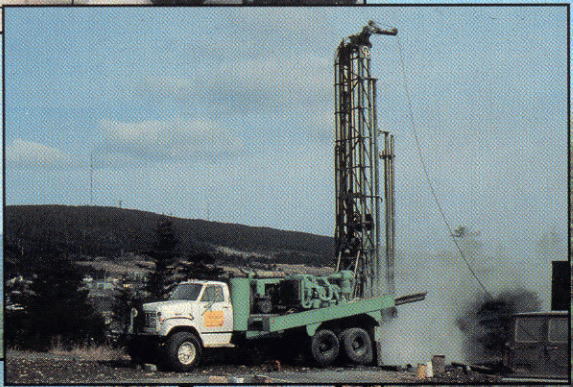
Regional Flood Frequency Analysis for the Island of Newfoundland



GOVERNMENT OF
NEWFOUNDLAND
AND LABRADOR

Department of Environment
Water Resources Division

Regional Flood Frequency Analysis for the Island of Newfoundland

Ken Rollings, P. Eng.

Hydrologic Modelling Section
Water Resources Management Division
Department of Environment and Labour
Government of Newfoundland and Labrador
St. John's, Newfoundland, Canada

December 1999

This page has been intentionally left blank.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The annual cost of flooding to public property in Newfoundland is estimated to be in the hundreds of thousands of dollars range. Accurate flood flow estimates are needed for the efficient design of instream structures (culverts, bridges, spillways, etc.) and for floodplain management. The locations for which flood flow estimates are required usually do not have streamflow data which could be used to directly estimate the flood flows. This study, like three previous studies (1971, 1984, 1990), will derive a set of equations for estimating return period flood flows in ungauged watersheds.

The following methodology was used: the flood, climatic and physiographic characteristics were examined; a database of flood flows was created, missing data were estimated, and the flood series were screened; a single station frequency analysis was conducted on each flood series; the best estimates of the 2-year to the 200-year flood flows were selected; hydrologically homogeneous regions were formulated based on watershed characteristics and other criteria; the physiographic database was updated and a new physiographic parameter was formulated; equations were formulated which predicted flood flows based on physiographic parameters; the equations were tested using an independent data set and the results were compared to previous studies. In addition, a note was provided on the application of the equations, the methodology and results were discussed, conclusions were drawn, and recommendations were made.

On the island of Newfoundland, floods can be characterized by their magnitude, timing and causative factors. Flood flows per unit area are highest in the South-west Region and lowest in the northern regions. On the Avalon and Burin Peninsulas, floods most frequently occur in February. In Central Newfoundland and Western Newfoundland floods most frequently occur in April and May respectively. On the South Coast (western half) floods most frequently occur in November. Floods in February are usually caused by rainfall and snowmelt. Typically, rainfall accompanied by above zero temperatures, for one or more days, combine to melt some of the snowpack which results in high flows. Floods in April

and May are the result of rainfall and snowmelt. Typically, rainfall on a snowpack, ripened by sustained above-zero temperatures, combine to produce high flows. Generally, the bulk of the snowpack disappears during this event. Floods in November are usually caused by heavy rainfall.

The most important climatic parameter in the magnitude and timing of floods is precipitation. The amount, form, and time distribution of precipitation are important as well as maximum rainfall intensities and maximum snow depths. Annual and monthly precipitation is highest along the south coast and decreases in a northerly direction. For the northern half of the island, monthly rainfall is highest in August. For the southern half of the island, monthly rainfall is highest in October or November. The heaviest rainfall intensities are experienced along the south coast. Snowmelt plays an important part in generating floods in Central and Western Newfoundland. Annual snowfall in this region is between 200 cm and 400 cm with a few local highs near 450 cm. Annual snowfall is generally less than 200 cm on the Avalon and Burin Peninsulas. Maximum snow depths occur near the end of March. Maximum snow depths increase with latitude.

Watershed characteristics such as: drainage area, the amount and location of natural storage, watershed slope, watershed shape, soils, vegetation, land use, etc., influence the flood response of drainage basins to hydrometeorological inputs. Watershed characteristics are sometimes difficult to quantify and thus are represented by a number of physiographic parameters. Physiographic parameters are measures of watershed characteristics which are relatively easy to extract from topographic maps.

A flood database was created and missing data were estimated. Data review consisted of: a review of the stability of the gauging datum and an examination of the stage discharge curve at high flows, a review of the history of physical changes (eg. fires, diversions, changes to channel geometry) to the watershed, and statistical testing of the annual maximum instantaneous discharge data.

Single station flood frequency analyses were conducted on 79 watersheds using the Generalized Extreme Value (GEV), and the Three Parameter Log-normal (LN3) distributions. The choice between the GEV or the LN3 distribution was based on the mean absolute deviation between the theoretical and empirical probabilities of the upper half of the data set. It quantifies the error associated with that portion of the curve with which we have the most interest: the 2-year return period up to T = 200. The LN3 distribution was the better fitting distribution on 42 (68%) of the 65 watersheds. About 25% of all watersheds with 10 or more years of record did not provide good estimates of return period flows due to the shortness of streamflow records. These watersheds had an upper 95% confidence limit on the 1:100 year return period which was double the estimate. Generally, the confidence in the estimate for long return periods is low when the sample size is small. It is clear that return period flows should not be used for some stations with small sample sizes.

The island was divided into 4 hydrologically homogeneous regions. The division was based on: previous studies, the availability of reliable data, regional flood characteristics, regional precipitation characteristics, regional physiographic characteristics, and the results of regression analyses on test regions. Fifty stations (50) were used for regional analysis and the remaining stations were used for testing. More than 100 test regressions were conducted and evaluated. Watersheds with large drainage areas (>1000 km^2) were readily dropped from the analyses if their residuals, leverage or influence was high. Watersheds with drainage areas less than 100 km^2 were retained for analysis unless their residuals indicated that they were an outlier. Leverage and influence considerations were relaxed slightly for these watersheds so that the regression equations would be applicable to watersheds in this range of drainage areas.

Physiographic data was compiled for the additional 26 watersheds used in this study. A new Lake Attenuation Factor (LAF) parameter was introduced. This parameter takes into account the size and the relative size of the area drained by large (>1% drainage area) lakes in a watershed.

Equations were developed for each of the hydrologically homogeneous regions which provided predictions of return period flood estimates based on physiographic data. A forward stepwise regression was performed. The coefficients and variables in the final regional equations were selected based on the following criteria: the coefficient of correlation between the dependent and independent variables had to be significantly high, the standard error of the estimate had to be a minimum, the final predictor variables had to be independent of each other, entry into the regression equation had to be significant at a 5% level using the F-ratio, the number of physiographic parameters in the regression equations had to be minimal.

The regression equations derived in this study cannot be used on all watersheds. Many ungauged watersheds have physiographic parameters which are outside the range of physiographic parameters which were used in the development of the regression equations. Extrapolation of results beyond the extremes of physiographic parameters used in the development of the regression equation is not generally recommended. The number of watersheds to which the regression equations may be applied has increased. This study had more watersheds available for analysis and thus a broader spectrum of physiography. The minimum drainage area in the North-west Region decreased from 93.2 km^2 to 33.5 km^2. The minimum FACLS was reduced in the other regions.

Drainage area (DA) was by far the most important physiographic parameter in all regions except the South-west Region where DA was forced to be the most important parameter. The (Lake Attenuation Factor) LAF parameter was selected as the second parameter in all regions except the South-west Region where the (Lakes and Swamps Factor) LSF was selected as the second parameter. The squared multiple R (SMR) statistic was in the 91-97% range in all regions except the South-west Region where SMR was in the 83-92% range. The standard error of the estimate (SEE) was in the 0.09-0.16 range in all regions except the South-west Region where SEE was in the 0.14-0.24 range. These statistics apply to log (to base 10) transformed data. Other parameter(s) could have been included in some of the regions, and at some of the return periods, based on the F statistic. It was decided to stop at this level for several reasons: to minimize the number of parameters in the equation, SEE was considered low, SMR was considered high, there was a negligible increase in the SMR with additional parameters, there was a negligible decrease in SEE with additional parameters, the chances of spurious correlation increase with an increase in the number of independent variables in the regression equations, because F statistics for significant third parameters while greater than 4.0 were much less than the F statistic for the first and second parameters, and to achieve some consistency across regions and return periods. After the inclusion of two parameters in the regional regression equations, nearly all of the variation in flood flows was explained.

Previous studies attempted to maximize the correlation and minimize the error by including all statistically significant variables in the regional regression equations. The result was, near-perfect correlations and very small errors associated with the flood estimates. The near-perfect correlations and small errors are unrealistic given the complexities of statistically modelling floods, and given the error associated with flood values. Previous studies may have included too many parameters. Since accurate estimates can be obtained with only 2 parameters, and because the parameters are consistent through all return periods in 3 of 4 regions, the regression equations developed in this study must be considered more robust. The equations for the NW, NE and SE Regions were of the form:

$$Q_T = C \times (DA)^{a1} \times (LAF)^{a2}$$

Where: Q_T is the return period flood estimate,
C, a1 and a2, are constants which depend on the desired return period,
DA is the drainage area of the watershed, and
LAF is a Lake Attenuation Factor.

The equation for the SW Region was of the form:

$$Q_T = C \times (DA)^{a1} \times (LSF)^{a2}$$

LSF is a Lakes and Swamps Factor.

The standard error of the estimate in the South-west Region was much higher than the error in the other regions. In addition, the LSF was sensitive to abstraction errors. An "Upper Envelope Curve" was developed which looked at only those watersheds which had high peak flows per unit area. This curve, while biased towards higher flood flows, had less error and an improved correlation. Within the applicable drainage area range, these floods represent the highest in magnitude on the island. The regression equation for the "Upper Envelope Curve" was of the form:

$$Q_T = C \times (DA)^{a1}$$

The accuracy of the regression equations was assessed using the data set that produced them and using an independent data set. The independent data set consisted of peak flow series that were not used in the formulation of the regional regression equations. The median absolute percentage difference between the frequency analysis estimates and the regression equation estimates for all stations used in the regional analysis ranged from 8.0% to 30.9%. The median absolute percentage difference between the flood estimates and the regression equation estimates for stations not used in the regional analysis ranged from 20.8% to 50.6%. The higher percentages were expected as the independent data set was composed primarily of stations which were removed from the main data set because of the low confidence in their flood estimates. A flood index method had to be used on some watersheds in the independent data set to obtain realistic flood estimates at high return periods. The LN3 and GEV distributions, which were used to model the high flows, provided unrealistic curve extensions at high return periods on some watersheds which had few data. The flood index method can provide a useful check at long return periods on watersheds with less than 15 years of data.

The regression equations provided in this study can provide two estimates of return period flood flows on ungauged watersheds. It is advisable to use several methods to estimate design floods. Previous regional flood frequency analysis can be used as checks. In addition, single station frequency analysis on nearby hydrologically similar watersheds may be useful. Regional flood index estimates can also be used as a check. Flood flows can also be simulated using computer intensive techniques on an event basis for a given design storm, or on a continuous basis from climatic data.

A *Users' Guide and Electronic Spreadsheet* for this regional flood frequency analysis is available under a separate cover.

The main recommendations were:

1.) The regional regression equations developed in this study are recommended for estimating return period flood flows on ungauged watersheds or on watersheds with less than 10 to 20 years of flood data.
2.) In the South-west Region, the "Upper Envelope Curve" is recommended where flooding may threaten life or cause severe flood damages.
3.) It is recommended that the regional flood frequency analysis be updated in 5 years.
4.) More streamflow gauges are required along the south coast from Isle aux Morts River to Bay du Nord River.
5.) There is a need for a separate model for floods on small (< 50 km^2) watersheds.

This page has been intentionally left blank.

1. INTRODUCTION

1.1 Background

Flood flow estimates are required in Newfoundland rivers for the design of instream structures (culverts, bridges, spillways, etc.) and for floodplain management (floodway zoning). The locations for which flood flow estimates are required usually do not have streamflow data which could be used to directly estimate the flood flows. Because it is impractical to gauge all river reaches in the province for peak flow, regional flood frequency analyses have been employed to provide estimates of peak flow where no gauge exists. Regional Flood Frequency Analyses are essential to flood flow estimation on ungauged rivers. This study, like three previous studies (1971, 1984, 1990), will derive a set of equations for estimating return period flood flows in ungauged watersheds.

Regular updates of the Regional Flood Frequency Analysis have been recommended in previous studies and are justified based on the expanded hydrometric and physiographic databases: more watersheds are available for analysis, periods of record are longer, and the range of physiographic parameter values has increased.

The annual cost of flooding to public property in the province is estimated to be in the hundreds of thousands of dollars range. Accurate flood estimation using regional flood frequency analysis will allow for the efficient design of instream structures by minimizing capital and flood damage costs. This is particularly relevant since many of the instream structures are constructed and repaired by the province.

Apart from the monetary losses to the province and individuals, homes have been washed away, personal possessions have been lost, and there has been loss of life. Floodway zoning, which is based on flood flows estimates, helps to minimize the monetary and non-monetary costs of flooding.

1.2 Objective

The objective of this study was to develop a set of equations to estimate return period flood flows in ungauged watersheds.

1.3 Methodology

The following methodology was used:

1. The characteristics of floods were examined along with the climatic considerations and the physiographic influences.
2. A database of flood flows was created, missing data were estimated, and the flood series were statistically and hydraulically screened.
3. A single station frequency analysis was conducted on each flood series using the Generalized Extreme Value and the Three Parameter Log-normal probability distributions.
4. The best estimates of the 2-, 5-, 10-, 20-, 50-, 100-, and 200-year flood flows were selected.
5. Regions were formulated based on flood, climatic and physiographic characteristics as well as previous studies, the availability of flood data and the results of regression analysis on test regions.
6. The physiographic database was updated and a new physiographic parameter was formulated: Lake Attenuation Factor (LAF).
7. Mathematical equations were formulated so that return period flood flows could be estimated on ungauged watersheds.
8. The equations for predicting return period flood flows were tested using an independent data set.
9. The results of this study were compared to previous studies.

10. A note was provided on the application of the regional flood equations and on flood flow estimation in general.
11. The methodology and the results of the study were discussed, conclusions were drawn, and recommendations were made.

This page has been intentionally left blank.

2. FLOODS IN NEWFOUNDLAND

2.1 Definition of a Flood

A flood is defined as the highest instantaneous river discharge in a year. On the island of Newfoundland, floods are caused by rainfall, snowmelt, or a combination of rainfall and snowmelt. For the purposes of this study, floods exclude high flows caused by ice jams, dam breaks, tidal surges, reservoir operations, diversions, debris jams, etc. Only natural river flows were considered.

2.2 Characteristics of Floods

On the island of Newfoundland, floods can be characterized by their magnitude, timing and causative factors.

The magnitude of floods are not easily comparable between watersheds since they are highly dependent on drainage area. This relationship, however, is nonlinear. In the 1990 study, power coefficients in the range of 0.627 to 0.999 have been estimated for regressions involving return period flows and drainage areas only. A power coefficient of 0.8 was selected to apply to the drainage area before it was divided into the flood magnitude. This procedure enabled regional comparison of flood magnitudes. Later, the power coefficient for drainage area was dropped since it did not improve our understanding of regional differences in flood magnitude and because flood magnitudes per unit area had an easily understood physical meaning. For most applications, return period flows longer than the mean annual flood are required. The 1:20 year return period flows (calculated in the next chapter) were selected to describe the magnitude of return period flows since this return period struck the right balance between accurate estimates and return periods which are commonly used in hydrological design (1:10, 1:25 and 1:100). The magnitudes of the 1:20 year return period flows per unit area are shown in Figure 2.1. The 1:20 year return period flows per unit area were highest in the south-west.

The modal months of all recorded floods by tertiary watersheds are shown in Figure 2.2. Four natural regions can be identified: Avalon and Burin Peninsulas, Central Newfoundland, Western Newfoundland, and the South Coast (western half). On the Avalon and Burin Peninsulas floods most frequently occur in February. In Central Newfoundland (from the south coast to the north coast) floods most frequently occur in April. In Western Newfoundland (includes Northern Peninsula and excludes south coast) floods most frequently occur in May. On the South Coast (western half) floods most frequently occur in November. The number of regions could be reduced to three by combining the Central and Western Regions. Floods can and do occur in any month of the year.

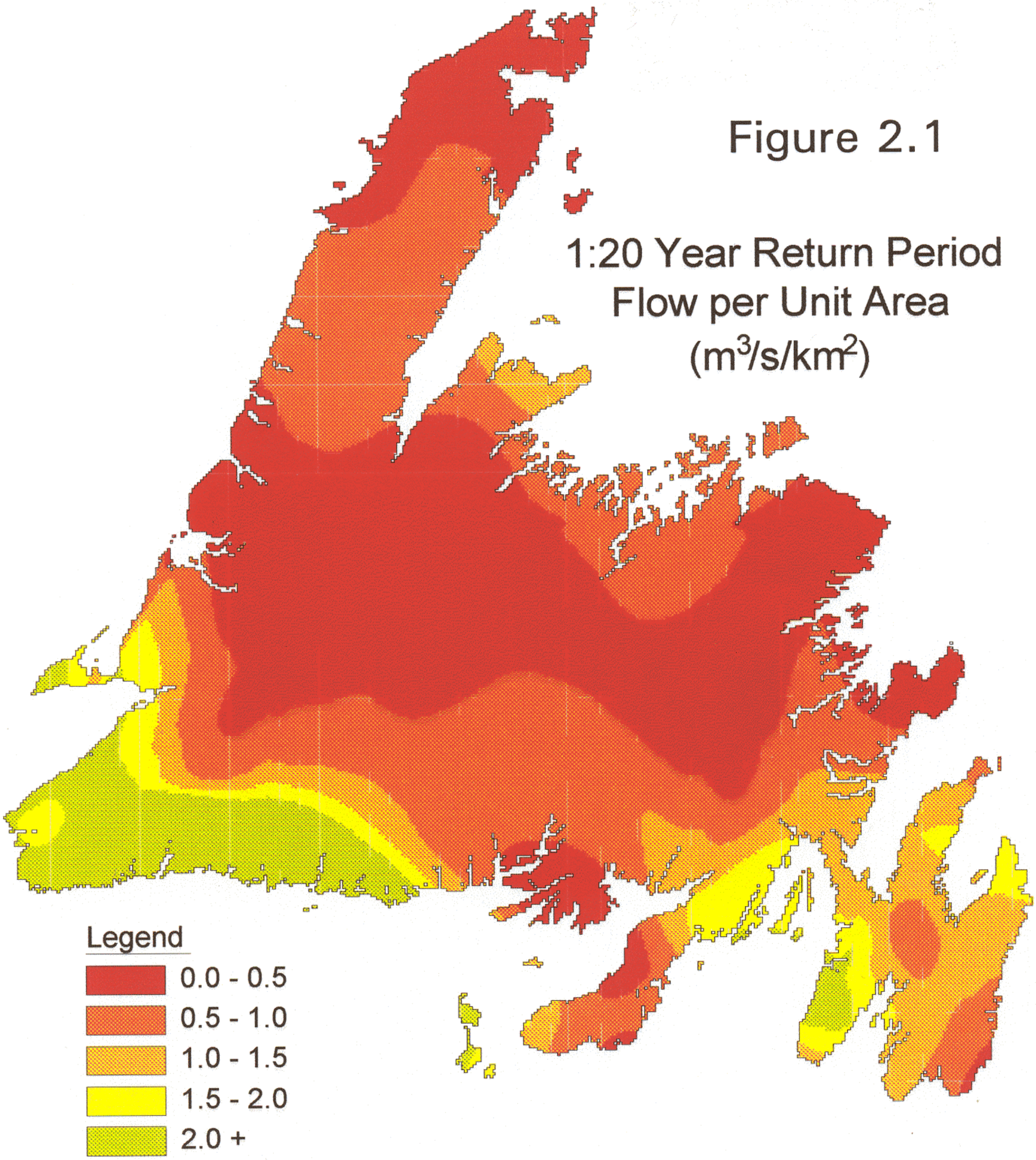
Floods in February are usually caused by rainfall and snowmelt. Typically, rainfall accompanied by above zero temperatures, for one or more days, combine to melt some of the snowpack which results in high flows. Floods in April and May are the result of rainfall and snowmelt. Typically, rainfall on a snowpack, ripened by sustained above-zero temperatures, combine to produce high flows. Generally, the bulk of the snowpack disappears during this event. Floods in November are usually caused by heavy rainfall.

2.3 Climatic Considerations

Climate plays an important role in the magnitude and timing of floods. The most important climatic parameter in the magnitude and timing of floods is precipitation. The amount, form, and time distribution of precipitation are important as well as maximum rainfall intensities and maximum snow depths.

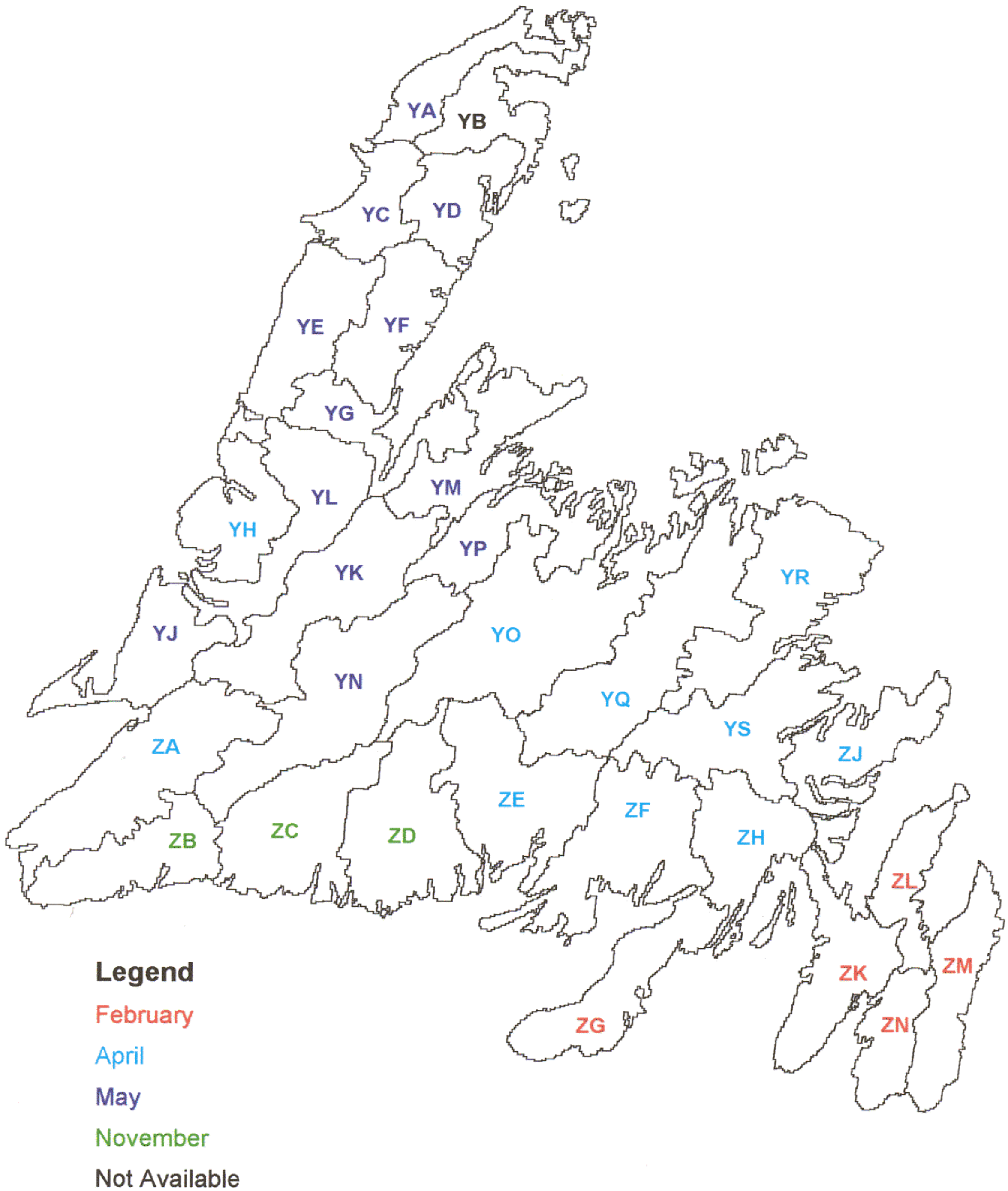
The mean annual total precipitation for the island is shown in Figure 2.3 along with the monthly distribution of rainfall and snowfall at select stations. Precipitation is highest along the south coast and decreases in a northerly direction. Most of the island's

Figure 2.1


1:20 Year Return Period
Flow per Unit Area
(m3/s/km2)
Legend
0.0 - 0.5
0.5 - 1.0
1.0 - 1.5
1.5 - 2.0
2.0 +

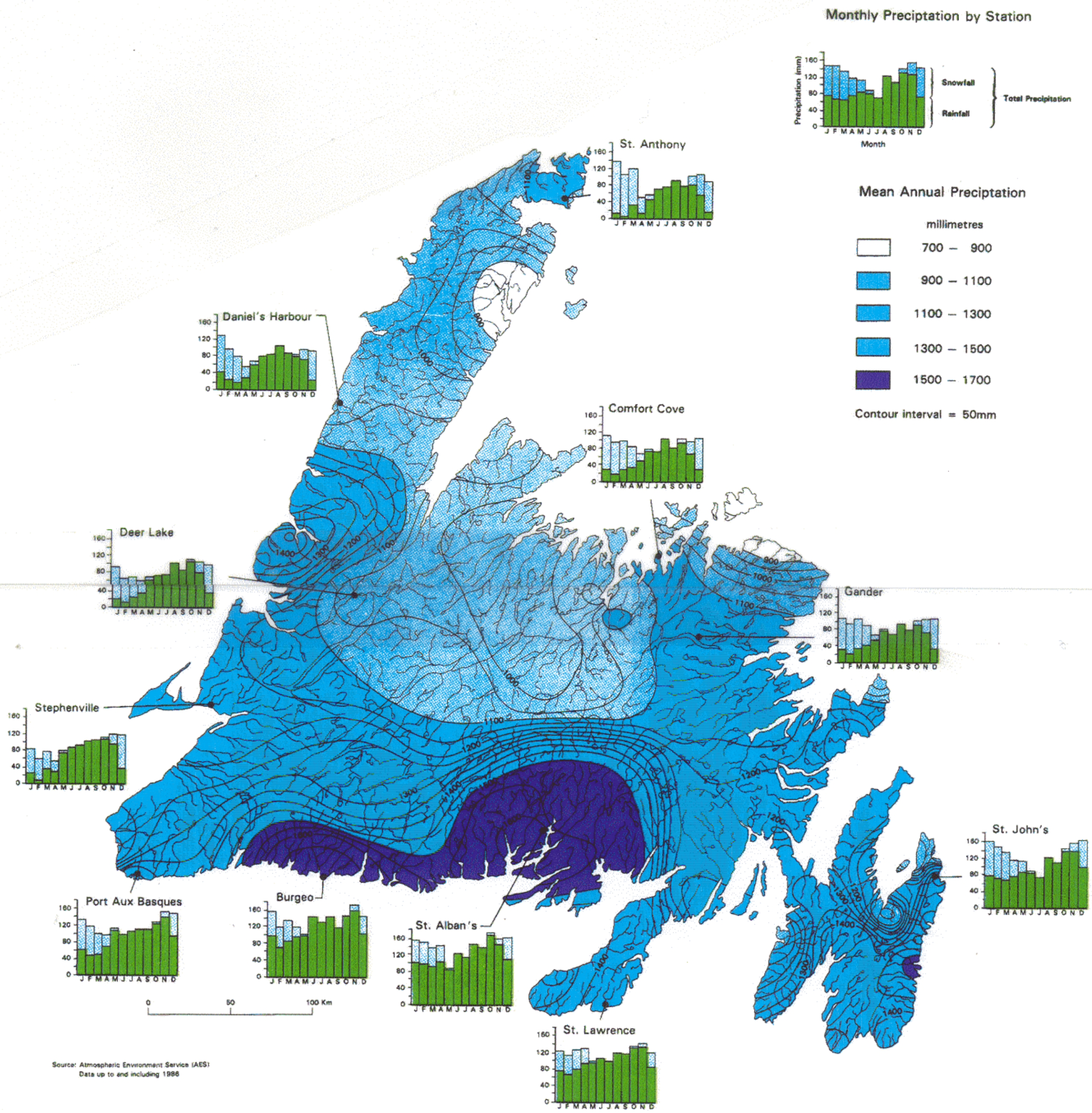
This page has been intentionally left blank.

Figure 2.2 - Modal Month of Recorded Floods by Tertiary Watershed



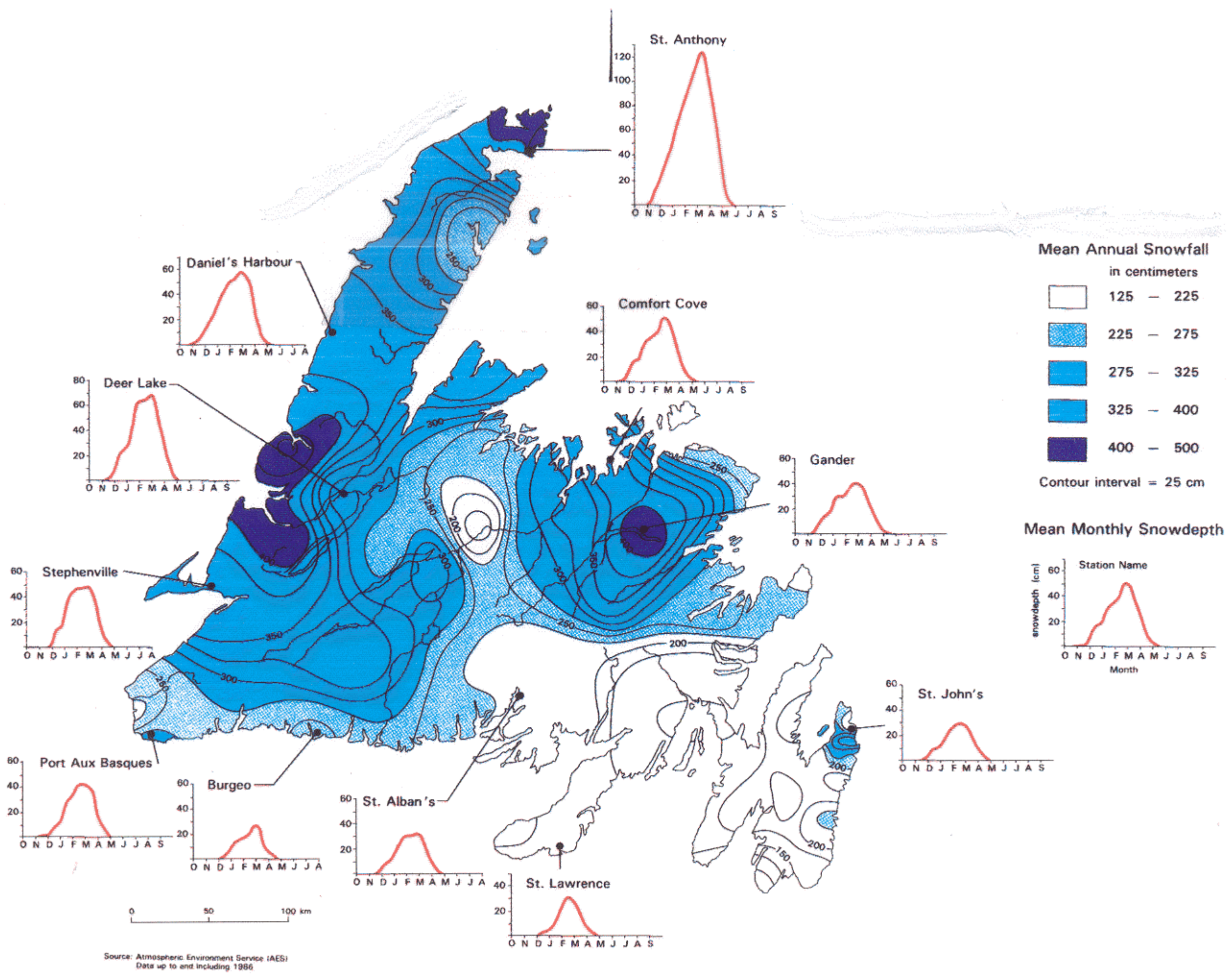
This page has been intentionally left blank.

Figure 2.3 Mean Annual and Monthly Distribution of Precipitation



This page has been intentionally left blank.

Figure 2.4 Mean Annual Snowfall and Mean Monthly Snow Depth



This page has been intentionally left blank.

precipitation is of the cyclonic type. Low pressure systems generally approach from the south-west. This accounts for the higher precipitation along the south and west coasts. An orographic effect is evident. Several precipitation regions could be delineated: South Coast - Burin Peninsula (1300 mm to 1600 mm), Avalon Peninsula (1100 mm to 1500 mm), Eastern Newfoundland (1100 mm to 1200 mm), Central Newfoundland (1000 mm to 1100 mm), Northern Peninsula (900 mm to 1100 mm), and Western Newfoundland (1100 mm to 1200 mm). The monthly distribution of total precipitation shows slightly less precipitation in late spring - early summer. For the northern half of the island, monthly rainfall is highest in August. For the southern half of the island, monthly rainfall is highest in October or November.

The 1:50 year return period 24 hour rainfall amounts, shown in Table 2.1, are sorted by latitude. The higher values along the south coast (lower latitudes) show why this region experiences floods due to rainfall. The 24-hour period was used since floods on medium sized watersheds in Newfoundland are associated with rainfalls near this duration. The 1:50 year return period was selected because it provided reliable estimates of heavy rainfalls.

Mean annual snowfall and mean monthly snow depth is shown in Figure 2.4. Snowmelt plays an important part in generating floods in Central and Western Newfoundland. Annual snowfall in this region is between 200 cm and 400 cm with a few local highs near 450 cm. Annual snowfall is generally less than 200 cm on the Avalon and Burin Peninsulas. Maximum snow depths occur near the end of March. Maximum snow depths increase with latitude. Mean maximum snow depths vary from a high of over 120 cm at St. Anthony in the north, to the 40 - 70 cm range at the mid latitudes, to the 30 - 40 cm range along most of the south coast.

2.4 Physiographic Influences

Watershed characteristics such as: drainage area, the amount and location of natural storage, watershed slope, watershed shape, soils, vegetation, land use, etc., influence the

Table 2.1 1:50 Year Return Period 24 Hour Rainfall Amounts

ID #	Station	Latitude	1:50 Year Rainfall Amount (mm)	n
8403401	St. Anthony	51º 22'	87.7	19
8401400	Daniel’s Harbour	50º 14'	122.9	22
8401259	Comfort Cove	49º 16'	81.4	23
8401501	Deer Lake Airport	49º 13'	68.7	25
8401700	Gander Int’l Airport	48º 57'	87.1	48
8403800	Stephenville Airport	48º 32'	109.7	23
8403290	St. Alban’s	47º 52'	166.2	14
8403506	St. John’s Airport	47º 37'	105.2	31
8400798	Burgeo	47º 37'	118.5	23
8402975	Port aux Basques	47º 34'	133.1	16
8403615	St. Lawrence	46º 55'	116.4	22

flood response of drainage basins to hydrometeorological inputs. Watershed characteristics are sometimes difficult to quantify and thus are represented by a number of physiographic parameters. Physiographic parameters are measures of watershed characteristics which are relatively easy to extract from topographic maps. Physiographic data was compiled for the gauged watersheds used in this study. The database is provided in section 4. In addition a narrative is provided on the influence of some of the more important parameters on flood flows. Details on physiographic parameter abstraction are provided in Appendix A. Single station frequency analyses of floods on gauged watersheds were used with their physiographic parameters to arrive at regional equations for flood estimation on ungauged watersheds.

3. SINGLE STATION FLOOD FREQUENCY ANALYSIS

3.1 Data Base

Initially, the data base for single station flood frequency analysis consisted of all flood data to 1995 for 114 gauged watersheds on the island of Newfoundland as listed in Environment Canada's HYDAT compact disc version 4.95. Preliminary data for 1996 was acquired directly from Environment Canada. Regulated watersheds were removed from the database as well as watersheds with less than 7 years of peak flow data. In addition, watersheds which drain urban areas were removed. There were 18 regulated watersheds, 9 watersheds with less than 7 years of record, and 8 watersheds which drain urban areas. Nine (9) watersheds which had 7 to 9 years of record were retained for testing regional regression equations. The remaining 70 watersheds which had 10 or more years of record were subjected to data review and statistical screening. Five (5) watersheds were removed from the data base as a result. In addition, portions of the records on several of the remaining 65 watersheds were deleted. The 70 watersheds which were considered for regional analysis are listed in Table 3.1. The locations of the hydrometric stations are shown in Figure 3.1.

3.2 Estimation of Missing Data

Missing annual maximum instantaneous discharge data were estimated from the available annual maximum daily discharge data using regression analysis techniques. A linear and a non-linear regression between the annual maximum instantaneous discharge data and the annual maximum daily discharge data was performed. Visual outliers were discarded. For the linear regression, the y-intercept was computed and was also forced through the origin. For the non-linear regression, the data were transposed by taking logarithms to base 10. Agreement was sought between the estimates calculated from the regression equations. One estimate was selected.

Peak flow series on watersheds which short periods of record were not extended using correlation analysis as had been the case in a previous regional flood frequency analysis (1990). Correlation analysis results in a peak flow series which has a lower variance than if the series were composed entirely of natural flows. In the 1990 analysis, 11 watersheds had their peak flow series artificially lengthened. Correlation coefficients ranged between 0.59 and 0.85. Peak flow records at least doubled in length for most watersheds. In the extreme case, one watershed had the length of its peak flow series increase by more than 5 times its natural record length.

3.3 Data Review

Data review consisted of: a review of the stability of the gauging datum and an examination of the stage discharge curve at high flows, a review of the history of physical changes (eg. fires, diversions, changes to channel geometry) to the watershed, and statistical testing of the annual maximum instantaneous discharge data. A station by station review is summarized in Appendix B.

A review of the stability of the gauging datum and an examination of the stage discharge curve, especially at high flows, was performed on 21 watersheds in a previous regional flood frequency analysis (1984). Modification to peak flow series in that study were included in the present study except as indicated in the station by station review. For the remaining watersheds the annual maximum instantaneous discharges as published/provided by Environment Canada were used.

The physical changes which have occurred in Newfoundland watersheds which had an effect on peak flows included fires which burnt a substantial portion of the watershed, diversions where a significant portion of the watershed has been diverted to or imported from another watershed, and outlet channels which have been excavated. Information on the

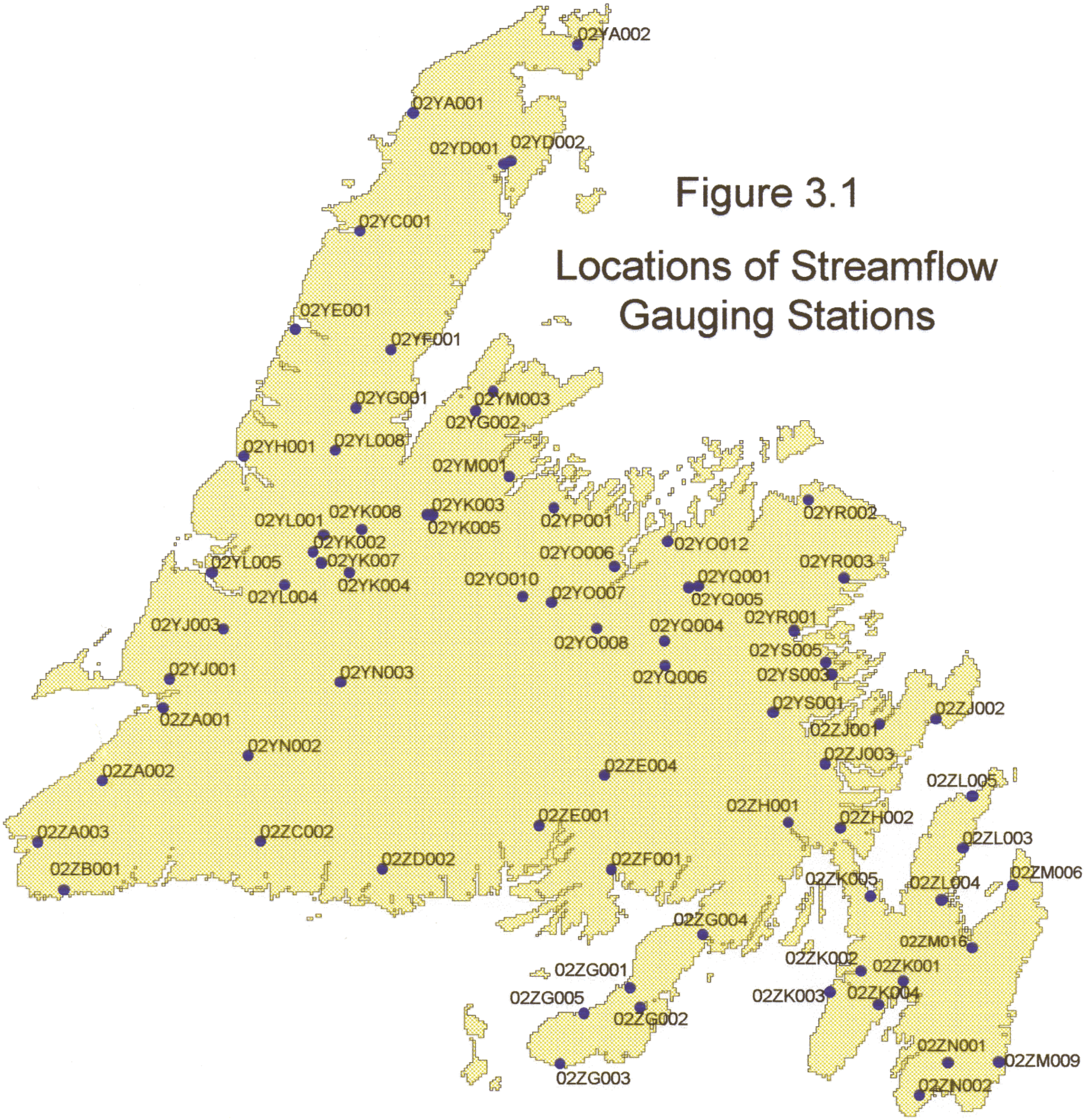


Figure 3.1

Locations of Streamflow Gauging Stations

Table 3.1 Data Base for Frequency Analysis

Station Number	Station Name	Area (km^2)	Start Year	Finish Year	Sample Size
02YA001	Ste. Genevieve River near Foresters Point	306	1970	1996	27
02YA002	Bartletts River near St. Anthony	33.6	1986	1996	11
02YC001	Torrent River at Bristols Pool	624	1959	1996	38
02YD001	Beaver Brook near Roddickton	237	1960	1978	19
02YD002	Northeast Brook near Roddickton	200	1980	1996	17
02YE001	Greavett Brook above Portland Creek Pond	95.7	1980	1996	17
02YF001	Cat Arm River above Great Cat Arm	611	1969	1982	14
02YG001	Main River at Paradise Pool	627	1986	1996	11
02YG002	Middle Arm Brook below Flat Water Pond	224	1987	1996	10
02YH001	Bottom Creek near Rocky Harbour	33.4	1985	1996	12
02YJ001	Harrys River below Highway Bridge	640	1969	1996	28
02YJ003	Pinchgut Brook at outlet of Pinchgut Lake	119	1986	1996	11
02YK002	Lewaseechjeech Brook at Little Grand Lake	470	1953	1996	35
02YK003	Sheffield River at Sheffield Lake	362	1956	1966	11
02YK004	Hinds Brook near Grand Lake	529	1957	1979	23
02YK005	Sheffield Brook near Trans-Canada Highway	391	1973	1996	24
02YK007	Glide Brook below Glide Lake	112	1984	1996	13
02YK008	Boot Brook at Trans-Canada Highway	20.4	1985	1996	12
02YL001	Upper Humber River near Reidville	2110	1929	1996	68
02YL004	South Brook at Pasadena	58.5	1983	1996	14
02YL005	Rattler Brook near McIvers	17.0	1985	1996	12
02YM001	Indian Brook at Indian Falls	974	1955	1996	41
02YM002	Indian Brook Diversion to Birchy Lake	?	1964	1978	15
02YM003	Southwest Brook near Baie Verte	93.2	1980	1996	17
02YN002	Lloyds River below King George IV Lake	469	1981	1996	16
02YO006	Peters River near Botwood	177	1981	1996	16
02YO007	Leech Brook near Grand Falls	88.3	1984	1995	12
02YO008	Great Rattling Brook above Tote River Confluence	823	1984	1996	13
02YO010	Junction Brook near Badger	61.6	1985	1996	12
02YP001	Shoal Arm Brook near Badger Bay	63.8	1982	1996	15
02YQ001	Gander River at Big Chute	4400	1959	1996	47
02YQ002	Gander River at outlet of Gander Lake	4160	1924	1939	14
02YQ004	Northwest Gander River near Gander Lake	2150	1983	1996	14
02YQ005	Salmon River near Glenwood	80.8	1987	1996	10
02YR001	Middle Brook near Gambo	275	1959	1996	38
02YR002	Ragged Harbour River near Musgrave Harbour	399	1977	1996	20
02YR003	Indian Bay Brook near Northwest Arm	554	1981	1996	16
02YS001	Terra Nova River at Eight Mile Bridges	1290	1953	1984	31
02YS003	Southwest Brook at Terra Nova National Park	36.7	1968	1996	29
02YS005	Terra Nova River at Glovertown	2000	1985	1996	12
02ZA001	Little Barachois Brook near St. Georges	343	1979	1996	18
02ZA002	Highlands River at Trans-Canada Highway	72	1982	1996	15
02ZA003	Little Codroy River near Doyles	139	1982	1996	15
02ZB001	Isle aux Morts River below Highway Bridge	205	1962	1996	35
02ZC002	Grandy Brook below Top Pond Brook	230	1982	1996	15
02ZD002	Grey River near Grey River	1340	1970	1996	19
02ZE001	Salmon River at Long Pond	2640	1944	1965	21
02ZF001	Bay du Nord River at Big Falls	1170	1951	1996	45
02ZG001	Garnish River near Garnish	205	1959	1996	38
02ZG002	Tides Brook below Freshwater Pond	166	1977	1996	20
02ZG003	Salmonier River near Lamaline	115	1980	1996	17
02ZG004	Rattle Brook near Boat Harbour	42.7	1981	1996	16
02ZH001	Pipers Hole River at Mothers Brook	764	1953	1996	44
02ZH002	Come by Chance River near Goobies	43.3	1961	1996	28
02ZJ001	Southern Bay River near Southern Bay	67.4	1977	1996	20
02ZJ002	Salmon Cove River near Champney's	73.6	1983	1996	14
02ZJ003	Shoal Harbour River near Clarenville	106	1986	1996	11
02ZK001	Rocky River near Colinet	301	1949	1996	48
02ZK002	Northeast River near Placentia	89.6	1979	1996	18
02ZK003	Little Barachois River near Placentia	37.2	1983	1996	14
02ZK004	Little Salmonier River near North Harbour	104	1983	1996	14
02ZK005	Trout Brook near Bellevue	50.3	1986	1996	11
02ZL003	Spout Cove Brook near Spout Cove	10.8	1979	1996	18
02ZL004	Shearstown Brook near Shearstown	28.9	1983	1996	14
02ZL005	Big Brook at Lead Cove	11.2	1985	1996	12
02ZM006	Northeast Pond River at Northeast Pond	3.63	1954	1996	43
02ZM009	Seal Cove Brook near Cappahayden	53.6	1979	1996	18
02ZM016	South River near Holyrood	17.3	1983	1996	14
02ZN001	Northwest Brook at Northwest Pond	53.3	1966	1995	30
02ZN002	St. Shotts River near Trepassey	15.5	1985	1996	12

Note: Sample size may not coincide with the start and finish years due to missing data.

physical changes which have occurred in gauged watersheds was obtained from previous reports and personal communication from Environment Canada.

All data sets with 10 or more years of data underwent non-parametric statistical testing to ensure that the preconditions of frequency analysis are met: that the data are independent of one another, that there is no trend in the data, that the data are random, and that the data are drawn from a homogeneous population. These tests are contained in the computer program Consolidated Frequency Analysis 88 (CFA88) (Environment Canada, 1985). Outliers exert a stronger influence on the position of the regression line than other data. CFA88 detects outliers. Their treatment was as follows. Low outliers were removed and the plotting positions were maintained. High outliers were deleted from the data set when the skewness of the natural log transform was greater than 0.4 and when it resulted in a reduction of the skewness to less than 0.4. Details on the statistical testing and the evaluation of the test statistics can be found in Appendix C. Test results are tabulated in Table 3.2. Stations which had less than 10 values did not undergo statistical testing since a minimum of 10 values were recommended for these tests. All data sets were visually inspected for any obvious trends or changes in the peak flow series.

The 02YH001 data set had a relatively short period of record (1985-96). The data set was highly skewed. After removal of a high outlier, skewness was reduced to an acceptable level.

The 02YK002 data set failed all statistical tests. This data set had a relatively long period of record (1954-96). The following years were missing 1955, 1967-72, and 1981. Environment Canada indicated that the outlet channel of this watershed had been excavated in 1962, causing a decrease in the water level on Little Grand Lake. The time series of annual maximum instantaneous discharges showed an increase in the magnitude of average peak flows of nearly 70% for the 1973-96 period over the 1954-66 period. Only the 1973-80 and 1982-96 portions of the records (n=23) were retained.

Table 3.2 Results of Statistical Screening

Station Number	Sample Size	I ?	T ?	R ?	H ?	High Skew?	High Outliers?
02YA001	27	Y	N	Y	*	**Y**	N
02YA002	11	Y	N	Y	*	**Y**	N
02YC001	38	Y	N	Y	*	N	N
02YD001	19	Y	N	Y	*	N	N
02YD002	17	Y	N	Y	*	N	N
02YE001	17	Y	N	Y	*	N	N
02YF001	14	Y	N	Y	*	N	N
02YG001	11	Y	N	Y	*	N	N
02YG002	10	Y	N	Y	*	N	N
02YH001	12	Y	N	Y	*	**Y**	**Y**
02YJ001	28	Y	N	Y	*	N	N
02YJ003	11	Y	N	Y	*	N	N
02YK002	35	**N(1%)**	**Y(1%)**	**N(5%)**	**N(1%)**	N	N
02YK003	11	Y	N	Y	*	N	N
02YK004	23	Y	N	Y	**N(5%)**	N	N
02YK005	24	Y	N	Y	*	N	N
02YK007	13	Y	N	Y	*	N	N
02YK008	12	Y	N	Y	*	N	N
02YL001	68	Y	N	Y	*	N	N
02YL004	14	Y	N	Y	*	**Y**	**Y**
02YL005	12	Y	N	Y	*	**Y**	N
02YM001	41	Y	N	Y	Y	N	N
02YM002	15	Y	N	Y	*	N	N
02YM003	17	Y	N	Y	*	N	N
02YN002	16	Y	N	Y	*	N	N
02YO006	16	Y	N	Y	*	**Y**	**Y**
02YO007	12	Y	N	Y	*	**Y**	N
02YO008	13	Y	N	Y	*	N	N
02YO010	12	Y	N	Y	*	**Y**	N
02YP001	15	Y	**Y(5%)**	Y	*	**Y**	**Y**
02YQ001	47	Y	N	Y	*	N	N
02YQ002	14	Y	N	Y	*	N	N
02YQ004	14	Y	N	Y	*	N	N
02YQ005	10	Y	**Y(5%)**	Y	*	N	N
02YR001	38	Y	N	Y	*	N	N
02YR002	20	Y	N	Y	*	**Y**	N
02YR003	16	Y	N	Y	*	N	N
02YS001	31	Y	N	Y	*	**Y**	N
02YS003	29	Y	N	Y	*	N	N
02YS005	12	Y	N	Y	*	N	N
02ZA001	18	Y	N	Y	*	N	N
02ZA002	15	Y	N	Y	*	**Y**	N
02ZA003	15	Y	N	Y	*	N	N
02ZB001	35	Y	N	Y	*	N	N
02ZC002	15	Y	N	Y	*	N	N
02ZD002	19	Y	N	Y	*	N	N
02ZE001	21	Y	N	Y	*	N	N
02ZF001	45	Y	**Y(5%)**	Y	*	**Y**	**Y**
02ZG001	38	**N(5%)**	N	Y	*	N	N
02ZG002	20	Y	N	Y	*	N	N
02ZG003	17	Y	N	Y	*	N	N
02ZG004	16	Y	N	Y	*	N	N
02ZH001	44	Y	**Y(1%)**	Y	*	N	N
02ZH002	28	Y	N	Y	*	N	N
02ZJ001	20	Y	N	Y	*	N	N
02ZJ002	14	Y	N	Y	*	**Y**	**Y**
02ZJ003	11	Y	N	**N(5%)**	*	**Y**	N
02ZK001	48	Y	N	Y	*	N	N
02ZK002	18	Y	N	Y	*	N	N
02ZK003	14	Y	N	Y	*	N	N
02ZK004	14	Y	N	Y	*	**Y**	N
02ZK005	11	Y	N	Y	*	**Y**	N
02ZL003	18	Y	N	Y	*	N	N
02ZL004	14	Y	N	Y	*	N	N
02ZL005	12	Y	N	Y	*	**Y**	N
02ZM006	43	Y	N	Y	*	N	N
02ZM009	18	Y	N	Y	*	**Y**	**Y**
02ZM016	14	Y	N	Y	*	N	N
02ZN001	30	Y	**Y(5%)**	Y	*	N	N
02ZN002	12	Y	**Y(1%)**	Y	*	N	N

Notes: I ? - Independent ?, T ? - Trend ?, R ? - Random ?, H ? - Homogeneous ?, Y - yes, N - no, * - no testing performed,
Y(X%) - yes at an "X" % level of significance, N(X%) - no at an "X" % level of significance

The 02YK004 data set had 23 values and a period of record from 1957 to 1979. Environment Canada indicated that the outlet channel of this watershed had been excavated in 1962. The data set failed the split sample homogeneity test by time at 5% but not at 1%. Values were retained only for the years 1963-79 (n=17).

The 02YL004 data set had a relatively short period of record (1983-96, n=14). One high outlier was detected and removed (January 1983 flood). The high skewness was reduced to an acceptable level.

In 1963, a partial diversion of the 02YM001 watershed split the data set in two parts: 1955-63 (n=9) and 1964-96 (n=32). A visual check on the annual maximum instantaneous discharges did not reveal any change in peak flow rates. In addition, a split sample test by time did not find any significant difference in the populations. The first data set was too small for frequency analysis. The second data set includes an ungauged amount of water contributed from another basin. This data set was discarded.

The 02YM002 data set had a period of record from 1964 to 1978. This station metered the diverted part of 02YM001. This diversion, however, is only partial, and therefore, this data set was discarded.

The 02YO006 data set had a period of record from 1981 to 1996. One high outlier was removed (January 1983 flood) which reduced the skewness to an acceptable level.

The 02YP001 data set had a relatively short period of record (1982-96). A trend was detected at 5%. After removal of a high outlier (January 1983 flood), the trend disappeared as well as the high skewness.

The 02YQ002 data set is relatively old (1924-39). All annual maximum instantaneous discharges had to be estimated from the annual maximum daily discharges.

This gauge was moved further downstream to the 02YQ001 location in 1950. The 02YQ002 data set was discarded in favour of the 02YQ001 data set.

The 02YQ005 data set had a relatively short period of record (1987-96). A significant trend was detected at 5% but not at 1%. The trend was assumed to be by chance.

The 02ZF001 data set had a relatively long period of record (1951-96). The January 1983 flood was a high outlier and was removed. The data set failed the test for trend at 5%. In addition, the data set failed the split sample homogeneity test by time (1951-74, 1975-96) at 1%! This station was discarded because of the unexplained trend and non-homogeneity.

The 02ZG001 data set had a relatively long period of record (1959-96). This series failed the test for independence at 5%. The 1984 study eliminated the years 1959 to 1962 because of problems with the rating curve. In addition, the 1962 value was a significant outlier which could not be confirmed by surrounding stations. After these values were removed, skewness was reduced to acceptable level and the data set passed the test for independence.

The 02ZH001 data set had a relatively long period of record (1953-96). The data set showed a significant (increasing) trend at 1%! A fire in 1961 had a significant effect on peak flows. Since most of this basin had low scrub trees, the vegetation was reestablished a few years later. When the values for 1953 to 1964 were deleted from the data set, the trend disappeared.

The 02ZJ002 data set had short period of record (1983-96, n=14). Four (4) missing values including the highest on record were estimated. The data set displayed high skewness even after the removal of 2 high outliers. Considering the small sample size, the number of missing data, and the high skewness, this data set was discarded.

The 02ZJ003 data set had short period of record (1986-96, n=11). The data set displayed high skewness. The data set also failed the test for randomness at 5%. Since there was no known physical reason for this cycle, it was assumed to be by chance. A certain number of test failures can be expected given the level of significance and the number of tests performed.

After removal of a high outlier in the 02ZM009 data set the skewness was reduced to an acceptable level.

The 02ZN001 data set had a long period of record (1966-95, n=30). This data set failed the test for trend at 5%. Peak flows appear to be increasing with time. Since there was no known physical reason for this trend it was assumed to be by chance.

The 02ZN002 data set had a short period of record (1985-96, n=12). This data set failed the test for trend at 1%! Contrary to 02ZN001, peak flows appear to be decreasing with time. Since there was no known physical reason for this trend it was assumed to be by chance.

3.4 Single Station Frequency Analysis

3.4.1 Probability Distributions

Similar to the 1990 study, the computer program CFA88 was used for frequency analysis. Sixty-five (65) data sets were available for frequency analysis. The available distributions were: the Generalized Extreme Value (GEV), the Three Parameter Log-normal (LN3), the Log-Pearson Type III (LP3), and the Wakeby distributions. The LP3 and Wakeby distributions were not used because the 1990 study concluded that the differences between these fits and the empirical distributions were often higher than those for the GEV and the

LN3. Distribution parameters were estimated using maximum likelihood theory.

3.4.2 Selection of a Distribution

The choice between the GEV or the LN3 distribution was based on the mean absolute deviation between the theoretical and empirical probabilities of the upper half of the data set. A smaller average absolute deviation would indicate a better fitting distribution. The upper half of the data set was defined as the median and all values above the median. Only the upper half of the data set was used because it is this portion of the curve with which we have the most interest: the 2-year return period and above. The mean absolute deviation of the upper half of each data set is listed in Table 3.3 along with selection of the best probability distribution. The LN3 distribution was the better fitting distribution on 42 (68%) of the 65 watersheds.

As a check on the suitability of these distributions for the flood data, the KS test was performed. This non-parametric test examines the maximum difference between the theoretical and empirical probabilities to see if the distributions are significantly different. Both distributions passed the test at a 20% level of significance for all flood series.

3.4.3 Results

The estimated flow rates for the 2-, 5-, 10-, 20-, 50-, 100-, and 200-year return periods are listed in Table 3.3. Generally, the confidence in the estimate for long return periods is low when the sample size is small. Return period flows were qualified by calculating the 95% confidence interval around the estimates of the 2-, 20- and 100-year return period flows. The confidence intervals were calculated assuming a LN3 distribution in all cases. The confidence intervals in Table 3.4 are given as a percentage of the LN3 return period flows. Typical upper and lower 95% confidence intervals are indicated in Table 3.4 as the medians of the upper and lower 95% confidence intervals of each station and return period. Large confidence intervals indicated small sample size and / or a poor LN3 fit to the data. Given the magnitude of some of the confidence intervals at high return periods, it is clear that return period flows should not be used for some stations with small sample sizes. The median sample size was 16. It was interesting to note that in some cases the confidence in the estimate was smaller at longer return periods. In these cases the LN3 distribution was upper bounded.

Table 3.3 Selection of the Distribution and Return Period Flows

Station Number	n	MAD GEV	MAD LN3	PDF	Area (km^2)	Q2	Q5	Q10	Q20	Q50	Q100	Q200
02YA001	27	0.0434	0.0386	LN3	306	30.1	40.4	48.3	56.8	69.1	79.2	90.1
02YA002	11	0.0322	0.0301	LN3	33.6	14.5	26.8	41.8	63.5	105	150	209
02YC001	38	0.0128	0.0128	LN3	624	187	251	294	338	395	439	485
02YD001	19	0.0358	0.0342	LN3	237	96.9	129	152	174	204	228	252
02YD002	17	0.0343	0.0334	LN3	200	38.0	49.4	57.4	65.5	76.6	85.4	94.4
02YE001	17	0.0458	0.0415	LN3	95.7	41.4	52.9	60.1	66.8	75.2	81.4	87.5
02YF001	14	0.0587	0.0478	LN3	611	260	337	397	460	552	627	708
02YG001	11	0.0400	0.0422	GEV	627	291	370	414	451	493	520	544
02YG002	10	0.0609	0.0575	LN3	224	45.6	65.2	79.0	92.7	111	126	141
02YH001a	11	0.0443	0.0482	GEV	33.4	4.52	5.73	6.46	7.11	7.89	8.43	8.93
02YJ001	28	0.0169	0.0187	GEV	640	312	415	483	549	634	698	762
02YJ003	11	0.0294	0.0269	LN3	119	31.3	36.6	38.7	40.2	41.7	42.5	43.2
02YK002a	23	0.0453	0.0438	LN3	470	119	151	175	201	238	267	299
02YK003	11	0.0197	0.0218	GEV	362	68.0	85.1	92.7	98.2	103	106	108
02YK004a	17	0.0440	0.0848	GEV	529	97.2	116	127	136	145	152	157
02YK005	24	0.0266	0.0241	LN3	391	73.8	93.2	106	118	133	145	157
02YK007	13	0.0242	0.0217	LN3	112	23.5	29.7	33.4	36.7	40.7	43.5	46.2
02YK008	12	0.0505	0.0399	LN3	20.4	9.77	18.5	27.2	38.3	57.0	74.8	96.3
02YL001	68	0.0166	0.0165	LN3	2110	582	709	787	859	948	1010	1080
02YL004a	13	0.0499	0.0444	LN3	58.5	43.4	64.6	82.0	101	130	154	181
02YL005	12	0.0458	0.1137	GEV	17.0	12.6	24.8	43.3	77.9	174	323	604
02YM003	17	0.0323	0.0334	GEV	93.2	39.3	56.3	67.3	77.7	90.8	100	110
02YN002	16	0.0226	0.0206	LN3	469	172	251	315	384	485	570	663
02YO006a	16	0.0822	0.0714	LN3	177	43.0	54.1	61.1	67.7	76.1	82.3	88.4
02YO007	12	0.0562	0.0452	LN3	88.3	27.3	37.3	48.3	62.8	88.7	114	147
02YO008	13	0.0296	0.0207	LN3	823	218	291	343	395	466	522	581
02YO010	12	0.0485	0.0426	LN3	61.6	9.22	15.5	22.6	32.6	51.0	69.8	94.0
02YP001a	14	0.0319	0.0288	LN3	63.8	22.5	26.2	27.8	29.0	30.1	30.8	31.4
02YQ001	47	0.0423	0.0421	LN3	4400	581	731	825	912	1020	1100	1180
02YQ004	14	0.0601	0.0518	LN3	2150	634	865	995	1110	1240	1330	1420
02YQ005	10	0.0371	0.0377	GEV	80.8	44.7	54.3	58.6	61.6	64.4	66.0	67.1
02YR001	38	0.0210	0.0205	LN3	275	27.4	34.4	38.7	42.7	47.5	50.8	54.0
02YR002	20	0.0172	0.0176	GEV	399	64.1	83.8	100	151	151	180	214
02YR003	16	0.0592	0.0626	GEV	554	59.3	71.3	76.7	80.6	84.1	86.0	87.4
02YS001a	20	0.0469	0.0528	GEV	1290	165	201	227	254	292	323	365
02YS003	29	0.0273	0.0283	GEV	36.7	13.0	17.0	19.9	22.8	26.8	29.9	33.3
02YS005	12	0.0421	0.0509	GEV	2000	239	297	323	342	359	368	375
02ZA001	18	0.0523	0.0593	GEV	343	115	156	182	205	234	254	273
02ZA002	15	0.0367	0.0331	LN3	72.0	49.9	85.2	121	165	241	313	399
02ZA003	15	0.0249	0.0202	LN3	139	149	211	252	292	344	384	424
02ZB001	35	0.0341	0.0315	LN3	205	340	509	635	765	947	1090	1250
02ZC002	15	0.0361	0.0345	LN3	230	357	486	577	668	790	886	984
02ZD002	19	0.0338	0.0353	GEV	1340	851	1190	1390	1580	1800	1950	2090
02ZE001	21	0.0651	na	GEV	2640	282	348	383	402	443	464	484
02ZG001a	34	0.0344	0.0310	LN3	205	55.9	74.1	86.8	99.4	116	129	143
02ZG002	20	0.0294	0.0311	GEV	166	46.6	64.7	77.4	90.0	107	120	134
02ZG003	17	0.0663	0.0557	LN3	115	55.5	79.9	97.2	115	138	157	176
02ZG004	16	0.0170	0.0148	LN3	42.7	35.1	49.0	58.0	66.5	77.4	85.5	93.6
02ZH001a	32	0.0166	0.0172	GEV	764	249	325	369	407	452	483	511
02ZH002	28	0.0243	0.0253	GEV	43.3	31.3	43.4	50.6	56.9	64.3	69.4	74.1
02ZJ001	20	0.0331	0.0365	GEV	67.4	21.7	29.3	34.6	39.9	47.1	52.7	58.6
02ZJ003	11	0.0337	0.0329	LN3	106	30.1	49.4	66.6	86.5	118	145	177
02ZK001	48	0.0223	0.0219	LN3	301	144	199	238	277	329	370	412
02ZK002	18	0.0367	0.0409	GEV	89.6	71.1	106	132	158	195	225	256
02ZK003	14	0.0898	0.0614	LN3	37.2	66.8	79.1	82.7	84.8	86.5	87.4	88.0
02ZK004	14	0.0362	0.0348	LN3	104	91.3	130	162	197	249	294	342
02ZK005	11	0.0695	0.0681	LN3	50.3	24.5	38.7	50.0	62.3	80.3	95.4	112
02ZL003	18	0.0189	0.0181	LN3	10.8	8.52	12.2	14.5	16.8	19.6	21.7	23.8
02ZL004	14	0.0389	0.0320	LN3	28.9	16.0	22.7	26.8	30.6	35.3	38.8	42.1
02ZL005	12	0.0308	0.0257	LN3	11.2	4.82	7.26	9.51	12.2	16.4	20.3	24.7
02ZM006	43	0.0245	0.0255	GEV	3.63	3.29	4.46	5.28	6.11	7.25	8.14	9.08
02ZM009a	17	0.0543	0.0635	GEV	53.6	26.5	29.5	30.8	31.8	32.6	33.1	33.5
02ZM016	14	0.0436	0.0417	LN3	17.3	12.3	16.7	19.2	21.6	24.4	26.4	28.4
02ZN001	30	0.0462	0.0404	LN3	53.3	37.9	47.4	52.8	57.5	63	66.8	70.5
02ZN002	12	0.0295	0.0223	LN3	15.5	8.07	11.7	14.4	17.3	21.4	24.7	28.3

Notes: n - sample size,
MAD - mean absolute deviation,
PDF - probability distribution function,
QT - "T" year return period flow in m^3/s.

Table 3.4 Ninety-five Percent (95%) Confidence Interval as a Percentage of the LN3 Return Period Flows

Station Number	n	Lower Limit	Q2	Upper Limit	Lower Limit	Q20	Upper Limit	Lower Limit	Q100	Upper Limit
02YA001	27	-10%	30.1	13%	-22%	56.8	32%	-30%	79.2	47%
02YA002	11	-29%	14.5	49%	-62%	63.5	218%	-75%	150	365%
02YC001	38	-10%	187	11%	-17%	338	21%	-21%	439	27%
02YD001	19	-13%	96.9	16%	-23%	174	31%	-29%	228	43%
02YD002	17	-12%	38.0	15%	-23%	65.5	33%	-30%	85.4	46%
02YE001	17	-15%	41.4	17%	-21%	66.8	27%	-26%	81.4	36%
02YF001	14	-12%	260	17%	-27%	460	44%	-36%	627	67%
02YG001	11	-17%	292	18%	-20%	449	22%	-23%	523	26%
02YG002	10	-22%	45.6	30%	-34%	92.7	53%	-41%	126	72%
02YH001a	11	-17%	4.53	19%	-22%	7.23	27%	-27%	8.68	34%
02YJ001	28	-12%	313	14%	-18%	554	22%	-22%	698	28%
02YJ003	11	-16%	31.3	12%	-10%	40.2	7%	-10%	42.5	6%
02YK002a	23	-9%	119	12%	-20%	201	30%	-28%	267	44%
02YK003	11	-20%	71.2	15%	-11%	93.8	6%	-9%	98.6	5%
02YK004a	17	-8%	90.0	13%	-29%	167	58%	-42%	255	99%
02YK005	24	-10%	73.8	11%	-16%	118	20%	-21%	145	27%
02YK007	13	-15%	23.5	17%	-20%	36.7	23%	-23%	43.5	28%
02YK008	12	-28%	9.77	51%	-54%	38.3	136%	-65%	74.8	206%
02YL001	68	-5%	582	6%	-8%	859	9%	-11%	1010	12%
02YL004a	13	-19%	43.4	28%	-37%	101	68%	-47%	154	100%
02YL005	12	-39%	16.5	50%	-39%	55.4	57%	-44%	83.9	70%
02YM003	17	-20%	39.2	24%	-26%	79.9	34%	-31%	105	42%
02YN002	16	-17%	172	23%	-33%	384	55%	-42%	570	80%
02YO006a	16	-13%	43.0	15%	-20%	67.7	25%	-24%	82.3	33%
02YO007	12	-11%	27.3	21%	-42%	62.8	119%	-59%	114	215%
02YO008	13	-15%	218	20%	-27%	395	41%	-34%	522	58%
02YO010	12	-18%	9.22	37%	-55%	32.6	168%	-69%	69.8	280%
02YP001a	14	-13%	22.5	11%	-9%	29.0	7%	-9%	30.8	7%
02YQ001	47	-7%	581	8%	-11%	912	13%	-14%	1100	16%
02YQ004	14	-21%	634	22%	-22%	1110	24%	-24%	1330	27%
02YQ005	10	-18%	44.7	16%	-15%	62.5	13%	-16%	66.0	13%
02YR001	38	-9%	28.5	10%	-13%	42.7	15%	-16%	50.8	19%
02YR002	20	-11%	64.4	15%	-25%	121	40%	-34%	170	60%
02YR003	16	-13%	58.9	12%	-12%	81.9	11%	-13%	80.3	12%
02YS001a	20	-7%	168	9%	-19%	254	27%	-26%	323	41%
02YS003	29	-10%	13.0	13%	-19%	23.4	26%	-25%	29.9	35%
02YS005	12	-19%	249	14%	-10%	327	6%	-9%	344	5%
02ZA001	18	-16%	115	21%	-28%	205	41%	-35%	254	57%
02ZA002	15	-20%	49.9	35%	-47%	165	107%	-59%	313	165%
02ZA003	15	-19%	149	23%	-27%	292	37%	-32%	384	47%
02ZB001	35	-14%	340	17%	-23%	765	30%	-28%	1090	41%
02ZC002	15	-16%	357	20%	-26%	668	38%	-33%	886	52%
02ZD002	19	-19%	846	21%	-23%	1620	27%	-26%	2060	43%
02ZG001a	34	-10%	55.9	12%	-17%	99.4	22%	-22%	129	30%
02ZG002	20	-16%	46.6	19%	-24%	91.6	32%	-30%	122	43%
02ZG003	17	-18%	55.5	23%	-28%	115	40%	-34%	157	53%
02ZG004	16	-18%	35.1	22%	-25%	66.5	32%	-30%	85.5	41%
02ZH001a	32	-11%	248	12%	-15%	417	17%	-18%	509	21%
02ZH002	28	-15%	31.0	17%	-19%	58.9	22%	-22%	74.8	27%
02ZJ001	20	-14%	21.7	17%	-23%	40.6	31%	-29%	53.5	42%
02ZJ003	11	-24%	30.1	40%	-47%	86.5	100%	-57%	145	149%
02ZK001	48	-10%	144	11%	-16%	277	20%	-21%	370	27%
02ZK002	18	-21%	70.9	27%	-30%	163	43%	-36%	230	56%
02ZK003	14	-21%	66.8	13%	-7%	84.8	3%	-5%	87.4	2%
02ZK004	14	-16%	91.3	23%	-34%	197	60%	-44%	294	90%
02ZK005	11	-25%	24.5	37%	-41%	62.3	77%	-50%	95.4	109%
02ZL003	18	-19%	8.52	22%	-24%	16.8	31%	-29%	21.7	39%
02ZL004	14	-22%	16.0	25%	-26%	30.6	32%	-30%	38.8	39%
02ZL005	12	-18%	4.82	30%	-42%	12.2	93%	-54%	20.3	146%
02ZM006	43	-10%	3.28	12%	-17%	6.20	21%	-21%	8.21	27%
02ZM009a	17	-7%	26.7	6%	-6%	31.3	4%	-6%	32.6	4%
02ZM016	14	-19%	12.3	21%	-23%	21.6	27%	-26%	26.4	33%
02ZN001	30	-10%	37.9	10%	-12%	57.5	13%	-14%	66.8	16%
02ZN002	12	-19%	8.07	27%	-35%	17.3	58%	-43%	24.7	83%
Median		-16%		21%	-24%		36%	-30%		46%

Notes: n - sample size,
QT - “T” year return period flow in m^3/s.

4. REGIONAL FLOOD FREQUENCY ANALYSIS

4.1 Regionalization

Division of the island into 4 hydrologically homogeneous regions was based on: previous studies, the availability of reliable data, regional flood characteristics, regional precipitation characteristics, regional physiographic characteristics, and the results of regression analyses on test regions.

Three regional flood frequency analyses have been conducted for streams in Newfoundland: Poulin 1971, Panu et al 1984, and Beersing 1990. Poulin had only one region with 17 stations. Panu et al had 21 station which were split in two regions: North and South. Panu also used the entire island as a region. Beersing used 4 regions: Avalon and Burin Peninsulas, Central, Humber Valley and Northern Peninsula, and Southwestern. Thirty-nine (39) stations were available for Beersing. A homogeneity test developed by Darymple (1960), which looks at the ratio of the 10 year flood to the average flood and compares it to the average ratio, was used to assist in the delineation of regions in the 1984 and 1990 studies. This test was not used in this study. The test assumes the shape of the dimensionless flood frequency curve is nearly constant within a region. The shape of this curve is dependent on climate (homogeneous within region) and physiography (varies within region).

Sixty-five (65) stations were available for frequency analysis. The confidence in the estimates of the return period flows at some stations was very low. Not all stations were retained for regional analysis. As a cutoff, only those stations which had an upper 95% confidence level of the 1:100 year flow (based on a three parameter log-normal distribution) which were less than 100% of the estimate were retained for regional analysis. Fifteen (15) stations were removed from the analysis and were retained for testing the regional regression equations. Fifty stations (50) remained for regional analysis and since at least 10 stations per region are desirable for multiple linear regression, 5 would be the maximum number of regions.

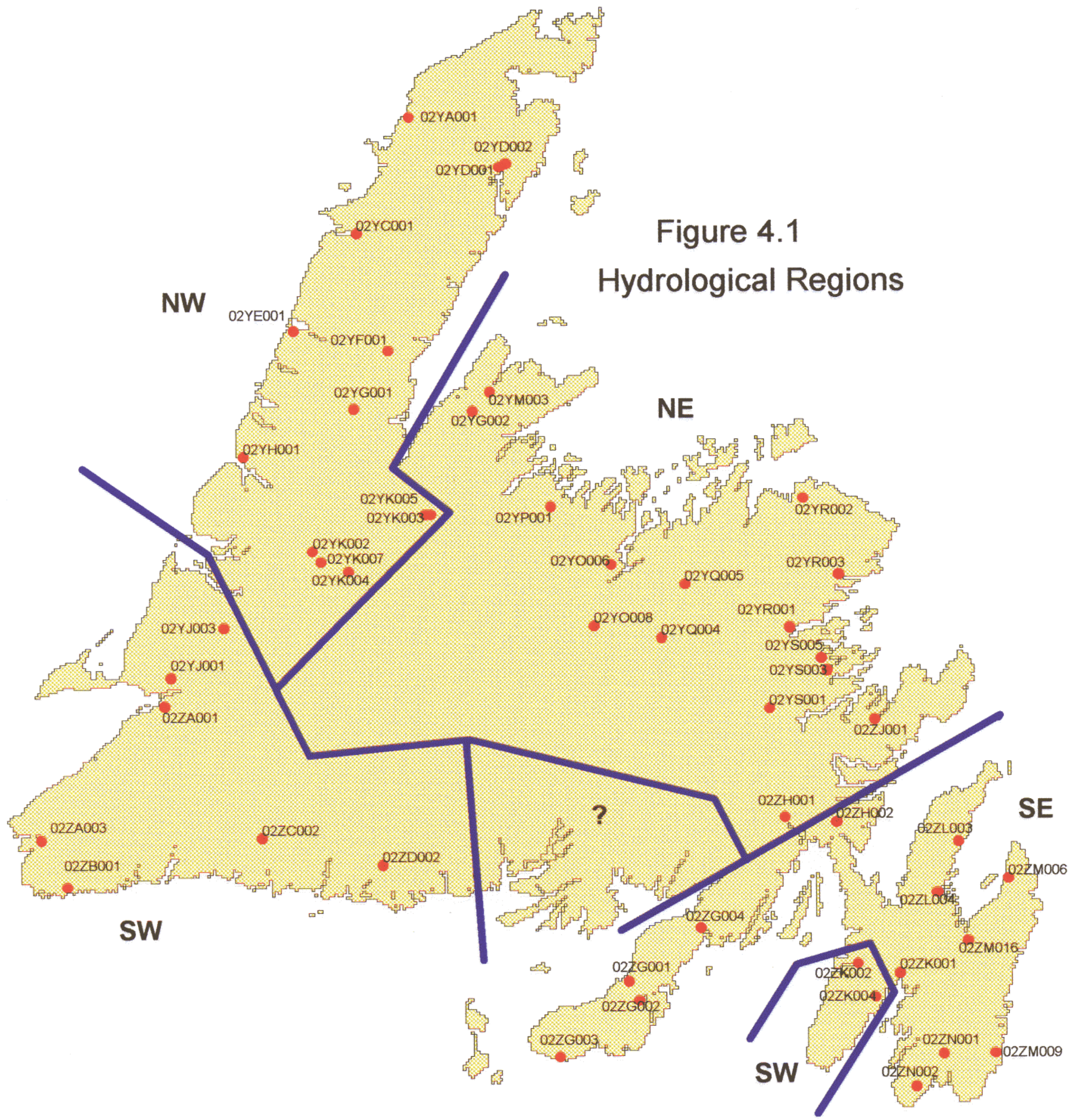
Regional flood characteristics are presented in Chapter 2. The 20-year flood per unit area provided an indication of the magnitude of floods on a regional basis. Local highs were identified in south-western Newfoundland and on the south-western part of Avalon Peninsula. The modal month of the flood provided an indication of the causative factors and the regional flood generating mechanisms. Floods most commonly occurred during February in the south-eastern region and are usually in the form of a rainfall on snow event. In the north-eastern and north-western regions, floods most commonly occurred during April or May and are caused primarily by snowmelt. In the south-western region, floods most commonly occurred during November and are caused by rainfall.

Regional climatic characteristics are presented in Chapter 2. The important climatic indices were: the mean annual precipitation, the 1:50 year 24-hour rainfall, mean annual snowfall, and snow depths. Mean annual precipitation and the 1:50 year 24-hour rainfall amounts were highest along the southern coast and lowest on the northern coast. Mean annual snowfall was lowest on the Avalon and Burin Peninsulas. Maximum snow depths increase with increasing latitude.

Physiography is regional and the selection of regions must be consistent with major physiographic features such as: major watershed divides, geology and relief. Regions were synthesized from adjoining tertiary watersheds where possible.

After consideration of the regional flood characteristics, precipitation indices, and physiography, the island was tentatively divided into 4 regions. These regions were similar to the regions used in the 1990 study. Over 100 test regressions were conducted and evaluated. Stations were moved in and out of regions based primarily on their studentized residuals, but also on their leverage and influence. The hydrologically homogeneous regions

Figure 4.1
Hydrological Regions



This page has been
intentionally
left blank.

identified in Figure 4.1 provided the best results. The delineation of the regions is approximate. The Bay d'Espoir region along the south coast remained unclassified due to a lack of gauged watersheds. Watersheds with large drainage areas (>1000 km^2) were readily dropped from the analyses if their residuals, leverage or influence was high. These watersheds were few in number and represented the extreme case (large drainage areas). Typically, return period flood estimates are required for watersheds which are much smaller than 1000 km^2 and quite often less than 100 km^2. Watersheds with drainage areas less than 100 km^2 were retained for analysis unless their residuals indicated that they were an outlier. Leverage and influence considerations were relaxed slightly for these watersheds so that the regression equations would be applicable to watersheds in this range of drainage areas.

Division of the island into 5 regions was attempted by splitting the North-western Region into a Humber Valley Region and a Northern Peninsula Region. Problems were encountered due to limited data. Due to similarities in the final regression equations, division of the island into 3 regions was attempted by combining the North-western and North-east regions. The error increased and the correlation decreased.

4.2 Physiographic Data

Physiographic data has been compiled for the 39 watersheds used in the 1990 study. Physiographic data was compiled for the additional 26 watersheds used in this study. The physiographic database is shown in Table 4.1. Details on the extraction procedures are given in Appendix A. The physiographic parameters which were selected for inclusion into the regression equations of 1984 and 1990 are reviewed below. A new Lake Attenuation Factor (LAF) parameter is introduced in the last paragraph.

The most important physiographic parameter for determining flood magnitudes is drainage area (DA). This was confirmed by the 1984 and 1990 studies. The main

Table 4.1 Physiographic Database

ID #	STATION NAME	GAUGE LOCATION						DRAIN AREA	FRAC TREE	FRAC SWAMP	FRAC LAKE	FRAC L+S	FRAC BAR'N	FRAC ACLS	LSF	LAF	LENGTH MAIN R	ELEV DIFF	SLOPE M1	SLOPE M2	DRAIN DENSITY	SHAPE FACTOR
		LATITUDE			LONGITUDE																	
		DEG	MIN	SEC	DEG	MIN	SEC	(KM^2)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(KM)	(M)	(%)	(%)	(KM^-1)	(-)
02YA001	STE. GENEVIEVE RIVER NEAR FORESTER'S POINT	51	8	18	56	47	32	306	0.64	0.14	0.22	0.35	0.00	0.96	1.78	1053	38.9	88	0.23	0.14	0.54	1.48
02YA002	BARTLETTS RIVER NEAR ST. ANTHONY	51	26	57	55	38	31	33.6	0.40	0.03	0.13	0.16	0.44	0.99	1.91	652	13.2	150	1.14	1.21	0.91	1.64
02YC001	TORRENT RIVER AT BRISTOL'S POOL	50	36	27	57	9	4	624	0.33	0.04	0.13	0.17	0.50	0.99	1.91	175	48.3	479	0.99	1.01	0.76	1.45
02YD001	BEAVER BROOK NEAR RODDICKTON	50	54	51	56	9	26	237	0.81	0.04	0.05	0.08	0.11	0.73	1.68	0	40.6	328	0.81	0.68	0.34	2.23
02YD002	NORTHEAST BROOK NEAR RODDICKTON	50	55	44	56	6	44	200	0.83	0.04	0.13	0.17	0.00	0.99	1.90	484	38.3	270	0.70	0.47	0.93	1.65
02YE001	GREAVETT BROOK ABOVE PORTLAND CREEK POND	50	9	37	57	34	45	95.7	0.49	0.06	0.06	0.12	0.39	0.88	1.82	134	24.5	700	2.86	3.09	0.75	1.64
02YF001	CAT ARM RIVER ABOVE GREAT CAT ARM	50	4	33	56	55	22	611	0.69	0.05	0.08	0.13	0.18	1.00	1.93	0	30.2	250	0.83	0.73	0.58	1.86
02YG001	MAIN RIVER AT PARADISE POOL	49	48	46	57	9	24	627	0.78	0.06	0.07	0.13	0.09	0.63	1.55	18.3	31.9	375	1.18	1.11	1.30	1.83
02YG002	MIDDLE ARM BROOK BELOW FLATWATER POND	49	48	22	56	20	33	224	0.83	0.06	0.09	0.15	0.02	0.96	1.88	299	26.4	255	0.96	0.75	0.45	1.84
02YH001	BOTTOM CREEK NEAR ROCKY HARBOUR	49	35	6	57	54	34	33.4	0.79	0.02	0.11	0.13	0.08	0.93	1.86	545	14.6	163	1.12	0.85	1.13	1.68
02YJ001	HARRYS RIVER BELOW HIGHWAY BRIDGE	48	34	31	58	21	48	640	0.79	0.09	0.06	0.14	0.07	0.75	1.67	141	60.0	509	0.85	0.35	1.12	1.81
02YJ003	PINCHGUT BROOK AT OUTLET OF PINCHGUT LAKE	48	48	21	58	0	56	119	0.86	0.05	0.05	0.10	0.04	1.00	1.95	290	16.6	164	0.99	0.78	1.73	1.54
02YK002	LEWASEECHJEECH BROOK AT LITTLE GRAND LAKE	49	9	43	57	25	28	470	0.55	0.06	0.10	0.16	0.29	1.00	1.92	274	54.9	561	1.02	0.59	0.63	2.32
02YK003	SHEFFIELD RIVER AT SHEFFIELD LAKE	49	20	10	56	37	57	362	0.67	0.07	0.11	0.18	0.15	1.00	1.91	688	37.0	351	0.95	1.13	0.43	1.85
02YK004	HINDS BROOK NEAR GRAND LAKE	49	4	21	57	10	46	529	0.35	0.24	0.12	0.36	0.29	0.95	1.77	666	49.3	320	0.65	0.32	0.64	1.78
02YK005	SHEFFIELD BROOK NEAR TRANS-CANADA HIGHWAY	49	20	11	56	39	56	391	0.68	0.08	0.10	0.17	0.15	0.94	1.85	590	38.1	378	0.99	1.07	0.19	1.98
02YK007	GLIDE BROOK BELOW GLIDE LAKE	49	6	50	57	22	9	112	0.87	0.09	0.04	0.13	0.00	0.98	1.91	132	26.8	234	0.88	0.90	1.28	1.61
02YK008	BOOT BROOK AT TRANS-CANADA HIGHWAY	49	15	59	57	6	14	20.4	0.75	0.22	0.02	0.24	0.01	0.65	1.50	0	10.1	137	1.35	1.16	1.28	1.47
02YL001	UPPER HUMBER RIVER NEAR REIDVILLE	49	14	26	57	21	45	2110	0.74	0.06	0.05	0.11	0.15	0.75	1.68	0	118.8	678	0.57	0.46	0.79	1.56
02YL004	SOUTH BROOK AT PASEDENA	49	0	43	57	36	47	58.5	0.94	0.01	0.01	0.02	0.05	0.08	1.06	0	13.2	130	0.99	1.04	1.34	1.54
02YL005	RATTLER BROOK NEAR MCIVERS	49	3	34	58	6	19	17	0.91	0.08	0.02	0.10	0.00	0.46	1.39	0	8.2	244	2.98	2.88	1.05	1.10
02YL008	UPPER HUMBER RIVER ABOVE BLACK BROOK	49	37	9	57	17	39	471	0.58	0.01	0.07	0.08	0.34	0.99	1.95	0	48.5	393	0.81	0.64	0.57	1.90
02YM001	INDIAN BROOK AT INDIAN FALLS	49	30	43	56	6	45	974	0.79	0.07	0.09	0.16	0.05	0.88	1.80	36.4	65.0	290	0.45	0.41	0.45	1.88
02YM003	SOUTH WEST BROOK NEAR BAIE VERTE	49	53	37	56	13	22	93.2	0.91	0.07	0.05	0.11	0.00	0.56	1.49	0	18.6	107	0.58	0.57	0.68	1.67
02YN002	LLOYD'S RIVER BELOW KING GEORGE IV LAKE	48	14	32	57	49	41	469	0.23	0.06	0.12	0.18	0.63	1.00	1.91	371	57.3	166	0.29	0.30	1.37	2.15
02YN003	STAR BROOK BELOW STAR LAKE	48	34	46	57	13	41	427	0.21	0.30	0.09	0.38	0.40	1.00	1.80	377	42.5	360	0.85	0.74	0.59	1.91
02YO006	PETER'S RIVER NEAR BOTWOOD	49	6	21	55	24	38	177	0.83	0.13	0.03	0.16	0.02	0.97	1.89	0	42.7	190	0.45	0.45	0.80	1.93
02YO007	LEECH BROOK NEAR GRAND FALLS	48	56	42	55	49	42	88.3	0.70	0.24	0.04	0.28	0.02	0.73	1.57	0	23.1	272	1.18	0.88	0.74	1.52
02YO008	GREAT RATTLING BR. ABOVE TOTE R. CONFLUENCE	48	49	38	55	31	50	823	0.73	0.19	0.05	0.24	0.03	0.55	1.40	0	69.0	221	0.32	0.30	0.69	1.80
02YO010	JUNCTION BROOK NEAR BADGER	48	58	23	56	1	21	61.6	0.81	0.09	0.10	0.19	0.00	0.89	1.79	601	18.2	187	1.03	0.62	0.77	1.55
02YO012	SOUTHWEST BROOK AT LEWISPORTE	49	12	58	55	3	11	58.7	0.80	0.08	0.12	0.20	0.00	0.67	1.55	128	22.7	134	0.59	0.42	0.54	1.87
02YP001	SHOAL ARM BROOK NEAR BADGER BAY	49	22	18	55	48	44	63.8	0.88	0.07	0.06	0.13	0.00	0.79	1.72	119	20.0	113	0.56	0.53	0.88	1.62
02YQ001	GANDER RIVER AT BIG SHUTE	49	0	55	54	51	13	4400	0.76	0.08	0.09	0.17	0.07	0.91	1.82	277	133.8	297	0.22	0.15	0.45	2.08
02YQ004	NORTHWEST GANDER RIVER NEAR GANDER LAKE	48	46	7	55	4	52	2150	0.66	0.25	0.06	0.31	0.03	0.44	1.22	0	104.2	265	0.25	0.17	0.45	1.63
02YQ005	SALMON RIVER NEAR GLENWOOD	49	0	41	54	55	3	80.8	0.85	0.11	0.04	0.15	0.00	0.87	1.79	0	22.5	372	1.65	1.03	1.09	1.78
02YQ006	SOUTHWEST GANDER RIVER BELOW LARSON FALLS	48	39	26	55	4	52	531	0.64	0.30	0.04	0.34	0.02	0.40	1.16	0	51.3	299	0.58	0.26	0.40	1.88
02YR001	MIDDLE BROOK NEAR GAMBO	48	48	28	54	13	28	267	0.75	0.07	0.18	0.24	0.01	0.98	1.86	881	49.3	177	0.36	0.32	0.26	1.93
02YR002	RAGGED HARBOUR RIVER NEAR MUSGRAVE HARBOUR	49	23	35	54	6	25	399	0.68	0.16	0.17	0.33	0.00	0.96	1.79	65.1	42.0	95	0.23	0.21	0.74	1.68
02YR003	INDIAN BAY BROOK NEAR NORTHWEST ARM	49	2	24	53	53	0	554	0.70	0.13	0.20	0.33	0.00	0.97	1.80	307	52.4	136	0.26	0.23	0.68	1.72
02YS001	TERRA NOVA RIVER AT EIGHT MILE BRIDGES	48	26	30	54	22	21	1290	0.55	0.21	0.09	0.30	0.15	0.92	1.76	138	105.0	207	0.20	0.12	0.73	2.35
02YS003	SOUTHWEST BROOK AT TERRA NOVA NATIONAL PARK	48	36	25	53	58	50	36.7	0.84	0.14	0.02	0.16	0.00	1.00	1.92	0	11.2	143	1.28	1.11	0.64	1.43
02YS005	TERRA NOVA RIVER AT GLOVERTOWN	48	39	43	54	1	5	2000	0.61	0.23	0.13	0.36	0.03	0.93	1.74	113	128.8	274	0.21	0.16	0.35	2.12
02ZA001	LITTLE BARACHOIS BROOK NEAR ST. GEORGE'S	48	26	44	58	23	55	343	0.60	0.02	0.07	0.09	0.30	0.83	1.78	0	65.5	463	0.72	0.68	1.04	2.45
02ZA002	HIGHLANDS RIVER AT TRANS-CANADA HIGHWAY	48	6	33	58	47	4	72	0.82	0.01	0.04	0.05	0.13	0.43	1.39	0	20.4	460	2.26	2.19	1.15	1.72
02ZA003	LITTLE CODROY RIVER NEAR DOYLES	47	49	19	59	11	40	139	0.66	0.07	0.04	0.11	0.16	0.73	1.66	131	25.2	450	1.78	1.46	1.46	1.68
02ZB001	ISLE AUX MORTS RIVER BELOW HIGHWAY BRIDGE	47	36	50	59	0	33	205	0.08	0.06	0.07	0.13	0.78	0.60	1.52	0	33.3	444	1.33	1.27	0.72	2.09
02ZC002	GRANDY BROOK BELOW TOP POND BROOK	47	51	27	57	44	0	230	0.20	0.01	0.05	0.06	0.82	0.34	1.30	38.4	28.9	360	1.24	1.08	0.96	1.84
02ZD002	GREY RIVER NEAR GREY RIVER	47	44	35	56	56	5	1340	0.04	0.16	0.04	0.20	0.75	0.63	1.51	0	60.0	310	0.52	0.34	0.15	5.31
02ZE001	SALMON RIVER AT LONG POND	47	56	40	55	54	50	2640	0.35	0.02	0.14	0.16	0.50	1.00	1.92	619	100.4	122	0.12	0.08	0.36	1.75
02ZE004	CONNE RIVER AT OUTLET OF CONNE RIVER POND	48	10	8	55	29	0	99.7	0.60	0.34	0.05	0.39	0.01	1.00	1.81	0	18.7	109	0.58	0.59	1.38	1.52
02ZF001	BAY DE NORD RIVER AT BIG FALLS	47	44	48	55	26	30	1170	0.32	0.05	0.18	0.24	0.44	0.96	1.84	401	68.1	282	0.41	0.34	0.61	2.15
02ZG001	GARNISH RIVER NEAR GARNISH	47	12	50	55	19	45	205	0.26	0.01	0.09	0.10	0.63	0.96	1.91	202	44.7	370	0.83	0.60	0.55	2.45
02ZG002	TIDES BROOK BELOW FRESHWATER POND	47	7	38	55	15	54	166	0.37	0.04	0.09	0.13	0.49	0.92	1.85	588	26.7	221	0.83	0.78	1.35	1.84
02ZG003	SALMONIER RIVER NEAR LAMALINE	46	52	29	55	46	39	115	0.16	0.06	0.07	0.13	0.73	0.92	1.85	42.8	24.5	136	0.55	0.34	1.55	1.62
02ZG004	RATTLE BROOK NEAR BOAT HARBOUR	47	27	2	54	51	18	42.7	0.34	0.03	0.14	0.16	0.46	0.92	1.83	123	10.0	107	1.07	1.10	1.62	1.53
02ZG005	LITTLE BARASWAY BROOK NEAR MOLLIERS	47	6	0	55	37	15	28.2	0.31	0.08	0.07	0.15	0.55	0.50	1.40	132	13.0	133	1.03	1.19	1.43	1.40
02ZH001	PIPERS HOLE RIVER AT MOTHERS BROOK	47	56	49	54	17	8	764	0.11	0.48	0.18	0.66	0.23	0.91	1.57	17.4	50.9	207	0.41	0.38	0.71	1.67
02ZH002	COME BY CHANCE RIVER NEAR GOOBIES	47	55	7	53	56	59	43.3	0.40	0.02	0.08	0.10	0.50	0.92	1.87	20.8	17.0	110	0.65	0.59	1.11	1.66
02ZJ001	SOUTHERN BAY RIVER NEAR SOUTHERN BAY	48	22	44	53	40	36	67.4	0.82	0.06	0.10	0.16	0.03	0.86	1.78	89.3	16.0	128	0.80	0.50	1.24	1.64
02ZJ002	SALMON COVE RIVER NEAR CHAMPNEYS	48	23	45	53	18	6	73.6	0.74	0.06	0.13	0.19	0.07	0.82	1.72	435	18.0	137	0.76	0.55	1.11	1.33
02ZJ003	SHOAL HARBOUR RIVER NEAR CLARENVILLE	48	12	24	54	2	23	106	0.65	0.10	0.07	0.17	0.18	0.68	1.58	166	25.1	250	0.99	0.91	0.66	1.66
02ZK001	ROCKY RIVER NEAR COLINET	47	13	29	53	34	6	301	0.51	0.02	0.10	0.12	0.37	0.57	1.49	8.79	45.2	165	0.37	0.23	0.96	1.95
02ZK002	NORTHEAST RIVER NEAR PLACENTIA	47	16	26	53	50	27	89.6	0.48	0.16	0.15	0.31	0.24	0.81	1.64	278	26.9	200	0.74	0.57	1.11	1.91
02ZK003	LITTLE BARACHOIS (BARASWAY) R NEAR PLACENTIA	47	10	53	54	2	27	37.2	0.86	0.11	0.02	0.13	0.01	0.34	1.24	0	14.6	228	1.56	1.77	1.16	1.48
02ZK004	LITTLE SALMONIER RIVER NEAR NORTH HARBOUR	47	7	15	53	44	0	104	0.23	0.38	0.08	0.46	0.31	0.91	1.67	116	28.5	236	0.83	0.66	1.50	1.85
02ZK005	TROUT BROOK NEAR BELLEVUE	47	36	21	53	45	53	50.3	0.36	0.00	0.07	0.07	0.57	0.50	1.45	91.6	14.8	137	0.92	0.88	1.18	1.90
02ZL003	SPOUT COVE BROOK NEAR SPOUT COVE	47	48	43	53	9	15	10.8	0.42	0.01	0.08	0.09	0.49	1.00	1.95	319	7.0	91	1.31	1.25	1.09	1.36
02ZL004	SHEARSTOWN BROOK AT SHEARSTOWN	47	35	2	53	18	29	28.9	0.70	0.00	0.04	0.04	0.27	0.39	1.36	0	13.4	122	0.91	1.03	1.14	1.73
02ZL005	BIG BROOK AT LEAD COVE	48	2	34	53	5	0	11.2	0.39	0.03	0.07	0.10	0.51	1.00	1.95	272	8.7	211	2.43	2.43	1.00	1.52
02ZM006	NORTHEAST POND RIVER AT NORTHEAST POND	47	38	6	52	50	14	3.9	0.75	0.17	0.04	0.21	0.04	1.00	1.89	265	2.6	64	2.44	2.42	1.04	1.24
02ZM009	SEAL COVE BROOK NEAR CAPPAHAYDEN	46	50	50	52	58	27	53.6	0.38	0.01	0.12	0.14	0.51	1.00	1.93	193	14.9	133	0.89	0.98	1.13	1.37
02ZM016	SOUTH (MAHERS) RIVER NEAR HOLYROOD	47	21	41	53	7	2	17.3	0.22	0.05	0.06	0.11	0.68	0.90	1.84	148	8.7	259	2.98	2.22	1.01	1.40
02ZN001	NORTHWEST BROOK AT NORTHWEST POND	46	51	8	53	18	11	53.3	0.09	0.00	0.13	0.13	0.79	1.00	1.94	132	14.6	93	0.63	0.61	1.09	2.06
02ZN002	ST. SHOTTS RIVER NEAR TREPASSEY	46	42	32	53	29	24	15.5	0.88	0.00	0.12	0.12	0.00	0.82	1.75	512	10.3	23	0.22	0.43	1.03	1.53
					Minimum			3.9	0.04	0.00	0.01	0.02	0.00	0.08	1.06	0	2.63	23	0.12	0.08	0.15	1.10
					Median			152.5	0.66	0.06	0.08	0.16	0.15	0.91	1.79	129.5	26.9	221	0.83	0.65	0.89	1.72
					Maximum			4400	0.94	0.48	0.22	0.66	0.82	1.00	1.95	1053	133.8	700	2.98	3.09	1.73	5.31

Abbreviations:

ID # - identification number
DEG - degrees
MIN - minutes
SEC - seconds
DRAIN - drainage
FRAC - fraction of watershed occupied by ...
TREE - forest area
SWAMP - swamp area
LAKE - lake area
L+S - lakes plus swamps area
BAR'N - barren area
ACLS - area controlled by lakes and swamps
LSF - lakes and swamps factor
LAF - lake attenuation factor
LENGTH MAIN R - length of main channel
ELEV DIFF - elevation difference between outlet and watershed divide
SLOPE M1 - slope method 1
SLOPE M2 - slope method 2

assumption is that during floods, nearly all of the watershed is contributing to the peak discharge. The peak flow to drainage area relationship is non-linear as indicated in section 2.2. Peak flows increase with drainage area at a decreasing rate.

The second most important physiographic parameter for determining flood magnitudes in 3 of the 4 regions of the 1990 study was a newly developed "lakes and swamps factor" (LSF). This factor is a combination of the "fraction of watershed area occupied by lakes and swamps" (FLSAR) parameter and the "fraction of watershed area controlled by lakes and swamps" parameter (FACLS). The LSF parameter replaced the FACLS parameter used in the 1984 study. The algorithm is:

$$LSF = (1 + FACLS) - (FLSAR) / (1 + FACLS)$$

Quoting from the 1990 study:

> The reasons for the transformation were: (1) When FLSAR and FACLS tend toward 0, $\log_{10}(LSF)$ tends towards 0 and at the limit drops out of the regression equation. (2) It is reasonable to assume that as FLSAR increases, the amount of water lost to infiltration decreases and to a slight extent compensates for the attenuating effects of lakes and swamps. However, the effect is reduced if a larger percentage of the watershed area is controlled by lakes and swamps. The fraction of drainage area occupied by lakes and swamps (FLSAR) ranges from 0.05 to 0.36 (not including the extreme case of Pipers Hole River watershed where FLSAR is 0.66). During this study it was found that combining the term FLSAR with FALCS in the form given improved the predictive capability of the regression equations, especially on watersheds with higher FLSARs.

Drainage density (DRD) was the next most important independent variable in the

1990 study followed by the slope of the main channel (SLOPE).

The second most important physiographic parameter for determining flood magnitudes in the 1984 study was mean annual runoff (MAR) expressed as an equivalent depth over the watershed. This parameter was eliminated in the 1990 study because return period flows were highly sensitive to MAR, and because abstraction errors from the computer-generated map in the report were significant.

The third most important physiographic parameter for determining flood magnitudes in the 1984 study was FACLS for 2 of the 3 regions. Watersheds with high FACLS had reduced peak flows. Latitude of the watershed centroid (LAT) was the third most important physiographic parameter for determining flood magnitudes in the other region. The forth most important physiographic parameter for determining flood magnitudes was the shape of the watershed (SHAPE). The fraction of the watershed occupied by barren areas (BAREA) or the slope of the main channel (SLOPE) was used as a fifth independent variable in some equations.

After the influence of drainage area is removed, and MAR is dropped, the next most important watershed characteristic appears to be associated with a flood attenuation factor related to the "fraction of the watershed area controlled by lakes and swamps" (FACLS) parameter. There is no doubt that lakes and swamps in a watershed attenuate flood flows. The amount of attenuation is related to the locations and sizes of the lakes and swamps. FACLS assumes all lakes with surface areas greater than 1% of their drainage areas attenuate floods the same amount. While this parameter is a good indicator of flood attenuation by lakes and swamp, improvement in the measure of flood attenuation was attempted by taking into consideration the sizes of large lakes (>1% drainage area) and the relative sizes of the areas drained. The influence of swamps (wetlands) was removed from this parameter because the attenuation by wetlands was assumed to be much less than attenuation by lakes. The new "Lake Attenuation Factor" (LAF) was defined as follows:

$$LAF = \sum_{i=1}^{n} \{(100 \text{ x } LAREA_i/DA) \text{ x } (100 \text{ x } CAREA_i/DA)\}$$

where: n is the number of lakes in the watershed with area greater than 1% of the watershed's drainage area, $LAREA_i$ is the area of a lake, DA is the drainage area of the watershed, and $CAREA_i$ is the drainage area which is controlled by a lake. LAF is defined graphically in Appendix A. LAF assumes that lake size and the size of the area drained, as percentages of the drainage area, are equally important in attenuating floods. LAF has no units and simplifies to:

$$LAF = 10^4 \text{ x } DA^{-2} \text{ x } \sum_{i=1}^{n} (LAREA_i \text{ x } CAREA_i)$$

4.3 Regression Analysis

Equations were developed for each region which provided predictions of return period flood estimates based on physiographic data. The equations were of the form:

$$Q_T = c\times(var1)^{a1}\times(var2)^{a2}\times(var3)^{a3}\times...$$

where, Q_T is the magnitude of the flood with return period T,
c, a1, a2, a3, ... are constants, and
var1, var2, var3, ... are variables which correspond to the significant physiographic parameters.

Taking the $\log_{10}$ of both sides of the equation linearizes the equation so that multiple linear regression can be performed.

$$\log_{10}(Q_T) = \log_{10}(c) + a_1 \times \log_{10}(var1) + a_2 \times \log_{10}(var2) + a_3 \times \log_{10}(var3) + ...$$

A forward stepwise regression was performed. The coefficients and variables in the final regional equations were selected based on the following criteria:

1) The coefficient of correlation between the dependent and independent variables had to be significantly high.
2) The standard error of the estimate had to be a minimum.
3) The final predictor variables had to be independent of each other.
4) Entry into the regression equation had to be significant at a 5% level using the F-ratio.
5) The number of physiographic parameters in the regression equations had to be minimal.

Since $\log_{10}(0) = -\infty$, cases where var = 0 needed to be remedied. The fraction of watershed occupied by lakes, forest, etc., were defaulted to 0.01 when zero. Defaulting LAF to 25, 50 and 100 when LAF = 0 was checked by plotting $\log_{10}$(LAF) against the $\log_{10}$(peak flow per unit area) for the entire island. The LAF was defaulted to 50.

4.4 Regression Results

The results of the regression analyses are shown by region in Tables 4.2(a), (b), (c) and (d). The significant parameters were drainage area (DA), lake attenuation factor (LAF) and lakes and swamps factor (LSF). DA was by far the most important variable in NW, North-east and South-east Regions. DA was forced to be the most important variable in the South-west Region by including stations on the south-west part of the Avalon Peninsula. This is justified based on the magnitudes of floods per unit area on the south-west part of the Avalon Peninsula. This move also increased the allowable range of parameters in the South-west Region to include drainage areas as small as 89.6 km^2.

Final and intermediate results were evaluated in terms of the squared multiple R (SMR) statistic and the standard error of the estimate (SEE) for log to base 10 transformed data. Drainage area alone accounted for 78% to 84% of the variation in flood flows (log to base 10 units) in the North-west and North-east Regions. In the South-east Region, drainage area accounted for 90% to 94% of the variation in flood flows. In the South-west Region, drainage area accounted for 40% to 61% of the variation in flood flows. The SEE (log to base ten units) varied from 0.20 to 0.24 in the North-west and North-east Regions. The SEE in the South-east Region was lowest ranging from 0.12 to 0.15. In the South-west Region, the SEE was highest and ranged from 0.29 to 0.41.

The addition of a second variable to the regression equations increased the SMR statistic and decreased the SEE in all regions. The selection of the second variable was based on the magnitude of the F statistic. The LAF parameter was selected as the second variable in the NW, North-east and South-east Regions. In the North-west Region, SMR increased significantly from the 79-82% range to the 96-97% range. In the North-east Region, SMR increased similarly from the 78-84% range to the 91-95% range. In the South-east Region, SMR increased marginally from the already high 90-94% range to the 93-97% range. In the North-west Region, SEE dropped from the 0.20-0.23 range to the 0.09-0.10 range. Similarly, in the North-east Region, SEE dropped from the 0.20-0.24 range to the 0.12-0.16 range. In the South-east Region where SEE was quite low, the SEE dropped from the 0.12-0.15 range to the 0.09-0.13 range. In the South-west Region, LSF was selected as the second variable. In the South-west Region, SMR increased dramatically from the 41-61% range to the 83-92% range. The SEE decrease from the 0.29-0.41 range to the 0.14-0.24 range.

Other parameter(s) could have been included in some of the regions, and at some of the return periods, based on the F statistic. It was decided to stop at this level for several reasons: to minimize the number of parameters in the equation, SEE was considered low,

Table 4.2a **Regression Coefficients - North-west Region**

T	C	DA	LAF	SMR	SEE	deltaSMR	deltaSEE
Q2	1.282	1.084	-0.392	0.964	0.093	0.141	0.105
Q5	1.750	1.084	-0.402	0.967	0.089	0.148	0.111
Q10	2.065	1.089	-0.413	0.969	0.087	0.153	0.116
Q20	2.323	1.098	-0.422	0.971	0.086	0.158	0.121
Q50	2.754	1.107	-0.435	0.968	0.092	0.163	0.123
Q100	3.034	1.116	-0.445	0.965	0.097	0.167	0.124
Q200	3.327	1.126	-0.455	0.960	0.104	0.169	0.124
Q2	0.126	1.105		0.823	0.198		
Q5	0.163	1.106		0.819	0.200		
Q10	0.182	1.111		0.816	0.203		
Q20	0.192	1.121		0.813	0.207		
Q50	0.211	1.131		0.805	0.215		
Q100	0.219	1.140		0.798	0.221		
Q200	0.226	1.150		0.791	0.228		

n=13

Table 4.2b **Regression Coefficients - North-east Region**

T	C	DA	LAF	SMR	SEE	deltaSMR	deltaSEE
Q2	4.365	0.780	-0.372	0.950	0.117	0.107	0.082
Q5	6.026	0.778	-0.386	0.942	0.127	0.113	0.082
Q10	7.211	0.776	-0.394	0.938	0.131	0.117	0.083
Q20	8.650	0.775	-0.410	0.940	0.130	0.127	0.089
Q50	10.046	0.769	-0.409	0.926	0.144	0.125	0.083
Q100	11.350	0.767	-0.415	0.918	0.152	0.128	0.081
Q200	12.647	0.766	-0.420	0.909	0.161	0.131	0.080
Q2	0.841	0.774		0.843	0.199		
Q5	1.094	0.772		0.829	0.209		
Q10	1.268	0.769		0.821	0.214		
Q20	1.413	0.768		0.813	0.219		
Q50	1.656	0.763		0.801	0.227		
Q100	1.820	0.760		0.790	0.233		
Q200	1.977	0.759		0.778	0.241		

n=15

Table 4.2c Regression Coefficients - South-east Region

T	C	DA	LAF	SMR	SEE	deltaSMR	deltaSEE
Q2	3.396	0.720	-0.157	0.967	0.088	0.031	0.028
Q5	5.070	0.708	-0.168	0.967	0.088	0.036	0.032
Q10	6.026	0.707	-0.170	0.963	0.092	0.036	0.032
Q20	6.887	0.706	-0.169	0.958	0.098	0.036	0.030
Q50	7.870	0.706	-0.167	0.948	0.110	0.035	0.025
Q100	8.570	0.707	-0.165	0.939	0.119	0.034	0.023
Q200	9.120	0.708	-0.162	0.929	0.129	0.032	0.020
Q2	1.265	0.783		0.936	0.116		
Q5	1.762	0.776		0.931	0.120		
Q10	2.070	0.775		0.927	0.124		
Q20	2.366	0.774		0.922	0.128		
Q50	2.748	0.773		0.913	0.135		
Q100	3.020	0.773		0.905	0.142		
Q200	3.296	0.774		0.897	0.149		

n=13

Table 4.2d Regression Coefficients - South-west Region

T	C	DA	LLSF	SMR	SEE	deltaSMR	deltaSEE
Q2	43.152	0.704	-5.112	0.924	0.140	0.311	0.152
Q5	77.983	0.687	-5.475	0.904	0.162	0.334	0.156
Q10	117.220	0.667	-5.743	0.889	0.177	0.354	0.159
Q20	169.044	0.648	-5.998	0.875	0.192	0.372	0.162
Q50	267.917	0.621	-6.306	0.857	0.210	0.394	0.167
Q100	374.973	0.598	-6.533	0.843	0.224	0.410	0.169
Q200	516.416	0.577	-6.750	0.829	0.237	0.424	0.172
Q2	1.416	0.876		0.613	0.292		
Q5	2.004	0.871		0.570	0.318		
Q10	2.523	0.860		0.535	0.336		
Q20	3.069	0.850		0.503	0.354		
Q50	3.954	0.833		0.463	0.377		
Q100	4.764	0.818		0.433	0.393		
Q200	5.675	0.804		0.405	0.409		

n=9

SMR was considered high, there was a negligible increase in the SMR with additional variables, there was a negligible decrease in SEE with additional variables, the chances of spurious correlation increase with an increase in the number of independent variables in the regression equations dramatically, because F statistics for significant third variables while greater than 4.0 were much less than the F statistic for the first and second variables and to achieve some consistency across regions and return periods.

The standard error of the estimate in the South-west Region was much higher than the error in the other regions. In addition, the correlations in the South-west Region were noticeably less than the correlations in the other regions. Return period flood flows per unit area were quite variable in this region. There was a concern that this region might not be hydrologically homogeneous. An “Upper Envelope Curve” was developed which looked at only those watersheds which had high peak flows per unit area. This curve, while biased towards higher flood flows, had less error and an improved correlation. Within the applicable drainage area range, these floods represent the highest in magnitude on the island. The results are shown in Table 4.2(e).

Table 4.2e Regression Coefficients - South-west Region - Upper Envelope Curve

T	C	DA	SMR	SEE	DA Range (km^2)	n
Q2	0.0256	1.765	0.995	0.027	72.0 - 230	6
Q5	0.0662	1.650	0.981	0.052		
Q10	0.1349	1.550	0.960	0.071		
Q20	0.2529	1.460	0.930	0.090		
Q20	0.822	1.225	0.927	0.103	37.2 - 230	7
Q50	0.841	1.262	0.928	0.105		
Q100	0.857	1.287	0.913	0.119		
Q200	0.855	1.314	0.888	0.140		

4.5 Testing of Regression Equations

The accuracy of the regression equations was assessed using the data set that produced them and using an independent data set. The independent data set consisted of peak flow series that were not used in the formulation of the regional regression equations. These watersheds included those that had low confidence in the flood estimates, watersheds with less than 10 years of record, watersheds with large drainage areas, redundant watersheds, watersheds that were removed during the regression analysis, and watersheds that were eliminated from the regional data set for other reasons such as failing prerequisite statistical tests.

The maximum absolute difference between the frequency analysis estimates and the regression equation estimates for all stations used in the regional analysis was 130%. Typical absolute percentage difference statistics can be represented by the median statistic. The median absolute percentage difference statistics for return period floods in the North-west Region was lowest and ranged from 8.0% to 11.7%. The median absolute percentage difference statistics for return period floods in the South-west Region was highest and ranged from 20.0% to 30.9%. The median absolute percentage difference statistics ranged from 18.2% to 28.0% in the North-east Region and from 10.6% to 14.3% in the South-east Region. The median absolute percentage difference statistics across all regions ranged from 8.0% to 30.9%. A detailed comparison between frequency analysis estimates and regression equation estimates for stations used in the regional analysis is provided in Table 4.3.

The maximum absolute difference between the frequency analysis estimates and the regression equation estimates for stations not used in the regional analysis was over 1000%. The 1000% difference occurred at the 1:200 year return period in the South-west Region at a station which had a drainage area which was less than 50% of the drainage area of the smallest watershed used in the development of the regression equations in the South-west Region. Further, LSF at 1.24, was below the applicable range for LSF. In addition,

Table 4.3 **Comparison between Frequency Analysis Estimates and Regression Equation Estimates**

Station	dA	LSF	LAF		Frequency Analysis Estimate							Regression Equation Estimates:							Percentage Difference:							
Number	(KM^2)	(-)	(-)	Region	Q2	Q5	Q10	Q20	Q50	Q100	Q200	Q2	Q5	Q10	Q20	Q50	Q100	Q200	Q2	Q5	Q10	Q20	Q50	Q100	Q200	n
02YH001	33.4	1.86	545	NW	4.52	5.73	6.46	7.11	7.89	8.43	8.93	4.86	6.23	6.99	7.66	8.64	9.22	9.83	7.6	8.8	8.1	7.8	9.5	9.4	10.1	11
02YE001	100.8	1.82	134	NW	41.4	52.9	60.1	66.8	75.2	81.4	87.5	27.9	36.3	41.5	46.6	54	59.1	64.6	-32.6	-31.4	-30.9	-30.3	-28.2	-27.4	-26.2	13
02YK007	112	1.91	132	NW	23.5	29.7	33.4	36.7	40.7	43.5	46.2	31.5	40.9	46.9	52.6	61.1	66.9	73.2	34.0	37.8	40.3	43.4	50.1	53.7	58.5	13
02YD002	200	1.9	484	NW	38	49.4	57.4	65.6	76.6	85.4	94.4	35.5	45.5	51.5	57.5	66	71.6	77.9	-6.7	-7.9	-10.2	-12.4	-13.9	-16.1	-17.5	17
02YD001	237	1.68	50	NW	96.9	129	152	174	204	228	252	104	136	158	181	214	238	265	7.1	5.6	4.1	3.7	4.8	4.3	5.1	19
02YA001	306	1.78	1053	NW	30.1	40.4	48.3	56.8	69.1	79.2	90.1	41.5	52.8	59.4	66	75.3	81.5	88.3	37.8	30.7	23.0	16.3	9.0	2.9	-2.1	27
02YK003	360	1.91	688	NW	68	85.1	92.7	98.2	103	106	108	58.4	74.7	84.5	94.5	109	118	129	-14.1	-12.2	-8.8	-3.8	5.4	11.4	19.1	11
02YK005	391	1.85	590	NW	73.8	93.2	106	118	133	145	157	67.9	86.9	98.5	110	127	139	151	-8.0	-6.8	-7.1	-6.5	-4.4	-4.4	-3.6	24
02YK002	470	1.92	274	NW	119	151	175	201	238	267	299	112	144	165	187	218	239	264	-5.9	-4.4	-5.6	-7.1	-8.6	-10.3	-11.7	23
02YK004	529	1.77	666	NW	97.2	116	127	136	145	152	157	89.8	115	130	146	169	184	201	-7.6	-1.0	2.6	7.5	16.2	21.1	28.2	17
02YF001	611	1.93	50	NW	260	337	397	460	552	627	708	290	380	444	511	610	684	769	11.5	12.8	11.8	11.0	10.4	9.1	8.6	14
02YC001	624	1.91	175	NW	187	251	294	338	395	439	485	181	235	271	308	362	401	445	-3.0	-6.3	-7.9	-8.9	-8.4	-8.6	-8.2	38
02YG001	627	1.55	18.3	NW	291	370	414	451	493	520	544	442	586	692	803	971	1101	1251	51.9	58.3	67.0	78.0	97.0	111.8	130.0	11
																		Absolute median	8.0	8.8	8.8	8.9	9.5	10.3	11.7	
																		Absolute maximum	51.9	58.3	67.0	78.0	97.0	111.8	130.0	
02YS003	36.7	1.92	50	NE	13	17	19.9	22.8	26.8	29.9	33.3	16.9	22	25.3	28.4	32.4	35.5	38.6	30.2	29.1	27.0	24.5	20.8	18.7	16.0	29
02YP001	63.8	1.72	119	NE	22.5	26.2	27.8	29	30.1	30.8	31.4	18.9	24.2	27.6	30.5	34.8	37.8	41	-16.2	-7.8	-0.7	5.3	15.5	22.9	30.6	14
02ZJ001	67.4	1.78	89.3	NE	21.7	29.3	34.6	39.9	47.1	52.7	58.6	21.9	28.2	32.2	35.8	40.8	44.5	48.2	1.0	-3.9	-6.8	-10.2	-13.4	-15.6	-17.7	20
02YQ005	80.8	1.79	50	NE	44.7	54.3	58.6	61.6	64.4	66	67.1	31.3	40.6	46.6	52.3	59.4	65	70.7	-29.9	-25.3	-20.4	-15.1	-7.7	-1.5	5.4	10
02YM003	93.2	1.49	50	NE	39.3	56.3	67.3	77.7	90.8	100	110	35	45.3	52.1	58.4	66.3	72.5	78.9	-10.9	-19.5	-22.6	-24.8	-27.0	-27.5	-28.3	17
02YO006	177	1.89	50	NE	43	54.1	61.1	67.7	76.1	82.3	88.4	57.7	74.7	85.7	96.1	109	119	129	34.3	38.0	40.3	41.9	42.7	44.1	45.9	15
02YG002	221	1.88	299	NE	45.6	65.2	79	82.7	111	126	141	35.3	44.5	50.3	54.8	62	66.9	72.1	-22.6	-31.8	-36.3	-33.7	-44.2	-46.9	-48.9	10
02YR001	267	1.86	881	NE	27.4	34.4	38.7	42.7	47.5	50.8	54	27.4	34	38.1	40.7	46.1	49.4	52.9	-0.1	-1.3	-1.6	-4.6	-3.0	-2.7	-2.0	30
02YR002	399	1.79	65.1	NE	64.1	83.8	100	120	151	180	214	98.7	127	145	162	182	198	215	53.9	51.5	45.1	34.9	20.6	10.2	0.5	20
02YR003	554	1.8	307	NE	59.3	71.3	76.7	80.6	84.1	86	87.4	71.6	90	102	111	124	134	144	20.7	26.3	32.5	37.1	47.8	55.8	65.0	16
02ZH001	764	1.57	17.4	NE	249	325	369	407	452	483	511	268	350	404	460	515	564	616	7.4	7.7	9.5	13.0	13.9	16.8	20.5	32
02YO008	823	1.4	50	NE	218	291	343	395	466	522	581	191	247	282	316	354	386	418	-12.2	-15.2	-17.7	-20.0	-24.0	-26.1	-28.0	13
02YS001	1290	1.76	138	NE	165	201	227	254	292	323	365	186	237	268	295	330	357	385	12.9	17.7	18.2	16.3	13.1	10.6	5.6	20
02YS005	2000	1.74	113	NE	239	297	323	342	359	368	375	283	360	408	450	502	543	587	18.2	21.0	26.3	31.7	39.9	47.6	56.4	12
02YQ004	2150	1.22	50	NE	634	865	995	1110	1240	1330	1420	405	521	595	665	741	805	873	-36.1	-39.8	-40.2	-40.1	-40.2	-39.5	-38.5	14
																		Absolute median	18.2	21.0	22.6	24.5	20.8	22.9	28.0	
																		Absolute maximum	53.9	51.5	45.1	41.9	47.8	55.8	65.0	
02ZM006	3.9	1.89	265	SE	3.29	4.46	5.28	6.11	7.25	8.14	9.08	3.77	5.2	6.11	7.01	8.1	8.93	9.68	14.5	16.7	15.7	14.7	11.8	9.8	6.6	43
02ZL003	10.8	1.95	319	SE	8.52	12.2	14.5	16.8	19.6	21.7	23.8	7.62	10.4	12.2	13.9	16.1	17.8	19.3	-10.6	-15.0	-16.1	-17.0	-17.7	-18.0	-18.8	18
02ZN002	16.6	1.75	512	SE	8.07	11.7	14.4	17.3	21.4	24.7	28.3	9.64	13	15.2	17.4	20.2	22.3	24.3	19.5	11.0	5.6	0.8	-5.7	-9.7	-14.3	12
02ZM016	17.3	1.84	148	SE	12.3	16.7	19.2	21.6	24.4	26.4	28.4	12.1	16.5	19.3	22.1	25.6	28.2	30.5	-1.9	-1.3	0.7	2.5	4.8	6.8	7.6	14
02ZL004	28.9	1.36	50	SE	16	22.7	26.8	30.6	35.3	38.8	42.1	20.7	28.4	33.4	38.2	44	48.5	52.4	29.4	25.3	24.7	24.9	24.7	24.9	24.4	14
02ZG004	42.7	1.83	123	SE	35.1	49	58	66.5	77.4	85.5	93.6	23.8	32.2	37.8	43.2	49.9	55.1	59.7	-32.2	-34.2	-34.8	-35.0	-35.5	-35.6	-36.2	16
02ZH002	43.3	1.87	20.8	SE	31.3	43.4	50.6	56.9	64.3	69.4	74.1	31.8	43.9	51.6	59	67.8	74.6	80.4	1.6	1.1	2.0	3.6	5.4	7.4	8.5	28
02ZN001	53.3	1.94	132	SE	37.9	47.4	52.8	57.5	63	66.8	70.5	27.6	37.3	43.7	50	57.7	63.7	69	-27.1	-21.4	-17.3	-13.1	-8.5	-4.7	-2.1	30
02ZM009	53.6	1.93	193	SE	26.5	29.5	30.8	31.8	32.6	33.1	33.5	26.1	35.1	41.1	47	54.3	60	65.2	-1.4	19.0	33.5	47.9	66.7	81.4	94.5	17
02ZG003	115	1.85	42.8	SE	55.5	79.9	97.2	115	138	157	176	57.4	77.6	91.1	104	120	132	143	3.3	-2.9	-6.3	-9.5	-13.2	-15.9	-18.9	17
02ZG002	166	1.85	588	SE	46.6	64.7	77.4	90	107	120	134	49.5	64.8	75.7	86.6	100	111	121	6.2	0.2	-2.3	-3.8	-6.3	-7.4	-9.6	20
02ZG001	205	1.91	202	SE	55.9	74.1	86.8	99.4	116	129	143	68.2	90	105	120	139	154	167	21.9	21.5	21.3	21.1	19.9	19.2	16.9	34
02ZK001	304	1.49	8.79	SE	144	199	238	277	329	370	412	148	202	237	270	310	341	367	2.8	1.3	-0.4	-2.5	-5.8	-7.9	-10.9	48
																		Absolute median	10.6	15.0	15.7	13.1	11.8	9.8	14.3	
																		Absolute maximum	32.2	34.2	34.8	47.9	66.7	81.4	94.5	
02ZK002	89.6	1.64	278	SW	71.1	106	132	158	195	225	256	81.5	114	137	160	193	218	245	14.6	7.6	3.9	1.3	-1.0	-3.2	-4.3	18
02ZK004	104	1.67	116	SW	91.3	130	162	197	249	294	342	82.5	114	137	158	189	211	236	-9.6	-12.0	-15.7	-19.7	-24.2	-28.1	-30.9	14
02YJ003	119	1.95	290	SW	31.3	36.6	38.7	40.2	41.7	42.5	43.2	41.1	53.7	61.3	68.1	77.3	83.2	89.7	31.2	46.7	58.5	69.5	85.3	95.9	107.7	11
02ZA003	139	1.66	131	SW	149	211	252	292	344	384	424	104	144	172	198	235	262	291	-30.0	-31.6	-31.9	-32.2	-31.7	-31.9	-31.4	15
02ZB001	205	1.52	50	SW	340	509	635	765	947	1090	1250	215	305	369	432	521	587	660	-36.7	-40.0	-41.9	-43.6	-45.0	-46.2	-47.2	35
02ZC002	230	1.3	38.4	SW	357	486	577	668	790	886	984	519	777	977	1188	1500	1746	2026	45.4	60.0	69.3	77.9	89.9	97.0	105.9	15
02ZA001	343	1.78	50	SW	115	156	182	205	234	254	273	138	183	210	234	265	285	306	20.0	17.4	15.3	14.1	13.3	12.0	12.0	18
02YJ001	640	1.67	141	SW	312	415	483	549	634	698	762	296	399	459	514	584	627	674	-5.0	-4.0	-5.0	-6.5	-7.9	-10.2	-11.5	28
02ZD002	1342	1.51	50	SW	851	1190	1390	1580	1800	1950	2090	836	1150	1341	1518	1745	1884	2040	-1.8	-3.3	-3.5	-3.9	-3.1	-3.4	-2.4	19
																		Absolute median	20.0	17.4	15.7	19.7	24.2	28.1	30.9	
																		Absolute maximum	45.4	60.0	69.3	77.9	89.9	97.0	107.7	

Table 4.4 **Testing of Regression Equations**

Station Number	dA (KM^2)	LSF	LAF	Region	Return Period Flood Estimates: Q2	Q5	Q10	Q20	Q50	Q100	Q200	Method	Regression Equation Estimates: Q2	Q5	Q10	Q20	Q50	Q100	Q200	Percentage Difference: Q2	Q5	Q10	Q20	Q50	Q100	Q200	n
02YL005	17	1.39	50	NW	12.6	16	18.3	20.3	22.9	24.8	26.8	FI	5.97	7.83	8.98	10	11.6	12.6	13.6	-52.6	-51.1	-50.9	-50.8	-49.5	-49.3	-49.2	12
02YK008	20.4	1.5	50	NW	9.77	12.4	14.2	15.8	17.7	19.2	20.8	FI	7.27	9.54	11	12.2	14.1	15.4	16.7	-25.6	-23.2	-22.8	-22.5	-20.3	-19.8	-19.5	12
02YA002	33.6	1.91	652	NW	14.5	18.4	21	23.4	26.3	28.5	30.8	FI	4.56	5.84	6.53	7.15	8.04	8.57	9.12	-68.5	-68.3	-69.0	-69.4	-69.5	-69.9	-70.4	11
02YL004	58.5	1.06	50	NW	43.4	55.2	63	70	78.8	85.3	92.3	FI	22.8	29.9	34.5	38.8	45.4	49.9	54.8	-47.5	-45.8	-45.2	-44.5	-42.4	-41.5	-40.6	13
02YN003	427	1.80	377	NW	99	132	156	181	217	246	277	FA	89	114	130	147	170	187	205	-10.0	-12.9	-16.4	-18.8	-21.5	-24.1	-26.0	9
02YL008	471	1.95	50	NW	250	295	317	332	348	357	364	FA	219	287	334	384	457	512	574	-12.4	-2.8	5.7	15.6	31.5	43.5	57.6	9
02YL001	2110	1.68	50	NW	582	709	787	859	948	1010	1080	FA	1111	1457	1712	1991	2404	2728	3105	90.8	105.5	117.5	131.8	153.6	170.1	187.5	68
																			Absolute median	47.5	45.8	45.2	44.5	42.4	43.5	49.2	
																			Absolute maximum	90.8	105.5	117.5	131.8	153.6	170.1	187.5	
02YO012	58.7	1.55	128	NE	14.5	18.6	21.7	25	29.7	33.6	37.9	FA	17.2	20.6	23.5	26.2	30.3	33	35.9	18.7	10.9	8.5	5.1	2.0	-1.9	-5.2	7
02YO010	61.6	1.79	601	NE	9.22	12	13.7	15.1	16.7	18	20.4	FI	10	12.6	14.2	15.3	17.4	18.8	20.2	9.0	4.5	3.8	1.5	4.2	4.2	-0.9	12
02ZJ002	73.6	1.72	435	NE	12.2	15.9	18.1	19.9	22.1	23.9	27	FI	13	16.4	18.5	20	22.8	24.7	26.5	6.7	2.8	2.3	0.5	3.1	3.2	-1.7	12
02YO007	88.3	1.57	50	NE	27.3	35.6	40.5	44.6	49.6	53.4	60.4	FI	33.6	43.5	50	56	63.6	69.6	75.7	22.9	22.0	23.5	25.6	28.4	30.2	25.3	12
02ZE004	99.7	1.81	50	NE	34.2	44.7	52.9	61.8	74.7	85.7	97.9	FA	36.9	53.3	61.6	69.8	81.9	90.5	99.9	7.9	19.2	16.5	12.9	9.7	5.6	2.0	7
02ZJ003	106	1.58	166	NE	30.1	39.3	44.6	49.2	54.6	58.9	66.6	FI	24.8	31.5	35.9	39.5	44.8	48.6	52.6	-17.7	-19.7	-19.6	-19.8	-18.0	-17.4	-21.0	11
02YN002	469	1.91	371	NE	172	224	255	281	312	337	380	FI	58.6	73.5	82.9	89.9	101	109	117	-65.9	-67.3	-67.5	-68.0	-67.6	-67.6	-69.2	16
02YQ006	531	1.16	50	NE	138	180	205	226	251	270	305	FI	136	327	381	438	522	585	657	-1.4	81.3	86.3	94.1	108.4	116.6	115.1	7
02YM001	982	1.80	36.4	NE	146	196	227	246	291	317	342	FA	247	723	849	983	1184	1338	1516	69.9	269.6	273.8	299.8	306.7	322.1	343.4	9
02ZF001	1170	1.84	401	NE	219	296	354	413	497	564	636	FA	116	145	163	177	198	213	229	-47.0	-50.9	-53.8	-57.2	-60.2	-62.3	-64.1	20
02ZE001	2640	1.92	619	NE	282	348	383	403	443	464	484	FA	186	231	259	278	310	332	355	-34.0	-33.5	-32.3	-30.9	-30.0	-28.6	-26.6	21
02YQ001	4400	1.82	277	NE	581	731	825	912	1020	1100	1180	FA	374	470	528	574	638	685	736	-35.6	-35.7	-35.9	-37.0	-37.4	-37.7	-37.6	47
																			Absolute median	20.8	27.8	27.9	28.3	29.2	29.4	26	
																			Absolute maximum	69.9	269.6	273.8	299.8	306.7	322.1	343.4	
02ZL005	11.2	1.95	272	SE	4.82	6.69	7.97	9.11	10.6	11.8	12.9	FI	8.02	10.9	12.8	14.7	17	18.8	20.3	66.4	63.5	61.0	61.4	59.8	59.5	58.1	12
02ZG005	28.2	1.40	132	SE	23.4	31	35.4	39.1	43.4	46.4	49.1	FA	17.5	23.7	27.9	31.9	36.8	40.6	44	-25.3	-23.4	-21.2	-18.5	-15.2	-12.4	-10.3	9
02ZK005	40.8	1.45	91.6	SE	24.5	34	40.5	46.3	54	59.8	65.4	FI	24.1	32.8	38.5	44	50.8	56	60.6	-1.5	-3.5	-5.0	-5.0	-6.1	-6.4	-7.4	11
02ZJ002	73.6	1.72	435	SE	12.2	15.2	18.2	21.9	28.2	34.6	42.7	FA	28.9	38.3	44.8	51.3	59.3	65.7	71.5	136.9	152.1	146.9	134.8	110.4	90.1	67.5	12
02ZE004	99.7	1.81	50	SE	34.2	44.7	52.9	61.8	74.7	85.7	97.9	FA	50.5	68.3	80.2	91.6	106	116	126	47.7	52.9	51.6	48.2	41.3	35.8	28.5	8
02ZJ003	106	1.58	166	SE	30.1	41.8	49.7	56.9	66.4	73.4	80.4	FI	43.7	58.3	68.3	78.1	90.2	99.7	108	45.2	39.7	37.3	37.3	35.8	35.7	34.6	11
02ZF001	1170	1.84	401	SE	219	296	354	413	497	564	636	FA	214	275	321	367	424	471	514	-2.1	-7.0	-9.3	-11.2	-14.7	-16.6	-19.3	20
																			Absolute median	45.2	39.7	37.3	37.3	35.8	35.7	28.5	
																			Absolute maximum	136.9	152.1	146.9	134.8	110.4	90.1	67.5	
02ZK003	37.2	1.24	50	SW	66.8	79.1	82.7	84.8	86.5	87.4	88	FA	183	288	380	485	652	800	974	174.4	264.2	359.8	471.4	653.7	814.8	1006.9	14
02ZA002	72	1.39	50	SW	49.9	69.8	81.5	93.4	110	124	138	FI	163	243	307	375	478	563	660	226.1	247.7	276.1	301.4	333.0	354.5	379.6	15
02ZE004	99.7	1.81	50	SW	34.2	44.7	52.9	61.8	74.7	85.7	97.9	FA	53.1	71.5	83.6	95	111	122	134	55.2	60.0	58.1	53.7	48.2	42.2	36.8	8
02YN002	469	1.91	371	SW	172	241	281	322	381	427	474	FI	120	154	172	188	206	216	228	-30.3	-35.8	-38.6	-41.7	-45.8	-49.3	-52.0	16
02ZF001	1170	1.84	401	SW	219	296	354	413	497	564	636	FA	276	355	393	424	461	477	496	26.1	19.9	11.1	2.8	-7.3	-15.4	-22.0	20
02ZE001	2640	1.92	619	SW	282	348	383	403	443	464	484	FA	394	492	530	557	584	588	596	39.6	41.2	38.5	38.4	31.8	26.6	23.1	21
																			Absolute median	47.4	50.6	48.3	47.7	47	45.7	44.4	
																			Absolute maximum	226.1	264.2	359.8	471.4	653.7	814.8	1007	

NOTES:

[heavy box] parameter out of range

[light box] parameter near range extremes

FA Frequency Analysis

FI Flood Index

Q200/Q2 at 1.32 was the smallest in the region and n = 14. Typical (median) absolute percentage difference statistics across all regions ranged from 20.8% to 50.6%. A detailed comparison between frequency analysis estimates and regression equation estimates for stations not used in the regional analysis is provided in Table 4.4. A flood index method (Appendix D) had to be used on some watersheds to obtain realistic flood estimates at high return periods. The LN3 and GEV distributions, which were used to model the high flows, provided unrealistic curve extensions at high return periods on some watersheds which had few data.

4.6 Comparison to Previous Studies

Regression equations can be evaluated in terms of the goodness of fit and / or the error associated with predicting the dependent variable. Previous regional flood frequency analyses (Beersing 1990, Panu et al 1984) used the correlation coefficient (R^2) as a measure of the goodness of fit and the standard error of the estimate (SEE) as a measure of the error associated with predicting the dependent variable. These studies attempted to maximize R^2 and minimize SEE. Towards this end, previous regression equations employed up to 5 independent variables with as few as 10 stations leaving only 4 degrees of freedom. The result was a near-perfect correlation (R^2=0.9998) and a very small standard error (<3%) which is unrealistic given the complexities of statistically modelling floods and given the very short periods of record. Improvements to R^2 and the SEE were only minimal when the fourth and fifth independent variables were forced into the regression equation. Increasing the number of independent variables and reducing the number of stations will generally increase R^2 and reduce SEE by decreasing the number of degrees of freedom available to the regression. This study used only 2 independent variables to explain 91% to 97% of the dependent variable in 3 of 4 regions and therefore must be considered to be more robust. This study used the Squared Multiple R (SMR) statistic to evaluate the goodness of fit and the SEE to assess the error associated with predicting the dependent variable. The SMR statistic, which is simply the square of Multiple R (R^2), was used because its value represents the proportion of the total variation in the dependent variable which can be explained by the independent variable(s). Regression statistics are provided in Table 4.5 by study.

Table 4.5 Regression Statistics by Study

Study	Region	Parameters	Range of R^2	Range of SMR	Range of SEE
1999	NW	DA, LAF	(0.980-0.982)	0.960-0.964	0.093-0.104
	NE	DA, LAF	(0.953-0.975)	0.909-0.950	0.117-0.161
	SE	DA, LAF	(0.964-0.983)	0.929-0.967	0.088-0.129
	SW	DA, LSF	(0.910-0.961)	0.829-0.924	0.140-0.237
1990	C (NW)	DA, LSF, SLP, DRD	0.93		0.081-0.096
	B (NE)	DA, DRD	0.97-0.98		0.079-0.098
	A (SE)	DA, LSF, DRD	0.96		0.098-0.111
	D (SW)	DA, LSF	0.94-0.97		0.120-0.160
1984	North	DA, MAR, LAT, SHAPE, BAREA,	0.9916-0.9998		2.6-19.9%
	South	DA, MAR, ACLS, SHAPE, SLOPE	0.9941-0.9982		12.5-24.2%
	Island	DA, MAR, ACLS, SHAPE	0.9883-0.9878		19.0-25.6%

Notes:
SEE in log_{10} units except for 1984 where they are given as a percentage.

4.7 Concurrent Evaluation of the 1984, 1990 and 1999 Regression Equations

The 1984, 1990 and 1999 regression equations were tested concurrently. The objective of the concurrent testing was to evaluate whether there has been an improvement in the predictive capabilities of successive regression equations. Three groups of watersheds were used: the independent watersheds which were used to evaluate the 1999 regression equations, watersheds which had calibrated and verified HYMO estimates, and watersheds with long periods of record. These groups of watersheds are discussed in more detail in the following paragraphs along with the results of the testing. The results of the testing for Q20 are shown graphically by watershed in: Figure 4.2 for the independent data set, Figure 4.3 for the HYMO data set, and Figure 4.4 for watersheds with long periods of record. These figures are located at the end of section 4.7.

4.7.1 Independent Watersheds

Twenty-three independent watersheds were used to assess the accuracy of the 1984, 1990 and 1999 regression equations. These watersheds were not used in the development of the 1999 equations. Some of them were used in the development of the 1984 and 1990 equations. These watersheds included those that had low confidence in the flood estimates, watersheds with less than 10 years of record, watersheds with large drainage areas, redundant watersheds, watersheds that were removed during the regression analysis, and watersheds that were eliminated from the regional data set for other reasons such as failing prerequisite statistical tests. The listing is provided in Table 4.4. Individual watersheds were sometimes placed into more than one region because these watersheds were near the boundary of 2 or more regions. It should be noted that the physiographic parameters for most of these watersheds are out of range for application of the regression equations. Despite the inadequacies of the database, it was considered sufficient for this purpose. The results are provided in Table 4.6. Return periods equal to 2, 10, 20 and 100 years were utilised in the 1984 analysis. The 1990 and 1999 analyses utilised return periods equal to 2, 5, 10, 20, 50, 100 and 200 years.

Table 4.6 Median Absolute Percentage Difference between Frequency Analysis Estimates and Regression Equation Estimates for the Independent Data Set

Region	1984	1990	1999
NW	69.5	52.7	68.1
NE	52.0	49.8	37.3
SE	36.1	35.7	25.3
SW	32.8	66.8	42.4

The application of regional regression equations is questionable when the physiography is outside the permissible range. Table 4.6 shows that the median absolute percentage difference between frequency analysis estimates and regional regression equation estimates have decreased with successive study. The table also shows that the 1999 equations were the better predictors in the NE and SE regions. The table also shows that the 1990 equations were best in the NW region and the 1984 equations were best in the SW region. It must be kept in mind that the flood data that were used for testing were less than ideal. The absolute percentage differences are somewhat bias between studies. Watersheds with drainage areas greater than 1000 km^2 were not used in the development of the 1999 regression equations. These watersheds were utilised in the development of the 1984 and 1990 regression equations. Since watersheds with drainage areas greater than 1000 km^2 were used for testing all equations, this would tend to inflate the absolute percentage difference for the 1999 equations. In addition, since the 1984 equations were based on shorter return periods, the median absolute percentage difference for the 1984 equations would be deflated somewhat.

4.7.2 Watersheds with HYMO Estimates

The 1984, 1990 and 1999 regression equations were also tested on watersheds which had calibrated and verified HYMO estimates. Over the past 15 years a number of flood risk mapping studies were conducted by consultants under the Canada-Newfoundland Flood Damage Reduction Program and the Canada-Newfoundland Agreement Respecting Water Resources Management. HYMO (HYdrologic MOdel) was used in these studies to simulate flood data from precipitation and other data. Estimates of Q20 and Q100 were obtained from frequency analysis of the simulated flood data.

Five flood studies were selected: Hydrotechnical Study of the Stephenville Area (1984), Flood Risk Mapping Study of the Codroy Valley Area (1990), Flood Risk Mapping Study - Trinity South Area (1995), Flood Risk Mapping Study: Goulds, Petty Harbour and Ferryland (1996), Flood Risk Mapping Study of Stephenville, Kippens and Cold Brook (1996). These flood studies covered four geographically diverse areas of the province. The Stephenville area was studied in 1984 and 1996. The individual watersheds are identified in Table 4.7

Regional flood frequency analysis estimates for 1984, 1990 and 1999 were compared to all HYMO estimates (return period flood estimates from the frequency analysis of flood data which was derived from HYMO modelling). The results are shown in Table 4.8. It should be noted that the physiographic parameters for most of these watersheds are out of range for application of the regression equations. Watersheds in the Stephenville area were tested in the NW and SW regions. All other watersheds were tested in the applicable region.

For this data set, the 1990 regression equation estimates had the lowest error, and the 1999 regression equation estimates provided an improvement over the 1984 regression equation estimates. This anomaly may be by chance or may be due to the fact that the 1990 regression equations used flood data up to 1988 which better coincides with the calibration

Table 4.7 Watersheds with Calibrated and Verified HYMO Estimates

Study	Watersheds	Drainage Area (km^2)
Stephenville Area (1984)	Warm Creek Blanch Brook	51.4 117
Codroy Valley Area (1990)	Little Codroy River South Brook North Brook Confluence of North and South Brooks Grand Codroy River near Upper Ferry	139 276.00 365.66 641.66 813.94
Trinity South Area (1995)	Hodges Cove Brook Hickman's Harbour River Shoal Harbour River	15.5 31 131
Goulds, Petty Harbour and Ferryland (1996)	Doyles River Cochrane Pond Brook	12.3 29.9
Stephenville, Kippens and Cold Brook (1996)	Gaudon Creek Cold Brook Blanch Brook	16.4 21.3 118.6

Table 4.8 Median Absolute Percentage Difference between HYMO Estimates and Regression Equation Estimates

Return Period	1984	1990	1999
Q20	74.4	45.5	65.2
Q100	75.4	49.1	66.2

and verification period of HYMO flood simulations than the period of record of the 1984 or 1999 regression equations. It was noted that flood estimates for Blanch Brook near Stephenville were obtained from calibrated and verified HYMO estimates in 1984 and 1996. Q20 went from 111 m^3/s in 1984 to 150 m^3/s in 1996 which was a 35% increase. Q100 went from 166 m^3/s in 1984 to 209 m^3/s in 1996 which was a 26% increase. Four of the five studies provided 1990 regression equation estimates as one of the alternative estimation techniques for Q20 and Q100. Agreement was sought between the 1990 regression equation estimates and the HYMO estimates.

4.7.3 Watersheds with Long Periods of Record

The 1984, 1990 and 1999 regression equations were also tested on watersheds which had long periods of record. The selection criteria for long term stations consisted of watersheds which were gauged for a period of at least 20 years and that the drainage areas of these watersheds had to be less than 1000 km^2. In addition, the selected watersheds had to pass all prerequisite statistical tests for frequency analysis. Fifteen (15) watersheds were identified. There were at least 2 stations per region. Almost all of the watersheds were utilized for regional analysis in the 1984, 1990 and 1999 studies. Regional flood frequency analysis estimates for 1984, 1990 and 1999 were compared to the single station flood frequency estimates. The results are shown in Table 4.9.

Table 4.9 Median Absolute Percentage Difference between Frequency Analysis Estimates and Regression Equation Estimates for Long Term Stations

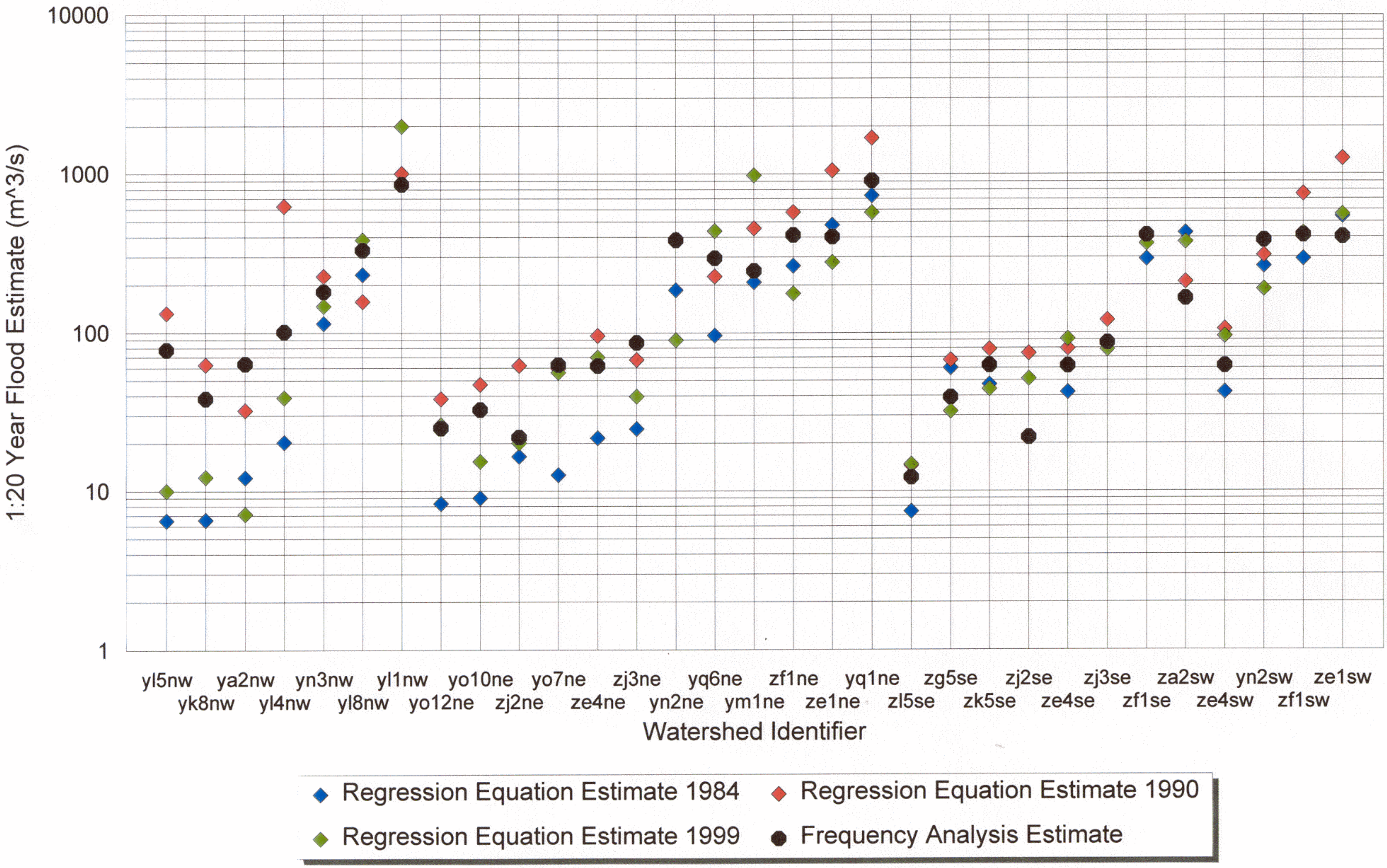
Return Period	1984	1990	1999
Q2	15.5	20.1	7.8
Q5	-	22	6.5
Q10	21.7	20.6	7.5
Q20	20.5	18.1	9.5
Q50	-	15.7	8.7
Q100	26.7	15.2	9.2
Q200	-	14.5	9

Table 4.9 shows is that successive regional flood frequency analyses provide

improved estimates of return period flood flows. The 1999 statistics may be somewhat deflated considering that only watersheds with drainage areas less than 1000 km^2 were considered for testing. The 1999 regression equations were based on watersheds with drainage areas less than 1000 km^2 while the 1984 and 1990 regression equations did not have an upper limit on the size of the drainage area.

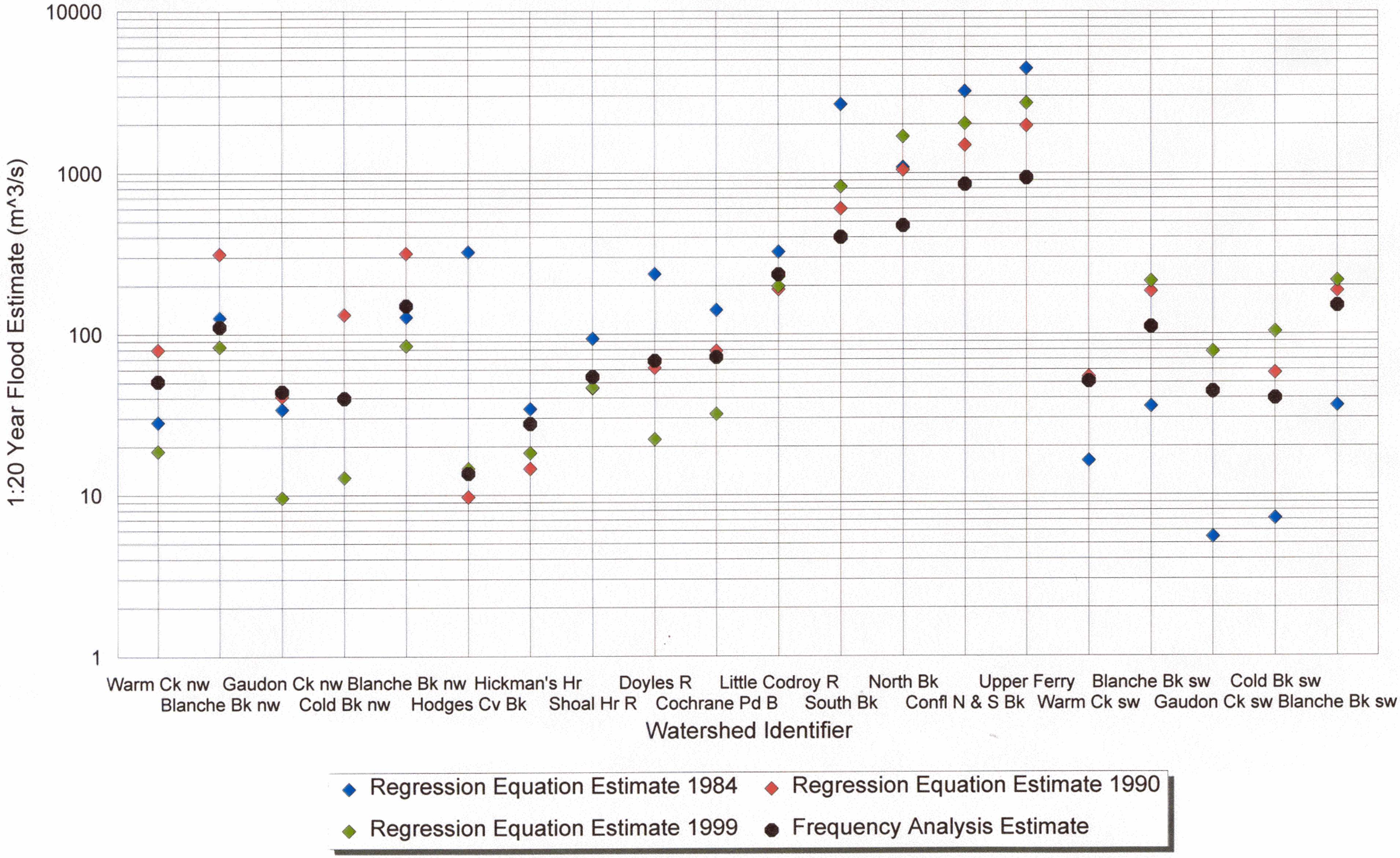
This page has been intentionally left blank.

Figure 4.2 Comparison between Regression Equation Estimates and Frequency Analysis Estimates for Q20
(Independent Data Set)



This page has been intentionally left blank.

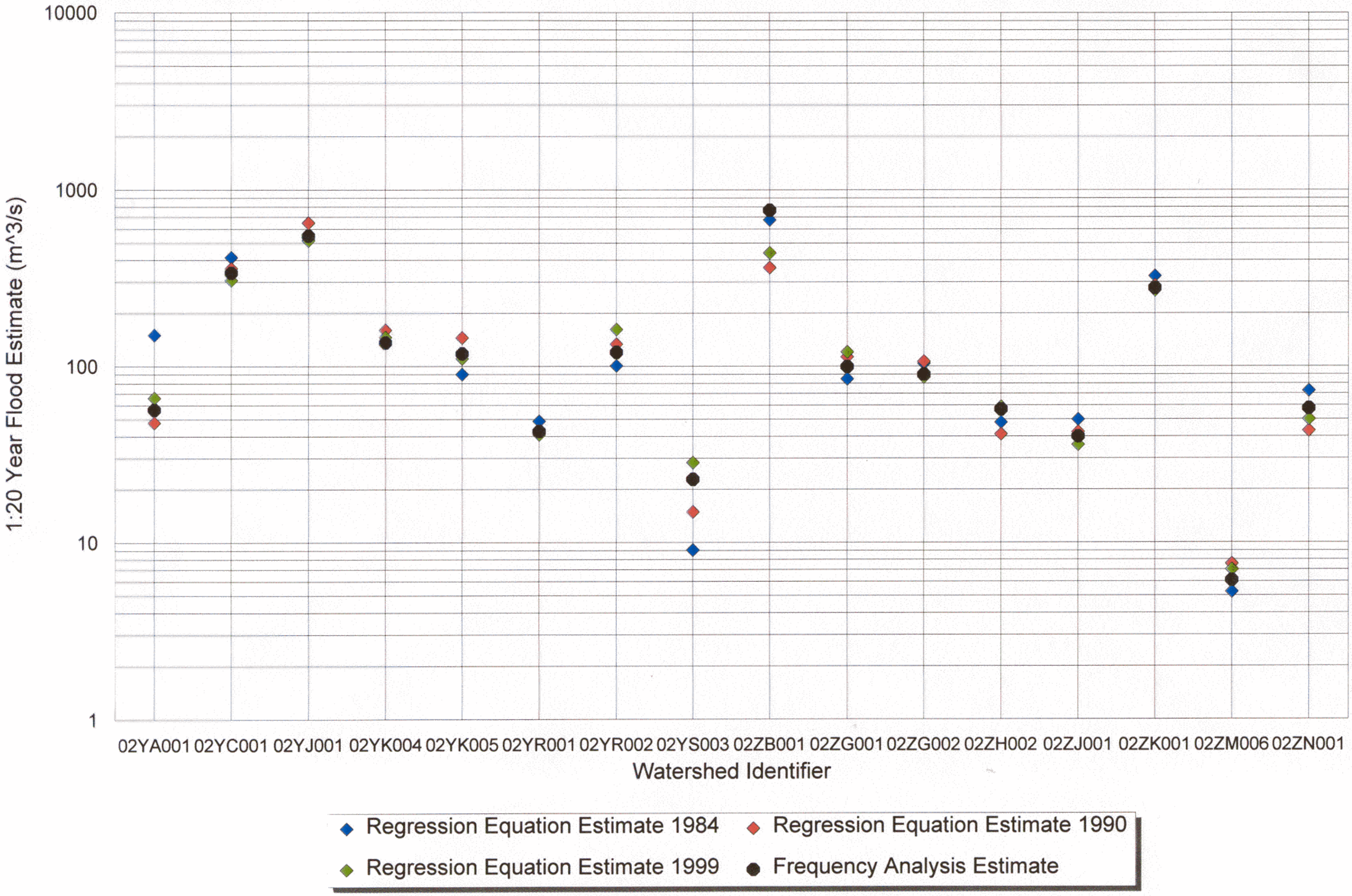
Figure 4.3 Comparison between Regression Equation Estimates and Frequency Analysis Estimates for Q20
(HYMO Data Set)



This page has been intentionally left blank.

Figure 4.4 Comparison between Regression Equation Estimates and Frequency Analysis Estimates for Q20

(Long Term Watersheds)



This page has been intentionally left blank.

5. APPLICATION OF REGRESSION EQUATIONS

5.1 General

While the regression equations derived in this study are recommended for estimating return period flood flows on ungauged watersheds, these equations cannot be used on all watersheds. Many ungauged watersheds have physiographic parameters which are outside the range of physiographic parameters which were used in the development of the regression equations. Section 5.2 provides the details. In addition, these equations cannot be used on urban watersheds or watersheds with significant regulation. Since regional boundaries are approximate, it may be necessary to obtain estimates assuming different regions. Where warranted, flood design should take on the upper 95% limit of the estimate.

Due care is required in the abstraction of physiographic parameters. This is particularly true for the Lakes and Swamps Factor (LSF) in the South-west Region where a 3% error in the abstraction of LSF may result in an error as high as 18% in the flood estimate. A sensitivity analysis is presented in Section 5.3

The confidence in the flood estimations in the South-west Region is lower than any other region. In addition, the equations are sensitive to the LSF parameter. The “Upper Envelope Curve” described in Section 4.4 may be warranted if the consequences of flooding are considerable.

Regional flood frequency analysis provides a method to estimate return period flood flows on ungauged watersheds based on the correlation between return period flood flows on gauged watersheds and their physiographic features. It is advisable to use several methods to estimate design floods. The regional equations provided in this study can be used with one or two physiographic parameters. Equations from previous regional flood frequency analysis should be used as checks. In addition, frequency analysis on nearby hydrologically similar watersheds may be transposed to the watershed under study. Regional flood index estimates (Appendix D) can also be used as a check. If a watershed fails to meet the physiographic prerequisites for regional equations or if a more thorough flood assessment is required, design flows could be simulated on an event basis for a given design storm or on a continuous basis from climatic data.

A *Users' Guide and Electronic Spreadsheet* for this regional flood frequency analysis is available under a separate cover. It documents a procedure for estimating flood flows using single station flood frequency analysis and regional flood frequency analysis. A worked example is provided. The electronic spreadsheet was created in Lotus and is available in a number of popular spreadsheet formats.

5.2 Allowable Range for Physiographic Parameters

The regression equations developed in this study cannot be used on all ungauged watersheds. Many ungauged watersheds have physiographic parameters which are outside the range of physiographic parameters which were used in the development of the regression equations. Caution is needed when applying regression equations to watersheds which have physiographic parameters near the extremes. Extrapolation of results beyond the extremes of physiographic parameters used in the development of the regression equation is not generally recommended. Physiographic extremes are listed by region in Table 5.1.

5.3 Sensitivity of Regression Equations

Small errors in the abstraction of physiographic parameters are inevitable. These errors may translate into larger errors in the estimation of return period floods. The sensitivity of the independent variables was assessed using the methodology described

Table 5.1 Extremes of Watershed Physiography by Region

Region	Statistic	DRAIN AREA (KM^2)	LAF (-)	LSF (-)	FRAC ACLS (-)	FRAC LAKE (-)	FRAC SWAMP (-)	FRAC L+S (-)	FRAC TREE (-)	FRAC BAR'N (-)	SLOPE M2 (%)	DRAIN DENSITY (KM^-1)	SHAPE FACTOR (-)	n
NW	Minimum	33.4	0	1.55	0.63	0.04	0.02	0.08	0.33	0.00	0.14	0.19	1.45	13
	2nd lowest	95.7	0	1.68	0.73	0.05	0.04	0.12	0.35	0.00	0.32	0.34	1.48	
	2nd highest	624	688	1.92	1.00	0.13	0.14	0.35	0.83	0.39	1.13	1.28	2.23	
	Maximum	627	1053	1.93	1.00	0.22	0.24	0.36	0.87	0.50	3.09	1.30	2.32	
NE	Minimum	36.7	0	1.22	0.44	0.02	0.06	0.11	0.11	0.00	0.12	0.26	1.43	15
	2nd lowest	63.8	0	1.40	0.55	0.03	0.06	0.13	0.55	0.00	0.16	0.35	1.62	
	2nd highest	2000	307	1.89	0.98	0.18	0.25	0.36	0.88	0.15	1.03	1.09	2.12	
	Maximum	2150	881	1.92	1.00	0.20	0.48	0.66	0.91	0.23	1.11	1.24	2.35	
SE	Minimum	3.9	0	1.36	0.39	0.04	0.00	0.04	0.09	0.00	0.23	0.55	1.24	13
	2nd lowest	10.8	8.79	1.49	0.57	0.04	0.00	0.09	0.16	0.04	0.34	0.96	1.36	
	2nd highest	205	512	1.94	1.00	0.13	0.06	0.16	0.75	0.73	2.22	1.55	2.06	
	Maximum	301	588	1.95	1.00	0.14	0.17	0.21	0.88	0.79	2.42	1.62	2.45	
SW	Minimum	89.6	0	1.30	0.34	0.04	0.01	0.06	0.04	0.04	0.34	0.15	1.54	9
	2nd lowest	104	0	1.51	0.60	0.04	0.02	0.09	0.08	0.07	0.35	0.72	1.68	
	2nd highest	640	278	1.78	0.91	0.08	0.16	0.31	0.79	0.78	1.27	1.50	2.45	
	Maximum	1340	290	1.95	1.00	0.15	0.38	0.46	0.86	0.82	1.46	1.73	5.31	

Abbreviations:

Region - hydrological region
DRAIN - drainage
LAF - lake attenuation factor
LSF - lakes and swamps factor
FRAC - fraction of watershed occupied by ...
ACLS - area controlled by lakes and swamps

LAKE - lake area
SWAMP - swamp area
L+S - lakes plus swamps area
TREE - forest area
BAR'N - barren area
SLOPE M2 - slope method 2

Table 5.2 Sensitivity of Regression Equations

Independent Variable (IV)	Anticipated error in IV	Region(s)	Maximum Anticipated Error in Dependent Variable (all return periods)
DA	3%	all	3.4%
LAF	3%	NW, NE, SE	1.3%
LSF	3%	SW	18.1%

in Appendix E and is shown in Table 5.2.

The anticipated error in the dependent variable ranged from 1.7% to 3.4% for all regions and return periods when the error in the abstraction of the most important independent variable (drainage area) was varied by 3%. Independent variable are considered insensitive to abstraction errors when the anticipated error in the dependent variable is near or less than the anticipated error in the independent variable. The LAF parameter was insensitive to abstraction errors. The LSF parameter was somewhat sensitive to abstraction errors.

6. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

6.1 Discussion

This study, like previous studies, derived a set of equations for estimating return period flood flows on ungauged watersheds. Flood flow estimates are required for the hydraulic design of instream structures and for floodplain management. Regular updates of the regional flood frequency analysis are required and justified on the basis of additional data, the capital costs of instream structures, and the annual cost of flooding.

The methodology for this study followed closely the methodology for previous studies. The characteristics of floods were examined along with climate and physiography. A flood database was created, missing data were estimated, and the flood series were screened. Single station frequency analysis conducted, hydrologically homogeneous regions were formulated, and multiple linear regressions were conducted. Since the results of the single station frequency analyses were used to examine the regional characteristics of floods, it may be prudent to conduct and report on the single station frequency analyses first. Examining the regional magnitude of return period floods which have been normalized for drainage area was particularly useful in the delineation of hydrologically homogeneous regions. The examination of the spatial distribution of the probability density function parameters was also illuminating.

The Generalized Extreme Value (GEV) and the Three Parameter Log-normal (LN3) distributions have been the two main probability distribution functions which were used for single station frequency analysis on the island of Newfoundland. These distributions provided unrealistic curve extensions at high return periods on some watersheds which had few data. A regional flood index method (Appendix D) had to be used on some watersheds in the independent data set to obtain realistic flood estimates at high return periods. This method can provide a useful check at long return periods on watersheds with less than 15 years of data. Consideration of other distributions is warranted where the fit is poor, or the distribution parameters indicate that another distribution may be better suited, or if there is an unrealistic curve extension at high return periods. The normal and log-normal distributions have been used with success on a few watersheds in the past.

The selection of a distribution was based on the mean absolute deviation between the theoretical and empirical probabilities of the upper half of the data set as defined in Section 3.4.2. This method of evaluating the relative goodness-of-fit was preferred over the KS test used in the 1990 study, or a comparison between the theoretical and actual values of skewness and kurtosis used in the 1984 study, because it quantifies the error associated with that portion of the curve with which we have the most interest: the 2-year return period up to T = 200.

Some watersheds had return period floods which were noticeably upper bounded. Flood flows on these watersheds seem to be attenuated by large lakes. The reasons for upper bounding need to be more fully understood. Upper bounding may be linked to, or provide a better understanding of, the Lake Attenuation Factor (LAF) parameter.

The upper 95% confidence limit of the 1:100 year return period flood estimate was greater than 100% of the estimate for 15 of the 65 watersheds which were considered for regional analysis. Confidence in return period flood estimates is generally low due to the shortness of flow records. These watersheds and others that were removed from the regional analysis were retained for testing the regression equations.

The regression equations developed in this study had a slightly higher standard error than the regression equations of 1984 or 1990. This study did not seek to minimize the error by including all statistically available physiographic parameters in the regression equations. This study used two physiographic parameters to explain 91% to 97% of the variation in flood flows for 3 of 4 regions. In the South-west Region 2 parameters explained 83% to 92%

of the variation in flood flows. The standard error in all regions was quite good considering the error associated with the annual maximum instantaneous discharges. Two physiographic parameters were favoured over only slight improvements to the standard error with the addition of a statistically significant third physiographic parameter.

Regional regressions went well for 3 of 4 regions. The higher spatial variability of flood characteristics and climatic indices in the South-west Region made modelling this region difficult. An “Upper Envelope Curve” was developed for this region due to the relatively larger standard error and the sensitive second parameter (LSF), and should be considered in situations where severe floods could be catastrophic.

The median absolute percentage difference between the frequency analysis estimates and the regression equation estimates for all stations used in the regional analysis ranged from 8.0% to 30.9%. The median absolute percentage difference between the frequency analysis estimates and the regression equation estimates for stations not used in the regional analysis ranged from 20.8% to 50.6%. These percentages are very good considering that the independent data set was composed primarily of stations which were removed from the main data set because of the low confidence in their flood estimates.

The regression equations derived in this study cannot be used on all watersheds. Many ungauged watersheds have physiographic parameters which are outside the range of physiographic parameters which were used in the development of the regression equations. Watersheds which have physiographic parameters which are out of range are subject to large errors in the flood estimates as demonstrated in Section 4.5.

The regression equations provided in this study can provide two estimates of return period flood flows on ungauged watersheds. It is advisable to use several methods to estimate design floods. Since this regional flood frequency analysis does not invalidate previous analysis, those equations may be used as checks. In addition, frequency analysis on nearby hydrologically similar watersheds may be useful. Regional flood index estimates can also be used as a check. Flood flows can also be simulated using computer intensive techniques on an event basis for a given design storm, or on a continuous basis from climatic data.

Previous studies attempted to maximize the correlation and minimize the error by including all statistically significant variables in the regional regression equations. The result was, near-perfect correlations and very small errors associated with the flood estimates. The near-perfect correlations and small errors are unrealistic given the complexities of statistically modelling floods, and give the error associated with flood values. Since accurate estimates can be obtained with only 2 parameters, and because the parameters are consistent through all return periods in 3 of 4 regions, the regression equations developed in this study must be considered more robust.

Concurrent testing of the 1984, 1990 and 1999 regression equations on the 23 independent watersheds pointed towards an increase in the accuracy of the regression equations with successive study. The flood data and physiographic data for these watersheds made them much less than ideal for testing the regression equations. The flood data sometimes had large confidence intervals. Physiography was quite often out of range. Including watersheds with drainage areas greater than 1000 km^2, and including all available return periods in the analysis purposefully biased the results towards earlier studies and against this study. The analysis still pointed towards an increase in accuracy with successive study. Testing on the 15 watersheds with calibrated and verified HYMO estimates indicated that the 1990 regression equations provided the best estimates. This may be due to the fact that the 1990 regression equations used flood data up to 1988 which better coincides with the calibration and verification period of HYMO flood simulations than the periods of record for 1984 or 1999 regression equations. The results did however indicate an improvement in predictive capability of the 1999 equations over the 1984 equations. Testing on the 15 watersheds which had long periods of record indicated that successive study improved the predictive capabilities of the regression equations. Almost all of these watersheds were used in the development of the 1984, 1990 and 1999 regression equations. This data set does not constitute an independent data set but can be used to assess the relative accuracy of successive study. The results may have been slightly biased towards agreement with the 1999 regression equations due to the 1000 km^2 maximum limit on drainage area for this test. This bias is removed if we accept that regional flood frequency estimation will not likely be the primary flood estimation technique on watersheds greater than 1000 km^2 in drainage area. This is acceptable since many of the larger watersheds are gauged.

6.2 Conclusions

1. The Three Parameter Log-normal distribution was the best fitting distribution for more than half of all gauged watersheds. The Generalized Extreme Value distribution was the better fitting distribution on the remaining gauged watersheds.

2. About 25% of all watersheds with 10 or more years of record did not provide good estimates of return period flows due to the shortness of streamflow records. These watersheds had an upper 95% confidence limit on the 1:100 year return period which was double the estimate.

3. The regression equations derived in this study can provide robust estimates of return period flood flows on the island of Newfoundland. Testing indicated that typical errors (median absolute percentage differences) ranged from 20.8% to 50.6%. Since the watersheds which were used for testing had flood estimates with low confidence, the typical error range given above is somewhat inflated. Typical errors for watersheds used in developing the regression equations ranged from 8.0% to 30.9%. Typical error for estimates on ungauged watersheds is suspected to be in the 15% to 40% range or about 30%.

4. The standard error of the estimate for regional regression equations was slightly larger in this study than in previous studies. This study used only two physiographic parameters to explain 91% to 97% of the variation in flood flows for 3 of 4 regions, and 83% to 92% of the variation in flood flows in the South-west Region. Considering the error associated with measuring flood flows, the standard error of the estimate is quite acceptable. Previous studies may have included too many parameters.

5. Concurrent testing of the 1984, 1990 and 1999 regression equations on the independent data set (23 watersheds), the HYMO data set (15 watersheds), and watersheds with long periods of record (15 watersheds), indicated that there was improvement in the predictive capability of the regression equations for successive regional flood frequency analysis.

6. Drainage area was the most important variable in the estimation of return period flood flows in all regions. Some problems were encountered in the South-west Region where drainage area was forced to be the most important variable by including watersheds on the south-west portion of the Avalon Peninsula that had similar flood characteristics. The Lake Attenuation Factor was the next most important variable in 3 of 4 regions. The Lakes and Swamps Factor (LSF) was the next most important variable in the South-west Region. After the inclusion of two variables in the regional regression equations, nearly all of the variation in flood flows was explained.

7. The LSF parameter was somewhat sensitive to abstraction errors. This variable needs to be abstracted with due care. A 3% abstraction error in LSF can result in an error which is as high as 18% in the flood estimate. The drainage area and LAF parameters were insensitive to minor abstraction errors.

8. The number of watersheds to which the regression equations may be applied has increased. This study had more watersheds available for analysis and thus a broader spectrum of physiography. The minimum drainage area in the North-west Region decreased from 93.2 km^2 to 33.5 km^2. The minimum FACLS was reduced in the other regions.

9. Problems were encountered in the South-west Region. Flood characteristics, climatic conditions and physiography are not quite homogenous in this region. The "Upper Envelope Curve" provides the best estimate of extreme floods in the South-west Region.

6.3 Recommendations

1. The regional regression equations developed in this study are recommended for estimating return period flood flows on ungauged watersheds or watersheds with less than 10 to 20 years of flood data. In the South-west Region, the "Upper Envelope Curve" is recommended where flooding may threaten life or cause severe flood damages.

2. It is recommended that the regional flood frequency analysis be updated in 5 years. The next analysis should consider improvements to the second parameter and a change in approach to the South-west Region.

3. More streamflow gauges are required along the south coast from Isle aux Morts River to Bay du Nord River.

4. Urban watersheds have been removed from the analysis, yet flood estimates are required in urban areas. A guideline needs to be developed for flood estimation in

urban areas.

5. Eighteen (18) regulated watersheds were removed from the analysis. Considering the low density of the hydrometric network, and the longer lengths of record on many of the regulated watersheds, extraction of flood information from these watersheds may be warranted.

6. Application of regression equations to small watersheds (< 50 km^2) is sometimes difficult because flood processes on small watersheds are different than those on larger watersheds. Generally, flood flows per unit area are greater on smaller watersheds due to the shorter travel time and the higher average precipitation rates. In addition, smaller watersheds are more likely to have physiographic parameters which are beyond the range of the physiographic parameters which were used in the development of the regression equations. A separate model, based on precipitation data and physiography, is recommended for watersheds with areas less than 50 km^2. The number of gauged watersheds (non-urban and unregulated) with drainage areas less than 50 km^2 is 15. If the drainage area threshold is moved from 50 to 70 km^2, then the number of gauged watersheds for analysis would increase to 23.

7. REFERENCES

Beersing, A. K., Regional Flood Frequency Analysis for the Island of Newfoundland, Government of Newfoundland and Labrador, Department of Environment and Labour, Water Resources Division, January 1990.

Condie, R., The Three Parameter Log Normal Distribution Applied to Regional Flood Frequency Analysis by the Index Flood Method, Technical Workshop Series No. 2, Inland Waters Directorate, Environment Canada, Ottawa, 1980.

Darymple, Flood Frequency Analysis, U. S. Geological Survey, Water Supply Paper 1543A, 1960.

Environment Canada, Consolidated Frequency Analysis Package 88 (User's Manual and Computer Program), Inland Waters Directorate, Ottawa, 1985.

Environment Canada, Surface Water Data Up To 1996 (CD-ROM), January 1998.

Environment Canada, Canadian Monthly Climate Data and 1961-1990 Normals (CD-ROM), 1994 Release, 1994.

Fisheries and Environment Canada, Hydrological Atlas of Canada, 1978.

Goulding, K., and Lye, L. M., Hydrometeorological and Physiographic Data Abstraction for the Island of Newfoundland: An Update, Memorial University of Newfoundland, for the Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division, April 1989.

Government of Newfoundland and Labrador (Department of Environment) and Environment Canada, Hydrotechnical Study of the Stephenville Area, 1984.

Government of Newfoundland and Labrador (Department of Environment and Lands) and Environment Canada, Flood Risk Mapping Study of the Codroy Valley Area, 1990.

Government of Newfoundland and Labrador (Department of Environment) and Environment Canada, Flood Risk Mapping Study - Trinity South Area, 1995.

Government of Newfoundland and Labrador (Department of Environment) and Environment Canada, Flood Risk Mapping Study: Goulds, Petty Harbour and Ferryland, 1996.

Government of Newfoundland and Labrador (Department of Environment) and Environment Canada, Flood Risk Mapping Study of Stephenville, Kippens and Cold Brook, 1996.

Grubbs, F. E., and Beck, G., Extension of Sample Sizes and Percentage Points for Significant Tests of Outlying Observations, Technometrics, Vol. 14, No. 4, pp. 847-8543, 1972.

Hydrology Subcommittee, Guidelines for Determining Flood Flow Frequency, Bulletin #178, Interagency Advisory Committee on Water Data, U. S. Geological Survey, Office of Water Data Coordination, Reston, Virginia 22092, 1982.

Kindervater, A. D., Flooding Events in Newfoundland and Labrador: An Historical Perspective, Environment Canada, Inland Waters Directorate, Water Planning and Management Branch, Atlantic Region, Halifax, Nova Scotia, July 1980.

Kite, G. W., Frequency and Risk Analysis in Hydrology, Water Resources Publications, Colorado, 1977.

MacPherson, A. G., and MacPherson, J. B. (editors), The Natural Environment of Newfoundland, Past and Present, Memorial University of Newfoundland, Department of Geography, 1981.

Poulin, R. Y., Flood Frequency Analysis for Newfoundland Streams, Water Planning and Operations Branch, Department of Environment, Ottawa, 1971.

Panu, U. S., Smith, D. A., and Ambler D. C., Regional Flood Frequency Analysis for the Island of Newfoundland, Canada-Newfoundland Flood Damage Reduction Program, Newfoundland and Labrador Department of Environment and Lands, St. John's; Environment Canada, Dartmouth, 1984.

Ullah, W. (editor), Water Resources Atlas of Newfoundland, Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division, January 1992.

Appendix A

Physiographic Parameters: Description and Abstraction

This page has been intentionally left blank.

Appendix A

Physiographic Parameters: Description and Abstraction

A number of physiographic parameters were selected for use in this study. This Appendix describes these parameters and indicates how they were determined.

The parameters selected and their units were:

l.	Drainage area (DA)	(km^2)
2.	Fraction of watershed occupied by forest (TREE)	(-)
3.	Fraction of watershed occupied by wetlands (SWAMP)	(-)
4.	Fraction of watershed occupied by lakes (LAKE)	(-)
5.	Fraction of watershed occupied by lakes and swamps (L+S)	(-)
6.	Fraction of watershed occupied by barrens (BAR'N)	(-)
7.	Fraction of drainage area controlled by lakes and swamps (FACLS),	(-)
8.	Lakes and swamps factor (LSF)	(-)
9.	Lake attenuation factor (LAF)	(-)
10.	Length of the main river (LENGTH)	(km)
11.	Elevation Difference (ELEVDIFF)	(m)
12.	Slope of the main channel method 1 (SLOPE1)	(%)
13.	Slope of the main channel method 2 (SLOPE2)	(%)
14.	Drainage Density (DD)	(km/km^2)
15.	Shape Factor (SHAPE)	(-)

Drainage area (DA), Fraction of watershed occupied by forest (TREE), Fraction of watershed occupied by swamps (SWAMP), and Fraction of watershed occupied by lakes

(LAKE), were determined from 1:50,000 scale National Topographic Series (NTS) maps using either a planimeter, a digitizer or a transparent grid with 0.01 km^2 blocks. Fraction of watershed occupied by barrens (BAR'N) was obtained by subtracting TREE, SWAMP and LAKE from DA. Fraction of watershed occupied by lakes and swamps (L+S) was calculated by summing LAKE and SWAMP. For basins with drainage areas greater than 2000 km^2, 1:250,000 scale NTS maps were used. Since less lakes and swamps are shown on 1:250,000 scale maps than on 1:50,000 scale maps, the values were adjusted upward based on comparisons of "representative" sample portions of each basin at each of the two scales. The area of forest and barren were then adjusted downward proportionally.

The fraction of the drainage area controlled by lakes and swamps (FACLS) was determined using 1:50,000 scale NTS mapping for all basins. A sub-basin was considered controlled if a lake or swamp at the outlet of the sub-basin had a surface area greater than 1% of the sub-basin. "Percentage of Basin Area Controlled by Lake and Swamp" is defined by Poulin (1971) in Figure A1.

Lakes and swamps factor (LSF) is a combination of the Fraction of drainage area occupied by lakes and swamps (L+S) parameter and the Fraction of watershed area controlled by lakes and swamps (FACLS) parameter. The algorithm is:

$$LSF = (1 + FACLS) - (L+S) / (1 + FACLS).$$

Lake attenuation factor (LAF) is a factor which sums the product of individual large (> 1% of DA) lake areas with their corresponding drainage areas. The algorithm is:

$$LAF = \sum_{i=1}^{n} \{(100 \text{ x } LAREA_i/DA) \text{ x } (100 \text{ x } CAREA_i/DA)\}$$

where, n is the number of lakes in the watershed with area greater than 1% of the watershed's drainage area, $LAREA_i$ is the area of a lake, DA is the drainage area of the watershed, and $CAREA_i$ is the drainage area which is controlled by a lake. Two lakes can be considered as one if they are hydraulically connected during high flows. LAF is defined in Figure A2.

Length of the main river (LENGTH) was determined using a map meter and 1:50,000 scale NTS mapping. The main river was the longest river in the watershed.

Elevation Difference (ELEVDIFF) was the difference in elevation between the outlet of the watershed and the highest point on the divide in the vicinity if the main channel.

Slope of the main channel method 1 (SLOPE1) was simply ELEVDIFF divided by LENGTH.

Slope of the main channel method 2 (SLOPE2) was the average slope of the curve that joins two points on the main river which are at 10% and 85% of LENGTH from the outlet. In effect, the slope of the main river was calculated over only 75% of its length.

Drainage Density (DD) was determined by dividing the total length of streams by the drainage area.

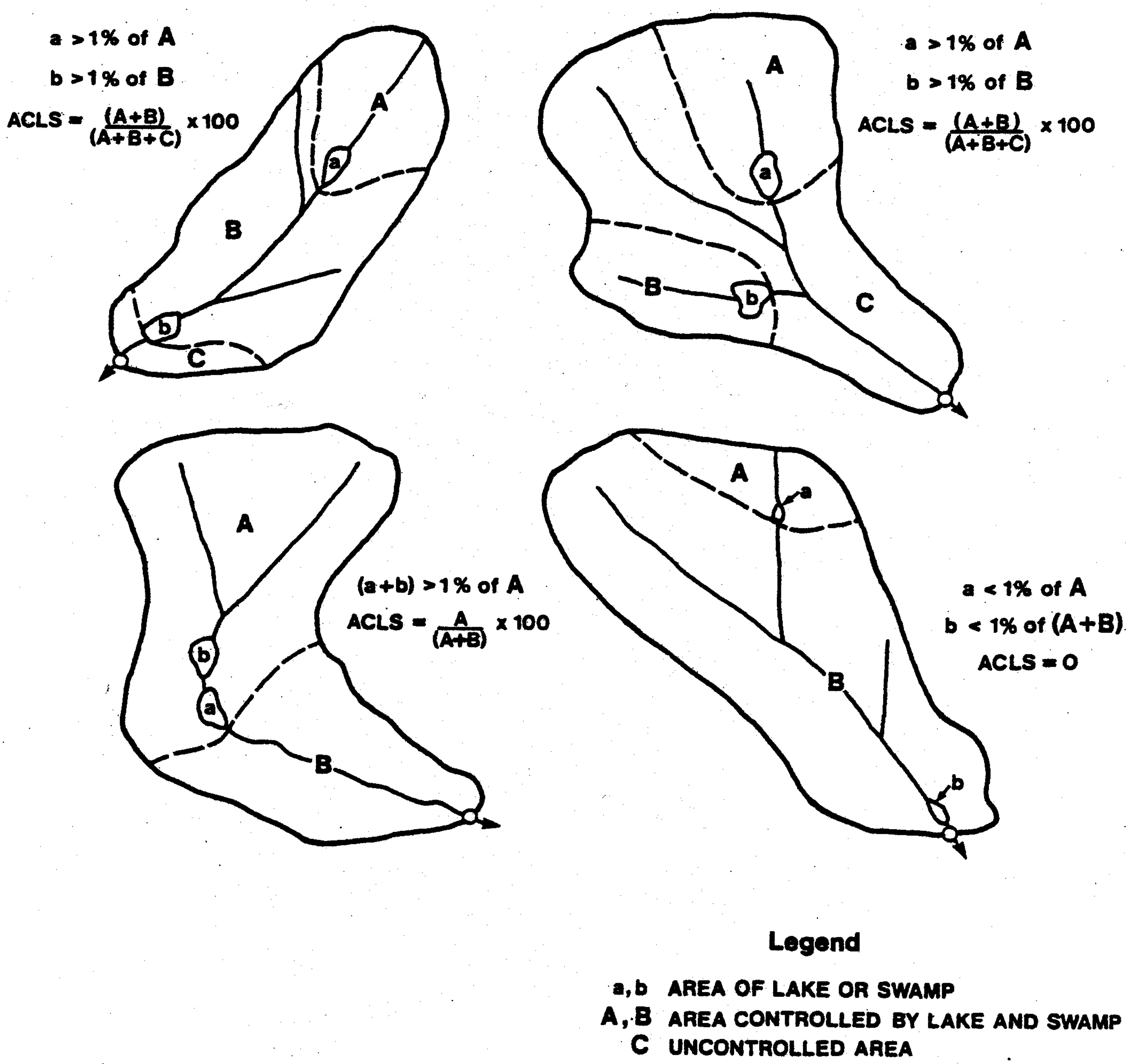
Shape factor (SHAPE) characterizes the physical shape of the watershed. The algorithm is:

$$\text{SHAPE} = 0.28 \times P / \sqrt{DA}$$

where P is the perimeter of the watershed, and DA is the drainage area. A circle would have a SHAPE of 1.00

Figure A1

PERCENTAGE OF BASIN AREA CONTROLLED BY LAKE AND SWAMP (ACLS) - DEFINITION

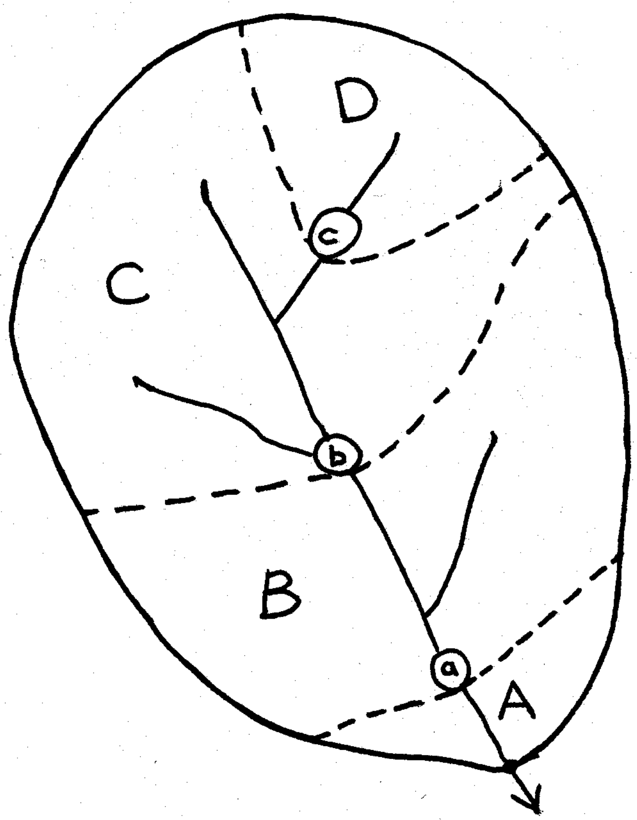


After Poulin (1971)

Figure A2 Lake Attenuation Factor (LAF) - Definition

$$LAF = \sum_{i=1}^{n} \{(100 \times LAREA_i/DA) \times (100 \times CAREA_i/DA)\}$$

where: n is the number of lakes in the watershed with area greater than 1% of the watershed's drainage area,
$LAREA_i$ is the area of a lake,
DA is the drainage area of the watershed, and
$CAREA_i$ is the drainage area which is controlled by a lake.



If area of lakes a, b, c > 1% of the drainage area (A+B+C+D), then

$$LAF = \{(100 \times a/(A+B+C+D)) \times (100 \times (B+C+D)/(A+B+C+D))\} + \{(100 \times b/(A+B+C+D)) \times (100 \times (C+D)/(A+B+C+D))\} + \{(100 \times c/(A+B+C+D)) \times (100 \times (D)/(A+B+C+D))\}$$

This page has been
intentionally
left blank.

Appendix B

Review of Data Sets for

Single Station Flood Frequency Analysis

This page has been intentionally left blank.

Appendix B

Review of Data Sets for Single Station Flood Frequency Analysis

Station Number	Comments:
02YA001	This data set had a relatively long period of record (1970-96, n=27) with only 2 missing values (1976, 1979) which were estimated with confidence. The data displayed high skewness (ln(x)>0.4). All values were retained for frequency analysis.
02YA002	This data set had a relatively short period of record (1986-96, n=11) with 2 missing values (1986, 1987) which were estimated. The data displayed high skewness. All values were retained for frequency analysis.
02YC001	This data set had a relatively long period of record (1959-96, n=36) with only 2 missing values (1984, 1994) which were estimated with confidence. All values were retained for frequency analysis.
02YD001	This data set had 19 values and a period of record from 1960 to 1978. There were no missing values. All values were retained for frequency analysis.
02YD002	This data set had 17 values and a period of record from 1980 to 1996. There was one missing value (1986) which was estimated. It was the third highest value. All values were retained for frequency analysis.
02YE001	This data set had a relatively short period of record (1984-96, n=13) with 3 missing values (1984, 1995, 1996) which were estimated. Two of the missing values were the largest on record. All values were retained for frequency analysis.
02YF001	This data set had a relatively short period of record (1969-82, n=14). There were no missing values. All values were retained for frequency analysis.
02YG001	This data set had a relatively short period of record (1986-96, n=11). There were 2 missing values (1989, 1996) which were estimated with confidence. All values were retained for frequency analysis.
02YG002	This data set had a relatively short period of record (1987-96, n=10). There were no missing values. All values were retained for frequency

analysis.

02YH001 — This data set had a relatively short period of record (1985-96). There was one missing value (1986) which was estimated with confidence. The data set was highly skewed. One high outlier (1993) was detected. After removal of the high outlier, skewness was reduced to an acceptable level. All values were retained for frequency analysis except for the high outlier (n=11).

02YJ001 — This data set had a relatively long period of record (1969-96, n=28) with only 2 missing values (1972, 1994) which were estimated with confidence. All values were retained for frequency analysis.

02YJ003 — This data set had a relatively short period of record (1986-96, n=11). There were no missing values. All values were retained for frequency analysis.

02YK002 — This data set had a relatively long period of record (1954-96). The following years were missing 1955, 1967-72, and 1981. The time series of annual maximum instantaneous discharges showed an increase in the magnitude of average peak flows of nearly 70% for the 1973-96 period over the 1954-66 period. Environment Canada indicated that the outlet channel of this watershed had been excavated in 1962, causing a decrease in the water level on Little Grand Lake. A decrease in the natural storage of this lake near the outlet lessened the attenuating effect this lake had on peak flows. Only the 1973-80 and 1982-96 portions of the records (n=23) were retained for frequency analysis.

02YK003 — This data set had a relatively short period of record (1956-66, n=11). There were no missing values. All values were retained for frequency analysis.

02YK004 — This data set had 23 values and a period of record from 1957 to 1979. There was one missing value (1979) which was estimated with confidence. Environment Canada indicated that the outlet channel of this watershed had been excavated in 1962. The data set failed the split sample homogeneity test by time at a 5% level of significance but not at a 1% level of significance. Values were retained only for the years 1963-79 (n=17).

02YK005 — This data set had 24 values and a period of record from 1973 to 1996. There were no missing values. A low outlier was detected and removed for curve fitting.

02YK007 This data set had a relatively short period of record (1984-96, n=13). There were 2 missing values which were estimated. All values were retained for frequency analysis.

02YK008 This data set had a relatively short period of record (1985-96, n=12). There was one missing value which was estimated. It was the highest value on record. All values were retained for frequency analysis.

02YL001 This data set had a relatively long period of record (1929-96, n=68). There were no missing values. All values were retained for frequency analysis.

02YL004 This data set had a relatively short period of record (1983-96, n=14). There was one missing value which was estimated. The data set displayed high skewness. One high outlier was detected and removed (January 1983 flood). The skewness was significantly reduced. All other values were retained for frequency analysis (n=13).

02YL005 This data set had a relatively short period of record (1985-96, n=12). There were 4 missing values which were estimated. Two of these values were the two highest on record. The data set displayed high skewness. All values were retained for frequency analysis.

02YM001 This data set had a relatively long period of record (1955-96, n=41). There was one missing value. In 1963, a partial diversion of the watershed split the data set in two parts: 1955-63 (n=9) and 1964-96 (n=32). A visual check on the annual maximum instantaneous discharges did not reveal any change in peak flow rates. In addition, a split sample test by time did not find any significant difference in the populations. The first data set was too small for frequency analysis. The second data set receives an ungauged amount of water from another basin. This data set was discarded.

02YM002 This data set had 15 values and a period of record from 1964 to 1978. This station metered the diverted part of 02YM001. This diversion, however, is only partial, and therefore, this data set was discarded.

02YM003 This data set had 17 values and a period of record from 1980 to 1996. There were 2 missing values which were estimated with confidence. All values were retained for frequency analysis.

02YN002	This data set had a period of record from 1981 to 1996. There were no missing values. All values were retained for frequency analysis.
02YO006	This data set had 16 values and a period of record from 1981 to 1996. There were no missing values. One high outlier was removed (January 1983 flood) which reduced the skewness to an acceptable level. All other values were retained for frequency analysis (n=15).
02YO007	This data set had a relatively short period of record (1984-95, n=12). There were 5 missing values which had to be estimated! Two of these values were the two highest on record. The data set displayed high skewness. All values were retained for frequency analysis.
02YO008	This data set had a relatively short period of record (1984-96, n=13). There were 7 missing values which had to be estimated! One of these values was the highest on record. All values were retained for frequency analysis.
02YO010	This data set had a relatively short period of record (1985-96, n=12). There were 7 missing values which had to be estimated! One of these values was the highest on record. All values were retained for frequency analysis.
02YP001	This data set had a relatively short period of record (1982-96, n=15). There were 2 missing values which were estimated with confidence. A trend was detected at a 5% level of significance. The data set displayed high skewness. A high outlier (January 1983 flood) was detected and removed. After removal of the high outlier, the trend disappeared as well as the high skewness. All values excluding the high outlier were retained for frequency analysis (n=14).
02YQ001	This data set had a relatively long period of record (1950-96, n=47). There were no missing values. One low outlier was detected and removed for curve fitting.
02YQ002	This data set had a relatively short period of record (n=14). In addition, the data set is relatively old (1924-39). Annual maximum instantaneous discharges were estimated from the annual maximum daily discharges. This gauged was moved further downstream to the 02YQ001 location in 1950. The 02YQ002 data set was discarded in favour of the 02YQ001 data set.
02YQ004	This data set had a relatively short period of record (1982-96, n=14).

There were 6 missing values which were estimated! Two of these values were the highest on record. All values were retained for frequency analysis.

02YQ005	This data set had a relatively short period of record (1987-96, n=10). There were 5 missing values which were estimated! Four of these values were the highest on record! A significant trend was detected at 5% level of significance but not at a 1% level of significance. All values were retained for frequency analysis.
02YR001	This data set had a relatively long period of record (1959-96, n=38). There were 5 missing values which were estimated. While there was a fire in 1961, there was no evidence to suggest that it had a significant effect on peak flows. Previous studies deleted the years 1959-66. All values were retained for frequency analysis.
02YR002	This data set had 20 values and a period of record from 1977 to 1996. There were 4 missing values which were estimated. One of the missing values was the highest on record. The data set displayed high skewness. All values were retained for frequency analysis.
02YR003	This data set had 16 values and a period of record from 1981 to 1996. There were 2 missing values which were estimated. All values were retained for frequency analysis.
02YS001	This data set had a relatively long period of record (1953-84). There were 2 missing values (1979, 1984). The 1979 value could not be easily estimated. The 1984 value was estimated. This watershed was regulated from 1951-62 for the purpose of driving pulpwood. All readings up to 1966 were of questionable quality due to a discharge measurement technique specific to this site. The values for 1967-78 and 1980-84 (n=16) were retained for frequency analysis.
02YS003	This data set had a relatively long period of record (1968-96, n=29). There were 2 missing values (1979, 1984) which were estimated with confidence. All values were retained for frequency analysis.
02YS005	This data set had a relatively short period of record (1985-96, n=12). There were no missing values. All values were retained for frequency analysis.
02ZA001	This data set had 18 values and a period of record from 1979 to 1996. There was one missing value (1979) which was estimated with

confidence. All values were retained for frequency analysis.

02ZA002 This data set had 15 values and a period of record from 1982 to 1996. There were no missing values. This data set displayed high skewness. There were three "high" values. All values were retained for frequency analysis.

02ZA003 This data set had 15 values and a period of record from 1982 to 1996. There was one missing value (1983) which was estimated with confidence. All values were retained for frequency analysis.

02ZB001 This data set had a relatively long period of record (1962-96, n=35). There were no missing values. All values were retained for frequency analysis.

02ZC002 This data set had 15 values and a period of record from 1982 to 1996. There were 3 missing value (1982, 1983, 1991) which were estimated. All values were retained for frequency analysis.

02ZD002 This data set had 19 values total. While the period of record went from 1970 to 1996, the following years were missing: 1972-78, 1981, 1982, 1993 and 1994. The 1982, 1993 and 1994 values were estimated. All values (n=19) were retained for frequency analysis.

02ZE001 This data set had 21 values total. The period of record went from 1944 to 1965. The years 1944-49 were missing. The 1944-48 values were estimated. All values (n=21) were retained for frequency analysis.

02ZF001 This data set had a relatively long period of record (1951-96). There were 2 missing values (1979, 1980). The 1979 value was estimated. The 1980 value could not be easily estimated. The 1983 flood was a high outlier and was removed. The data set failed the test for trend at a 5% level of significance. In addition, the data set failed the split sample homogeneity test by time (1951-74, 1975-96) at a 1% level of significance! Peak flows appear to be increasing from 1951 to 1984 and then appear to decrease with time. This station was not retained for frequency analysis because of the trend and non-homogeneity. The reasons for the trend / non-homogeneity are not known.

02ZG001 This data set had a relatively long period of record (1959-96). There were 4 missing values (1990, 1991, 1994, 1996) which were estimated with confidence. This data set displayed high skewness and failed the test for independence at a 5% level of significance. The 1959 to 1962 values were deleted from the data set. A previous hydrometric data review saw these values increase from 40% to 105%. In addition, the 1962 value was a significant outlier which could not be confirmed by surrounding stations. After these values were removed, skewness was reduced to acceptable level and the data set passed the test for independence. The values for 1962-96 (n=16) were retained for frequency analysis.

02ZG002 This data set had 20 values and a period of record from 1977 to 1996. There was one missing value (1994) which was estimated with confidence. All values were retained for frequency analysis.

02ZG003 This data set had 17 values and a period of record from 1980 to 1996. There were 2 missing values (1987, 1989) which were estimated with confidence. All values were retained for frequency analysis.

02ZG004 This data set had 16 values and a period of record from 1981 to 1996. There were 2 missing values (1993, 1996) which were estimated with confidence. All values were retained for frequency analysis.

02ZH001 This data set had a relatively long period of record (1953-96). There were 3 missing values (1988, 1990, 1994) which were estimated. The data set showed a significant (increasing) trend at a 1% level of significance! A fire in 1961 had a significant effect on peak flows. Since most of this basin had low scrub trees, the vegetation was reestablished a few years later. When the values for 1953 to 1964 were deleted from the data set, the trend disappeared. The 1965-96 (n=32) values were retained for frequency analysis.

02ZH002 This data set had a relatively long period of record (1961-96). The years 1962-69 were missing and could not be easily estimated. There were 3 other missing values (1961, 1970, 1984) which were estimated from the maximum daily discharge records. Twenty-eight values (n=28) were retained for frequency analysis.

02ZJ001 This data set had 20 values and a period of record from 1977 to 1996. There were 2 missing values (1978, 1988) which were estimated with confidence. All values were retained for frequency analysis.

02ZJ002 This data set had short period of record (1983-96, n=14). There were 4 missing values (1987, 1988, 1991, 1992) which were estimated. The 1992 estimated value was the highest on record. The data set displayed high skewness and a high outlier was detected. After removal of the high outlier, the data set still displayed high skewness and another high outlier was detected. After the removal of the second outlier the data set still displayed high skewness. Considering the small sample size, the number of missing data, and the high skewness, this station was not retained for frequency analysis.

02ZJ003 This data set had short period of record (1986-96, n=11). There was one missing value (1988) which was estimated. The data set displayed high skewness. The data set also failed the test for randomness at a 5% level of significance. The number of runs above and below the median was 10 which indicated a very short term cycle. Since there was no known physical reason for this cycle, it was assumed to be by chance. A certain number of test failures can be expected given the level of significance and the number of tests performed. This data set was retained for frequency analysis.

02ZK001 This data set had a relatively long period of record (1949-96, n=48). There were 3 missing values (1961, 1989, 1993) which were estimated with confidence. All values were retained for frequency analysis.

02ZK002 This data set had 18 values and a period of record from 1979 to 1996. There were no missing values. All values were retained for frequency analysis.

02ZK003 This data set had short period of record (1983-96, n=14). There was one missing value which was estimated. All values were retained for frequency analysis.

02ZK004 This data set had short period of record (1983-96, n=14). There were no missing values. The data set displayed high skewness. All values were retained for frequency analysis.

02ZK005 This data set had short period of record (1986-96, n=11). There were 5 missing values which were estimated. The data set displayed high skewness. All values were retained for frequency analysis.

02ZL003 This data set had 18 values and period of record from 1979 to 1996. There were no missing values. All values were retained for frequency analysis.

02ZL004 This data set had a short period of record (1983-96, n=14). There were 2 missing values which were estimated. All values were retained for frequency analysis.

02ZL005 — This data set had a short period of record (1985-96, n=12). There were no missing values. The data set displayed high skewness. All values were retained for frequency analysis.

02ZM006 — This data set had a long period of record (1954-96, n=43). There were 16 values which were estimated from the annual maximum daily discharges. Two of these values were the highest on record. All values were retained for frequency analysis.

02ZM009 — This data set had a period of record from 1979 to 1996. There were no missing values. The data set displayed high skewness and a high outlier (1994) was detected. After removal of the high outlier the skewness was reduced to an acceptable level. All values were retained for frequency analysis except for the 1994 value (n=17).

02ZM016 — This data set had a short period of record (1983-96, n=14). There were no missing values. All values were retained for frequency analysis.

02ZN001 — This data set had a long period of record (1966-95, n=30). There were 2 missing values (1971, 1986) which were estimated with confidence. This data set failed the test for trend at a 5% level of significance. Peak flows appear to be increasing with time. Since there was no known physical reason for this trend it was assumed to be by chance. A certain number of test failures can be expected given the level of significance and the number of tests performed. All values were retained for frequency analysis.

02ZN002 — This data set had a short period of record (1985-96, n=12). There were 5 missing values which were estimated. This data set failed the test for trend at a 1% level of significance! Contrary to 02ZN001, peak flows appear to be decreasing with time. Since there was no known physical reason for this trend it was assumed to be by chance. A certain number of test failures can be expected given the level of significance and the number of tests performed. All values were retained for frequency analysis.

This page has been intentionally left blank.

Appendix C

Statistical Tests Performed on the Flood Data

This page has been
intentionally
left blank.

Appendix C

Statistical Tests Performed on the Flood Data

C1 INTRODUCTION

Statistical frequency analysis assumes that the sample to be analyzed is a reliable set of measurements of independent random events from a homogeneous population which is free from erroneous data. The validity of this assumption can be verified using statistical tests. A computer program CFA88 (Environment Canada, 1985) was used to test the peak streamflow database developed for this study for independence, trend, homogeneity, general randomness and outliers. These tests are briefly discussed in the following sections.

C2 DESCRIPTION OF STATISTICAL TESTS

Brief descriptions of the rationale for each test are given here. The theory of the tests is not given; however, the required functions to be evaluated and the determination of their significance is given in Section C4.

C2.1 Spearman Rank Order Serial Correlation Coefficient Test for Independence

Two events can be considered independent only if the probability of occurrence of either is unaffected by the occurrence of the other, and this definition can be extended to a sample of size N. Practically, in a time series, independence can be measured by the significance of the correlation coefficient between the N-1 pairs of the ith and (i+l)th members of the series and, if the correlation coefficient is not significantly greater than zero, then independence is assumed. It is noted here that, in the strict mathematical sense, this does not necessarily define independence. To avoid the assumptions made in the derivation of the sampling distribution of the Pearson correlation coefficient, the nonparametric Spearman rank order serial correlation coefficient is used.

C2.2 Spearman Rank Order Correlation Coefficient Test for Trend

If successive measurements of members of a time series have been made during a period of gradually changing conditions, then there will be a more or less noticeable trend in the magnitude of the members of the series when arranged in chronological order. As an example, it would be expected that gradual land use changes in a drainage basin would affect the magnitude of the annual flood. Similarly, long term climatic changes will be reflected in the hydrology of a basin, although it is customary to assume climatic time invariance. A very simple test for presence of trend and its significance can be made on the Spearman rank order correlation coefficient. The computation and determination of its significance are given in Section C4.

C2.3 Mann-Whitney Split Sample Test for Homogeneity

If some more or less abrupt change occurred during the sampling period, then some difference could be expected between the means of the subsamples before and after the change. Examples could include the construction of an ungated reservoir in the basin, or a forest fire which denuded a substantial portion of the basin. Assuming a normal distribution, and that the two subsamples have the same variance, then the difference in the subsample means can be tested for significance using the Student's t distribution. These assumptions are not commonly met in hydrology and so the Mann-Whitney nonparametric test is used. With two subsamples of approximately the same size, it would be expected that if there were no change in conditions, the sums of the ranks of the two subsamples would not differ by too much. For a given data set, the question to be answered is whether the difference is significant or not. The Mann-Whitney U statistic is a function of the subsample sizes and their sums of ranks. The distribution of U statistics is known and critical values at various levels of significance have been tabulated. Hence, a decision can be made on whether the means of the subsamples differ significantly. Computation methods are shown in Section C4.

The Mann-Whitney test can also be used to decide whether hydrologic events occurring in different seasons have a significant difference of means. The sample is simply split by seasons rather than by time spans, although the computations are identical. When splitting by seasons, two runs of the program are needed. The program CFA88 provides a histogram of the data sample by months of occurrence, and the user may then choose the most sensible split from the hydrologic point of view. It was found impossible to make any arbitrary seasonal split because, in many cases, one of the subsamples would contain no data.

C2.4 Wald-Wolfowitz Split Sample Test for Homogeneity

This test can determine whether two samples have significantly different means, variances, skewness or kurtosis but it is not as powerful as the Mann-Whitney test in detecting a significant difference in means. On the other hand, the Mann-Whitney test cannot detect differences between the other statistical parameters of the two subsamples. The program ranks the entire sample in descending order and determines and lists the subsample (either 1 or 2) from which the corresponding data item came. If the sample is well mixed, the 1's and 2's will be well mixed and the number of runs will be relatively high. A run is defined as a sequence of identical symbols preceded and followed by a different symbol or no symbol at all. Consider the case where the mean of subsample 2 is significantly greater than that of subsample 1, then 2's will tend to cluster towards the top end of the scale and 1's toward the bottom, thus reducing the number of runs. Suppose now that the variance of subsample 2 was substantially greater than that of subsample 1. The 2's will then tend to cluster towards both ends of the scale, again reducing the number of runs. Similarly, differences in skewness and kurtosis will cause clustering, reducing the number of runs. The distribution of the number of runs is known and lower critical values have been tabulated. Thus, it can be determined if the subsamples differ significantly in any respect.

Like the Mann-Whitney test, the Wald-Wolfowitz test can be used to decide whether hydrologic events occurring in different seasons differ significantly.

C2.5 Runs Above and Below the Median for General Randomness

This is a very simple nonparametric test. Data are ranked in chronological order and the median determined. Identifiers A or B or * are assigned to each data point according to whether the corresponding data item is above or below or equal to the median, and the number of runs are counted. Theoretically, the number of runs, RUNAB, could be as high as N, indicating an extreme short term cyclic pattern, or as low as 2, indicating an abrupt change halfway through the period over which the sample was collected. Notice that the median is used since the probability of exceeding the median is always 0.5, regardless of the probability distribution from which the sample was drawn, thus making the test nonparametric or distribution free. The distribution of RUNAB is known and upper and lower critical values have been tabulated, thus enabling a decision to be made on whether the data are random or not.

C2.6 Grubbs and Beck Test for Outliers

The presence of outliers in a data sample will cause difficulties in satisfactorily fitting a frequency distribution to the sample. Depending on whether the outliers are high or low, and on the chosen frequency distribution, the estimates of the T-year event will often be underestimated or overestimated. Techniques are available for approximately dealing with these outliers; but, these outliers must first be detected.

C3 LIMITATIONS

C3.1 Sample Sizes

For each data sample to be analyzed, the lower limit to the sample size is 10 and the upper limit is 200. There is no limit to the number of samples which can be processed in one computer run.

For the Mann-Whitney test, the program will automatically split a sample size N into equal subsamples of size N/2 for N even, or two subsamples of (N-1)/2 and (N+1)/2 for N odd. The user has the option of splitting the sample into other subsample sizes, subject to the following limitations:

n_2 greater than 20	n_1 not less than 5
n_2 between 8 and 20	n_1 not less than 2
$n_2 = 7$	n_1 not less than 3
$n_2 = 6$	n_1 not less than 4
$n_2 = 5$	$n_1 = 5$

In the Wald-Wolfowitz test, the sample will be automatically split as in the Mann-Whitney test and again the user has the option of other subsample sizes. Where both subsamples are less than 21, the smaller subsample must not be less than 3; and if one of the subsamples is greater than 20, the smaller subsample must not be less than 5.

C3.2 Significance Levels

The first two tests both use the Spearman rank correlation coefficient, and its significance is tested at both the five and one percent levels.

The significance of the Mann-Whitney U statistic is tested at both the five and one percent levels.

If both subsamples in the Wald-Wolfowitz test are less than 21, testing is at the five percent level only. If one of the subsamples is greater than 20, the test is at both the five and one percent levels.

For the general randomness test, if the total sample size is greater than 40 after deleting sample members equal to the median, then testing is at both the five and one percent levels. Otherwise, the test is at the five percent level only.

The significance of the Grubbs and Beck statistic is tested at the 10 percent level.

C4 FUNCTIONS EVALUATED IN PROGRAM CFA88

This section briefly summarizes the functions evaluated in program CFA88 and gives the methods used to determine their statistical significance.

C4.1 General

Any statistical test of significance will generally be made using the following steps.

a) State the null hypothesis, H_o. For instance, in split sample tests, the null hypothesis may be that there is no difference between the sample means.
b) Choose a significance level α.
c) Choose an appropriate statistical test.
d) Compute the test statistic.
e) The sampling distribution of the test statistic is known and has been tabulated, and the chosen significance level then defines the region of rejection.
f) If the computed test statistic lies in the region of rejection, then the null hypothesis is rejected.

C4.2 Evaluation of Test Statistics

C4.2.1 Spearman Rank Order Serial Correlation Coefficient for Independence

If the series Q_i with i ranging from 1 to N is put in chronological order, ranks assigned, and denoting the series:

$Q_1, Q_1 \;\ldots\; Q_{N-1}$ by s_i, the rank of Q_i
and $Q_2, Q_3 \;\ldots\; Q_N$ by y_i, the rank of Q_i,

then the Spearman rank order serial correlation coefficient is:

$$S_1 = \tfrac{1}{2}(\Sigma x_i^2 + \Sigma y_i^2 - \Sigma d_i^2)(\Sigma x_i^2 . \Sigma y_i^2)^{-\frac{1}{2}} \qquad \text{C4.1}$$

where $\Sigma x_i^2 = (m^3 - m)/12 - \Sigma T_x$
$\Sigma y_i^2 - (m^3 - m)/12 - \Sigma T_y$
d_i is the difference in rank between x_i and y_i.
$m = N-1$

and the summations are over the m pairs of x_i, y_i.

Ignoring for the moment the terms in T and putting them at zero, equation C4.1 becomes:

$$S_1 = 1-(6\Sigma d_i^2)/(m^3-m)$$

the more familiar form of the Spearman rank correlation coefficient.

The terms in T adjust for tied ranks and are computed as follows. If for instance, three observations in the x series were tied for ranks 17, 18, and 19, then each observation is given the rank 18; if two were tied for ranks 24 and 25, then each is ranked 24.5.

For each tied set, T is computed from:

$$T_x - (t^3-t)/12$$

where t is the number or observations tied at a given rank.
ÓT$_x$ and ÓT$_y$ are defined by extension of the foregoing.

For N less than 10, special tables are available for defining the region of rejection for a computed S_1 at given significance level á. When N is 10 or greater, then the function:

$$t = s_1[(m-2)/(1-s_1^2)]^{½}$$

is distributed like Student's t with m-2 degrees of freedom. A one-tail test must be used.

C4.2.2 <u>Spearman Rank Order Correlation Coefficient Test for Trend</u>

If the series Q_i with i ranging from 1 to N is put in chronological order, ranks assigned and denoting the series:

Q_1, Q_2 . . . Q_N by y_i, the rank of Q_i
and 1, 2, . . . N by x_i, the sequential order of Q_i,

then the Spearman rank order correlation coefficient r_s is calculated as in equation C4.1, except that m = N, $T_x = 0$, and the summations are taken over the N pairs of x_i, y_i.

For N less than 10, special tables define the region of rejection for a computed value of r_s at a given significance level á.

For N = 10 or greater, then the function:

$$t = r_s[(N-2)/(1-r_s^2)]^{½}$$

is distributed like Student's t with N-2 degrees of freedom. The null hypothesis is that there is no trend, either upward or downward with time, and so a two-tail test is used.

C4.2.3 Mann-Whitney Split Sample Test for Homogeneity

As described in Section A.2, the sample is split into two subsamples, and ranks assigned. Then the Mann-Whitney U statistic is defined by the smaller of:

$$U_1 = n_1n_2 + n_1(n_1 + 1)/2\text{-}R_1$$
$$U_2 = n_1n_2 - U_1$$

where n_1 is the size of the smaller subsample,
n_2 is the size of the larger subsample,
and R_1 is the sum of the ranks in subsample n_1.

For both n_1 and n_2 less than 21, the critical values of U have been tabulated which define the region of rejection. For n_1 greater than 4 and n_2 greater than 20, the sampling distribution of U rapidly tends to normality with:

$$z = \frac{U\text{-}n_1n_2/2}{\{[(n_1n_2)/(N(N\text{-}1))][(N^3\text{-}N)/12 - \text{Ó}T]^{½}}$$

T = $(t^3 - t)/12$, where t is the number of observations tied at a given rank. The summation of T is over all groups of tied observations in both subsamples.

z is an N (0, 1) variate and in the applications of the Mann-Whitney test used in this program, the region of rejection is:

z less than - 1.645 for á = 0.05
z less than - 2.326 for á = 0.01

C4.2.4 Wald-Wolfowitz Split Sample Test for Homogeneity

Having determined the number of runs, R_{ww}, as explained in Section C2, the method by which its significance is determined depends on the subsample sizes, n_1 and n_2. When both n_1 and n_2 are less than 21, the critical values of R_{ww} which define the region of rejection have been tabulated. For n_1 greater than 4 and n_2 greater than 20, the sampling distribution of R_{ww} tends to normality with

$$z = \frac{* \; R_{ww}\text{-}[(2n_1n_2)/(n_1+n_2)+1] - 0.5 \; *}{\{2n_1n_2(2n_1n_2 - n_1 - n_2)/[(n_1 + n_2)^2(n_1 + n_2 - 1)]\}^{½}}$$

z is an N (0, 1) variate, and in the applications of the Wald-Wolfowitz test used herein, the region of rejection is:

z greater than 1.645 for $\alpha = 0.05$
z greater than 2.326 for $\alpha = 0.01$

Theoretically, ties cannot occur in the Wald-Wolfowitz test, since in its derivation the samples are assumed to be drawn from continuous distributions. In hydrology practice, published values of flows have been rounded to comply with some rule for significant figures and ties are very common. If for any tied group, all members are in the same subsample, there is no problem; but if members of one subsample are tied with members of the other subsample, then there is no unique ordered series and hence, no unique value of R_{ww}. For instance, if a quartet of ties had two members in each subsample and a duo of ties had one member in each subsample, then there are 12 possible ordered series and the test becomes meaningless. In this program, if ties are split between subsamples, the Wald-Wolfowitz statistic is not computed.

C4.2.5 Runs Above and Below the Median for General Randomness

Section C2 explains how the number or runs, RUNAB is determined, and for n_1 A's and n_2 B's with n_1 and n_2 both less than 21, the region of rejection is defined by tables. For n_1 and n_2 both greater than 20, the sampling distribution of RUNAB tends to normality with:

$$z = \frac{\left| RUNAB - [(2n_1n_2)/(n_1 + n_2) + 1] \right|}{\{2n_1n_2(2n_1n_2 - n_1 - n_2)/[(n_1 + n_2)^2(n_1 + n_2 - 1)]\}^{1/2}}$$

z is an N (0, 1) variate and, as used in this program, the region of rejection is:

z greater than 1.96 for $\alpha = 0.05$
z greater than 2.326 for $\alpha = 0.01$

C4.2.6 Grubbs and Beck Test for Outliers

The theory of outliers is till incomplete and has only been satisfactorily developed for a normal distribution. Application of the test is simple, requiring only the mean and standard deviation of the sample, and tabulated values of the Grubbs and Beck (1972) statistic for various sample sizes and significance levels.

The Grubbs and Beck outlier test has been adopted in modified form by the Hydrology Subcommittee (1982) of the United States. Since the test is applicable only to samples from a normal population, the assumption is made that the logarithms of the sample members are normally distributed. Rearranging the Grubbs and Beck test as done by the Hydrology Subcommittee (1982), the two following equations are obtained:

$$X_H = \exp\,(\mu + K_N\ l)$$

$$X_L = \exp\,(\mu - K_N\ l)$$

where: X_H is the lower limit of the high outliers,
X_L is the higher limit of the low outliers,
μ is the mean of the natural logarithms of the sample,
l is the standard deviation of the natural logarithms of the sample,
and K_N is the Grubbs and Beck statistic.

Typical values of K_N range from 2.036 for n = 10 to 3.017 for n = 100.

This page has been intentionally left blank.

Appendix D

Regional Flood Index

This page has been intentionally left blank.

Appendix D

Regional Flood Index

Poulin (1971) used an index flood technique for regional flood frequency analysis on the island of Newfoundland. This technique uses an average dimensionless flood frequency curve for all watersheds in a hydrologically homogeneous region. This analysis method is inferior to multiple linear regression between return period flood flows and physiographic parameters, because flood index techniques do not take into account the varying amount of influence that the physiographic parameters have on floods of different magnitudes. None the less, the application of these ratios to hydrologically similar watersheds with few data may provide better estimates of flood flows at the longer return periods than single station frequency analysis. The table below provides summary statistics which indication of the ratios between the return period flood flows (Q5, Q10, Q20, Q50, Q100 and Q200) and the median flood (Q2).

Table D1 Flood Ratios

Region	Statistic	Q5/ Q2	Q10/ Q2	Q20/ Q2	Q50/ Q2	Q100/ Q2	Q200/ Q2	n
NW	minimum	1.19	1.31	1.40	1.49	1.56	1.59	13
	median	1.27	1.45	1.61	1.82	1.97	2.13	
	maximum	1.34	1.60	1.89	2.30	2.63	2.99	
NE	minimum	1.16	1.24	1.29	1.34	1.37	1.40	15
	median	1.31	1.48	1.63	1.82	1.96	2.21	
	maximum	1.43	1.73	1.98	2.43	2.81	3.34	
SE	minimum	1.11	1.16	1.20	1.23	1.25	1.26	13
	median	1.39	1.65	1.89	2.21	2.44	2.67	
	maximum	1.45	1.78	2.14	2.65	3.06	3.51	
SW	minimum	1.17	1.24	1.28	1.33	1.36	1.38	9
	median	1.40	1.63	1.87	2.21	2.48	2.76	
	maximum	1.50	1.87	2.25	2.79	3.22	3.75	

Appendix E

Derivation of Equation to

Determine Sensitivity of Physiographic Parameters

This page has been intentionally left blank.

Appendix E

Derivation of Equation to Determine Sensitivity of Physiographic Parameters

The sensitivity of each accepted regression equation to a specified change in each parameter is summarized in Table 5.2. The development of the equation used to analyze parameter sensitivity is outlined in this Appendix.

Determine the effect on y_1 as a result of a factor ß applied to a parameter x_n:

$$\log_{10}y_1 = k + a_1\log_{10}x_1 + a_2\log_{10}x_2 + \ . \ . \ . \ + a_n\log_{10}x_n \qquad D.1$$

$$\log_{10}y_2 = k + a_1\log_{10} ßx_1 + a_2\log_{10}x_2 + \ . \ . \ . \ + a_n\log_{10}x_n \qquad D.2$$

Rewrite the two equations as:

$$y_1 = 10^{(k + a_1\log_{10}x_1 + a_2\log_{10}x_2 + \ . \ . \ . \ + a_n\log_{10}x_n)}$$

$$y_2 = 10^{(k + a_1\log_{10}ßx_1 + a_2\log_{10}x_2 + \ . \ . \ . \ + a_n\log_{10}x_n)}$$

Determine the relative difference between y_2 and y_1 from:

$$\frac{y_2 - y_1}{y_1} = \frac{10^{(k + a_1\log_{10}ßx_1 + a_2\log_{10}x_2 + ... + a_n\log_{10}x_n)} - 10^{(k + a_1\log_{10}x_1 + a_2\log_{10}x_2 + ... + a_n\log_{10}x_n)}}{10^{(k + a_1\log_{10}x_1 + a_2\log_{10}x_2 + ... + a_n\log_{10}x_n)}}$$

or $$\frac{y_2 - y_1}{y_1} = 10^{a_1(\log_{10}ßx_1 - \log_{10}x_1)} - 1$$

or $$\frac{y_2 - y_1}{y_1} = 10^{a_1(\log_{10}ß + \log_{10}x_1 - \log_{10}x_1)} - 1$$

or $$\frac{y_2 - y_1}{y_1} = 10^{a_1\log_{10}ß} - 1$$

or similarly for any parameter n and expressed in percent:

$$\frac{y_2 - y_1}{y_1} \text{ x } 100\% = (10^{a_n \log_{10} ß} - 1) \text{ x } 100\%$$