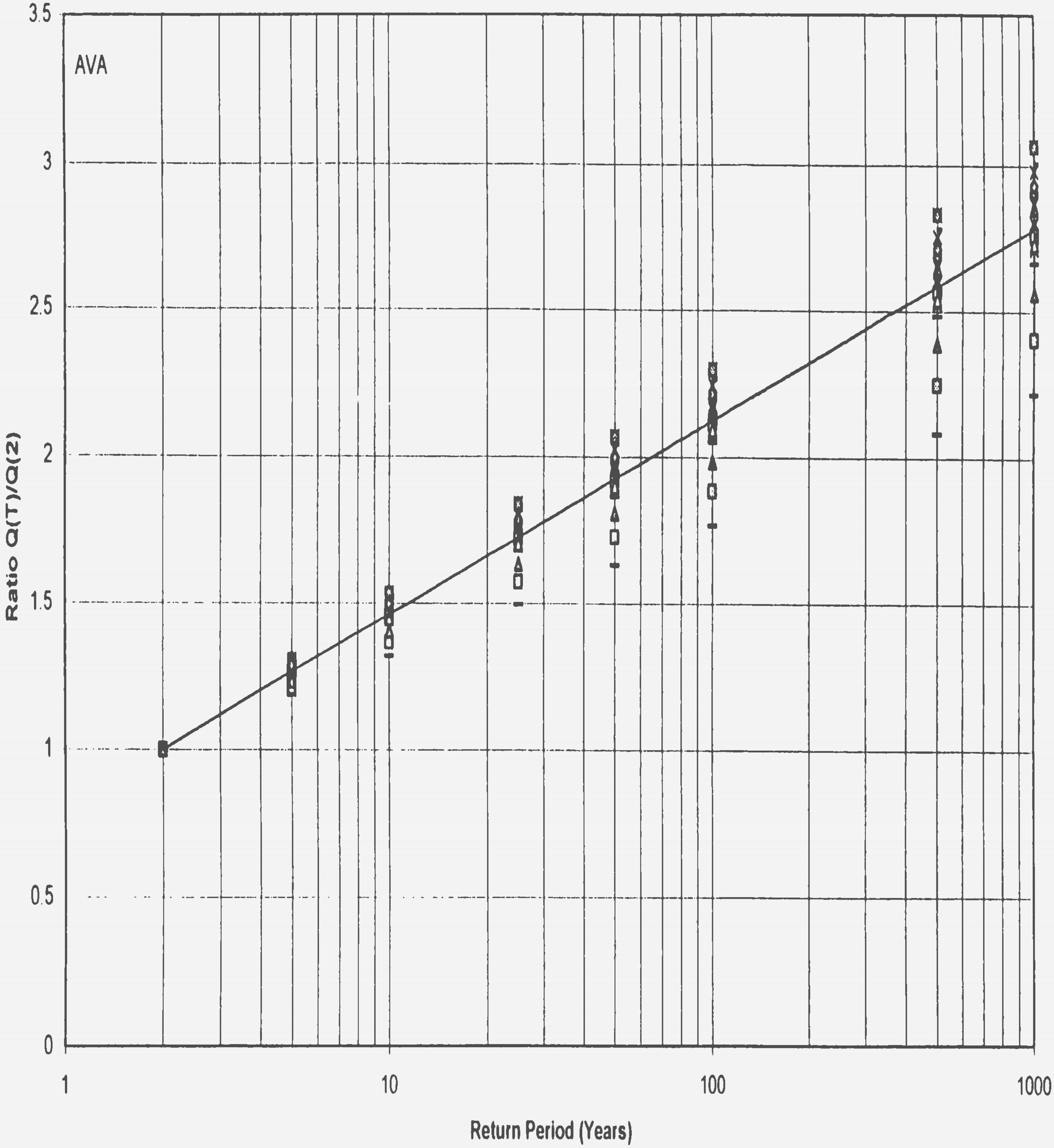

Figure 6.6b Quantile Estimation Chart For Avalon Region, Q(T)/Q(2) with Data Scatter Shown

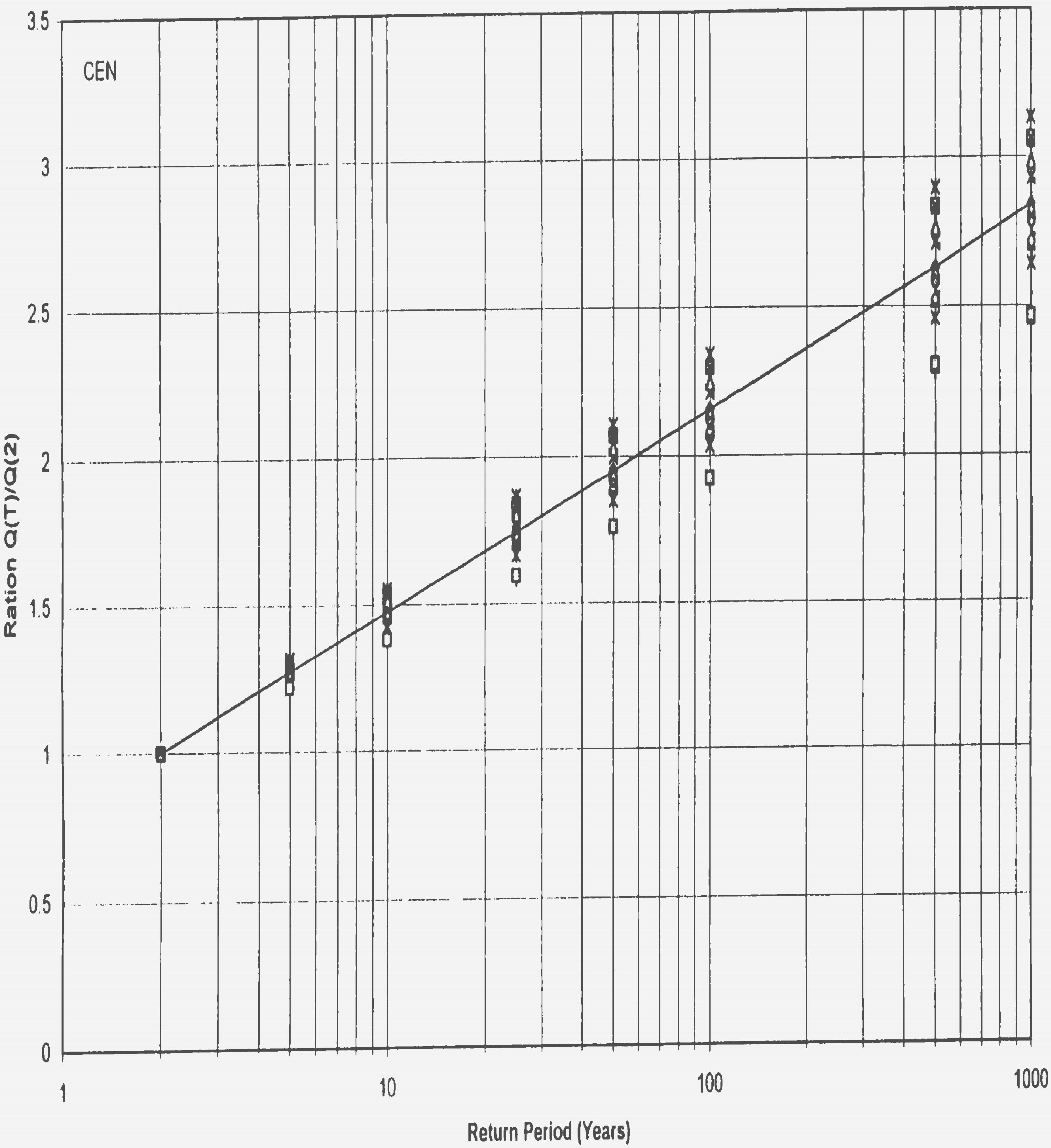

Figure 6.6c Quantile Estimation Chart For Central Region, Q(T)/Q(2) with Data Scatter Shown

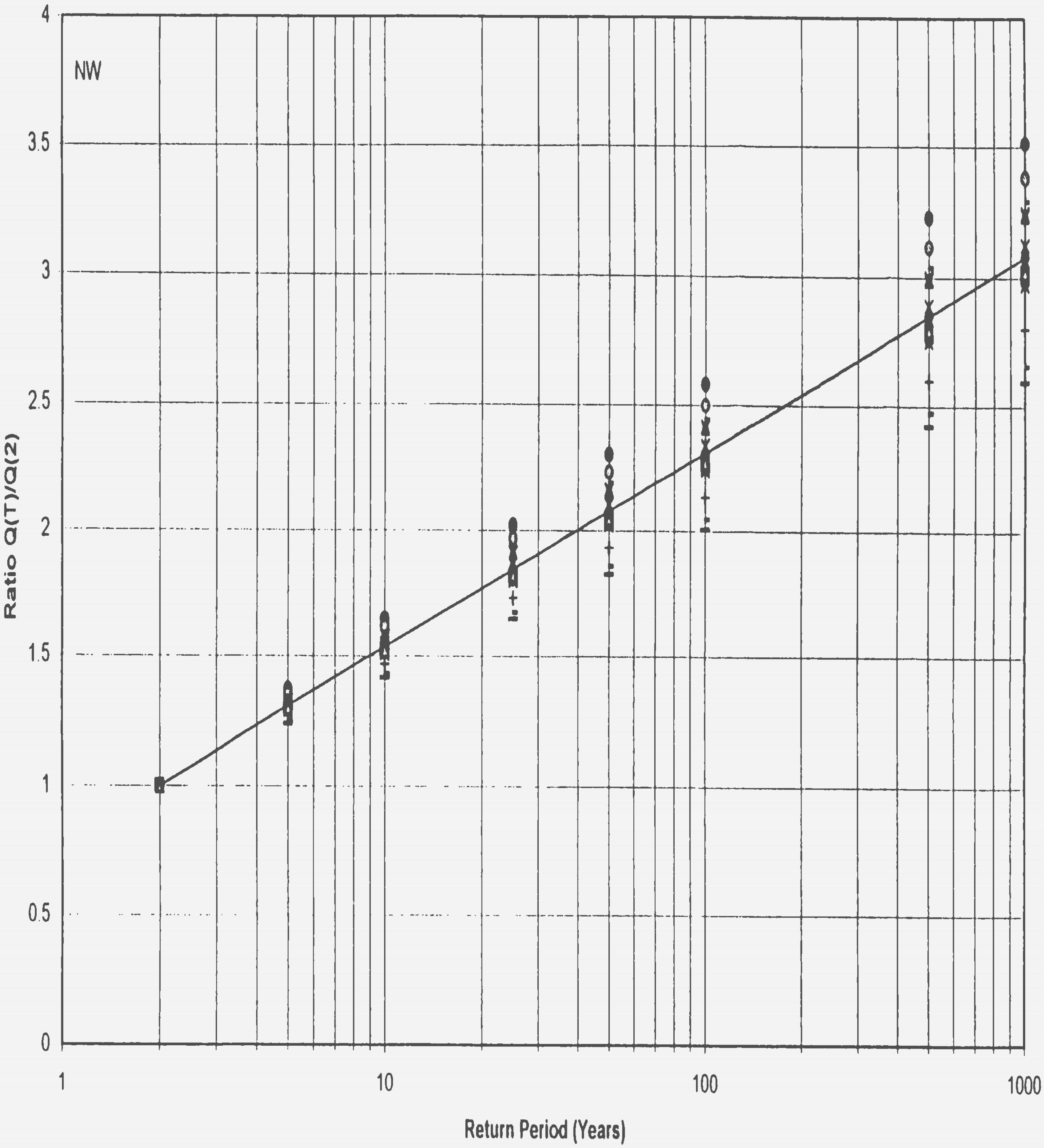

Figure 6.6d Quantile Estimation Chart For Northwest Region, Q(T)/Q(2) with Data Scatter Shown

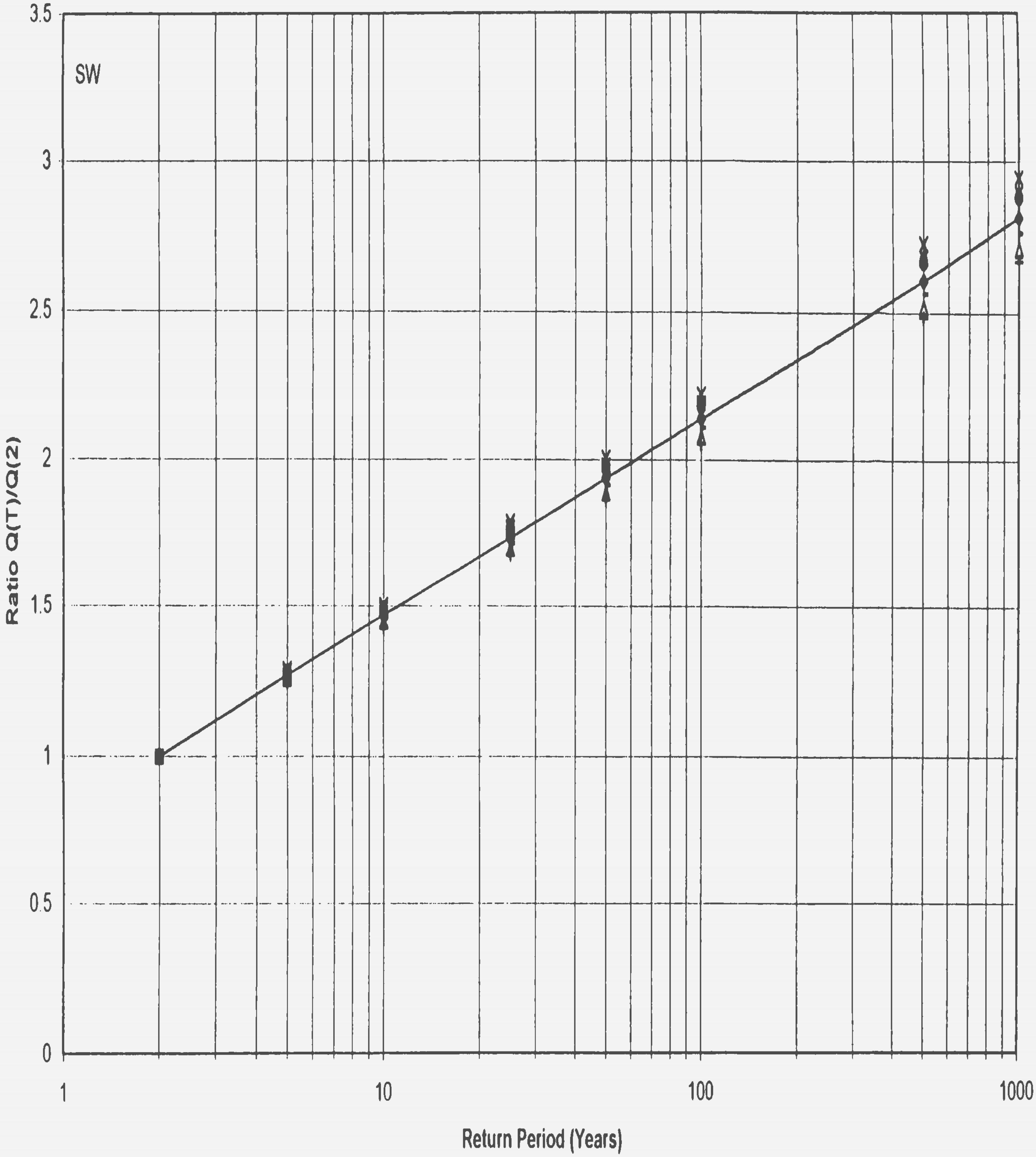

Figure 6.6e Quantile Estimation Chart For Southwest Region, Q(T)/Q(2) with Data Scatter Shown

7.0 CONCLUSIONS

In this chapter, some conclusions are presented based on the expected and obtained results from application of the peak-over-threshold method to the development of regional flood frequency models for Newfoundland.

1. For the quantile estimates generated for the 63 data series analysed, there is no statistically significant difference between the central position of the results of the 3LN, GEV, PED and PPD models.

2. For the standard error of quantile estimates generated by resampling of the 63 data series analysed, the Poisson-Exponential Distribution model exhibited comparable standard error for lower quantiles and lower standard error for higher quantiles. Because of this, the PED model was determined to be the most robust for a variety of quantiles.

3. Regional models for estimation of the 2-year quantile developed using nonlinear regression exhibited better fit to the underlying data than did the models produced using the traditional log-linear method. The nonlinear models exhibited lower bias as measured by mean error, ME, and also less estimation error as measured by root mean squared error, RMSE.

4. Using the 2-year quantile as the index flood, the ratios of Q(T)/Q(2) were easily calculated, and allowed estimation of flood quantiles for stations in the regions with a reasonably good fit to the expected values. For most regions RMSE was less than 10% of the mean of the expected values.

5. The estimated values from application of the index flood technique tended to overestimate the quantile slightly and results were somewhat positively skewed from expected values. This will tend to produce more conservative (higher) estimates of flood quantiles.

6. Quantile estimates using the index flood method produced the poorest results in the Northwest Region. Results were still reasonable and at lower quantiles, the RMSE was less than 10% of the mean expected value. The Q(T)/Q(2) estimators derived for the whole island were tried for this region but did not produce significantly better results.

7. With the exception of the Northwest Region, the use of regional index floods produced improved quantile estimates when compared to the estimates produced by equations developed for the whole island.

8. In the Southwest Region the equation which performed best (generated estimates with the lowest error) relied on three descriptors. The number of gauge records available in this

region was only six. The coefficients developed for this equation are also somewhat suspect as they suggest a significant scaling of the result. In this region, the use of the whole island equation may provide a more reliable result and is recommended.

9. The regional models developed in this thesis, based on a POT approach, the fitting of the Poisson-Exponential model to the at site data, and the development of regional models using nonlinear regression on basin descriptors provides regional models with relatively low error when compared to similar models developed for this region using AMF data. However, because this thesis includes more data sets, uses POT data, and uses non-linear regression methods, it is difficult to attribute the improved performance one source.

8.0 REFERENCES

Ashkar, Fahim, 1994, Regional Flood Frequency Estimation by the Peak-over-threshold Method, Proceedings of the 1st Workshop on Regional Estimation of Floods (GREHYS-NSERC), C. E. Delisle and M. A. Bouchard, *eds.*, November 15-17, Montreal, Quebec, Canada.

Associate Committee on Hydrology (ACH), 1989, *Hydrology of Floods in Canada*, National Research Council of Canada, Ottawa.

Atlantic Development Board (ADB), 1969, *Water Resources of The Atlantic Provinces: Background Study No 6*, Queen's Printer, Ottawa

Beersing, A. K., 1990, *Regional Flood Frequency Analysis for the Island of Newfoundland*, Department of Environment and Lands, Government of Newfoundland and Labrador, St. John's, Newfoundland.

Bobée, Bernard and Peter Rasmussen, Recent Advances in Flood Frequency Analysis, Review of Geophysics, Volume 33 Supplement, American Geophysical Union, 1995

Caissie, Daniel, and Nassir El-Jabi, 1991a, A Stochastic Study of Floods in Canada: Frequency Analysis and Regionalisation, *Canadian Journal of Civil Engineering*, Canadian Society of Civil Engineers, Vol. 18.

Caissie, Daniel, and Nassir El-Jabi, 1991b, A Stochastic Study of Floods in Canada: Truncation Level by Region, *Canadian Journal of Civil Engineering*, Canadian Society of Civil Engineers, Vol. 18.

Dalrymple, T., (1960), Flood Frequency Analysis, Water Supply Paper 1543-A, pp.11-51, United States Geological Survey.

Department of Environment and Lands (DOE), 1992, *Water Resources Atlas of Newfoundland*, Government of Newfoundland and Labrador, St. John's, Newfoundland

Department of Environment and Lands (DOE), 1984, *Regional Flood Frequency Analysis for the Island of Newfoundland*, Canada- Newfoundland Flood Damage Reduction Program, Government of Newfoundland and Labrador, St. John's, and Environment Canada, Dartmouth, Nova Scotia.

Ekanayake, S. T., and J. F. Cruise, 1994, Comparative Evaluation of Poisson Partial Duration Series for Mixed Flood Populations, *Stochastic Hydrology and Hydraulics*, Volume 8, Number 3, pp 207-218, Springer International.

Fill, H.D., and Stedinger, J.R., (1995) Homogeneity Tests Based Upon Gumbel Distribution and a Critical Appraisal of Dalrymple's Test, *Journal of Hydrology*, Vol. 166, Elsevier Science Publishers, Amsterdam.

Hann, C. T., H. P. Johnson, and D. L. Brakensiek, eds., 1982, *Hydrologic Modelling of Small Watersheds*, American Society of Agricultural Engineers, St. Joseph, Michigan, U.S.A.

Hebson, C.S. and C. Cunnane, 1987, Assessment of Use of At-Site and Regional Flood Data for Flood Frequency Estimation, *Hydrologic Frequency Modelling: Proceedings of the International Symposium on Flood Frequency and Risk Analysis - May 1986*, Edited by V.P. Singh, published by D. Reidel Publishing Company

Lye, L. M., and E. Moore, 1991, Discussion: "Instantaneous Peak Flow Estimation Procedures for Newfoundland Streams" by U. S. Panu and D. A. Smith, *Water Resources Bulletin*, American Water Resources Association, Maidment, David R., Ed., 1992, *Handbook of Hydrology*, McGraw-Hill Inc., New York.

Martins, E.S. and J.R. Stedinger, 2000, Generalized Maximum Likelihood Generalized Extreme Value Quantile Estimators for Hydrologic Data, *Water Resources Research*, Volume 36, Number 3, Pages 737-744, American Geophysical Union, Washington, USA, March 2000

Mrawira, M.D., *Generalized Pareto Distribution*, a paper submitted to Memorial University as part of the course Engineering 9714: Statistical Methods in Engineering, Fall 1991.

Natural Environment Research Council (NERC), 1975, Flood Studies Report, Volume 1, *Hydrological Studies*, London, United Kingdom, pages 185-213

Neill, Charles R., 1986, Unusual Canadian Floods and the Creager Diagram, *Canadian Journal of Civil Engineering*, Volume 13, Ottawa, Canada

Pilon, P. J., and K. David Harvey, 1994, *Consolidated Frequency Analysis Version 3.1: Reference Manual*, Environment Canada, Ottawa

Ouarda, T.B.M.J. and Fahim Ashkar, 1995, The Peaks-Over-Threshold Method for Regional Flood Frequency Estimation, *Proceedings of the 48th Canadian Water Resources Association Conference*, Canadian Water Resources Association.

Richter, Susan A., 1994, *Relationships of Flow and Basin Variables on The Island of Newfoundland, Canada, With a Regional Application*, A thesis submitted to Memorial University of Newfoundland, St. John's, Newfoundland.

Riggs, H.C., 1973, "Regional Analysis of Streamflow Characteristics", Techniques of Water Resource Investigations, Book 4, Chapter B3, USGS, Washington, DC

Rollings, K., (1999), *Regional Flood Frequency Analysis for the Island of Newfoundland*, Department of Environment and Labour, St. John's, Newfoundland

Rosbjerg, D., H. Madsen and P.F. Rasmussen 1992, Prediction in Partial Series with Generalized Pareto Distribution Exceedances, *Water Resources Research*, Vol.28, No.11, American Geophysical Union, Washington, USA, November 1992

Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L., 1988, *Applied Modelling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colorado

Soil Conservation Service (SCS), 1972, *National Engineering Handbook*, United States Department of Agriculture, Washington, D.C.

Taesombut and Yevjevich, V. And V. Yevjevich, 1978, Use of Partial Flood Series for Estimating Distribution of Maximum Annual Flood Peak, *Hydrology Papers No. 97*, Colorado State University, Fort Collins, Colorado

Wang, Q.J., 1991, The POT Model Described by the Generalized Pareto Distribution with Poisson Arrival Rate, *Journal of Hydrology*, Vol. 129, Elsevier Science Publishers, Amsterdam.

Appendix A

Data and Error Analysis for Nonlinear Regression Models

No.	Station	Years	Q(2)	Q(5)	Q(10)	Q(25)	Q(50)	Q(100)	Q(500)	Q(1000)	Station
1	02YA001	20	30.032	38.858	47.486	57.394	64.803	72.412	89.848	97.357	02YA001
2	02YA002	12	16.525	22.005	26.15	31.620	35.774	39.919	49.544	53.689	02YA002
3	02YC001	30	177.085	232.829	274.081	330.225	372.257	414.29	511.686	553.818	02YC001
4	02YD001	20	81.857	125.802	161.882	185.848	211.389	237.13	296.897	322.838	02YD001
5	02YD002	18	38.0701	51.437	61.5487	74.8156	85.0273	95.139	118.818	128.728	02YD002
6	02YE001	14	38.3833	47.78	54.8532	64.2299	71.3232	78.4164	94.8883	101.878	02YE001
7	02YF001	15	285.185	343.331	402.48	480.826	539.758	598.686	735.181	795.311	02YF001
8	02YG001	12	258.983	334.205	381.4	487.01	524.207	581.404	714.211	771.408	02YG001
9	02YH001	13	5.40227	7.01112	8.22817	9.83703	11.0541	12.2711	15.097	16.3141	02YH001
10	02YJ001	30	201.204	251.836	289.313	341.045	379.422	417.799	506.909	545.286	02YJ001
11	02YJ003	11	28.8783	37.5326	44.0036	52.5578	59.0288	65.4999	80.5251	86.9961	02YJ003
12	02YK002	23	110.495	142.080	167.008	189.181	223.52	247.856	304.37	328.709	02YK002
13	02YK004	24	82.25	109.447	130.022	157.218	177.794	198.369	246.14	266.714	02YK004
14	02YK005	20	60.8237	82.3185	98.503	119.898	136.082	152.267	189.848	206.031	02YK005
15	02YK007	13	23.1046	28.2414	33.0837	40.0204	44.6627	49.305	60.0841	64.7264	02YK007
16	02YK008	13	8.75148	11.7114	13.9508	16.9106	19.1497	21.3889	26.588	28.8271	02YK008
17	02YL001	70	514.83	635.087	727.004	848.74	940.38	1032.01	1244.79	1336.42	02YL001
18	02YL004	19	26.7432	34.8101	40.9124	48.9703	55.0617	61.164	75.3533	81.4556	02YL004
19	02YL005	13	10.1766	13.188	15.46	18.4984	20.7834	23.0675	28.3709	30.6540	02YL005
20	02YM003	18	35.2427	45.7504	53.6982	64.2080	72.1557	80.104	98.561	106.51	02YM003
21	02YN002	17	174.18	225.229	263.834	314.87	353.479	392.085	481.729	520.338	02YN002
22	02YO006	17	44.3773	58.9783	69.9871	79.1081	88.6369	98.1876	120.297	129.828	02YO006
23	02YO007	13	24.3846	30.2822	34.7611	40.6800	45.1378	49.6067	59.9634	64.4524	02YO007
24	02YO008	14	201.851	254.833	294.712	347.094	387.773	427.893	520.914	560.893	02YO008
25	02YO010	12	12.5346	16.2847	19.1213	22.8712	25.7078	28.5444	35.1309	37.9675	02YO010
26	02YP001	18	19.1766	24.5802	28.0725	34.0822	38.1745	42.2667	51.7687	55.861	02YP001
27	02YQ001	49	548.678	707.217	827.3	986.04	1106.12	1226.21	1505.03	1625.11	02YQ001
28	02YQ004	15	525.105	657.78	758.162	890.848	991.218	1091.59	1324.65	1425.02	02YQ004
29	02YQ005	11	29.9844	39.1149	46.0142	55.1347	62.034	68.9334	84.953	91.853	02YQ005
30	02YR001	39	29.0824	37.5414	43.9559	52.4345	58.8486	65.2627	80.1550	86.5690	02YR001
31	02YR002	20	65.895	86.001	101.17	121.218	136.384	151.55	186.765	201.931	02YR002
32	02YR003	17	52.3766	68.8541	81.3186	97.7862	110.261	122.725	151.687	164.132	02YR003
33	02YS001	34	177.082	224.474	259.868	306.890	342.11	377.522	459.746	495.158	02YS001
34	02YS003	31	10.3352	12.5234	14.1787	16.3889	18.0222	19.6775	23.521	25.1763	02YS003
35	02ZA001	19	107.010	137.708	160.923	191.612	214.827	238.043	291.947	315.162	02ZA001
36	02ZA002	16	38.6873	48.848	54.71	64.9706	72.7329	80.4848	98.5170	106.28	02ZA002
37	02ZA003	15	98.402	125.880	146.29	173.953	194.177	214.601	262.889	283.313	02ZA003
38	02ZB001	38	172.434	214.906	247.034	289.508	321.634	353.763	428.363	460.491	02ZB001
39	02ZC002	16	224.258	282.421	326.418	384.58	428.577	472.575	574.73	618.73	02ZC002
40	02ZE001	22	269.201	363.28	418.223	493.202	549.185	605.129	735.071	791.034	02ZE001
41	02ZF001	48	173.737	218.848	249.638	292.848	325.934	358.221	434.12	466.907	02ZF001
42	02ZG001	40	66.888	72.6285	84.4505	100.081	111.905	123.729	151.184	163.008	02ZG001
43	02ZG002	20	47.8045	60.6371	70.2680	83.0014	92.6332	102.265	124.629	134.261	02ZG002
44	02ZG003	19	50.0150	63.449	73.1580	85.8801	95.600	105.408	127.847	137.455	02ZG003
45	02ZG004	17	27.2263	34.0488	40.015	48.2378	54.0038	59.77	73.1587	78.9249	02ZG004
46	02ZH001	48	191.017	240.557	278.033	327.573	365.048	402.524	489.539	527.015	02ZH001
47	02ZH002	30	22.7407	29.4688	34.5504	41.2888	46.3782	51.4658	63.2836	68.3732	02ZH002
48	02ZJ001	22	19.8882	25.2381	29.2837	34.4446	38.4092	42.3737	51.5782	55.5438	02ZJ001
49	02ZJ002	15	12.1615	14.8148	16.8068	19.4402	21.4322	23.4242	28.0466	30.0410	02ZJ002
50	02ZJ003	12	22.8753	29.0183	33.687	39.811	44.4586	49.1065	59.8962	64.5458	02ZJ003
51	02ZK001	50	107.886	134.422	154.496	181.033	201.107	221.181	267.792	287.866	02ZK001
52	02ZK002	18	44.8812	57.826	67.2686	79.8715	89.4521	99.0327	121.270	130.850	02ZK002
53	02ZK003	15	30.1742	38.03	43.8729	51.6284	57.7711	63.7138	77.5123	83.4540	02ZK003
54	02ZK004	15	72.1584	88.8258	104.986	123.083	137.033	151.003	183.441	197.411	02ZK004
55	02ZK005	11	17.6885	22.0747	25.8481	30.8363	34.6898	38.3832	47.1448	50.9183	02ZK005
56	02ZL003	18	6.04606	7.66784	9.28856	11.1210	12.5143	13.9089	17.1408	18.5333	02ZL003
57	02ZL004	19	10.8401	14.1364	16.8306	19.8273	22.4212	24.815	30.7055	33.1894	02ZL004
58	02ZL005	13	3.77804	4.70742	5.57006	6.50144	7.38408	8.13872	9.93074	10.7034	02ZL005
59	02ZM008	48	2.10878	2.73578	3.18308	3.72987	4.15814	4.58831	5.58848	6.00865	02ZM008
60	02ZM009	19	21.5868	25.4404	29.3707	32.2443	35.1746	38.1049	44.9088	47.8391	02ZM009
61	02ZM010	15	8.75295	11.2472	13.134	15.6282	17.515	19.4018	23.7829	25.6697	02ZM010
62	02ZN001	30	30.2478	36.5177	41.2806	47.5304	52.2733	57.0162	68.0288	72.7718	02ZN001
63	02ZN002	13	6.00186	8.49395	9.89624	11.2902	12.4945	13.6988	16.495	17.8993	02ZN002

Station	DRAIN. AREA (KM^2)	FRAC TREE (-)	FRAC SWAMP (-)	FRAC LAKE (-)	FRAC L&S (-)	FRAC BARW (-)	FRAC ACLS (-)	LSF	LENGTH MAIN R (KM)	ELEV. DIFF. (M)	SLOPE M1 (%)	SLOPE M2 (%)	DRAIN. DENSITY (KM^-1)	SHAPE FACTOR (-)	CENTROID LAT. (DEG.)	LONG (DEG.)
02YA001	308.0	0.64	0.14	0.22	0.35	0.00	0.98	1.79	38.8	88	0.23	0.14	0.84	1.40	51.100	56.658
02YA002	33.8	0.40	0.03	0.13	0.16	0.44	0.98	1.81	13.2	158	1.14	1.21	0.81	1.84	51.600	55.747
02YC001	824.0	0.33	0.04	0.13	0.17	0.50	0.98	1.81	40.3	479	0.98	1.01	0.78	1.45	50.820	56.014
02YD001	237.0	0.81	0.04	0.05	0.08	0.11	0.73	1.88	40.6	328	0.61	0.68	0.34	2.23	50.836	56.418
02YD002	200.0	0.83	0.04	0.13	0.17	0.00	0.98	1.80	38.3	270	0.70	0.47	0.93	1.85	50.947	56.016
02YE001	95.7	0.49	0.08	0.08	0.12	0.38	0.84	1.82	24.5	700	2.88	3.08	0.75	1.84	50.208	57.583
02YF001	811.0	0.89	0.05	0.06	0.13	0.18	1.00	1.83	30.2	298	0.83	0.73	0.58	1.88	50.070	57.048
02YG001	827.0	0.78	0.08	0.07	0.13	0.08	0.83	1.55	300.0	375	1.16	1.11	1.30	1.83	49.858	57.400
02YH001	33.4	0.79	0.02	0.11	0.13	0.06	0.93	1.88	14.6	183	1.12	0.85	1.13	1.88	49.672	57.883
02YJ001	840.0	0.79	0.09	0.08	0.14	0.07	0.75	1.87	80.0	909	0.85	0.35	1.12	1.81	49.783	58.109
02YJ003	118.0	0.80	0.05	0.05	0.10	0.04	1.00	1.85	16.6	164	0.88	0.78	1.73	1.54	48.847	58.044
02YK002	470.0	0.55	0.06	0.10	0.18	0.29	1.00	1.87	54.8	981	1.02	0.58	0.83	2.32	48.987	57.654
02YK004	520.0	0.35	0.24	0.12	0.38	0.29	0.95	1.77	49.3	329	0.85	0.32	0.84	1.70	48.858	57.018
02YK005	381.0	0.68	0.08	0.10	0.17	0.18	0.94	1.85	38.1	370	0.88	1.07	0.19	1.88	49.208	56.502
02YK007	112.0	0.87	0.08	0.04	0.13	0.00	0.98	1.81	26.8	234	0.88	0.80	1.20	1.81	48.211	57.500
02YK008	20.4	0.75	0.22	0.02	0.24	0.01	0.85	1.50	10.1	137	1.35	1.10	1.28	1.67	49.414	57.103
02YL001	2110.0	0.74	0.08	0.05	0.11	0.15	0.75	1.60	110.8	878	0.57	0.40	0.78	1.58	49.920	57.207
02YL004	58.5	0.94	0.01	0.01	0.02	0.05	0.08	1.08	13.2	130	0.99	1.04	1.34	1.54	48.030	57.831
02YL005	17.0	0.81	0.08	0.02	0.10	0.00	0.40	1.38	8.2	244	2.88	2.88	1.09	1.10	48.858	56.108
02YM003	93.2	0.81	0.07	0.05	0.11	0.00	0.56	1.49	18.8	107	0.58	0.57	0.68	1.67	48.894	56.217
02YN002	480.0	0.23	0.09	0.12	0.18	0.63	1.00	1.81	57.3	188	0.28	0.30	1.37	2.18	48.137	57.881
02YO006	177.0	0.83	0.13	0.03	0.16	0.02	0.87	1.89	42.7	190	0.45	0.40	0.88	1.83	48.032	55.586
02YO007	68.3	0.70	0.24	0.04	0.28	0.02	0.73	1.97	23.1	272	1.18	0.88	0.74	1.52	48.158	55.844
02YO008	823.0	0.73	0.18	0.05	0.24	0.03	0.55	1.40	89.0	221	0.32	0.30	0.89	1.80	48.887	55.808
02YO010	81.8	0.81	0.09	0.10	0.18	0.00	0.89	1.79	18.2	187	1.03	0.82	0.77	1.55	49.114	55.887
02YP001	83.6	0.88	0.07	0.08	0.13	0.00	0.78	1.72	20.0	113	0.56	0.93	0.86	1.82	49.201	55.830
02YQ001	4400.0	0.78	0.08	0.09	0.17	0.07	0.81	1.82	133.6	287	0.22	0.15	0.45	2.08	48.829	55.101
02YQ004	2150.0	0.88	0.25	0.06	0.31	0.03	0.44	1.22	104.2	265	0.25	0.17	0.45	1.83	48.587	54.558
02YQ005	90.8	0.85	0.11	0.04	0.15	0.00	0.87	1.79	22.8	372	1.85	1.03	1.09	1.78	49.044	55.122
02YR001	267.0	0.75	0.07	0.18	0.24	0.01	0.88	1.86	49.3	177	0.38	0.32	0.28	1.93	48.807	54.480
02YR002	398.0	0.68	0.16	0.17	0.33	0.00	0.96	1.79	42.0	85	0.23	0.21	0.74	1.68	48.277	54.184
02YR003	554.0	0.70	0.13	0.20	0.33	0.00	0.97	1.80	52.4	138	0.28	0.23	0.68	1.72	48.078	54.181
02YS001	1290.0	0.55	0.21	0.09	0.30	0.15	0.82	1.78	105.0	207	0.20	0.12	0.73	2.35	48.383	54.778
02YS003	38.7	0.84	0.14	0.02	0.16	0.00	1.00	1.92	11.2	143	1.28	1.11	0.84	1.43	48.983	54.037
02ZA001	343.0	0.60	0.02	0.07	0.09	0.30	0.83	1.78	68.5	483	0.72	0.68	1.04	2.45	48.416	54.082
02ZA002	72.0	0.82	0.01	0.04	0.05	0.13	0.43	1.38	20.4	480	2.28	2.18	1.18	1.72	48.841	54.719
02ZA003	138.0	0.88	0.07	0.04	0.11	0.18	0.73	1.86	29.2	450	1.78	1.66	1.40	1.84	47.813	56.070
02ZB001	205.0	0.08	0.06	0.07	0.13	0.78	0.80	1.52	33.3	444	1.33	1.27	0.72	2.09	47.745	56.885
02ZC002	230.0	0.20	0.01	0.05	0.06	0.82	0.34	1.30	21.8	380	1.24	1.08	0.88	1.84	47.932	57.642
02ZE001	2640.0	0.35	0.02	0.14	0.16	0.50	1.00	1.82	100.4	172	0.12	0.08	0.38	1.75	48.222	56.110
02ZF001	1170.0	0.32	0.08	0.18	0.24	0.44	0.86	1.64	68.1	282	0.41	0.34	0.61	2.15	48.038	55.288
02ZG001	205.0	0.26	0.01	0.09	0.10	0.83	0.88	1.81	44.7	370	0.83	0.80	0.59	2.45	47.270	55.158
02ZG002	188.0	0.37	0.04	0.09	0.13	0.48	0.92	1.83	28.7	221	0.83	0.78	1.35	1.84	47.882	55.339
02ZG003	115.0	0.16	0.06	0.07	0.13	0.73	0.92	1.85	24.5	138	0.85	0.34	1.55	1.82	48.854	55.782
02ZG004	42.7	0.34	0.03	0.16	0.16	0.48	0.92	1.83	10.0	107	1.07	1.10	1.62	1.53	47.480	54.888
02ZH001	784.0	0.11	0.40	0.18	0.08	0.23	0.81	1.57	58.8	207	0.41	0.38	0.71	1.87	48.053	54.483
02ZH002	43.3	0.40	0.02	0.08	0.10	0.50	0.82	1.87	17.0	110	0.85	0.58	1.11	1.88	47.952	53.953
02ZJ001	67.4	0.82	0.08	0.18	0.16	0.03	0.88	1.78	18.8	128	0.80	0.50	1.24	1.64	48.333	53.717
02ZJ002	73.6	0.74	0.08	0.13	0.19	0.07	0.82	1.72	18.8	137	0.76	0.55	1.11	1.33	48.475	53.414
02ZJ003	108.0	0.85	0.10	0.07	0.17	0.18	0.84	1.58	25.1	250	0.88	0.81	0.88	1.88	48.275	54.239
02ZK001	285.0	0.51	0.02	0.10	0.12	0.37	0.55	1.47	45.2	185	0.37	0.23	1.01	2.00	47.357	53.533
02ZK002	89.6	0.40	0.16	0.15	0.31	0.24	0.81	1.64	28.8	200	0.74	0.57	1.11	1.81	47.340	53.732
02ZK003	37.2	0.86	0.11	0.02	0.13	0.01	0.34	1.24	14.6	228	1.58	1.77	1.18	1.40	47.183	54.100
02ZK004	104.0	0.23	0.38	0.08	0.40	0.31	0.81	1.87	28.5	238	0.83	0.68	1.50	1.85	47.158	53.833
02ZK005	50.3	0.38	0.00	0.07	0.07	0.57	0.90	1.45	14.8	137	0.82	0.68	1.18	1.80	47.647	53.788
02ZL003	10.8	0.42	0.01	0.06	0.08	0.48	1.00	1.95	7.0	81	1.31	1.25	1.09	1.38	47.808	53.108
02ZL004	28.8	0.70	0.00	0.04	0.04	0.27	0.38	1.38	13.4	122	0.91	1.03	1.14	1.73	47.558	53.431
02ZL005	11.2	0.38	0.03	0.07	0.10	0.51	1.00	1.85	8.7	211	2.43	2.43	1.00	1.52	48.089	53.183
02ZM008	3.8	0.75	0.17	0.04	0.21	0.04	1.00	1.88	2.6	84	2.44	2.42	1.04	1.24	47.646	52.826
02ZM009	53.6	0.38	0.01	0.12	0.14	0.81	1.00	1.83	14.8	133	0.89	0.88	1.13	1.37	46.887	53.014
02ZM010	17.3	0.22	0.05	0.06	0.11	0.68	0.98	1.84	8.7	258	2.88	2.22	1.01	1.40	47.431	53.200
02ZN001	53.3	0.09	0.00	0.13	0.13	0.79	1.00	1.84	14.8	83	0.83	0.81	1.09	2.08	48.013	53.289
02ZN002	15.5	0.88	0.00	0.12	0.12	0.00	0.82	1.75	10.3	23	0.22	0.43	1.03	1.53	48.817	53.803

Avalon

	Station	Q(2)	DA km^2	SLP M2 (%)	FRAC ACLS (-)	LSF	DRAIN. DENSITY km^-1	SHAPE FACTOR (-)
1	02zg001	56.996	205.0	0.60	0.96	1.91	0.55	2.45
2	02zg002	47.9045	166.0	0.78	0.92	1.85	1.35	1.84
3	02zg003	50.6158	115.0	0.34	0.92	1.85	1.55	1.62
4	02zg004	27.2263	42.7	1.10	0.92	1.83	1.62	1.53
5	02zh001	191.017	764.0	0.38	0.91	1.57	0.71	1.67
6	02zh002	22.7407	43.3	0.59	0.92	1.87	1.11	1.66
7	02zk001	107.886	285.0	0.23	0.55	1.47	1.01	2.00
8	02zk002	44.9612	89.6	0.57	0.81	1.64	1.11	1.91
9	02zk003	30.1742	37.2	1.77	0.34	1.24	1.16	1.48
10	02zk004	72.1584	104.0	0.66	0.91	1.67	1.50	1.85
11	02zk005	17.0865	50.3	0.88	0.50	1.45	1.18	1.90
12	02zl003	6.04686	10.8	1.25	1.00	1.95	1.09	1.36
13	02zl004	10.8401	28.9	1.03	0.39	1.36	1.14	1.73
14	02zl005	3.77604	11.2	2.43	1.00	1.95	1.00	1.52
15	02zm006	2.16978	3.9	2.42	1.00	1.89	1.04	1.24
16	02zm009	21.5668	53.6	0.98	1.00	1.93	1.13	1.37
17	02zm016	8.75295	17.3	2.22	0.90	1.84	1.01	1.40
18	02zn001	30.2479	53.3	0.61	1.00	1.94	1.09	2.06
19	02zn002	6.90198	15.5	0.43	0.82	1.75	1.03	1.53

Estimates Calculated using DA only

Output	Error	Mean Erro	Square Err	MSE	RMSE
73.69738	16.70138	0.541017	278.9361	95.97557	9.796712
63.19045	15.28595		233.6602		
48.35677	-2.25903		5.103205		
23.48732	-3.73898		13.97994		
192.2655	1.248518		1.558796		
23.72743	0.986729		0.973634		
93.70248	-14.1835		201.1723		
40.31364	-4.64756		21.59986		
21.24141	-8.93279		79.79469		
44.93971	-27.2187		740.8572		
26.46595	9.379454		87.97416		
8.623374	2.576514		6.638425		
17.67108	6.83098		46.66229		
8.855023	5.078983		25.79606		
4.104324	1.934544		3.742461		
27.72062	6.153815		37.86944		
12.157	3.404051		11.58756		
27.60744	-2.64046		6.972036		
11.22141	4.31943		18.65747		

Estimates Calculated using DA & SLP

Output	Error	Mean Erro	Square Err	MSE	RMSE
69.64164	12.64564	-0.1721	159.9123	79.12005	8.894945
57.17186	9.267362		85.884		
52.99113	2.375328		5.642184		
21.15356	-6.07274		36.87817		
187.3726	-3.64441		13.28172		
24.29035	1.549646		2.401403		
106.6287	-1.2573		1.580804		
40.1907	-4.7705		22.75765		
17.41503	-12.7592		162.7965		
43.05874	-29.0997		846.7902		
24.75274	7.666236		58.77117		
8.092628	2.045768		4.185166		
16.43818	5.59808		31.3385		
7.211766	3.435726		11.80421		
3.5264	1.35662		1.840417		
25.26651	3.699714		13.68788		
9.874136	1.121186		1.257057		
27.78625	-2.46165		6.059739		
12.9362	6.034221		36.41182		

Estimates Calculated using DA & DRD

Output	Error	Mean Erro	Square Err	MSE	RMSE
59.50561	2.509614	0.184984	6.298161	74.65818	8.640496
70.73379	22.82929		521.1766		
55.34405	4.728251		22.35636		
25.15875	-2.06755		4.274743		
191.2753	0.258312		0.066725		
22.047	-0.6937		0.481226		
98.09069	-9.79531		95.94814		
39.77326	-5.18794		26.91477		
19.8192	-10.355		107.226		
50.39422	-21.7642		473.6795		
25.48893	8.402427		70.60078		
7.096979	1.050119		1.102751		
16.03756	5.197464		27.01364		
7.063343	3.287303		10.80636		
3.041129	0.871349		0.759249		
26.40814	4.841336		23.43854		
10.09357	1.340621		1.797264		
25.91102	-4.33688		18.8085		
9.301156	2.399176		5.756045		

Estimates Calculated using DA & FRAC

Output	Error	Mean Erro	Square Err	MSE	RMSE
70.17106	13.17506	0.315833	173.5821	85.70844	9.257885
60.60427	12.69977		161.2842		
46.05368	-4.56212		20.81292		
21.94939	-5.27691		27.8458		
190.4029	-0.61411		0.377131		
22.17968	-0.56102		0.314741		
104.0571	-3.82892		14.66061		
39.52196	-5.43924		29.58533		
25.77194	-4.40226		19.37989		
42.8411	-29.3173		859.5039		
29.15986	12.07336		145.7661		
7.6784	1.63154		2.661923		
20.5754	9.735303		94.77612		
7.890143	4.114103		16.92584		
3.584149	1.414369		2.000441		
25.44984	3.883039		15.07799		
11.2318	2.478851		6.144704		
25.34322	-4.90468		24.05593		
10.60397	3.701992		13.70475		

Estimates Calculated using DA & LSF

Output	Error	Mean Erro	Square Err	MSE	RMSE
64.13961	7.143607	0.710325	51.03112	73.40004	8.567382
57.07824	9.173743		84.15755		
44.19989	-6.41591		41.16385		
22.42997	-4.79633		23.00474		
193.5797	2.562697		6.567415		
22.25122	-0.48948		0.239588		
103.5365	-4.34946		18.91777		
41.8107	-3.1505		9.925661		
29.69118	-0.48302		0.233311		
45.58398	-26.5744		706.2		
31.47517	14.38867		207.0338		
8.129276	2.082416		4.336457		
22.82507	11.98497		143.6395		
8.354271	4.578231		20.9602		
4.130298	1.960518		3.843632		
24.99687	3.430068		11.76537		
11.93534	3.182394		10.12763		
24.82297	-5.42493		29.42986		
11.59489	4.692907		22.02338		

Estimates Calculated using DA & LSF & drd

Output	Error	Mean Erro	Square Err	MSE	RMSE
53.0522	-3.9438	0.313434	15.55358	57.97161	7.613909
64.10801	16.20351		262.5537		
50.69916	0.083365		0.00695		
24.14092	-3.08538		9.519548		
191.3752	0.358229		0.128328		
20.97216	-1.76854		3.12772		
106.7658	-1.12018		1.254803		
41.18563	-3.77557		14.25493		
27.00475	-3.16945		10.04544		
50.77108	-21.3873		457.4174		
29.95008	12.86358		165.4718		
6.859748	0.812888		0.660786		
20.40159	9.561494		91.42216		
6.842856	3.066816		9.405361		
3.145975	0.976195		0.952957		
24.20021	2.633411		6.934852		
10.09819	1.345236		1.80966		
23.70262	-6.54528		42.84068		
9.748026	2.846046		8.09998		

Estimates Calculated using DA & SHP

Output	Error	Mean Erro	Square Err	MSE	RMSE
70.01072	13.01472	0.815673	169.3829	94.36929	9.714386
63.25905	15.35455		235.7623		
49.63319	-0.98261		0.965519		
24.47997	-2.74633		7.542323		
194.031	3.013952		9.083908		
24.38033	1.639629		2.688383		
92.11016	-15.7758		248.8771		
40.21717	-4.74403		22.50585		
22.29636	-7.87784		62.06033		
45.06056	-27.0978		734.2929		
26.52489	9.438388		89.08317		
9.261071	3.214211		10.33115		
18.06737	7.227272		52.23346		
9.316537	5.540497		30.69711		
4.505559	2.335779		5.455864		
29.4328	7.865996		61.8739		
12.94597	4.19302		17.58142		
27.26424	-2.98366		8.902205		
11.7699	4.867921		23.69665		

Central

STATION	Q(2)	DA km^2	M2 (%)	Frac ACLS (-)	LSF	DENSITY km^-1	Shape FACTOR (-)
02yn002	174.19	469.0	0.30	1.00	1.91	1.37	2.15
02yo006	44.3773	177.0	0.45	0.97	1.89	0.80	1.93
02yo007	24.3845	88.3	0.88	0.73	1.57	0.74	1.52
02yo008	201.651	823.0	0.30	0.55	1.40	0.69	1.80
02yo010	12.5348	61.6	0.62	0.89	1.79	0.77	1.55
02yp001	19.1705	63.8	0.53	0.79	1.72	0.88	1.62
02yq001	548.476	4400.0	0.15	0.91	1.82	0.45	2.08
02yq004	525.105	2150.0	0.17	0.44	1.22	0.45	1.63
02yq005	29.9944	80.8	1.03	0.87	1.79	1.09	1.78
02yr001	29.0624	267.0	0.32	0.98	1.86	0.26	1.93
02yr002	65.955	399.0	0.21	0.96	1.79	0.74	1.68
02yr003	52.3768	554.0	0.23	0.97	1.80	0.68	1.72
02ys001	177.662	1290.0	0.12	0.92	1.76	0.73	2.35
02ys003	10.3352	36.7	1.11	1.00	1.92	0.64	1.43
02ze001	289.281	2640.0	0.08	1.00	1.92	0.36	1.75
02zf001	173.737	1170.0	0.34	0.96	1.84	0.61	2.15
02zj001	19.9982	67.4	0.50	0.86	1.78	1.24	1.64
02zj002	12.1815	73.6	0.55	0.82	1.72	1.11	1.33
02zj003	22.8753	106.0	0.91	0.68	1.58	0.66	1.66

Estimates Calculated using DA only

Output	Error	Mean Erro	Square Err	MSE	RMSE
103.1394	-71.0506	3.251113	5048.194	3407.845	58.37675
49.07994	4.702641		22.11483		
28.88952	4.505015		20.29516		
158.3251	-43.3259		1877.132		
21.9565	9.421698		88.76839		
22.5516	3.381104		11.43187		
568.0661	19.59012		383.7729		
329.1357	-195.969		38403.98		
26.99988	-2.99452		8.967161		
67.13791	38.07551		1449.745		
91.18535	25.23035		636.5703		
117.0988	64.72199		4188.936		
222.9996	45.33757		2055.495		
14.79638	4.461185		19.90217		
384.8823	95.60132		9139.612		
207.0084	33.27144		1106.989		
23.51502	3.516818		12.36801		
25.14613	12.96463		168.0817		
33.20544	10.33014		106.7117		

Estimates Calculated using DA & SLPonly

Output	Error	Mean Erro	Square Err	MSE	RMSE
107.9333	-66.2567	4.707915	4389.956	2951.302	54.32589
52.57591	8.198607		67.21715		
35.28152	10.89702		118.7451		
174.2858	-27.3652		748.8557		
23.46527	10.93047		119.4752		
23.09074	3.920244		15.36832		
594.3642	45.88823		2105.73		
336.2084	-188.897		35681.93		
34.21615	4.22175		17.82317		
67.78039	38.71799		1499.083		
85.14516	19.19016		368.2624		
114.4675	62.09072		3855.257		
196.7096	19.04764		362.8127		
17.82005	7.484848		56.02295		
323.9149	34.63394		1199.51		
243.7408	70.0038		4900.532		
23.7688	3.770603		14.21745		
26.394	14.2125		201.9951		
41.63573	18.76043		351.9538		

Estimates Calculated using DA & FACLS

Output	Error	Mean Erro	Square Err	MSE	RMSE
76.60538	-97.5846	-1.28021	9522.757	695.9943	26.3817
34.97408	-9.40322		88.42053		
24.71519	0.330686		0.109353		
198.1362	-3.51477		12.3536		
15.61581	3.081006		9.492599		
17.70448	-1.46602		2.149226		
530.0121	-18.4639		340.9173		
526.5684	1.463361		2.141425		
19.92216	-10.0722		101.4499		
48.78787	19.72547		389.0942		
69.23746	3.282465		10.77458		
90.15546	37.77866		1427.227		
189.7682	12.10617		146.5595		
9.246329	-1.08887		1.185641		
321.3972	32.11617		1031.449		
169.0757	-4.6613		21.7277		
17.29984	-2.69836		7.281164		
19.34153	7.16003		51.26603		
30.46056	7.585256		57.53611		

Estimates Calculated using DA & LSF

Output	Error	Mean Erro	Square Err	MSE
74.45143	-99.7386	-2.23373	9947.782	737.9308
33.26017	-11.1171		123.5905	
24.39794	0.013443		0.000181	
191.1793	-10.4717		109.6575	
14.77955	2.24475		5.038902	
16.17003	-3.00047		9.002815	
528.1317	-20.3443		413.889	
522.405	-2.69998		7.289913	
18.57787	-11.4165		130.3371	
48.21916	19.15676		366.9813	
71.36403	5.409027		29.25757	
93.17667	40.79987		1664.629	
196.6011	18.93911		358.6897	
8.589739	-1.74546		3.046634	
316.8205	27.53947		758.4225	
170.184	-3.55297		12.62362	
16.12787	-3.87033		14.97947	
18.2922	6.110702		37.34068	
28.17882	5.303516		28.12728	

	Estimates Calculated using DA & DRD					
RMSE	Output	Error	Mean Erro	Square Err	MSE	RMSE
27.16488	126.3153	-47.8747	3.469568	2291.989	3284.476	57.31035
	48.16624	3.788939		14.35606		
	26.50597	2.121465		4.500614		
	164.4892	-37.1618		1381.002		
	19.91093	7.376129		54.40727		
	21.28223	2.111727		4.459389		
	581.0668	32.59079		1062.16		
	320.9601	-204.145		41675.13		
	27.60022	-2.39418		5.732087		
	48.38688	19.32448		373.4355		
	92.26536	26.31036		692.2353		
	118.2685	65.89169		4341.715		
	242.1325	64.47048		4156.443		
	12.29349	1.958287		3.834888		
	357.0746	67.79363		4595.976		
	212.4098	38.67283		1495.588		
	24.64517	4.64697		21.59433		
	25.68697	13.50547		182.3978		
	29.80944	6.934143		48.08234		

Estimates Calculated using DA & SHP

Output	Error	Mean Erro	Square Err	MSE	RMSE
83.03911	-91.1509	2.037112	8308.484	2875.458	53.6233
42.4031	-1.9742		3.89746		
31.51792	7.133424		50.88574		
160.5471	-41.1039		1689.533		
22.98835	10.45355		109.2768		
22.51829	3.347789		11.20769		
534.5016	-13.9744		195.2833		
392.842	-132.263		17493.51		
24.52336	-5.47104		29.93229		
59.27971	30.21731		913.086		
95.95436	29.99936		899.9619		
122.821	70.44415		4962.379		
171.2843	-6.37769		40.6749		
16.51135	6.176147		38.14479		
428.5445	139.2635		19394.31		
174.9236	1.186559		1.407923		
23.14396	3.145756		9.895779		
31.6005	19.419		377.0975		
33.10905	10.23375		104.7297		

Estimates Calculated using DA & SLP &FACLS

Output	Error	Mean Erro	Square Err	MSE	RMSE
78.86622	-95.3238	-3.90085	9086.622	887.4464	29.79004
35.15721	-9.22009		85.01008		
25.67201	1.287513		1.657689		
190.4217	-11.2293		126.0969		
14.70392	2.169122		4.705089		
15.56652	-3.60398		12.98867		
584.2506	35.77465		1279.825		
464.0179	-61.0871		3731.634		
22.28083	-7.71357		59.49922		
47.17635	18.11395		328.1153		
62.74879	-3.20621		10.27979		
86.80141	34.42461		1185.054		
168.9091	-8.75294		76.61398		
9.847735	-0.48747		0.237622		
286.868	-2.41302		5.82268		
200.7425	27.00554		729.2994		
15.35845	-4.63975		21.52732		
17.67891	5.497407		30.22148		
32.16356	9.28826		86.27177		

Northwest			DA	ACLS	LSF	M2	DENSITY	FACTOR
		Q(t)	km^2	(-)		(%)	km^-1	(-)
	STATION	2.00						
1	02ya001	30.032	306.0	0.96	1.78	0.14	0.54	1.48
2	02ya002	16.525	33.6	0.99	1.91	1.21	0.91	1.64
3	02yc001	177.065	624.0	0.99	1.91	1.01	0.76	1.45
4	02yd001	91.857	237.0	0.73	1.68	0.68	0.34	2.23
5	02yd002	38.0701	200.0	0.99	1.90	0.47	0.93	1.65
6	02ye001	38.3833	95.7	0.88	1.82	3.09	0.75	1.64
7	02yf001	265.165	611.0	1.00	1.93	0.73	0.58	1.86
8	02yg001	258.593	627.0	0.63	1.55	1.11	1.30	1.83
9	02yh001	5.40227	33.4	0.93	1.86	0.85	1.13	1.68
10	02yj003	28.9783	119.0	1.00	1.95	0.78	1.73	1.54
11	02yk002	110.495	470.0	1.00	1.92	0.59	0.63	2.32
12	02yk004	82.25	529.0	0.95	1.77	0.32	0.64	1.78
13	02yk005	60.9237	391.0	0.94	1.85	1.07	0.19	1.98
14	02yk007	23.1046	112.0	0.98	1.91	0.90	1.28	1.61
15	02yk008	8.75146	20.4	0.65	1.50	1.16	1.28	1.47
16	02yl001	514.83	2110.0	0.75	1.68	0.46	0.79	1.56
17	02yl004	26.7432	58.5	0.08	1.06	1.04	1.34	1.54
18	02yl005	10.1766	17.0	0.46	1.39	2.88	1.05	1.10
19	02ym003	35.2427	93.2	0.56	1.49	0.57	0.68	1.67

Estimates Calculated using DA only

Output	Error	Mean Erro	Square Err	MSE	RMSE
89.75634	59.72434	1.46528	3566.996	1533.348	39.15798
11.81278	-4.71222		22.20497		
172.6443	-4.42068		19.54241		
70.98913	-20.8679		435.4682		
60.7461	22.676		514.2008		
30.87806	-7.50524		56.32863		
169.3397	-95.8253		9182.497		
173.4061	-85.1869		7256.803		
11.74822	6.345951		40.27109		
37.71594	8.737641		76.34638		
133.0941	22.59906		510.7176		
148.356	66.10601		4370.005		
112.4064	51.48272		2650.47		
35.67426	12.56966		157.9964		
7.471595	-1.27986		1.638054		
528.28	13.44996		180.9014		
19.65269	-7.09051		50.27538		
6.320115	-3.85649		14.87248		
30.13677	-5.10593		26.07056		

Estimates Calculated using DA only

Output	Error	Mean Erro	Square Err	MSE	RMSE
79.12415	49.09215	-13.3491	2410.04	2414.798	49.14059
11.91567	-4.60933		21.2459		
145.7199	-31.3451		982.5123		
63.56311	-28.2939		800.5441		
54.9577	16.8876		285.1909		
29.22048	-9.16282		83.95727		
143.1143	-122.051		14896.37		
146.3201	-112.273		12605.2		
11.85486	6.452593		41.63596		
35.22002	6.241717		38.95903		
114.2967	3.801697		14.4529		
126.4874	44.23742		1956.949		
97.62056	36.69686		1346.66		
33.43687	10.33227		106.7559		
7.769604	-0.98186		0.964042		
413.9576	-100.872		10175.24		
19.16454	-7.57866		57.43609		
6.645697	-3.5309		12.46728		
28.56507	-6.67763		44.59078		

Estimates Calculated using DA & SLP

Output	Error	Mean Erro	Square Err	MSE	RMSE
37.27307	7.24107	-0.51712	52.4331	922.3541	30.37028
10.56291	-5.96209		35.54651		
212.1939	35.12889		1234.039		
62.45914	-29.3979		864.2341		
43.32358	5.253483		27.59909		
50.93767	12.55437		157.6123		
176.1961	-88.9689		7915.472		
223.0994	-35.4936		1259.792		
8.803726	3.401456		11.5699		
32.30964	3.331339		11.09782		
120.2928	9.797764		95.99618		
99.87022	17.62022		310.4722		
132.7922	71.8685		5165.082		
32.54259	9.437987		89.0756		
6.103726	-2.64773		7.010496		
517.1549	2.324943		5.405361		
17.60282	-9.14038		83.54657		
7.917313	-2.25929		5.104377		
21.32723	-13.9155		193.6404		

Estimates Calculated using DA & SLP &FACLS

Output	Error	Mean Erro	Square Err	MSE	RMSE
38.69688	8.664882	1.166213	75.08019	875.1627	29.58315
10.96539	-5.55961		30.90928		
202.2032	25.1382		631.9293		
67.25994	-24.5971		605.0153		
43.70454	5.634436		31.74687		
51.45096	13.06766		170.7638		
169.1102	-96.0548		9226.525		
239.1132	-19.4798		379.4642		
9.386508	3.984238		15.87416		
32.56308	3.584775		12.85061		
117.0829	6.587873		43.40007		
99.91223	17.66223		311.9544		
129.9835	69.05985		4769.263		
32.9045	9.7999		96.03805		
7.205608	-1.54585		2.389659		
522.4418	7.611759		57.93887		
35.22211	8.478914		71.89198		
10.04596	-0.13064		0.017068		
25.49384	-9.74886		95.04029		

Estimates Calculated using DA LSF & SLP

Output	Error	Mean Erro	Square Err	MSE	RMSE
39.50955	9.477547	0.413683	89.8239	893.817	29.89677
10.98674	-5.53826		30.67233		
201.454	24.389		594.8232		
66.6812	-25.1758		633.821		
43.48854	5.418442		29.35952		
51.67747	13.29417		176.7349		
167.1218	-98.0432		9612.475		
239.8416	-18.7514		351.6143		
9.373645	3.971375		15.77182		
32.13783	3.159533		9.982647		
116.1723	5.677328		32.23206		
102.4762	20.22618		409.0983		
130.1717	69.24804		4795.291		
32.6975	9.592898		92.02369		
7.475487	-1.27597		1.628107		
517.9114	3.081411		9.495092		
25.84369	-0.89951		0.809122		
10.02579	-0.15081		0.022745		
25.40172	-9.84098		96.84483		

Estimates Calculated using DA SLP DRD

Output	Error	Mean Erro	Square Err	MSE	RMSE
39.58734	9.555338	0.496702	91.30448	783.4308	27.98983
12.29672	-4.22828		17.87834		
211.5546	34.48957		1189.53		
55.37291	-36.4841		1331.089		
49.54684	11.47674		131.7155		
51.60739	13.22409		174.8764		
168.844	-96.321		9277.728		
250.0457	-8.5473		73.05638		
10.97362	5.571349		31.03993		
42.21708	13.23878		175.2654		
119.8905	9.395537		88.27611		
102.8637	20.61371		424.925		
98.83888	37.91518		1437.561		
39.52547	16.42087		269.6449		
7.852932	-0.89853		0.807353		
516.4407	1.610689		2.594317		
22.02322	-4.71998		22.27817		
9.339256	-0.83734		0.701144		
23.20467	-12.038		144.9143		

Estimates Calculated using DA SLP LSF DRD

Output	Error	Mean Erro	Square Err	MSE
39.35864	9.326642	0.626506	86.98625	781.8229
12.3024	-4.2226		17.83035	
213.6692	36.60424		1339.87	
54.6583	-37.1987		1383.743	
49.83062	11.76052		138.3098	
51.61981	13.23651		175.2052	
170.2293	-94.9357		9012.793	
249.2569	-9.33608		87.1624	
10.96856	5.56629		30.98359	
42.74301	13.76471		189.4671	
120.7002	10.20519		104.1459	
102.749	20.49895		420.2071	
98.16013	37.23643		1386.552	
39.8445	16.7399		280.2241	
7.700816	-1.05064		1.103853	
517.8673	3.037269		9.225003	
21.03439	-5.70881		32.59046	
9.088096	-1.0885		1.18484	
22.71074	-12.532		157.0501	

RMSE	Estimates Calculated using DA only Output	Error	Mean Erro	Square Err	MSE	RMSE
27.9611	29.53867	-0.49333	-1.76325	0.243371	856.2264	29.26135
	8.326884	-8.19812		67.20911		
	192.468	15.40301		237.2527		
	66.7714	-25.0856		629.2874		
	37.5083	-0.5618		0.31562		
	45.43836	7.055062		49.77391		
	182.2753	-82.8897		6870.701		
	233.5742	-25.0188		625.9384		
	6.921541	1.519271		2.308183		
	26.45514	-2.52316		6.366327		
	137.8109	27.31588		746.1573		
	95.71849	13.46849		181.4001		
	140.3183	79.39462		6303.506		
	27.41988	4.315279		18.62164		
	4.331248	-4.42021		19.53827		
	517.7652	2.935242		8.615647		
	13.84186	-12.9013		166.4445		
	4.864363	-5.31224		28.21986		
	17.73832	-17.5044		306.4032		

Southwest

	STATION	Q(t) 2.00	DA km^2	M2 (%)	ACLS (-)	LSF	DENSITY km^-1	FACTOR (-)
1	02yj001	201.204	640.0	0.35	0.75	1.67	1.12	1.81
2	02za001	107.019	343.0	0.68	0.83	1.78	1.04	2.45
3	02za002	36.6873	72.0	2.19	0.43	1.39	1.15	1.72
4	02za003	98.402	139.0	1.46	0.73	1.66	1.46	1.68
5	02zb001	172.434	205.0	1.27	0.60	1.52	0.72	2.09
6	02zc002	224.259	230.0	1.08	0.34	1.30	0.96	1.84

Estimates Calculated using DA only

Output	Error	Mean Erro	Square Err	MSE	RMSE
208.5195	7.315524	0.448753	53.51689	2475.8597496	49.75801
162.5786	55.55956		3086.865		
87.20805	50.52075		2552.346		
113.3819	14.97992		224.3981		
132.3947	-40.0393		1603.143		
138.615	-85.644		7334.89		

Estimates Calculated using DA SLP

Output	Error	Mean Erro	Square Err	MSE	RMSE
180.45	-20.7536	0.325199	430.7117	1159.311	34.04866
167.891	60.87215		3705.418		
39.7003	3.013001		9.078177		
90.1577	-8.24428		67.96816		
189.134	16.70046		278.9055		
174.622	-49.6365		2463.786		

Estimates Calculated using DA SLP FACLS

Output	Error	Mean Erro	Square Err	MSE	RMSE
188.5159	-12.6881	-0.47748	160.9879	343.067	18.52207
139.3429	32.32388		1044.834		
47.1876	10.5003		110.2564		
73.81365	-24.5883		604.5868		
161.074	-11.36		129.05		
227.2064	2.947404		8.687189		

Estimates Calculated using DA SLP LSF

Output	Error	Mean Erro	Square Err	MSE	RMSE
193.69390055	-7.5101	0.298883	56.40159	243.7508	15.61252
130.48710161	23.4681		550.7518		
48.994624247	12.30732		151.4702		
72.899647083	-25.5024		650.37		
166.79936628	-5.63463		31.7491		
228.92396123	4.664961		21.76186		

Estimates Calculated using DA SLP SHP

Output	Error	Mean Erro	Square Err	MSE
204.1245	2.920531	-0.17804	8.529502	57.30543
100.7164	-6.30258		39.72251	
31.90511	-4.78219		22.86937	
104.9305	6.528518		42.62154	
183.4399	11.0059		121.1298	
213.8206	-10.4384		108.9599	

RMSE
7.570035