RELATIONSHIPS OF FLOW AND BASIN VARIABLES ON THE ISLAND OF NEWFOUNDLAND, CANADA WITH A REGIONAL APPLICATION

SUSAN RICHTER

Relationships of flow and basin variables on the island of Newfoundland, Canada with a regional application

Prepared by

© Susan Richter, B.A., B.Eng.

Faculty of Engineering and Applied Science
Memorial University of Newfoundland

October 1994

St. John's, Newfoundland

ISBN 0-315-96065-5

Abstract

A hydrological study of the island of Newfoundland (Canada) was carried out to identify the key basin characteristics associated with a range of flow measures and to assess the potential for improving flow estimates at ungauged sites using various regionalization methods. The data set was the natural flow records of 40 stations on the island with record lengths of more than ten years. The research included a detailed assessment of the flow records, selection and computation or abstraction of appropriate flow and basin variables, analysis of the relationships of the flow measures to basin characteristics, grouping of the basins (for flood analysis only) into regions of geographic and basin characteristic dataspace, development of predictive equations for all groups, and assessment of the effectiveness of the regionalization methods. A procedure was developed in this work for estimating the effective precipitation in ungauged basins from geographic and topographic variables.

The most important explanatory variables were found to be drainage area, area controlled by lakes and swamps, fraction of barren area in the basin, and distance of the basins north and/or southwest of defined lines. A detailed assessment of five methods of regional subdivision carried out using the mean annual maximum daily flow as the measure of interest found that dividing the island into regions generally improves the estimates at ungauged sites. Clustering based on basin characteristics is a promising method of regionalization.

Acknowledgements

The assistance of the following people in the preparation of this thesis is gratefully acknowledged.

- Dr. Leonard Lye, for his enthusiasm and guidance;
- Acres International Ltd., for support and educational leave;
- Sheila Hunt and Paul Mease, for their care with the presentation of the text, tables and figures.

Financial assistance to carry out this work was provided by the Natural Sciences and Engineering Council of Canada (NSERC).

Table of Contents

List of Figures

List of Tables

1 Introduction

In this work, the hydrology of the island of Newfoundland (Canada) was examined, with two purposes

- to identify the key basin characteristics associated with a range of flow measures;
- to assess whether flow estimates at ungauged sites can be improved through regionalization, and if so, to identify the preferred method of defining regions.

The island of Newfoundland is a large, roughly triangular island about 111,000 km^2 in area lying off the east coast of North America, between latitudes 46° 30' and 51° 30' North. The island has a cool, moist, maritime climate characterized by unsettled weather conditions with few extremes of temperature and precipitation. It lies in the belt of the westerly trade winds, and weather systems from continental North America cross the island in a generally southwest-to-northeast direction. Runoff is consequently high in the southwest, at over 2000 mm a year, reducing to just over 700 mm on the northeast coast.

Surface water is much more important than groundwater in Newfoundland. Most of the island consists of bedrock overlain by a thin veneer of glacial till, so subsurface aquifer storage is negligible. Most of the population obtains its water from surface supplies, and about two-thirds of the island's energy comes from hydroelectric generation from surface sources. The abundance of good quality water in lakes, streams and ponds also sustains important recreational and fisheries uses.

Despite the importance of surface water, the hydrometric data base is quite limited. There are only thirteen records of unregulated streams with lengths of over 30 years, for example. At present, there are about 40 stations with record lengths of 10 years or more (on average one station per 2800 km^2). Installation of additional gauges in the mid-1980's means that within a few years the useable data base will be about 50 percent larger.

The climate network is also sparse, and the stations are not well located for hydrological analysis. In general, the stations are in communities which are situated along the coastline. As a result, there are almost no precipitation records representative of the central or upstream parts of drainage basins, gauged or ungauged.

Nevertheless, estimates of flows at ungauged basins are required throughout the range of flows from droughts to floods. Flood flows are obviously needed for design of structures, and low flows are required to estimate maintenance flows for fisheries and possible failures in water supplies. Daily flow sequences are required for design of new hydroelectric and water supply facilities, and for improved water management at existing facilities.

Previous studies have provided regression equations to enable hydrotechnical engineers to estimate some flows at ungauged sites. But with the limited data base, appropriate regression equations are frequently not available or are not suitable for application to the ungauged basin of interest. The purpose of the present study is therefore not simply to provide equations, but to enhance the understanding of basin response. Estimates resulting from equations can then be rationally assessed and adjusted if required.

This thesis consists of six chapters. Chapter 2, which follows this Introduction, provides a review of the literature, much of which has been in the field of regional flood frequency analysis. Chapter 3 presents a review of the data, selection and computation of the flow measures, and selection and abstraction of the basin characteristics. Chapter 4 presents the analysis of relationships of all the flow measures to the basin characteristics, paying particular attention to variables representing hydrologic input. Graphical techniques, principal components analysis, and multiple regression are the tools used for the quantitative analysis. In Chapter 5 one of the high flow measures, Qavgfld (the mean of the annual maximum daily flow series) is used to assess whether regionalization can improve predictions. Qavgfld was selected because it is frequently used as an index flood in regional flood frequency analysis, and many investigators have noted that the estimates of the index flood are a large source of error in estimates of more remote events. Conclusions and recommendations are provided in Chapter 6.

2 Literature Review

The definition of homogenous regions for hydrological analyses and the relating of basin characteristics to inflows within those regions has long been of interest to engineers seeking to improve estimates of hydrological variables at both gauged and ungauged sites. It continues to be an active field of research, being revisited each time there are advances in either techniques of estimating the variables at gauged sites or in statistical multivariate techniques.

This section reviews the relevant literature in three categories, depending on the focus of the particular study. These categories are briefly described below.

1. **Regional Flood Frequency Analysis**

   Most of the interest in regional hydrology has been in the delineation of regions for the purpose of improving estimates of flood flows. Until recently, most regions have usually been defined geographically, with perhaps a simple statistical test to confirm groupings. The purpose of the regionalization is usually to improve flood flow estimates, especially at ungauged sites. Recent improvements in flood frequency analysis techniques have improved the definition of regions. The statistics obtained using the L-moments method in particular have been shown to be very useful in identifying basins with similar flood responses.

2. Relating Hydrologic Response to Basin Characteristics

Estimates of hydrologic response variables, whether for floods or other flow or meteorological variables, are often required at ungauged sites. These estimates are usually obtained using regression equations relating to the hydrologic response to basin variables. Regionalization techniques are used to improve the estimates.

3. Regional Analysis in Non-Geographic Dataspace

Many investigators have noted that for the purpose of predicting hydrologic response it may be more helpful to group basins together in a data space which is not necessarily geographical. Such grouping can be done using multivariate techniques such as cluster analysis. A variation on this approach has been called the region of influence approach (ROI), in which each basin has its own unique region. The regions have flexible boundaries, and consist of basins particularly similar to the basin of interest. Multivariate analyses can be carried out without computers, but there is no doubt that access to powerful multivariate statistical programs has led to substantial development in regional hydrology.

The remainder of this chapter reviews the relevant literature for each of these categories in turn, in roughly chronological order to show the development in each category. The studies specific to Newfoundland are described in a final section.

2.1 Regional Flood Frequency Analysis

One of the most influential of early investigators into the use of regionalization in flood frequency analysis was Dalrymple (1960). Like many others, he was principally concerned with improving estimates of flood quantiles. He proposed a method of using data from many gauges to compensate for the fact that individual records were too short to produce satisfactcry estimates of upper quantile flood flows (i.e., he substituted spatial information for temporal).

His method, the index flood method, has become widely used, with relatively minor variations. It depends on the fundamental assumption that all the flood series within a region come from the same parent distribution. Dalrymple also published Langbein's test for homogeneity, based on the premise that the slope of an individual frequency curve shall be no more at variance with the regional curve than can be explained by errors in sampling. The test is applied by plotting the flow having an estimated return period of 10 years on a plot which also shows the control curves representing a range of variation equal to two standard deviations of the reduced variate for the extremal distribution. If the flow for the basin falls within the control curves, the basin passes the test for homogeneity. Because of the funnel shape of the control curves, this test is sometimes referred to as the funnel test. Its publication by Dalrymple and shortly thereafter by

Chow (1964) led to its wide use and acceptance. Condie (1979) expanded the concept to include lognormal distributions.

A growing recognition by hydrologists of the improvements in statistical techniques, aided by computers which allowed more sophisticated analysis and plots, led to increased use of three parameter distributions. It soon became clear to those working with three parameter distributions that skew cannot be reliably estimated from the short records typically found in most hydrometric networks. Matalas, et al (1975) and Wallis, et al (1974, 1977) were among the first and most influential to call attention to some of the problems related to this issue. Regional analysis becomes important because it has the potential to provide regional estimates of a shape parameter (skew), which can be combined with at-site estimates of the location and scale parameters (mean and standard deviation) to improve quantile estimates.

Maclaren Atlantic (1980) used a regional approach to estimating skew in a regional flood frequency study of Nova Scotia streams. The investigators first identified mainland Nova Scotia as a homogeneous region different from Cape Breton Island, based on experience and judgment. They then estimated a regional skew coefficient based on the average of the skew coefficients of stations with record lengths longer than 20 years. Multiple regression was used to develop the relationships between the average flood flow, as well as flows with return periods of 20 and 100 years, and basin characteristics.

Gabriele and Arnell (1981) took this type of approach a step further, using a hierarchical approach to regi‚nalization. They assumed that the shape parameter (skew) would be constant over a larger region than the scale coefficient (represented by coefficient of variation Cv). In other words, a basin would be classified into a large region for skew, and a smaller region for Cv. They did not have much success with this approach on the Italian basins on which it was tested.

Fiorentino, et al (1986) continued this approach, using simulation to consider three levels. The most general level was Cs constant in a region, the next level was Cv constant in the region, and the most specific level was multiple regression to obtain the location parameter (the index flood). The observed variance of Cs was similar to the sampling variance from the simulation; however, the observed variance of Cv was nearly twice the Cv in the simulations. They concluded that further subdivision would be required to obtain good stimulation results for Cv.

Arnell and Beran (1987) used a concept similar to regional skew in testing the suitability of the two-component extreme value (TCEV) distribution proposed by Rossi (1984) for regional flood estimation. Two parameters of the TCEV distribution control skewness and kurtosis and the ratio of outliers. In their regional analysis, they assumed that these two parameters were constant in a region. This approach produces realistically variable estimates of sample skewness, but it needs long records for successful calibration, and is not robust in non-TCEV worlds.

Other investigators continued to explore other ways using regional information to improve flood estimates. Wood and Rodriguez-Iturbe (1975), and later Kuczera (1982, 1983) considered Bayesian approaches to infer probabilities of extreme hydrologic events. The Bayesian approach is one method to combine at-site and regional information. In a case study by Wu (1988), quantile estimates from the Bayesian approach were much lower than from the index flood method. Care is required with Bayesian approaches because at sites with short records the prior distribution will dominate, and must be correctly specified (which may be difficult).

Heo, et al (1990) developed some regional frequency models and estimation techniques for gauged sites with short records, assuming independence in space and time. They concluded that regional flood frequency analysis was preferable to single site analysis even when both intersite dependence and heterogeneity appeared.

Guo, et al (1990) investigated a regional maximum likelihood estimation (MLE) method as an alternative to index flood, multiple regression or empirical Bayesian techniques for predicting floods and gauged and ungauged catchments. They found that the regional MLE method performed about the same for predicting flood quantiles at ungauged basins. They noted (as have many others) that finding the proper basin characteristics to be included in a regional model is more important than finding the proper regional model to improve the performance of regional flood frequency analysis.

Cavadid, et al (1991) examined the operation of the Box-Cox transformation for regional flood frequency analysis, especially its performance when the underlying distribution is lognormal, Gumbel or gamma. They concluded that its use is not very promising.

Caissie and El-Jabi (1991) analyzed 237 stream records across Canada using the theory of stochastic processes applied to extreme values. They preselected the regions based on Government of Canada (1984). Using a partial duration series (peaks over threshold), they found that they needed to have four peaks per year above a truncation level to obtain reasonable results. They noted again that the weakness of the method is in estimating the index flood.

The most promising work of the last few years in regional flood frequency analysis is the L-moments approach, which provides powerful tests of regional homogeneity in flood response. L-moments are mathematically equivalent to probability weighted moments (PWMs), brought to the attention of flood frequency analysts by Greis and Wood (1981). The technique of PWM's offers an alternative to the method of moments (MOM) or maximum likelihood (ML) for estimating parameters of a distribution. As in the conventional index flood method, regional quantiles are scaled using an index flood, e.g., the average flood flow. This approach is robust, especially for short, highly skewed or highly kurtotic records.

The first important publication in this area was by Hosking, et al (1985). In their appraisal of the U.K. Flood Studies Report, they demonstrated the superiority of the PWM method. Similarly, Wallis and Wood (1985) showed by Monte Carlo simulation that an index flood approach using PWMs was superior to the Log-Pearson III approach recommended by the U.S. Water Resources Council. Hosking and Wallis (1988) later showed by a Monte Carlo experiment that regional flood frequency methods give better results than at-site analysis, even if both the assumptions of intersite independence and homogeneity are violated. They point out that the error in quantile estimate is often due more in estimating the at-site index flood than in estimating the regional growth factor.

Both Potter (1987) and Cunnane (1988) discussed the contemporary situation in regional flood frequency analysis. Potter reviewed the research from 1983 to 1986, and noted that the index flood methods based on PWMs performed very well in a variety of situations. The main source of variability appeared to be the uncertainty in estimating the at-site index flood. Cunnane (1988) provides a good discussion on the methods and merits of regional flood frequency analysis. He notes that the dimensionless scale and shape parameters Cv and Cs are commonly used as measures to judge regional homogeneity, as well as the funnel test. He prefers across-region averages using PWM's and L-moments, although he cautions that there may still be unusual catchments for which at-site analyses are preferable.

The introduction of L-moments appears to have led in the last few years to the consensus Potter felt was lacking in the mid-1980's. Like PWM's, they allow use of a regional averaging to obtain growth curves. Hosking (1990) and Hosking and Wallis (1992) present convenient and efficient methods for using L-moment statistics in regional analysis. The L-moments approach also provides a method of selecting an appropriate regional distribution, testing for homogeneity and identifying discordant gauges.

The L-moments procedure has now become well-accepted. Vogel and Fennessey (1993) concluded from a case study that L-moment diagrams should replace product moment diagrams. Using large samples of daily streamflow in Massachusetts, they found that conventional moment diagrams based on estimated product moments (Cv, skew, kurtosis) revealed almost no information about the distributional properties of daily streamflow, whereas L-moments diagrams enabled them to discriminate among alternate distributional hypotheses.

L-moments are now used in Environment Canada's Consolidated Frequency Analysis (CFA) package, and the use of L-moments to estimate flood quantiles within regions is recommended in Maidment's *Handbook of Hydrology* (1993). In New Zealand, Pearson (1991) estimated L-moments for 275 annual maximum flood peak series to identify the most suitable parent distribution and compare it with previous assumptions. This work is continuing, as well as application of L-moments to low flows and extreme rainfalls. In Canada, Pilon and Adamowski (1992) demonstrated the value of regional information to flood frequency analysis using the method of L-moments in Nova Scotia through simulation. Pilon (1991) also used L-moments in a simulation study to show that all the gauges on the island of Newfoundland can be considered to be part of a homogeneous region. L-moments have been applied to other data besides flows; Cong, et al (1993), for example, used the L-moments approach to identify the underlying distribution form of precipitation using regional data.

2.2 Relating Hydrologic Response to Basin Characteristics

In order to make flow estimates at ungauged sites, flows at gauged sites must first be related to basin characteristics. In the analysis described in Section 2.1, the concern was with flood flows. Other investigators have been interested in relating not only flood flows but other flow measures to basin characteristics.

A landmark study was carried out by Thomas and Benson (1970), in a large scale multiple regression analysis. They preselected four regions in continental United States, with quite different climatic regimes and physiography (eastern, central, southern and western continental U.S.) and assessed the relationships among a wide variety of basin topographical and meteorological characteristics and 71 flow indices. Their flow indices did not include extremely rare events; they ranged from the seven day low flow with a return period of 20 years at the low flow end to an instantaneous flood peak with a return period of 50 years at the high flow end. They found that

1. streamflow characteristics can be defined more accurately in humid regions;
2. low flows can be only weakly defined;
3. medium flows can be more accurately defined than high;
4. standard deviation of monthly and annual flows are significantly related to basin characteristics; and
5. some indices of flow distribution in time can be better described by regional averages than by basin characteristics.

These conclusions have generally been reproduced by others in one way or another. They also noted first, that some basin characteristics, such as basin geology, cannot yet be satisfactorily represented by simple numerical indices (which may partly explain why low flows are difficult to predict); and second, that the basin characteristic indices most highly related to streamflow are drainage basin size (a physiographic characteristic) and mean annual precipitation (a meteorological characteristic). The usefulness of other variables (forest cover, snow, and surface storage) varied from region to region.

Another influential regional study also carried out by the USGS a few years later was a regional analysis of streamflow characteristics (Riggs, 1973). It is intended as a manual describing ways of generalizing streamflow characteristics and evaluating their applicability under various hydrologic conditions. It is mostly concerned with regional flood frequency analysis, and notes that four basin variables, three physiographic (drainage area, slope, and percent lake and swamps) plus one meteorologic (mean annual precipitation) will ordinarily reduce standard error to a minimum. Like Thomas and Benson, he observes that the application of regional analysis to low flows is less successful because of the greater dependence of low flows on basin characteristics that are imperfectly known and that cannot be described by simple indices.

Schaefer (1983) took an interest in regionalization from a meteorological perspective, in considering storm analysis for spillway design. He argues that extreme storms can be analyzed more reliably on a regional rather than on a point basis. His regions are geographical, and homogeneity is assessed by judgment and experience. Regions can be considered homogeneous if events originate from the same storm type, or if the precipitation data at all stations with the region share homogeneous statistical characteristics. He concluded that large events are rare for any given station but are quite commonplace for a region, a result important for regulators responsible for all structures in a region.

Schaefer continued his work in this field with a regional analysis of precipitation annual maxima in Washington State (Schaefer, 1990). He looked for site-to-site similarity of Cv as the mark of a homogeneous region. He defined his subregions using a continuously varying mean annual precipitation (in geographic space).

Mimikou and Kaemaki (1985) sought relationships which could be used to predict the shape of flow-duration curves in western and northwestern Greece. They described the flow duration curve as a cubic polynomial with four parameters. Assuming the area to be one homogeneous region, they developed regression equations using precipitation, drainage area, hypsometric fall and channel length to predict the values of the parameters.

Fennessey and Vogel (1990) also developed a regional hydrologic model for estimating flow duration curves, but they approximated the lower half with a two parameter lognormal function. They developed regional regression equations using drainage area and a basin relief parameter, with good results. Vogel (1992) took a different approach a few years later in estimating low flow statistics. He approximated low flow behaviour with a simple stream-aquifer model, and then estimated modified model parameters (area, slope and base flow regression constant) using multivariate regression procedures. In the region considered, he found that the low flow statistics were highly correlated with these parameters.

Acreman (1985) reviewed the U. K. Flood Studies Report (FSR) in the Scottish context. Taking Scotland as one region, he used the FSR data base expanded to include new data to develop a prediction equation for the mean annual flood Qavg, based on basin characteristics. He found that the standard average annual rainfall (SAAR) is a better predictor of the mean flood flow than the extreme rainfalls used in the FSR; that the fraction of lake storage (LOCH) is a better predictor than the fraction draining through lakes (LAKE), and that drainage area, stream frequency and a soil parameter are also important variables. Slope, whether average basin slope or main channel slope, was not statistically significant, although examination of residuals suggested that it is an important explanatory variable in some basins. It is either not well represented in the data set or is not adequately specified by the indices. Unlike other investigators, he found no correlation between average valley slope and basin slope, possibly because the region was glaciated.

Pilon (1990) outlined the extension of the index flood method to low flow analysis when the regional distribution is assumed to be Weibull. He derives a homogeneity test similar to the funnel test, and suggests using non-geographic regions. He also recommends extending L-moments techniques to investigate the distribution.

2.3 Regional Analysis in Non-Geographic Dataspace

The studies relating hydrologic response to basin characteristics showed the importance of defining regions. Another line of research dealt with methods of grouping other than geographical for establishing regions. In New Zealand, Blake, et al (1973) took a different and then-novel approach to regionalization. The hydrologic network in New Zealand had been established with the aim of ensuring that all regions would have representative basins, i.e., that a representative gauged basin could be assumed to be representative of the region in which it was located. The regional subdivision had been made on the basis of rock type, slope and precipitation, and the regions were geographically contiguous. Blake et al (1973) used a principal components analysis to test the regional subdivision. They found that they could reduce their original 39 characteristics to seven.

White (1975) took a related approach, using data for basins in Pennsylvania. She used factor analysis of basin geomorphological characteristics, including drainage density, slope, shape, and geometry.

Mosley (1981) continued the evaluation of New Zealand hydrological regions. He developed clusters based on specific mean annual flood and coefficient of variation, and concluded that where a number of factors are equally important in controlling hydrologic regime (e.g., climate, lithology and topography), a complex mosaic of hydrologically homogeneous areas results, and no broad scale regions can realistically be identified. Where one factor is dominant, a regional system may be identifiable, as in the South Island, where differences in climatic regime appear to dominate. He stressed that the groupings should be based on sound principles of classification, and in particular should be based on attributes of the basin, not on factors that supposedly influence the attributes.

Tasker (1982a) published a useful paper comparing methods of regionalization. Using data splitting techniques on 221 basins in Arizona, he compared the effect of clustering on basin characteristics with clustering based on flow characteristics. Had he only based his results on the estimation data, he would have concluded that clustering on hydrologic characteristics was better. In fact he found that better prediction results were obtained using basin characteristics. Elsewhere (Tasker 1982b) he suggested using the Wilcoxon Signed Ranks Test to test whether the apparent clustering of plotting residuals from regression is real or the result of chance. Several years later (Tasker, 1986) he reviewed the issue of regional homogeneity in the context of regional floods, concluding that more work was required to define homogeneity, and pointing out that the estimates for some sites may be adversely affected by regionalization.

Although regionalization for the purposes of improving flood flow estimates has received the most attention in the literature, the interest in other flow characteristics has continued. In the mid-eighties, Hughes (1987) used cluster analysis to group 77 rivers in Tasmania into four groups based on 12 hydrological indices of monthly, annual, peak and low flows. She found that the groups were distinctive and spatially significant, and that drainage area, mean annual rainfall and Cv of annual flows could be used as indicator variables to extrapolate other indices. No basin characteristics were used except drainage area and mean annual rainfall.

Acreman and Sinclair (1986) took the opposite approach to data from 168 basins in Scotland, classifying the basins independently of the discharge data. They then tested homogeneity using flood flow data, using the likelihood ratio test based on Generalized Extreme Value (GEV) parameters. Where homogeneity was rejected, they found it was due to a small number of badly fitting basins.

Hawley and McCuen (1982) used regions defined by cluster analysis for the purpose of estimating water yield in the western United States. The separation of the study area into five regions using cluster analysis of 18 characteristics (17 basin characteristics and one flow characteristic, yield). The clusters were geographically contiguous. They used principal component analysis to identify important factors and stepwise regression to develop the equations.

An important contributor to the research on regional analysis is Wiltshire. In the mid-eighties, he carried out multivariate studies of drainage basins in Britain and developed homogeneity statistics to test their classification. In Wiltshire (1986a), he notes that although the funnel is much used, it is weak, and developed two statistical tests of regional homogeneity and examined properties of the test statistic. He also notes that the U.K. Flood Studies Report did not use any statistical test to define regions.

The first test Wiltshire developed was based on Cv of the flood series, on the assumption that Cv is related to the slope of the flood frequency curve. (Similarity of Cv is often considered an indicator of a regional grouping.) Simulations, however, showed that this test is not very good; it accepted homogeneity too often. The second test, based on the distribution function of the regional parent, performed better. Its power depends on region size, record length and choice of parent distribution. He found that a few long records will characterize the data better than many short ones. Also, if the series are ten years or less in length there will rarely be sufficient information present to detect heterogeneity even where there are gross differences.

In Wiltshire (1986b), he tests clusters in a flow statistic dataspace of average flood specific runoff and Cv. The clusters were developed using a partitioning clustering scheme rather than an agglomerative hierarchical scheme. The clusters were interpreted in terms of basin characteristics through the use of a multivariate linear discriminant analysis. He also considered fractional membership using weighting. Clusters in a flood statistic dataspace were not particularly well mapped onto a basin characteristic dataspace.

In Wiltshire (1986c) he reports a procedure for grouping basins by an iterative search through the basin characteristic database to optimize statistics that describe the efficiency of the grouping. When he applied the procedure to U.K. data, five groups resulted, based on drainage area, mean annual precipitation, and fraction urban. He tested the ten geographical regions identified in the FSR with the same statistics, and found that only five were homogenous. Overall the regions were not significantly different in terms of their mean Cv's.

Wiltshire and Beran (1987a) continued this investigation into multivariate techniques for the identification of homogeneous flood frequency regions. In their study the regions were identified using a plot of Cv as a function of the average flood specific runoff. Basins which plotted near each other were recombined until the total sum of squares of the distances from the centroid was minimized (a partitioning scheme). They then used discriminant analysis to determine which basin characteristics should be used to describe each cluster. They found that some basins did not fit into any of the flow clusters based on their basin characteristics. This was particularly a problem for basins/clusters with relatively small mean floods and high Cv's. They addressed this problem by using the scores from the discriminant analysis to weight the growth curves, and suggested that other statistics could be used to define the clustering dataspace. They concluded that the effectiveness of the discriminant analysis, as well as the regression estimate of the standardizing parameter, are limited by the available basin characteristics. Wiltshire and Beran (1987b) also present a significance test for homogeneity of flood frequency regions.

Burn (1988) addressed the delineation of groups for regional flood frequency analysis. He was concerned with both the technique for determining a homogeneous group, and with the robustness of the method for estimating regional parameters with respect to the distribution. He used a principal components analysis of the correlation structure of annual flows at 41 stations in Manitoba as the flow dataspace. The number of principal components was the number of groups for the analysis; three groups were finally selected. He then evaluated the technique using Monte Carlo simulation, considering four alternative estimators. He found that at-site estimates were improved using these regions.

Burn continued his regional analyses in an appraisal of the region of influence (ROI) approach, similar to the fractional membership approach of Wiltshire and Beran (1987a). In this approach, each site has a unique set of stations which constitute its region. Considering the flow dataspace only, he found that the ROI approach performs better than traditional methods of grouping, probably because the available information is used more efficiently. In 1989 Burn used cluster analysis (by partitioning) of flow variables as a method of grouping basins (Burn, 1989), and followed up on this work with a report on a large study using the U.K. FSR data set (Burn and Boorman, 1993).

In this report, Burn and Boorman present a procedure for estimating various hydrologic variables (high and low flows, mean flow, and rainfall-runoff model parameters) at ungauged catchments. They grouped the catchments using cluster analysis by partitioning of three principal components obtained from seven flow measures. They then identified three groups, those with flashy, big floods, those with slow, sustained, low runoff production, and those with slow, sustained, but high runoff production. They used stepwise discriminant analysis to select the key basin variables, and canonical analysis to identify canonical variables for discriminating between clusters. These canonical variables separated the clusters based on flows reasonably well. They applied the technique to two parameters of a rainfall-runoff model used in the U.K. for ungauged catchments. Burn's most recent published work uses Newfoundland data, and is described in Section 2.4.

Haines, et al (1988) used cluster analysis for descriptive purposes, to prepare a preliminary global classification of seasonal flow regimes, based on a data set of mean monthly flows. Their results were presented as a map of regime types drawn for the first time on the basis of streamflow characteristics alone.

Bhasker and O'Connor (1989) present a comparison of cluster analysis and the method of residuals for flood regionalization. The pattern of residuals has been used in conjunction with regression analysis to identify geographic regions in regional flood frequency analysis. Like Wiltshire they choose the mean specific flood and Cv as the clustering variables (although they use log Cv). They found that clusters were in no way similar and were not coincident with geographical boundaries, but were better discriminating in terms of the hydrological characteristics than region based on patterns of residuals. They then used discriminant analysis of the basin characteristics to compare assignment of basins to clusters. They selected drainage area, slope, sinuosity and shape as the basin characteristics for the discriminant analysis.

Nathan and McMahon (1990) present the results of a thorough investigation into the identification of homogeneous regions for the purposes of regionalization. They used low flow as the catchment response for grouping. They discuss previous approaches and note problems with them. They conclude that cluster analysis is a promising approach, but the classification should be based on catchment characteristics; variables must be selected and possibly weighted, and results will be highly dependent on scale. In addition, there are a plethora of linkage algorithms and distance measures, which will lead to different results with the same data. Once regions have been selected, the final problem is how to allocate new catchments. Computerized procedures such as discriminant analysis, will always assign a variable to some region, even if it is quite dissimilar.

The purpose of their study was to develop regression equations for predicting low flows and yield characteristics. They addressed the problems identified above by using multiple regression to select and weight the independent variables (basin characteristics) before carrying out the cluster analysis. Once the preliminary groups were selected using cluster analysis, they used multidimensional plotting (Andrews plots) to identify group signatures and minimize heterogeneity. The multidimensional plots were also an aid to assigning new catchments to the appropriate group.

Gingras and Adamowski (1993) took an unusual approach to homogeneous region delineation, basing their analysis of New Brunswick basins on annual flood generation mechanism. Their approach combines geography and flood data characteristics through the use of the shape of the probability density function, which is dependent on the processes that generate floods in the watershed. Once they identified three regions, they used multiple regression to develop equations for flood quantiles. For all regions and all quantiles, drainage area and mean annual precipitation were the only significant independent variables, and they were significant in all equations.

2.4 Regional Analyses in Newfoundland

Since the late 1960's a number of studies have been done in Newfoundland which include some form of regionalization, although the regional boundaries rarely agree. The earliest large study was undertaken for the Atlantic Development Board by Shawinigan-Maclaren (Gov't of Canada, 1968). It reviewed the island hydrology and water uses. It was followed two years later by a study done by Ingledow to develop a hydrometric plan for the Atlantic Provinces (Gov't of Canada, 1970). An important component of the work was to delineate hydrologic zones for the purpose of ensuring that each zone was adequately sampled in the hydrometric network. Two types of zone delineation were identified

- physiographic, having generally uniform topography, geology and vegetative cover, and subject to similar climate variations;
- statistically similar, having similar runoff parameters, so that within each zone computation models (regression equations) can be developed to relate physiographic characteristics to hydrologic response.

In Newfoundland they did not have sufficient data to undertake the second approach, so they defined the zones using geographic and climate features as well as the annual runoff distribution, They finally ended up with four zones on the island. In the process they identified two flood zones, the first consisting of the west coast and the Great Northern Peninsula, plus the Avalon peninsula, and the second the remaining central area. They also identified two average runoff zones, eastern and western. These were similar to the flood zones, with the exception that the Avalon peninsula was included in the eastern zone. In the final regionalization they subdivided the large eastern region into three zones, one north and one south of the Water Survey of Canada (WSC) divide between subregions Y and Z (plus subsubregion YS), and the third consisting of the Avalon peninsula. Because of the lack of data in Newfoundland, they were able to undertake only very preliminary statistical analyses. They recommended the addition of about 17 new gauges, many of which were installed.

The first flood frequency analysis relating flood flows to basin characteristics in Newfoundland was undertaken by Poulin (Gov't of Canada, 1971). He treated the island as one region, using 17 stations with an average record length of 15.8 years. Poulin found that the average flood flow based on these records was a function of drainage area, area controlled by lakes and swamps, and slope.

In the early 1980's the provincial Water Resources Branch, in association with the Inland Waters Directorate of Environment Canada, undertook a major regional flood frequency analysis (Gov't of Nfld, 1984). Regression analysis techniques were used to develop relationships between flood flows with return periods of 20 and 100 years and relevant basin and climate characteristics. Equations were provided for two regions, as well as for the island as one region. The regional division was north-south, based on the observation that in most years the maximum daily flows in the north region occurred in the spring in conjunction with snowmelt, whereas in the south region the maximum daily flows could occur in almost any season. Five rivers in the two regions failed the funnel test. Plots of the residuals showed no need to have two regions, but the regions were maintained because the standard errors of the estimates were smaller.

Using logarithmically transformed variables, they found that for the island treated as one region, and in the south, the peak flows were a function of drainage area, mean annual runoff, area controlled by lakes and swamps, and shape. In the north region, treated separately, the best explanatory variables were drainage area, mean annual runoff, and latitude. The report noted the lack of snowcourse and inland precipitation data, and indicated that they had used location (latitude and longitude) as a pseudohydrologic index of exposure to major storms.

Sharp and Moore (1988) used the data from the regional flood frequency study to explore the possibility of improving the regression results using principal components analysis. They found five components that explained 82 percent of the variance, and plots of the first two components suggested possible east-west clustering (not north-south as the regional flood frequency study had found).

Some problems with the 1984 regional frequency analysis were identified by Lye and Moore (1991). The particular items they identified were as follows.

1) Fitting a multiplicative model using logarithmic transforms leads to biased partial regression coefficients and biased estimates of the flood quantiles, as well as goodness of fit statistics that do not reflect the accuracy of the predictions.

2) The use of mean annual runoff (MAR) as an explanatory variable is problematic, both because it is difficult to estimate at an ungauged site, and because the correlation is spurious. A correlation between drainage area and MAR is implied by the formulation of the equation, but such a correlation does not exist.

3) The high correlation between latitude and MAR may lead to problems of multicollinearity.

4) The use of the variable LAT (latitude) is suspect because of its small coefficient of variation (Cv) together with the very large coefficient associated with it in the regression equation.

Some of these problems were addressed in a thorough revision to the provincial regional flood frequency analysis carried out by the provincial Water Resources Division in the late 1980's (Gov't of Nfld, 1990a). Neither of the variables MAR or LAT was used. This study identified four regions, based on specific flood runoff (m^3/s per km^2) and time of occurrence of maximum flows. These four regions gave better results in homogeneity testing than the previous regions or than one region alone. The report provided regression equations for estimating instantaneous peak flows having return periods ranging from 2 to 200 years. The significant variables were drainage area, a lakes and swamps factor (combining the area controlled by lakes and swamps with the fraction of the basin consisting of lakes and swamps), drainage density, and slope. Not all variables were important in all regions. As with the 1984 study, the limited range of the data was noted.

The provincial Water Resources Division also carried out a low flow frequency analysis (Gov't of Nfld, 1991). Three regions were identified, based on annual precipitation, runoff potential, and the fractions of barren and forest areas. These boundaries were different from the flood regions. It was difficult to obtain significant regression parameters; the only significant explanatory variable was drainage area, with fraction of forest coming in as a second variable in one region.

In the early to mid-1980's, two other studies were carried out by Acres with a somewhat different focus. These were hydrologic design methodologies for small scale hydro projects at ungauged sites (Gov't of Canada, 1984, 1985). The first was a general study for Canada as a whole; the country was subdivided into 12 regions, of which the island of Newfoundland was one. In the second project, the methodology was developed for the Atlantic region (except PEI) in more detail.

In the Atlantic study, the three provinces were divided into 14 regions based on physiographic and climate characteristics. Three methods of synthesizing flow records were developed, and an index flood approach was recommended for floods. In Newfoundland, four regions were identified, corresponding with minor modifications to Ingledow's. The regions turned out to be the same for both floods and streamflow; when the stations used in the 1984 regional flood frequency analysis were retested using the streamflow regions, no new nonhomogeneous stations were introduced.

All the above studies used a geographic approach to regionalization. The advent of relatively easy-to-use multivariate techniques along with a good data set provided by the provincial 1984 regional flood frequency study has led to several interesting papers using alternative regionalization approaches.

Cavadias took a canonical correlation approach to regional flood estimation, presenting the results in two related papers (Cavadias, 1988, 1990). Like principal components analysis, canonical correlation reduces the dimensionality of the problem, in this application in the spaces of both the basin and the flood characteristics. Using the data set of flood quantiles and basin characteristics from the 1984 regional flood frequency analysis, he identified canonical flood variables and canonical basin variables. He found that the locations of the gauged basins in a plot of the first two canonical basin variables were similar to their locations in a plot of the first two canonical flood variables. An ungauged basin can be located on the plot of the canonical basin variables, and the neighbouring basins can be identified. The flood characteristics of the ungauged basin can then be assumed to be similar to the flood characteristics of those basins, as represented on the plot of the canonical flood variables.

In a preliminary study, Sceviour and Lye (1993) also used graphical techniques to represent multivariate flow and basin data in two dimensional space. Rather than reducing the data to canonical variables, however, they presented it in Andrews Fourier plots (Andrews, 1972). Four key basin characteristics were taken from the 1990 regional flood frequency study; for the flood characteristics they took the L-moment ratios computed from the annual maximum discharge series. Well-defined clusters could be identified using Andrews Fourier plots in both flow and basin characteristic dataspace, but the same basins were not always clustered together in both dataspaces. The clusters did not coincide with the geographical regions identified in the regional flood frequency study.

Pilon et al (1990) used Newfoundland data to test an approach to regionalization similar to the funnel test but based on L-moments. They were particularly interested in addressing the problem of how much of the difference in hydrological response is due to heterogeneity (that is, where there are true regional differences) and how much is simply noise.

Starting with the assumption that the island of Newfoundland is one region, they generated 1000 replications of the hydrometric network, computing the variance of the L-moments for each. The advantage of the L-moments approach is that no assumptions on the distribution are required. They computed the expected value of the variance based on the simulation results, then tested to see whether the differences were statistically significant. They found that the differences were not significant at the 95 percent level, and concluded that all the basins in the network could be grouped together in one region. Based on their work, they noted that the amount of information in the flood statistics that could be related to basin characteristics may be relatively small.

Zrinji and Burn (1994) present a homogeneity test which allows them to define a region of influence for any basin. As noted previously by Burn (e.g., Burn, 1990), the ROI approach allows flexible cluster boundaries, and for weighting of results, based on the proximity (in dataspace) of an ungauged basin to gauged basins. Gauged stations are added sequentially to the region of influence of an ungauged basin, in order of similarity of catchment characteristics. As the gauged basins are added to the region of influence, their suitability for inclusion in the region is tested using the extreme flow characteristics for gauged basins on the ROI only. The homogeneity test used is a chi-squared test to determine if the at-site L-moment ratios are similar to the L-moment ratios of the regional parent distribution.

They applied the approach to Newfoundland data from the 1984 regional flood frequency analysis, and found a somewhat different result from Pilon et al. Whereas Pilon's test had identified all the stations as being in one region, based on the variance of the L-moment ratios, Zrinji and Burn found that the numbers of stations in the regions of influence tended to be small, because additional stations did not pass the homogeneity test for inclusion in the region. Zrinji and Burn do note, however, that the set of gauging stations they used forms a nearly homogeneous region. Based on a comparison of the results of the region of influence approach with other methodologies (e.g., regression approach, one region, and regression approach, north-south geographical regions), they conclude that the regionalization component in regional flood frequency analysis is an important factor for efficient extreme flow quantile estimation.

3 Data Preparation and Preliminary Analysis

The data set used for the analysis consisted of flow and physiographic characteristics for all basins

- on the island of Newfoundland
- gauged by Water Survey of Canada (WSC)
- with 10 or more years of record.

Regulated basins or basins with unknown drainage boundaries or diversions were excluded. Table 3.1 lists all the basins meeting the above conditions, together with the number of years of record and drainage areas. This list shows 45 stations, but a few stations were removed as discussed below, for a final list of 40 basins. The locations of the basins are shown in Figure 3.1.

The record for Indian Brook at Indian Falls (02YM001) was not included because a culvert diverts water from the basin into Birchy Lake. The amount of water diverted, although probably small, is unknown. Similarly, a set of fisheries culverts at the upstream divide of the Grey River basin (02ZD001) allows the release of water from Meelpaeg Reservoir into the basin. Again, the volume of water is probably very small,

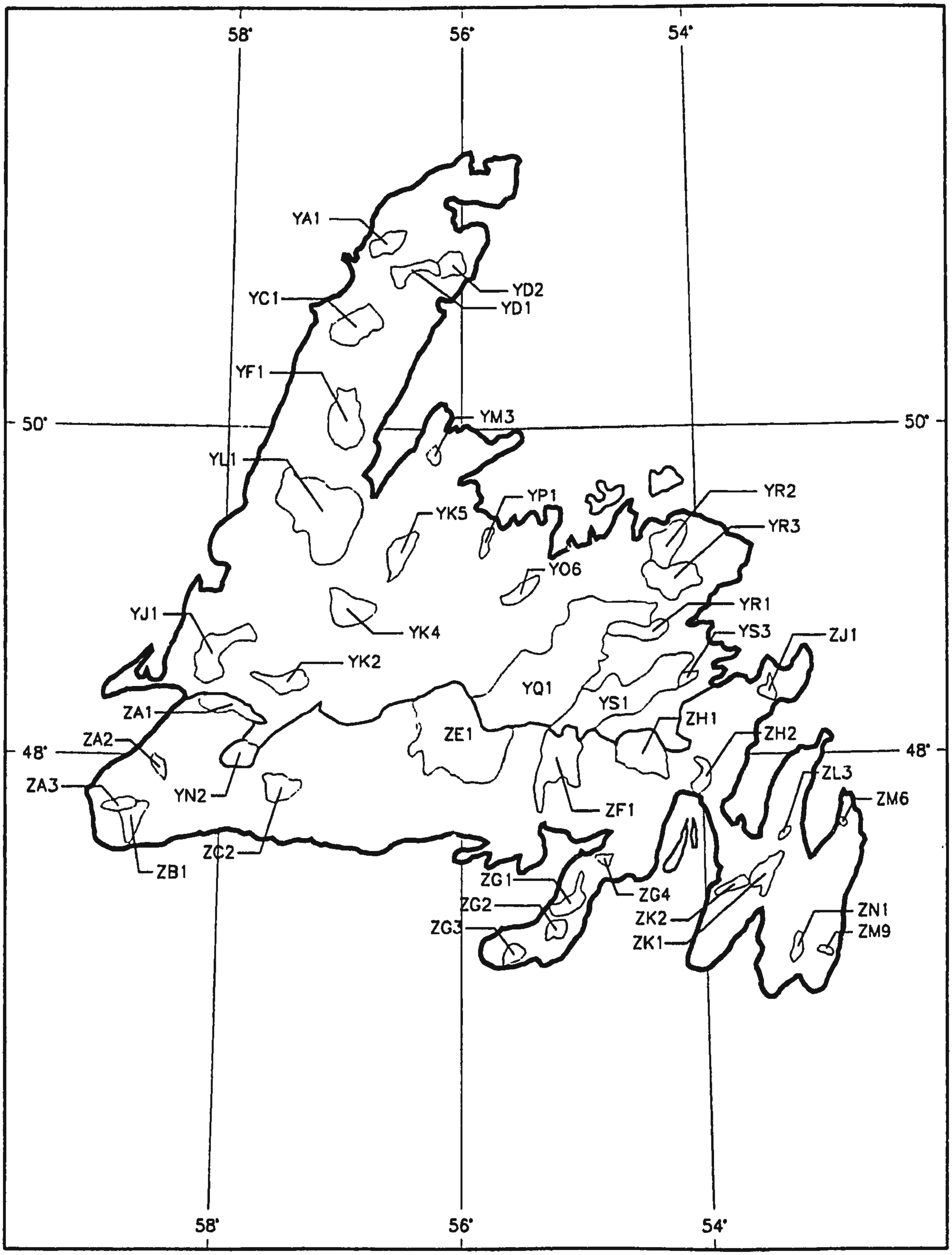


Fig. 3.1 – Locations of Gauged Basins Used in Study

Table 3.1

Unregulated Basins on the Island of Newfoundland with 10 or more years of record

	Basin ID	WSC Gauge #	Station Name	WSC Count	Drainage Area (km²)
YA1	SteG	02YA001	Ste. Genevieve R near Forresters Point	24	306
YC1	Tor	02YC001	Torrent R at Bristol's Pool	33	624.0
YD1	BvrB	02YD001	Beaver Brk near Roddickton	20	237.0
YD2	NeRk	02YD002	Northeast Brk near Roddickton	12	200.0
YF1	Cat	02YF001	Cat Arm R above Great Cat Arm	15	611.0
YJ1	Hrys	02YJ001	Harrys R below Highway Bridge	24	640.0
YK2	LewB	02YK002	Lewaseechjeech Brk at Little Grand Lake	36	470.0
YK3	ShfL	02YK003	Sheffield R at Sheffield Lake	12	362.0
YK4	Hnds	02YK004	Hinds Brk near Grand Lake	24	529.0
YK5	Shfd	02YK005	Sheffield Brk near Trans Canada Highway	20	391.0
YL1	UpHm	02YL001	Upper Humber R near Reidville	64	2110.0
YM3	SwBV	02YM003	South West Brk near Baie Verte	12	93.2
YN2	Lds	02YN002	Lloyds R below King George IV Lake	11	469.0
YO6	Ptrs	02YO006	Peters R near Botwood	11	177.0
YP1	ShlA	02YP001	Shoal Arm Brk near Badger Bay	10	63.8
YQ1	GdBC	02YQ001	Gander R at Big Chute	43	4444.0
YQ2	GdGL	02YQ002	Gander R at Outlet of Gander Lake	17	4160.0
YR1	MdlB	02YR001	Middle Brk near Gambo	33	275.0
YR2	RgdH	02YR002	Ragged Harbour R near Musgrave Harbour	15	399.0
YR3	IndB	02YR003	Indian Bay Brk near Northwest Arm	11	554.0
YS1	TerN	02YS001	Terra Nova R at Eight Mile Bridges	34	1365.0
YS3	SwTN	02YS003	Southwest Brk at Terra Nova National Park	25	36.7
ZA1	LtlB	02ZA001	Little Barachois Brk near St. George's	14	343.0
ZA2	Hlds	02ZA002	Highlands R at Trans Canada Highway	10	72.0
ZA3	LtCd	02ZA003	Little Codroy R near Doyles	10	139.0
ZB1	IaM	02ZB001	Isle aux Morts R below Highway Bridge	30	205.0
ZC2	Grdy	02ZC002	Grandy Brk below Top Pond Brook	10	230.0
ZD2	Gry	02ZD002	Grey R near Grey River	20	1340.0
ZE1	SmLP	02ZE001	Salmon R at Long Pond	22	2640.0
ZF1	BdN	02ZF001	Bay du Nord R at Big Falls	42	1170.0
ZG1	Grn	02ZG001	Garnish R near Garnish	34	205.0
ZG2	Tds	02ZG002	Tides Brk below Freshwater Pond	15	166.0
ZG3	SmLm	02ZG003	Salmonier R near Lamaline	12	115.0
ZG4	Rtl	02ZG004	Rattle Brk near Boat Harbour	11	42.7
ZH1	PH	02ZH001	Piper's Hole R at Mothers Brook	40	764.0
ZH2	CbC	02ZH002	Come by Chance R near Goobies	24	43.3
ZJ1	SthB	02ZJ001	Southern Bay R near Southern Bay	16	67.4
ZK1	Rky	02ZK001	Rocky R near Colinet	44	300.0
ZK2	NE-P	02ZK002	Northeast R near Placentia	13	89.6
ZL3	SpCv	02ZL003	Spout Cove Brk	13	10.8
ZM10	WtMP	02ZM010	Waterford R at Mount Pearl	11	16.6
ZM6	NEP	02ZM006	Northeast Pond R at Northeast Pond	39	3.63
ZM8	WatK	02ZM008	Waterford R at Kilbride	18	52.7
ZM9	SlCv	02ZM009	Seal Cove Brk near Cappahayden	13	53.6
ZN1	NwNP	02ZN001	Northwest Brk at Northwest Pond	26	53.3

but is unknown. The record of the Gander River at the outlet of Gander Lake (02YQ002) for the period 1923 - 1939 was also not included in the data set. Although the record provides useful information from a historical perspective, it was a manual gauge perhaps maintained to different standards, and the data therefore may not be directly comparable to that from the remaining gauges. In addition, the record from the gauge on the Gander River at Big Chute (02YQ001) provides very good coverage of the same basin. The records of the two Waterford River basins (02ZM008 and 02ZM010) were not included because the basin has been subject to increasing urbanization.

In general the complete records were used as provided in WSC's computer database HYDAT. All basins on the island of Newfoundland are in WSC hydrologic region 02, and in subregion Y or Z. (The Y-Z boundary is shown in Figure 3.1) The station code is therefore simplified for the remainder of this report to two letters and one digit, consisting of the subregion identifier Y or Z, the subsubregion code letter A to R and the station number 1 to 9. Station 02ZH001, for example, is simplified to ZH1.

Ten years of records for the calculation of the hydrologic response measures was considered sufficient because no extreme events are estimated in this study. It is a standard record length accepted by many agencies and authors. It also happens that the last ten years of record are reasonably representative of the longer term hydrology on the island of Newfoundland.

3.1 Flow Characteristics

The flow characteristics were selected to represent average, high and low flow regimes, as well as flow availability, e.g., for hydroelectric generation (not just low flow or reliable yield). These are summarized below. The derivation of the characteristics and the preliminary analysis leading to the final selection is described in more detail in the remainder of this chapter.

High flow

- average maximum daily flow (m^3/s);
- linear coefficient of variation (Lcv);
- 10th flow exceedance percentile, i.e., daily flow exceeded 10 percent of the time.

Average flow

- average daily flow (m^3/s).

Low flow

- median low flow (m^3/s);
- 90th flow exceedance percentile, i.e., daily flow exceeded 90 percent of the time.

Available Flow

- median flow (50th percentile on daily flow duration curve).

All values were taken from the complete period of record, except as noted below. Values were also converted to specific runoff flow per unit area, and in the case of the flow duration characteristics, to fractions of average flows, for some analyses.

3.1.1 High Flows

The selected high flow measures were the mean maximum daily flow, the linear coefficient of variation (Lcv) and the 10th flow exceedance percentile, i.e., daily flow exceeded 10 percent of the time.

Since most of the literature on regional analysis relates to flood frequency analysis, at site frequency analyses were carried out before selecting the final high flow measures. Environment Canada's Consolidated Frequency Analysis software package CFA3 was used for this work. This package estimates parameters of the candidate distributions using the method of L-Moments developed by Hosking (1990).

In addition, the series were tested for trend, randomness and independence. Two records showed significant trend at the one percent level, Lewaseechjeech Brook (YK2) and Pipers Hole (ZH1). The gauge at Lewaseechjeech Brook was removed in 1967 because the outlet of Little Grand Lake (just upstream of the gauge) was blasted out, and was reinstalled in 1973. There is a significant difference in the annual maximum flow series before and after this change, no doubt due to the reduced hydraulic control at the lake outlet as a result of the blasting. The early years were therefore omitted from the final analysis.

There is no satisfactory explanation for the apparent trend at Pipers Hole. A fire in the basin in the early 1960's burned about 65 km^2 of wooded area, but since the total area of the basin is 764 km^2, largely lakes, swamps and barrens, it does not seem plausible that this fire would cause a marked permanent change. There are not many hydrometric records in the province extending back to the late 1950's, nor are there any precipitation records in the basin, so it is difficult to make comparisons. In general the late 1950's and early 1960's were dry, but no other records show the statistical difference exhibited by this one series from Piper's Hole. The annual series and the low flow series do not show this trend. Since there is no satisfactory physical explanation, the apparent trend in maximum flows is assumed to have occurred by chance, and the record was maintained with no adjustment.

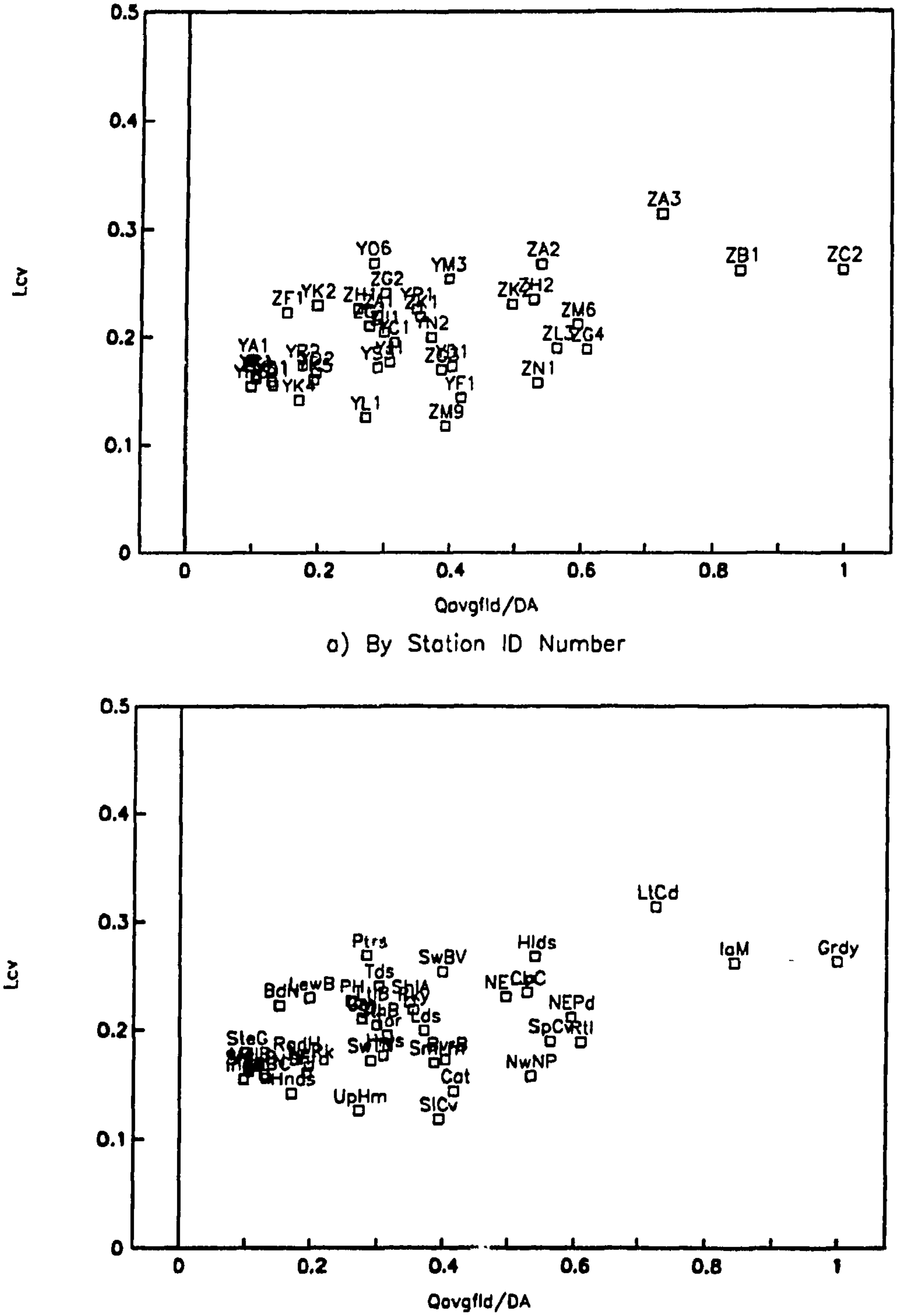
Seventeen of the at-site frequency distributions were upper bounded. In all cases where the record length was 15 years or longer, the upper bounds were sufficiently high that they would not be likely to influence estimates of floods at high quantiles, and the records could be used in a regional flood frequency analysis. Six of the stations with upper bounded distributions had only 10-15 years of record and of these, 4 were upper bounded at flows that were unreasonably low (only slightly above the highest recorded flow). In only two cases this result could be attributed to negative skew (-0.365 for ZL3, and -0.057 for YR3).

The means of the daily maximum flows were calculated for the period of record for each gauge, [as well as the linear second, third and fourth order moments]. These linear moments are referred to as Lcv, L-skewness (Ls), and L-kurtosis (Lk), and are presented in Table 3.2. Similarity of CV is often taken to be indicative of a regional cluster (e.g., Morley, 1981, and Wiltshire, 1986). Figure 3.2 shows a plot of L-cv and specific annual mean maximum daily flow ($Q_{sp\text{-}fld}$). This plot suggests some geographical clustering of basins into Y and Z WSC regions. Plots of Ls and Lk can also be used to suggest clusters, and to identify preferred regional parent distributions (Hosking, 1990). Figure 3.3 is an L-moment diagram for the basins in this study. Like the CV plots, it suggests geographical clusters of Y's and Z's. Although this study is not concerned with selection of flood frequency distributions, it does appear that different parent distributions might be appropriate for different regions. Figure 3.3 suggests that a good candidate for a parent distribution for flood flows in WSC hydrologic

Table 3.2

Mean, standard deviation and L–moment ratios for selected gauges

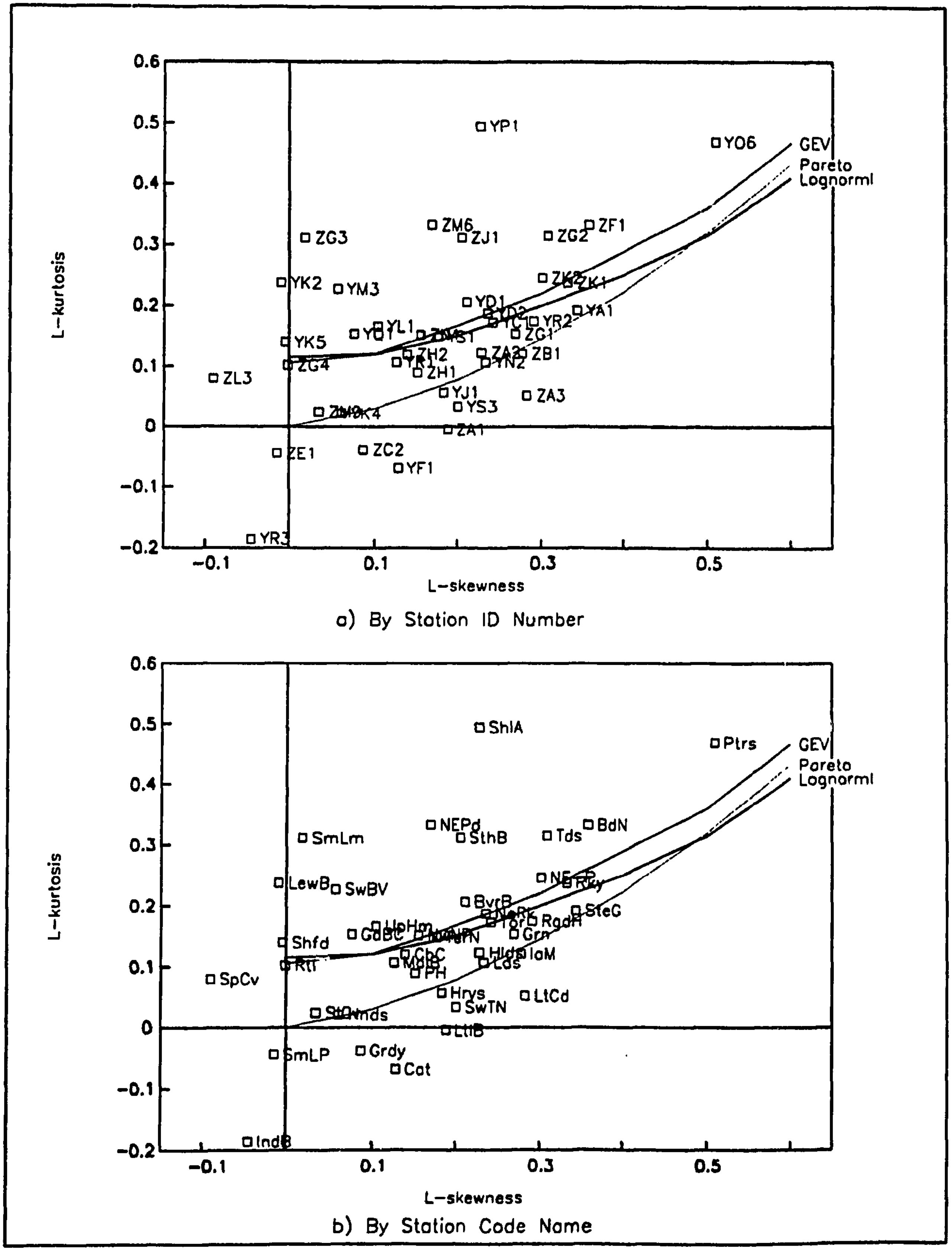
	Station Name	Mean (m^3/s)	L–std.dev (m^3/s)	Lcv	Ls	Lk
YA1	Ste. Genevieve R near Forresters Point	31.19	5.581	0.179	0.344	0.192
YC1	Torrent R at Bristol's Pool	197.08	38.428	0.195	0.242	0.173
YD1	Beaver Brk near Roddickton	95.81	16.566	0.173	0.212	0.207
YD2	Northeast Brk near Roddickton	39.46	6.573	0.167	0.236	0.188
YF1	Cat Arm R above Great Cat Arm	255.00	36.703	0.144	0.128	−0.068
YJ1	Harrys R below Highway Bridge	197.61	35.043	0.177	0.183	0.057
YK2	Lewaseechjeech Brk at Little Grand Lake	93.17	21.435	0.230	−0.009	0.238
YK4	Hinds Brk near Grand Lake	90.49	12.892	0.142	0.061	0.022
YK5	Sheffield Brk near Trans Canada Highway	75.69	12.152	0.161	−0.004	0.140
YL1	Upper Humber R near Reidville	573.53	72.429	0.126	0.104	0.166
YM3	South West Brk near Baie Verte	37.17	9.430	0.254	0.057	0.[illegible]28
YN2	Lloyds R below King George IV Lake	174.24	34.864	0.200	0.234	0.107
YO6	Peters R near Botwood	50.13	13.460	0.269	0.510	0.469
YP1	Shoal Arm Brk near Badger Bay	22.31	5.048	0.226	0.228	0.495
YQ1	Gander R at Big Chute	589.67	91.787	0.156	0.076	0.154
YR1	Middle Brk near Gambo	29.19	4.821	0.165	0.126	0.107
YR2	Ragged Harbour R near Musgrave Harbour	70.59	12.255	0.174	0.291	0.175
YR3	Indian Bay Brk near Northwest Arm	54.66	8.498	0.155	−0.045	−0.186
YS1	Terra Nova R at Eight Mile Bridges	177.56	28.311	0.159	0.177	0.149
YS3	Southwest Brk at Terra Nova National Park	10.65	1.833	0.172	0.201	0.034
ZA1	Little Barachois Brk near St. George's	99.19	21.822	0.220	0.188	−0.004
ZA2	Highlands R at Trans Canada Highway	38.89	10.408	0.268	0.229	0.123
ZA3	Little Codroy R near Doyles	100.86	31.658	0.314	0.283	0.052
ZB1	Isle aux Morts R below Highway Bridge	172.75	45.186	0.262	0.278	0.121
ZC2	Grandy Brk below Top Pond Brook	229.71	60.423	0.263	0.087	−0.038
ZE1	Salmon R at Long Pond	280.05	45.314	0.162	−0.014	−0.044
ZF1	Bay du Nord R at Big Falls	178.01	39.741	0.223	0.358	0.334
ZG1	Garnish R near Garnish	56.71	11.967	0.211	0.269	0.154
ZG2	Tides Brk below Freshwater Pond	50.12	12.[illegible]67	0.241	0.309	0.316
ZG3	Salmonier R near Lamaline	44.51	7.558	0.170	0.019	0.312
ZG4	Rattle Brk near Boat Harbour	26.06	4.918	0.189	−0.001	0.102
ZH1	Piper's Hole R at Mothers Brook	198.47	45.009	0.227	0.151	0.090
ZH2	Come by Chance R near Goobies	22.88	5.366	0.235	0.139	0.120
ZJ1	Southern Bay R near Southern Bay	20.19	4.126	0.205	0.206	0.313
ZK1	Rocky R near Colinet	106.53	23.364	0.219	0.333	0.238
ZK2	Northeast R near Placentia	44.51	10.263	0.231	0.302	0.246
ZL3	Spout Cove Brk	6.08	1.154	0.190	−0.091	0.080
ZM6	Northeast Pond R at Northeast Pond	2.16	0.457	0.212	0.169	0.334
ZM9	Seal Cove Brk near Cappahayden	21.10	2.492	0.118	0.035	0.025
ZN1	Northwest Brk at Northwest Pond	28.50	4.493	0.158	0.155	0.152
	Mean	114.8	21.4	0.199	0.169	0.152
	Standard Dev	130.3	21.8	0.199	0.164	0.151
	Maximum	589.7	91.8	0.314	0.510	0.495
	Minimum	2.161	0.457	0.118	−0.091	−0.186



a) By Station ID Number

b) By Station Code Name

Fig. 3.2 – Lcv and Specific Mean Maximum Daily Flow



a) By Station ID Number

b) By Station Code Name

Fig. 3.3 – L-moment Diagram

region Y might be three parameter log normal (3PLN), whereas a generalized extreme value distribution (GEV) might be more appropriate for the Z region.

The results of the regional analysis using L-moments by Pilon et al (1991) referred to in Chapter 2 may be compared with the present results. Using Hosking's L-moments procedure, they selected a parent distribution with Lcv = 0.1839 and Ls = 0.1843, compared with (unweighted) means of 0.199 and 0.189 from Table 3.2.

3.1.2 Average Flows

The average flow in a basin is important because

- it gives an upper limit on flow availability, e.g., for water supply or hydroelectric generation;
- instream or other requirements may be set as fractions of the average flow, or flow duration percentiles may be expressed as a fraction of the annual average flow;
- proration factors used to develop flow series for ungauged basins are developed as a ratio of the estimated average flow at the ungauged basin to the measured average flow at the gauged basin (not only drainage area);

- converted to an average runoff depth over the basin, it is the effective precipitation (Eff-P) on the basin. It may therefore serve as a proxy for hydrologic input (precipitation).

This last use is of considerable interest, because in most regional studies relating flow characteristics to basin characteristics, precipitation is a significant explanatory variable. In Newfoundland, the network of climate stations collecting precipitation data is extremely limited, not only in number but also in space; the stations tend to be located near the coast, at low elevations. The maps of runoff and precipitation presented in the Water Resources Atlas for the province of Newfoundland illustrate this problem clearly - over parts of the island the runoff exceeds the precipitation, which is physically impossible (Gov't of Nfld, 1992).

In the 1984 provincial regional flood frequency analysis (Gov't of Nfld, 1984), mean annual runoff (MAR) was used as an independent variable in the regression equations, and in all cases was the second most important explanatory variable after drainage area. The use of MAR was criticized (Lye and Moore, 1991) on several grounds, the most important of which was the difficulty of estimating MAR for ungauged basins. In fact MAR was not used in the subsequent regional flood frequency analysis (Gov't of Nfld, 1990). More regions were defined, which perhaps implicitly took account of MAR, in the sense that the MAR was relatively homogeneous in the region. The homogeneity of MAR was not taken into account explicitly - the regions were defined on the basis of flood generating mechanisms, and the regression equations were developed using basin physiographic characteristics only.

Since it is the interaction of the hydrologic input with the physiographic characteristics of a basin which define its hydrologic response, it is clearly desirable to include the hydrologic input. Otherwise the basin physiographic characteristics are expected to completely explain hydrologic response, with all other variation assumed to be random. Precipitation, however, is random, and yet it is not explained by basin-specific physiographic characteristics; rather, it results from atmospheric processes interacting with topography. Eff-P may be a suitable variable as a proxy for precipitation, since it is essentially total precipitation minus losses. In any event, it is the only variable related to hydrologic input for which there is any information. It should be possible to explain a large proportion of the variation in Eff-P using measurable independent variables such as distance to the sea in the direction of the prevailing storms, and elevation and orientation of the basin.

The approach in this study is to treat Eff-P as a variable dependent on location and topographical characteristics. Eff-P can then be used directly with drainage area to estimate average flow volume for an ungauged basin. It may also be used to improve estimates of other flow variables in one of two ways

- by estimating Eff-P first from regression equations, and then incorporating it as an independent variable representing basin hydrologic input in regression equations developed to estimate other flow variables (recognizing the associated error); or
- by incorporating the independent variables used to estimate Eff-P directly into the regression equations for other flow variables.

The average flow is thus represented in three ways,

1) as the average flow for the period of record;
2) as specific runoff (m^3/s per km^2); and
3) as Eff-P (m or mm) assumed to occur at the basin centroids. (Although it is actually integrated over the basin area, the difficulties of estimation are compounded, with insufficient data to justify this refinement.)

In only one case, that of Cat Arm (YF1), was the average flow adjusted. An examination of the records for other rivers in the area shows that the period of record of the Cat Arm basin, 1969 to 1982, was wetter than the long term average. The average flow was thus reduced by about 7 percent.

Figure 3.4 presents a plot of Eff-P for the study basins. Like the high flow plots, this figure also indicates that for the same drainage area, basins in WSC region Z are likely to have higher runoff than basins in region Y. This result is not surprising because region Z is closer to the sea in the direction of incoming weather systems.

3.1.3 Low Flows

Low flows are generally difficult to estimate and in many cases even difficult to measure, particularly if the low flows occur during ice conditions. The median daily low flow for the period of record at each basin was therefore selected as a robust variable to represent the low flow regime.

In addition, the 90th exceedance percentile of daily flows (the flow which is exceeded 90 percent of the time) was extracted from the flow duration curve as an additional measure of low flow.

In only one case was the record adjusted. At Hinds Brook (YK4), the minimum low flows in 1967 and 1968 are reported as zeros. These zero flows occurred because of some construction in the watershed which required damming of the brook. The lowest flow in any other year was 1.4 m^3/s, so including zeros in two

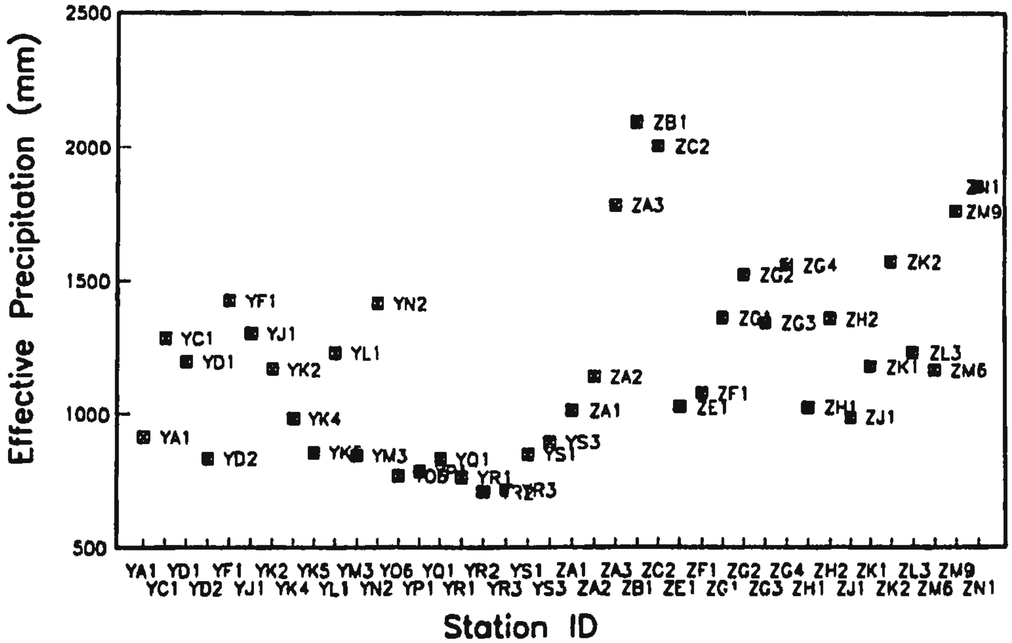


Fig. 3.4 — Effective Precipitation and Station Identification

years could have a large effect. Those two years were consequently omitted in identifying the median.

3.1.4 Available Flow

Two related but slightly different types of measures were considered to represent flow availability. The first type was based on exceedance values obtained from the daily flow duration curve, and included the 10th, 50th (median) and 90th percentile. These are the flows that are exceeded 10, 50 and 90 percent of the time. (The 10th and 90th percentiles were also used as measures of high and low flow, as described above.) The flow duration measures are abbreviated here as $Q_{fd\text{-}10}$, $Q_{fd\text{-}50}$, and $Q_{fd\text{-}90}$. As a fraction of average annual flow (Qavg), these measures are abbreviated as FD-10, FD-50 and FD-90.

The alternative measures are based on the area under the flow duration curve at various flow levels. Since these areas represent the flows which are theoretically available to turbines having the appropriate flow capacities, they are referred to as turbinable flows. Three flow levels were selected, multiples of 4.0, 1.0 and 0.2 times the average annual flow, to represent high, medium and low flows. As fractions of Qavg they are abbreviated as Qt4, Qt1 and Qt2. The relationship between these turbinable flow measures at these flow capacities and the flow duration curve measures is shown in Figure 3.5, for a very flashy river, Isle aux Morts (ZB1) and for a non-flashy river, Bay du Nord (ZF1).

The turbinable flow measure is in some ways preferable, because it directly describes the amount of water available. The turbinable flow is the integral of the flow duration curve, and may in turn be plotted as a turbinable flow curve. The turbinable flow curve is more amenable to mathematical description (as a polynomial function) than the flow duration curve as a function. Acres (Gov't of Canada, 1986) has related basin characteristics to the coefficients of the polynomial function for basins in Newfoundland. Others (e.g., Mimikou and Kaemaki, 1985, Vogel, 1993) have attempted to relate basin characteristics to parameters describing the flow duration curve, and have generally been successful only when using a selected portion of the curve.

The advantages of the flow duration curve measures are that they are easily obtained using available software, and are widely understood. For the purpose of this thesis, similar for more than one flow characteristic, the flow duration measures are quite suitable. In any event, since one measure is simply the integral of the other, there is a high correlation between the two at each of the flow levels (high, medium, and low). Such differences as these arise from the fact that the flow levels are not identical; for example, the medium level flow measures are

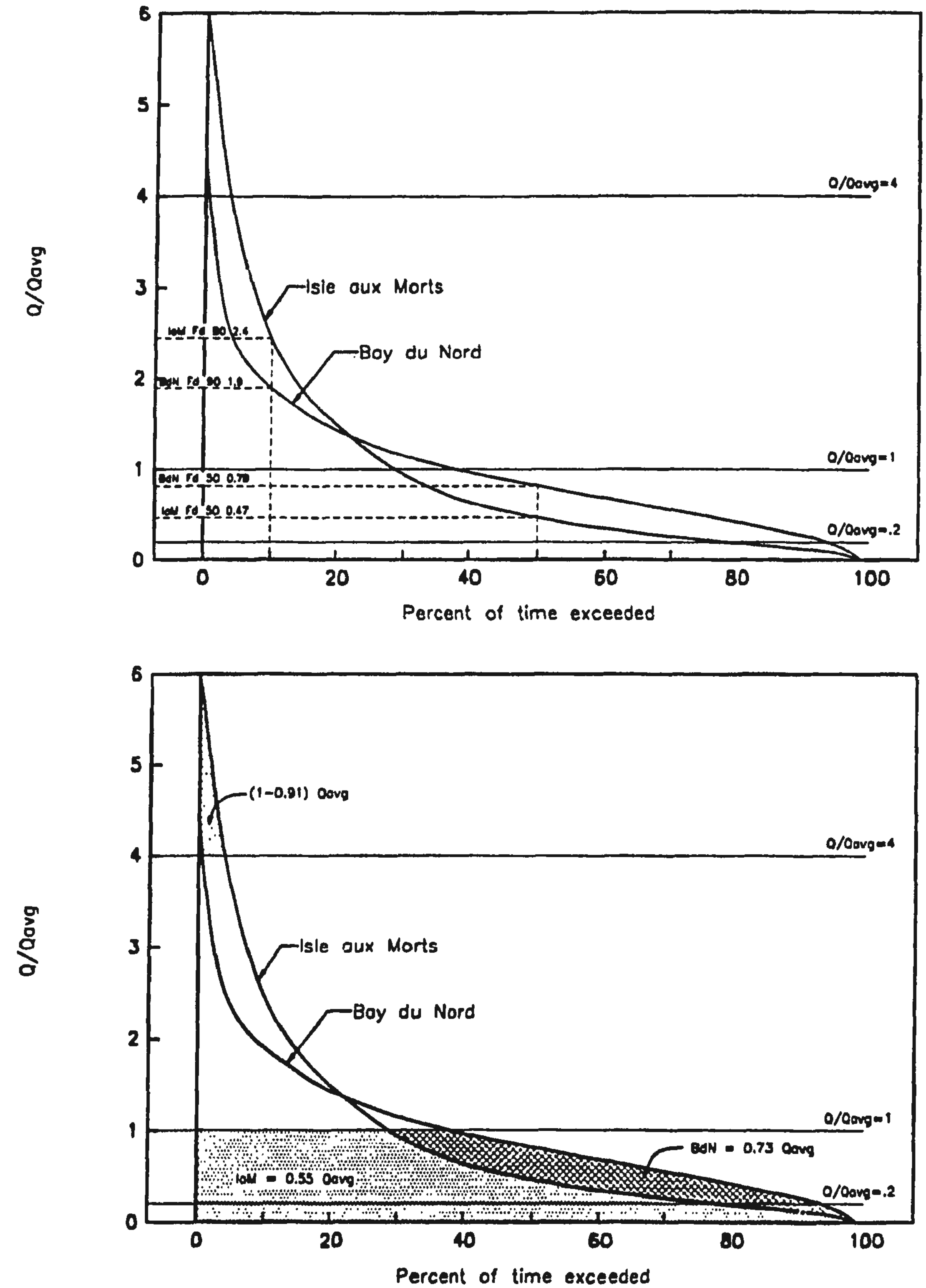


Fig. 3.5 — Examples of Available Flow Measures

in one case the *median* flow (FD-50) and in the other, the area under the flow duration curve at the *mean* flow. In the context of this work, these differences do not matter. If useful groupings can be defined, it might then be reasonable to put more effort into estimating turbinable flows within the regions.

The flow duration curve measures were therefore selected for this study.

3.1.5 Summary of Flow Measures

Tables 3.3 and 3.4 present the final selected flow measures to be used in the multivariate analysis for each of the basins used in the study.

3.2 Basin Physiographic Characteristics

The previous section described the selection of flow characteristics of the gauged basins. This section describes the physiographic characteristics used in the analysis for the same basins.

The physiographic characteristics of the basins were selected from those which have been shown in other studies, or which could reasonably be expected, to have an important influence on hydrological response. The selected basin characteristics are presented below, with a brief description of each and how it was obtained.

Table 3.3

Flow Measures for Study Basins

ID	Code Name	Average Qflood m³/s	Lcv	FD 10	FD 50	FD 90	Average flow m³/s	Effective Precip. (mm)	Median low flow m³/s
			as fractions of average flow						
YA1	SteG	31.2	0.179	2.041	0.749	0.359	8.87	915	2.62
YC1	Tor	197.1	0.195	2.240	0.602	0.206	25.58	1284	3.74
YD1	BvrB	95.8	0.173	2.558	0.412	0.099	9.01	1197	0.57
YD2	NeRk	39.5	0.167	2.652	0.513	0.090	5.20	833	0.37
YF1	Cat	255.0	0.144	2.736	0.417	0.084	27.50	1425	1.38
YJ1	Hrys	197.6	0.177	2.155	0.644	0.256	26.24	1302	4.02
YK2	LewB	120.0	0.140	2.190	0.632	0.226	17.39	1168	2.75
YK4	Hnds	90.5	0.142	2.152	0.679	0.252	16.40	984	3.26
YK5	Shfd	75.7	0.161	2.245	0.606	0.206	10.56	855	1.42
YL1	UpHm	573.5	0.126	2.482	0.624	0.163	82.29	1229	7.66
YM3	SwBV	37.2	0.254	2.478	0.364	0.070	2.52	843	0.06
YN2	Lds	174.2	0.200	2.386	0.543	0.191	21.11	1413	2.34
YO6	Ptrs	50.1	0.269	2.575	0.469	0.133	4.25	768	0.33
YP1	ShlA	22.3	0.226	2.363	0.461	0.156	1.79	785	0.16
YQ1	GdBC	589.7	0.156	2.171	0.697	0.227	115.54	831	19.10
YR1	MdlB	29.2	0.165	2.150	0.729	0.189	6.60	763	0.91
YR2	RgdH	70.6	0.174	2.386	0.570	0.089	8.78	709	0.58
YR3	IndB	54.7	0.155	2.238	0.679	0.263	12.60	718	2.73
YS1	TerN	180.8	0.159	2.308	0.733	0.250	36.86	846	6.68
YS3	SwTN	10.6	0.172	2.404	0.529	0.156	1.03	894	0.06
ZA1	LtlB	99.2	0.220	2.273	0.561	0.188	10.98	1012	1.55
ZA2	Hlds	38.9	0.268	2.358	0.454	0.169	2.59	1140	0.25
ZA3	LtCd	100.9	0.314	2.293	0.535	0.163	7.78	1782	0.73
ZB1	IaM	172.8	0.262	2.441	0.473	0.118	13.53	2093	0.80
ZC2	Grdy	229.7	0.263	2.411	0.406	0.108	14.49	2003	0.60
ZE1	SmLP	280.1	0.162	2.028	0.808	0.256	85.80	1026	16.65
ZF1	BdN	178.0	0.223	1.897	0.794	0.311	39.78	1076	9.23
ZG1	Grn	56.7	0.211	2.005	0.758	0.237	8.82	1359	1.28
ZG2	Tds	50.1	0.241	1.925	0.748	0.229	7.97	1521	1.16
ZG3	SmLm	44.5	0.170	2.372	0.583	0.129	4.95	1342	0.28
ZG4	Rtl	26.1	0.189	2.175	0.583	0.171	2.09	1559	0.15
ZH1	PH	198.5	0.227	2.153	0.673	0.164	24.45	1024	2.55
ZH2	CbC	22.9	0.235	2.226	0.597	0.125	1.86	1356	0.10
ZJ1	SthB	20.1	0.205	2.346	0.531	0.099	2.09	988	0.07
ZK1	Rky	112.1	0.219	2.045	0.682	0.187	11.10	1178	1.05
ZK2	NE-P	44.5	0.231	2.027	0.722	0.218	4.48	1571	0.50
ZL3	SpCv	6.1	0.190	1.971	0.648	0.173	0.42	1230	0.03
ZM6	NEP	2.2	0.190	2.276	0.485	0.104	0.13	1165	0.01
ZM9	SlCv	21.1	0.118	2.140	0.676	0.228	3.00	1760	0.42
ZN1	NwNP	28.5	0.158	2.035	0.645	0.239	3.14	1853	0.48

Table 3.4

Descriptive Statistics of Flow Characteristics

Flow Char	Units	Minimum	Maximum	Range	Mean	Std Dev	Median
Qavgfld	m^3/s	2.22	589.67	587.45	115.70	131.98	63.65
Spec flood	$m^3/s/km^2$	0.099	0.999	0.900	0.366	0.206	0.312
Lcv	-	0.118	0.314	0.196	0.196	0.045	0.190
Med Low Flow	m^3/s	0.006	19.10	19.09	2.47	4.17	0.86
Spec low flow	$m^3/s/km^2$	0.001	0.009	0.008	0.004	0.002	0.004
Qavg	m^3/s	0.134	115.544	115.41	17.239	24.692	8.845
EffP	m	0.709	2.093	1.384	1.195	0.362	1.167
FD_10	-	1.897	2.736	0.839	2.258	0.199	2.243
FD_50	-	0.364	0.808	0.444	0.600	0.116	0.604
FD_90	-	0.070	0.359	0.289	0.182	0.066	0.180

All variables are measurable on continuous scales (not rank order, categorical or ordinal). Two descriptive variables were also included in the data set, the basin name, reduced to a three or four letter diminutive, and the WSC subregion identifier.

1. Drainage Area

In general the drainage areas (DA) were taken from HYDAT. For several basins, the drainage area was adjusted to account for discrepancies. About 44 km^2 was added to the Gander River basin as a result of a changed drainage boundary in the area of Lake Miguel. (The exact area is presently being determined by the Water Resources Division (WRD) of the provincial Department of Environment and Lands and WSC.) An additional 16 km^2 was added to the Rocky River basin, based on a recent change made by the WRD and WSC. An adjustment of 75 km^2 for Terra Nova River was made based on a review by Acres of 1:50 000 scale mapping and flyovers of the area. This adjustment has been called to the attention of the WRD and WSC and is under review.

2. Fractions of Drainage Area covered by Lakes and Swamps, Forests, and Barrens; Fraction of Area Controlled by Lakes and Swamps

These were obtained from WRD's database of physiographic characteristics. Definitions are provided in Gov't of Nfld, 1986 and Gov't of Nfld, 1990. They are abbreviated in this study as Fr-LSw, Fr-Forst, Fr-Barrn, and FACLS.

3. Length and Slope of the Main Channel

These were also obtained from the WRD database. The elevations are taken at the 10 percent and 85 percent points on the hypsometric curve, and the length is the difference between the two points along the main channel. The elevation difference used to calculate slope excluded waterfalls. This measure of slope is different from that used in the 1986 regional flood frequency analysis (Gov't of Nfld, 1986) which was an overall basin slope. It is abbreviated here as SLP1085. The measure used here for length is length of the main channel per unit of drainage area (LMC-Sp).

4. Drainage Density, Shape Factor

The drainage density (DrDens) and shape factor (Shape) were taken directly from the WRD database, with no adjustments.

5. Elevations of the Divide, Gauge and Basin Centroid

The elevations of the divide (El-Divde), gauge (El-Gauge) and centroid (El-Cntrd) were taken from 1:50 000 scale mapping specifically for this study. The elevation of the divide was taken as the lowest elevation along the drainage boundary in the vicinity of the origin of the main channel. The elevation of the gauge was taken from 1:50 000 scale mapping at the location of the gauge as identified by the WRD.

For a few basins, the locations and elevations of basins were obtained from various provincial Water Resources Studies (Gov't of Nfld, 1987, 1988, 1989, 1991, 1992,1993). For the remaining basins, the centroid had to be located before its elevation could be obtained. The drainage boundaries were outlined on 1:250 000 scale mapping (or 1:500 000 for very large basins), traced onto heavy card, and cut out. The location of the centroid was taken as the intersection of verticals marked when the card was suspended from three suspension points. The location was then transferred to 1:50 000 scale maps and the elevation was estimated from contours. The contour interval on the National Topographic Service (NTS) 1:50 000 scale mapping used for this and all similar projects is 50 ft (10 m on a few newer maps).

6. Distance Southwest

Weather systems approach the island from the southwest, so the amount or type of precipitation over the basin nay be related to the distance of the basin from the sea. The selected measure was the distance in kilometers of the centroid to the south or west coast (Dist-SW) in a southwesterly direction (a compass direction of 225 degrees), taken from an appropriate scale map (1:250 000 or 1:500 000). If this measure proves promising, other similar measures (e.g., shortest distance) could be tried and the results compared. A smoothed coastline was assumed, ignoring indentations for small bays. Figure 3.6 gives an example of how the distance is defined.

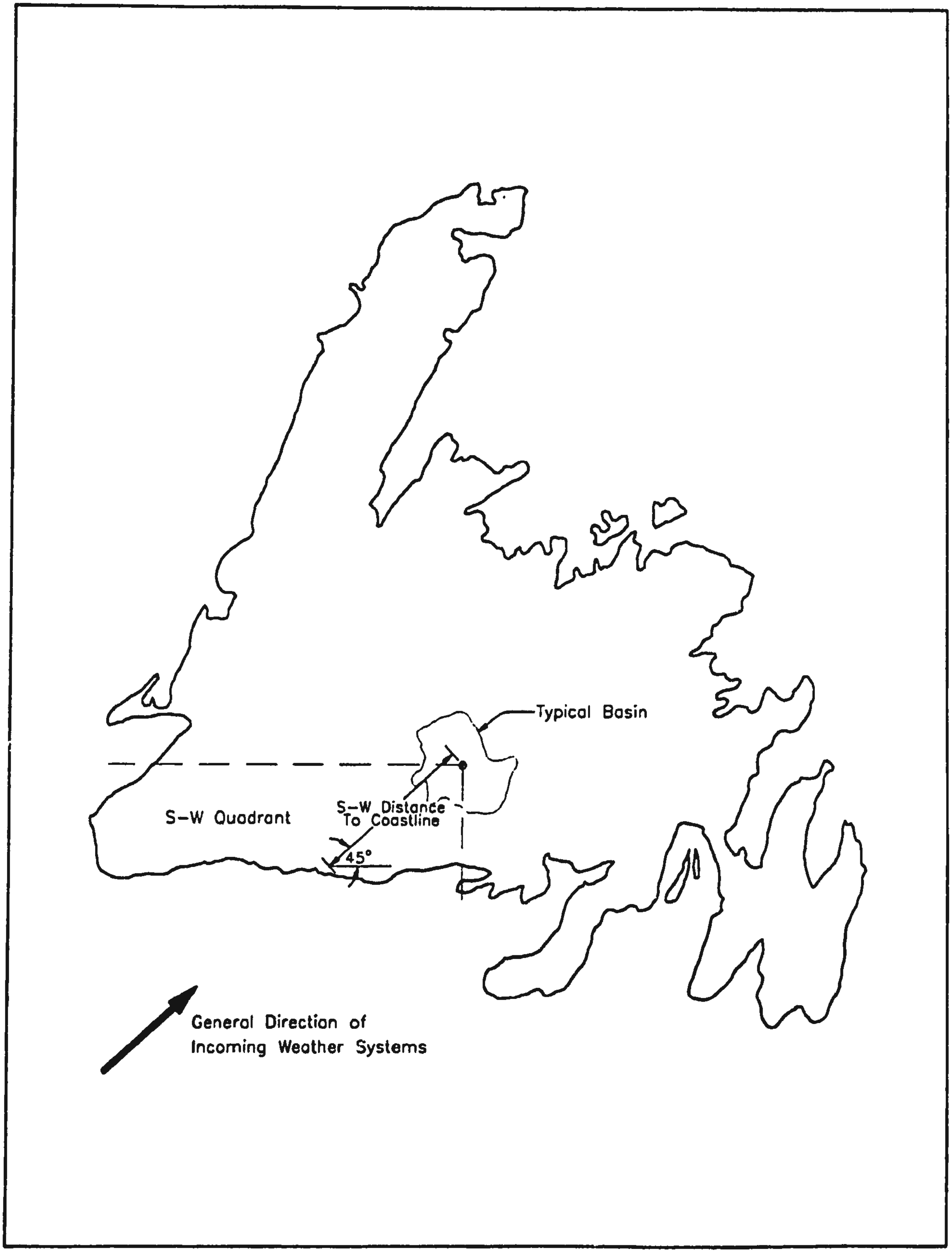


Fig. 3.6 – Example of Definition of Distance to the Sea

7. Distance North

The distance north (Dist-N) is the distance in kilometers from the centroid of the basin to latitude 46° 30′. It may be taken from appropriate scale mapping or by converting the difference in latitude to kilometers.

The complete list of basin characteristics is presented in Table 3.5, with the associated descriptive statistics in Table 3.6.

Table 3.5

Basin Physiographic Characteristics

ID	Code Name	Drainage Area km²	Length main chnl km	Elev Gauge m	Elev Divide m	Elev Centroid m	Slope 10/85	Drainage Density
YA1	SteG	306	39.8	12.2	107	81	0.23	0.540
YC1	Tor	624	49.9	7.6	488	309	1.01	0.755
YD1	BvrB	237	40.3	7.6	335	351	0.67	0.339
YD2	NeRk	200	38.0	22.9	290	110	0.47	0.930
YF1	Cat	611	30.6	350.5	594	549	0.73	0.582
YJ1	Hrys	640	57.6	15.2	541	274	0.35	1.120
YK2	LewB	470	56.4	137.2	655	351	0.59	0.627
YK4	Hnds	529	47.6	198.1	518	290	0.32	0.637
YK5	Shfd	391	39.1	109.7	488	290	1.07	0.191
YL1	UpHm	2110	126.6	22.9	701	183	0.40	0.786
YM3	SwBV	93.2	18.6	68.6	168	152	0.27	0.680
YN2	Lds	469	56.3	327.7	488	442	0.22	1.370
YO6	Ptrs	177	42.5	22.9	213	137	0.45	0.800
YP1	ShlA	63.8	15.2	83.8	198	152	0.53	0.880
YQ1	GdBC	4444	133.3	22.9	305	168	0.14	0.452
YR1	MdlB	275	49.5	22.9	198	122	0.32	0.255
YR2	RgdH	399	43.9	27.4	122	76	0.21	0.740
YR3	IndB	554	52.4	7.6	137	107	0.22	0.680
YS1	TerN	1365	109.2	83.8	290	244	0.12	0.726
YS3	SwTN	36.7	11.0	22.9	168	107	1.11	0.641
ZA1	LtlB	343	65.2	7.6	472	137	0.68	1.040
ZA2	Hlds	72	20.2	68.6	533	244	2.19	1.150
ZA3	LtCd	139	25.0	7.6	457	274	1.46	1.460
ZB1	IaM	205	32.8	7.6	457	335	0.84	0.720
ZC2	Grdy	230	29.9	83.8	442	335	1.06	0.960
ZE1	SmLP	264	100.4	182.9	305	274	0.08	0.360
ZF1	BdN	1170	70.2	7.6	274	152	0.29	0.612
ZG1	Grn	205	45.1	10.7	381	152	0.60	0.547
ZG2	Tds	166	26.6	7.6	229	213	1.35	1.350
ZG3	SmLm	115	24.2	1.5	137	91	0.34	1.550
ZG4	Rtl	42.7	9.8	45.7	152	122	1.10	1.620
ZH1	PH	764	53.5	38.1	244	213	0.35	0.709
ZH2	CbC	43.3	16.9	53.3	152	168	0.59	1.110
ZJ1	SthB	67.4	16.2	7.6	137	91	0.50	1.240
ZK1	Rky	300	45.0	14.6	168	60	0.50	1.005
ZK2	NE-P	89.6	26.9	15.2	213	120	0.55	1.110
ZL3	SpCv	10.8	7.0	76.2	168	170	1.25	1.090
ZM6	NEP	3.63	2.6	121.9	191	168	2.42	1.038
ZM9	SlCv	53.6	15.0	35.1	168	168	0.62	1.130
ZN1	NwNP	53.3	14.4	114.3	213	183	0.61	1.089

Table 3.5 Continued

ID	Code Name	Shape factor	Lakes & Swamps	cntrolld by L&S	Fraction Barren	Fraction Forest	from sea SW
YA1	SteG	1.48	0.354	1.000	0.002	0.644	17
YC1	Tor	1.45	0.167	0.990	0.498	0.335	74
YD1	BvrB	2.23	0.082	0.730	0.112	0.806	112
YD2	NeRk	1.65	0.170	0.990	0.001	0.829	184
YF1	Cat	1.86	0.131	0.000	0.180	0.689	123
YJ1	Hrys	1.81	0.142	0.750	0.069	0.789	34
YK2	LewB	2.32	0.161	1.000	0.290	0.549	150
YK4	Hnds	1.78	0.355	0.950	0.292	0.353	218
YK5	Shfd	1.98	0.172	0.940	0.152	0.676	266
YL1	UpHm	1.56	0.115	0.750	0.145	0.740	145
YM3	SwBV	1.67	0.100	0.560	0.001	0.900	225
YN2	Lds	2.15	0.160	1.000	0.620	0.220	89
YO6	Ptrs	1.93	0.160	0.970	0.018	0.822	222
YP1	ShlA	1.62	0.130	0.790	0.001	0.872	255
YQ1	GdBC	2.08	0.172	0.910	0.068	0.760	160
YR1	MdlB	1.93	0.245	0.980	0.008	0.747	180
YR2	RgdH	1.68	0.320	0.960	0.001	0.680	260
YR3	IndB	1.72	0.310	n.970	0.001	0.690	225
YS1	TerN	2.35	0.327	0.920	0.136	0.537	115
YS3	SwTN	1.43	0.159	1.000	0.005	0.836	175
ZA1	LtlB	2.45	0.100	0.830	0.299	0.601	128
ZA2	Hlds	1.72	0.050	0.430	0.129	0.820	65
ZA3	LtCd	1.67	0.130	0.730	0.190	0.680	26
ZB1	laM	2.09	0.134	0.600	0.782	0.084	27
ZC2	Grdy	1.84	0.040	0.340	0.790	0.170	40
ZE1	SmLP	1.75	0.160	1.000	0.490	0.350	105
ZF1	BdN	2.15	0.236	0.960	0.442	0.322	76
ZG1	Grn	2.45	0.101	0.960	0.634	0.265	66
ZG2	Tds	1.84	0.130	0.920	0.488	0.382	33
ZG3	SmLm	1.62	0.120	0.920	0.722	0.158	13
ZG4	Rtl	1.53	0.180	0.920	0.470	0.350	97
ZH1	PH	1.67	0.659	0.910	0.234	0.107	68
ZH2	CbC	1.66	0.099	0.920	0.496	0.405	20
ZJ1	SthB	1.43	0.150	0.860	0.033	0.817	125
ZK1	Rky	2.00	0.113	0.550	0.377	0.510	70
ZK2	NE-P	1.91	0.300	0.810	0.230	0.470	40
ZL3	SpCv	1.36	0.090	1.000	0.490	0.420	76
ZM6	NEP	1.24	0.226	1.000	0.039	0.738	100
ZM9	SlCv	1.37	0.130	1.000	0.500	0.370	28
ZN1	NwNP	2.06	0.126	1.000	0.788	0.086	38

Table 3.6

Descriptive Statistics of Basin Characteristics

Variable	Units	Mini-mum	Maxi-mum	Range	Mean	Std Dev	C.V.	Median
Drainage Area	km²	3.63	4444	4440.4	517.7	840.0	1.6	233.5
Dist SW	km	13	266	253	111.75	75.829	0.679	98.5
Dist N	km	35	445	410	194.75	113.04	0.58	180
Dr Density		0.19	1.62	1.43	0.87	0.35	0.40	0.79
El Centroid	m	60	549	489	204	107	1	168
El Divide	m	107	701	594	312	167	1	259
El Gauge	m	1.5	350.5	349.0	62.0	81.2	1.3	22.9
FACLS		0.34	1	0.66	0.871	0.168	0.194	0.93
FR Barren		0.00	0.79	0.79	0.28	0.26	0.92	0.21
Fr Forest		0.08	0.90	0.82	0.54	0.25	0.46	0.58
Fr Lakes&Swmps		0.04	0.66	0.62	0.18	0.11	0.62	0.16
Length Mn Chnl	km	0.03	0.72	0.69	0.19	0.14	0.75	0.17
Effective Precipitation	m	0.709	2.093	1.384	1.195	0.362	0.303	1.167
Shape		1.24	2.45	1.21	1.81	0.31	0.17	1.77
Slope 10/85	%	0.08	2.42	2.34	0.67	0.52	0.78	0.54

4 Assessment of Relative Importance of Basin Characteristics to Flow Variables

Having selected the basins, flow measures, and physiographic/geographic variables to include in the data set, the next step was to investigate the relationships among the flow variables and the basin variables.

Multivariate techniques such as cluster analysis, discriminant analysis and Andrews Fourier plots are sensitive to the variables selected, the transformations used (if any), and the weightings assigned to the variables. These techniques can only be used with a clear understanding of the key variables and of their relative importance. In this chapter, therefore, each of the flow variables is examined qualitatively and quantitatively to identify the important basin variables. The relationships can then be used directly in a one-region analysis, or to identify the key variables and weightings for further analysis of regions in geographical or basin characteristic dataspace.

The Island was first assumed to be one region. A brief comparison was also done for most of the flow variables assuming geographic subdivision along the boundary of the WSC hydrologic regions Y and Z.

The analysis proceeded as follows.

1. Plot the data and review the correlation matrix of all variables.

2. Identify principal components, to see whether there are any natural groupings of flow and basin variables, and to consider the possibility of reducing dimensionality using surrogate variables.

3. Explore the relationship between Eff-P and topographic or geographic characteristics, based on a close examination of the data set followed by multiple regression analysis.

4. Explore the relationships between flow variables in the three flow categories, high, low, and available flow, and basin characteristics. For each of the flow measures examine the data set carefully to identify the most likely explanatory variables for use in the follow-up multiple regression analysis. Consider these relationships both with and without Eff-P as a basin characteristic.

Depending on the results, mathematical and graphical cluster techniques can be used to assess whether these characteristics have the potential to identify similar clusters for all flow categories, and if so, what weightings of basin characteristics are likely to give good results.

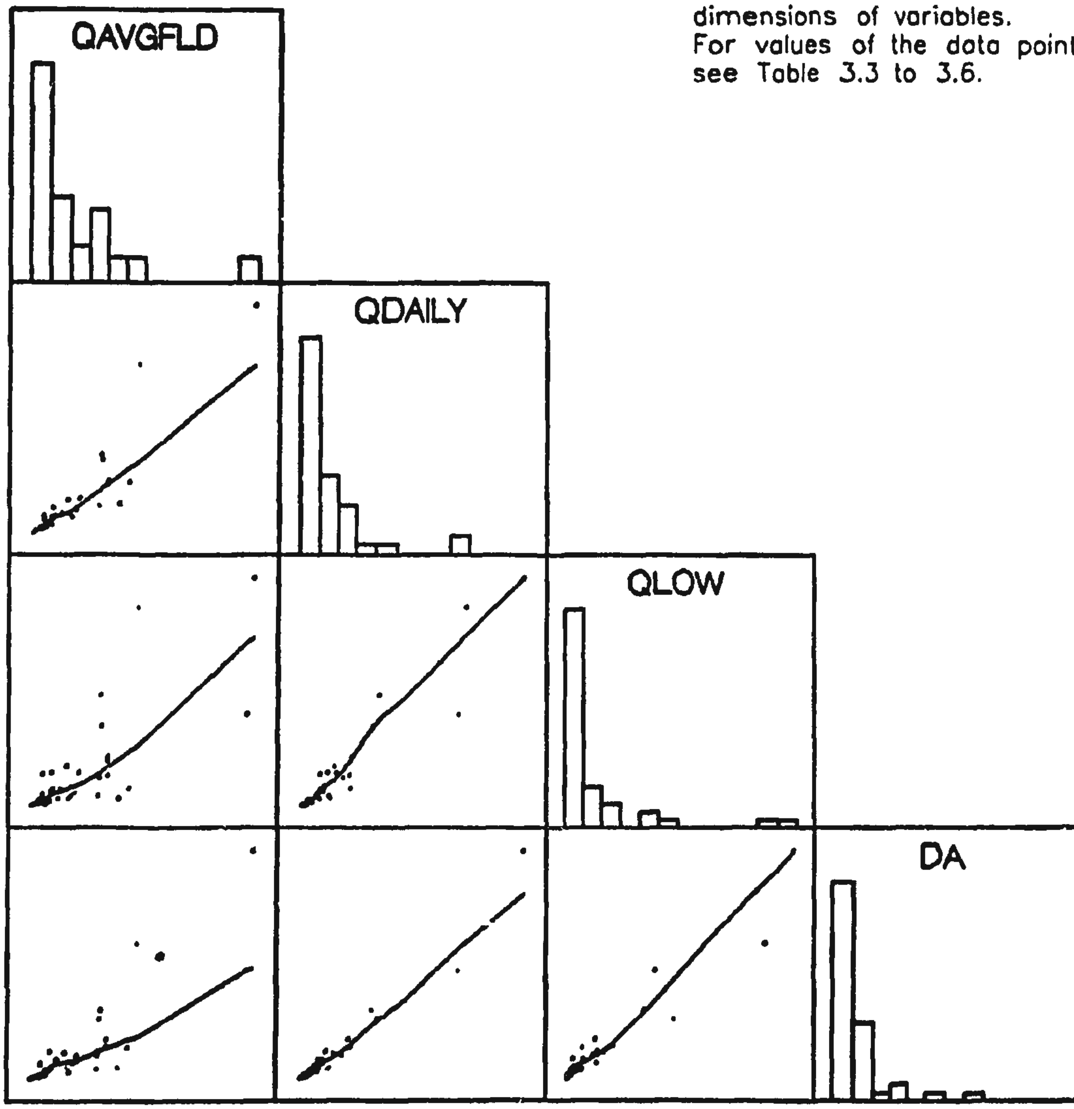
4.1 Initial Review of Relationships

4.1.1 Scatter Plots and Correlation Matrices

The first step in examining the relationships among the variables was to plot the data. The scatter plots for all pairs of variables were examined. Figures 4.1 to 4.8 show the scatter plots for the variables which have the potential to be important. A smoothed line is provided on the scatter plots, produced using locally weighted scatter plot smoothing (LOWESS). In addition, the correlation matrices for both untransformed and logarithmically transformed variables were examined, and are presented in Tables 4.1 and 4.2. In general the variables shown on the plots correspond to the variables with shaded correlation coefficients in the matrices (r greater than 0.5). A brief discussion follows of the relationships observed in the figures and tables.

4.1.2 Qavgfld, Qdaily, Qlow

The three flow variables Qavgfld, Qdaily, and Qlow (all with natural dimensions of m^3/s) are plotted in Figure 4.1, along with the basin variables DA and LMC-Sp. The flow variables are all related to each other, as expected since they are all strongly related to DA. Although the LOWESS smoothed lines appear nearly



Note: Scatter plots and histograms are provided to give a visual representation of the relationships among variables. The plots are smoothed using the LOWESS Procedure. See text for discussion and dimensions of variables. For values of the data points see Table 3.3 to 3.6.

Fig. 4.1 – Scatter Plots of Flow Variables and Drainage Area

Note: Scatter plots and histograms are provided to give a visual representation of the relationships among variables. The plots are smoothed using the LOWESS Procedure.
See text for discussion and dimensions of variables.
For values of the data points see Table 3.3 to 3.6.

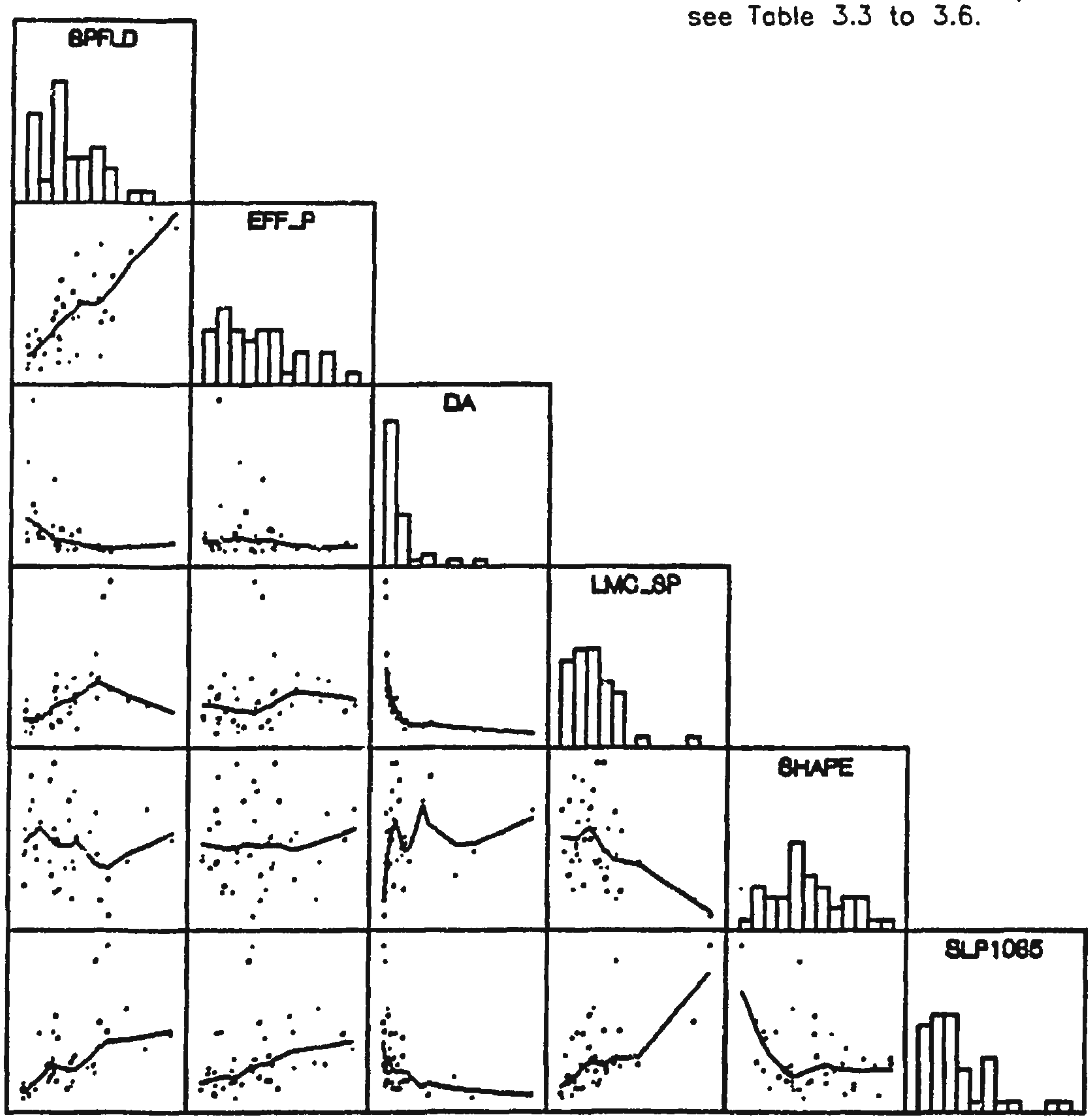
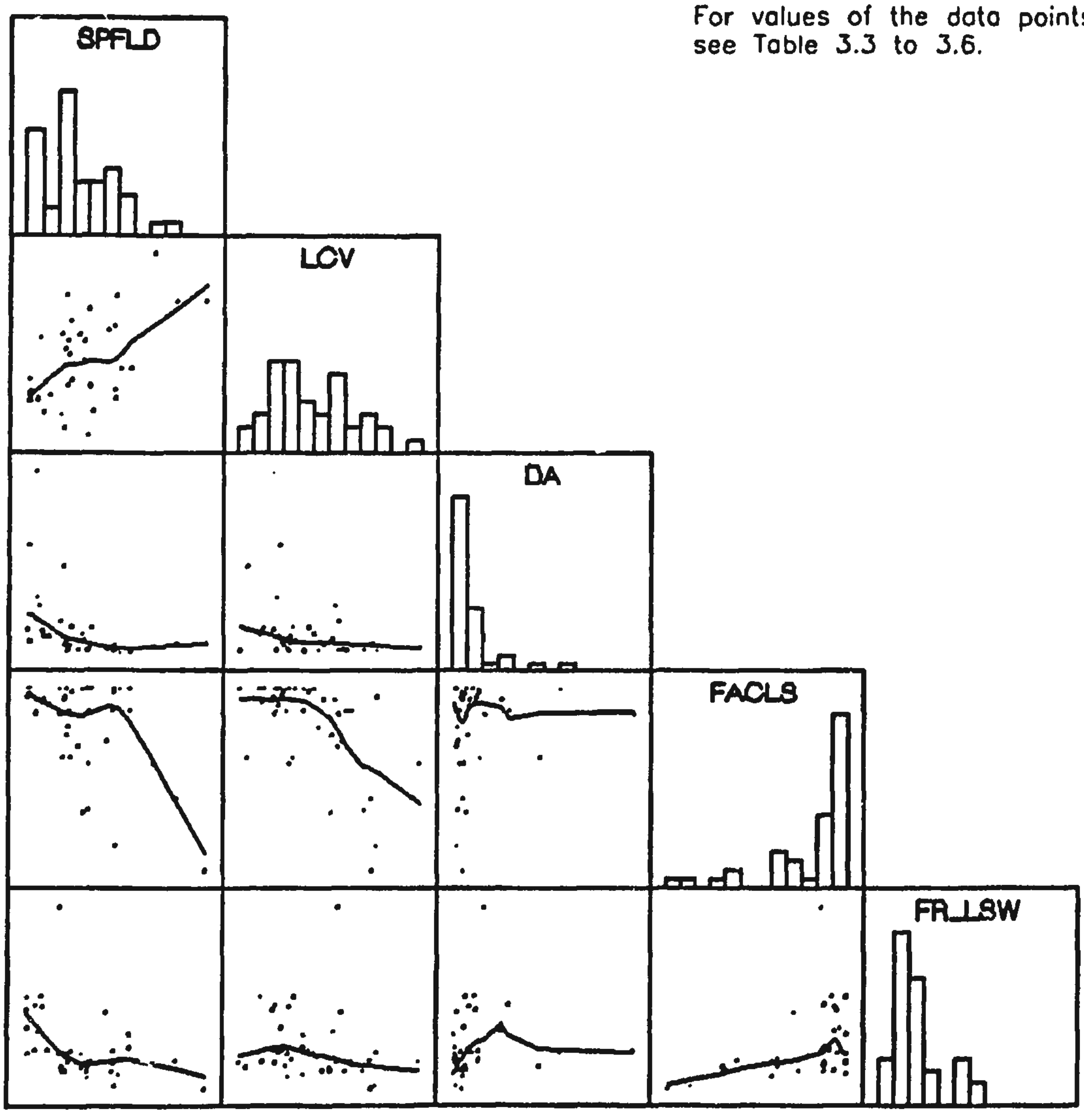


Fig. 4.2 — Scatter Plots of Sp—Fld and Basin Characteristics



Note: Scatter plots and histograms are provided to give a visual representation of the relationships among variables. The plots are smoothed using the LOWESS Procedure. See text for discussion and dimensions of variables. For values of the data points see Table 3.3 to 3.6.

Fig. 4.3 – Scatter Plots of Sp--Fld, Lcv and Basin Variables

Note: Scatter plots and histograms are provided to give a visual representation of the relationships among variables. The plots are smoothed using the LOWESS Procedure. See text for discussion and dimensions of variables. For values of the data points see Table 3.3 to 3.6.

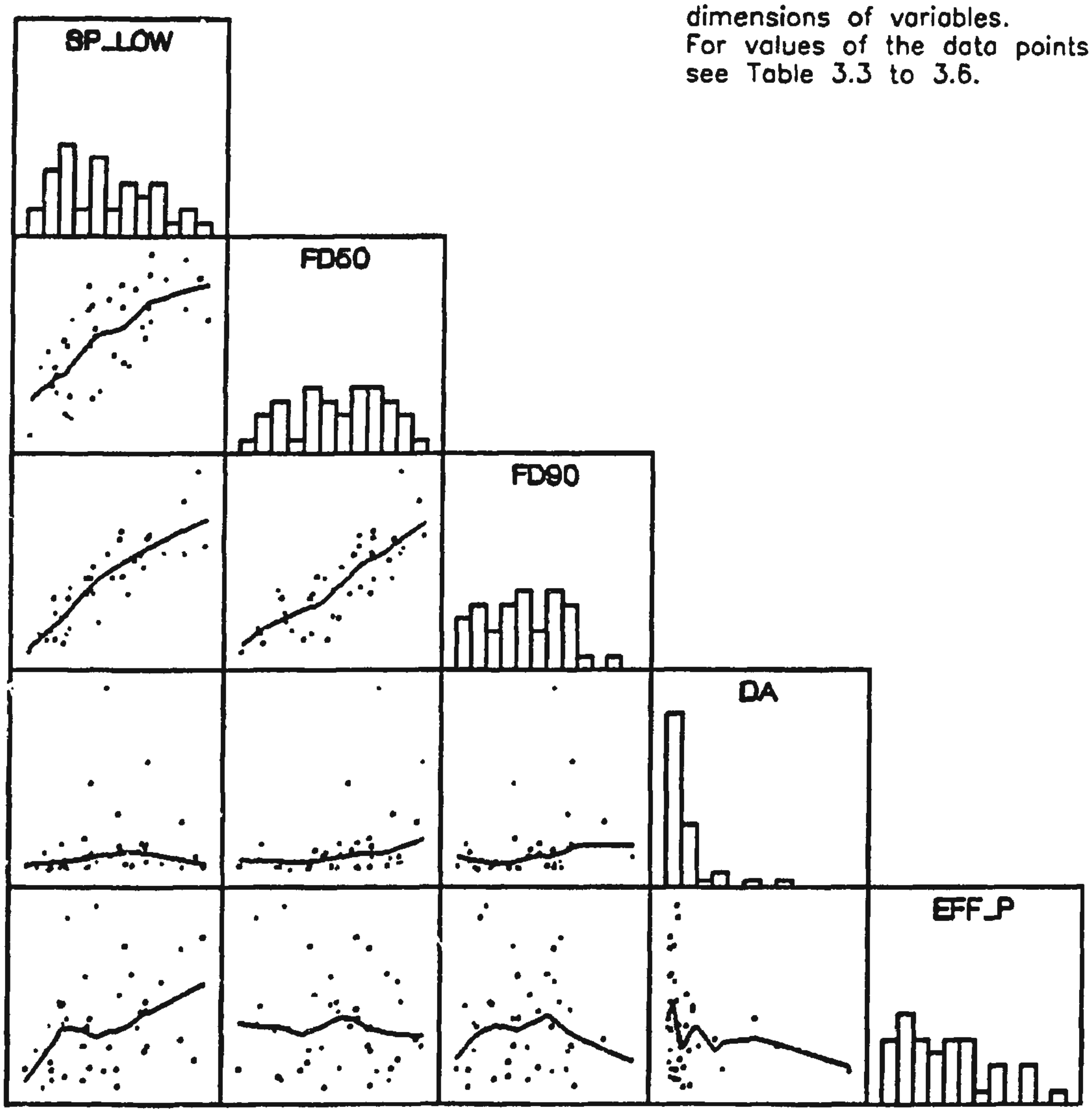


Fig. 4.4 – Scatter Plots of Low Flow, Flow Duration Measures, Drainage Area and Effective Precipitation

Note: Scatter plots and histograms are provided to give a visual representation of the relationships among variables. The plots are smoothed using the LOWESS Procedure. See text for discussion and dimensions of variables. For values of the data points see Table 3.3 to 3.6.

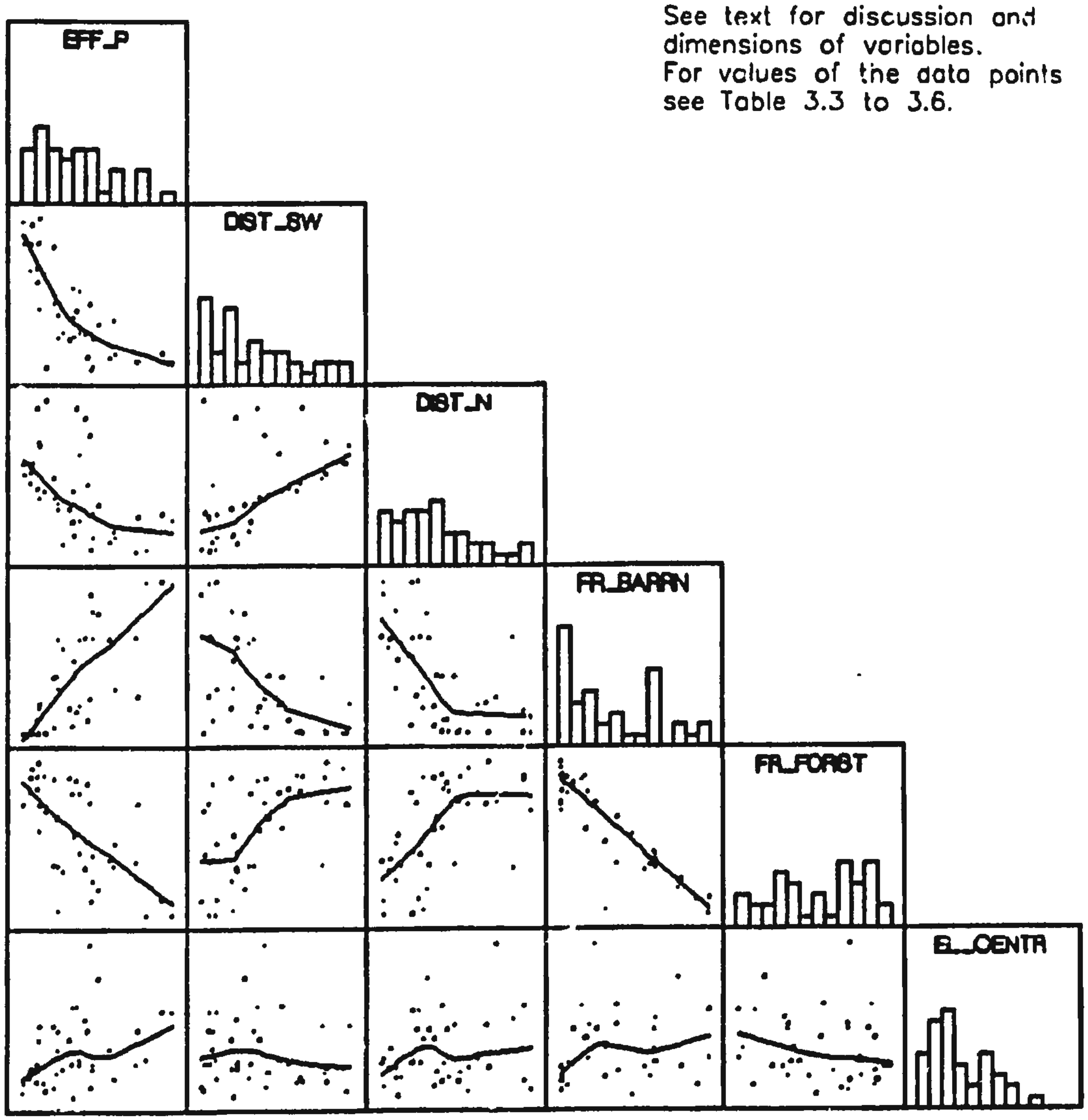


Fig. 4.5 – Scatter Plots of Effective Precipitation, Distance, and Basin Variables

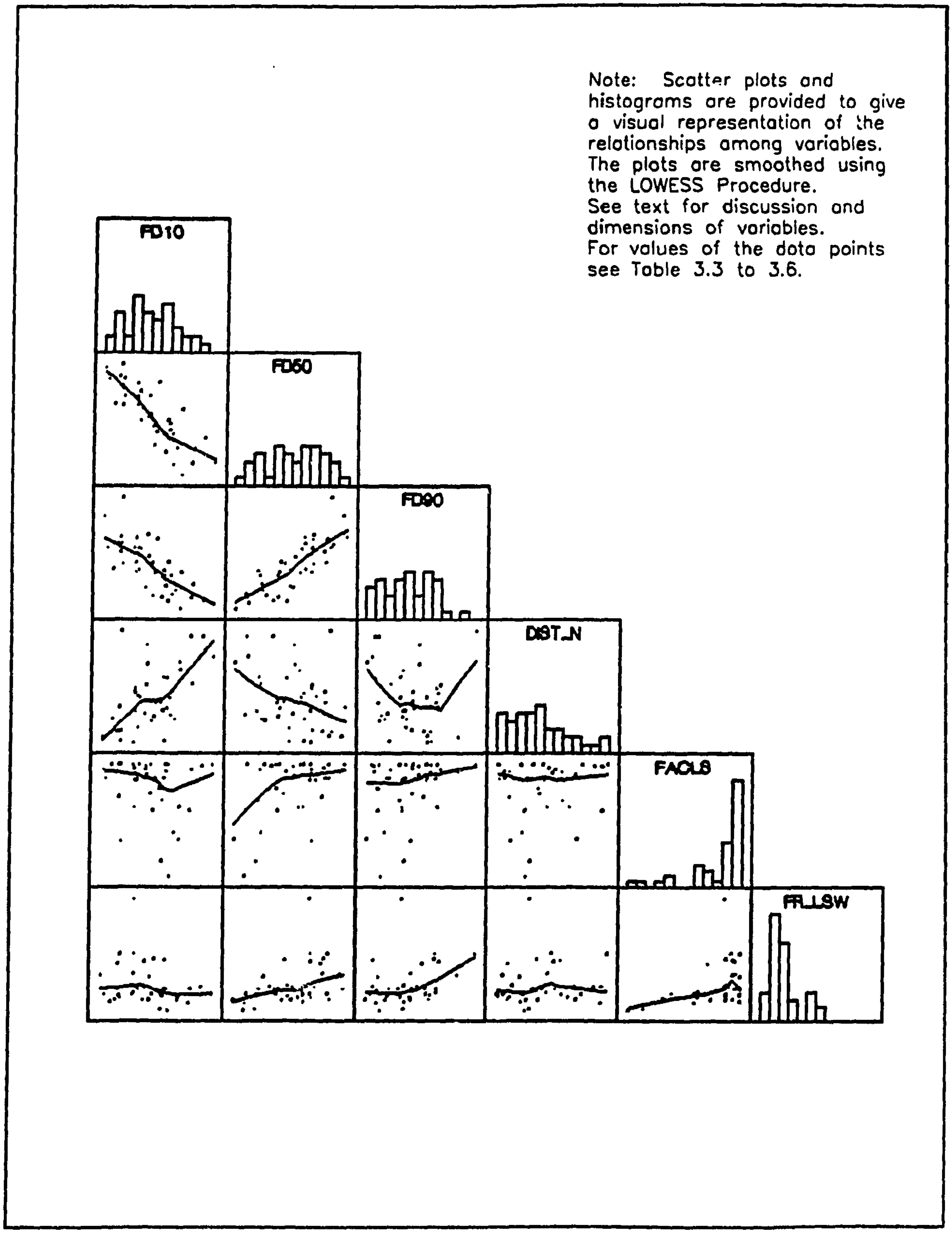


Fig. 4.6 – Scatter Plots of Flow Duration and Basin Variables

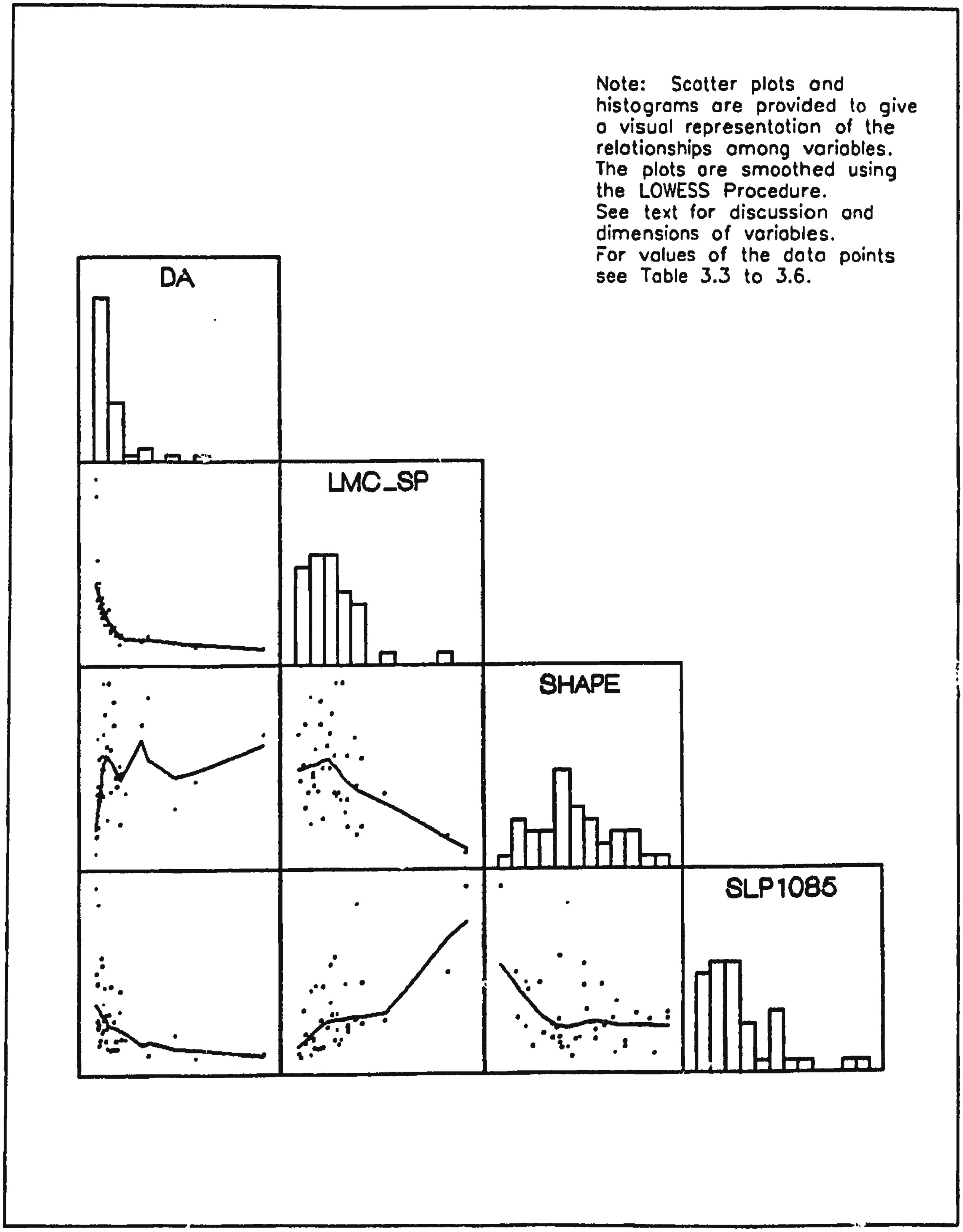


Fig 4.7 – Scatter Plots of Basin Variables Related to Drainage Area

Note: Scatter plots and histograms are provided to give a visual representation of the relationships among variables. The plots are smoothed using the LOWESS Procedure. See text for discussion and dimensions of variables. For values of the data points see Table 3.3 to 3.6.

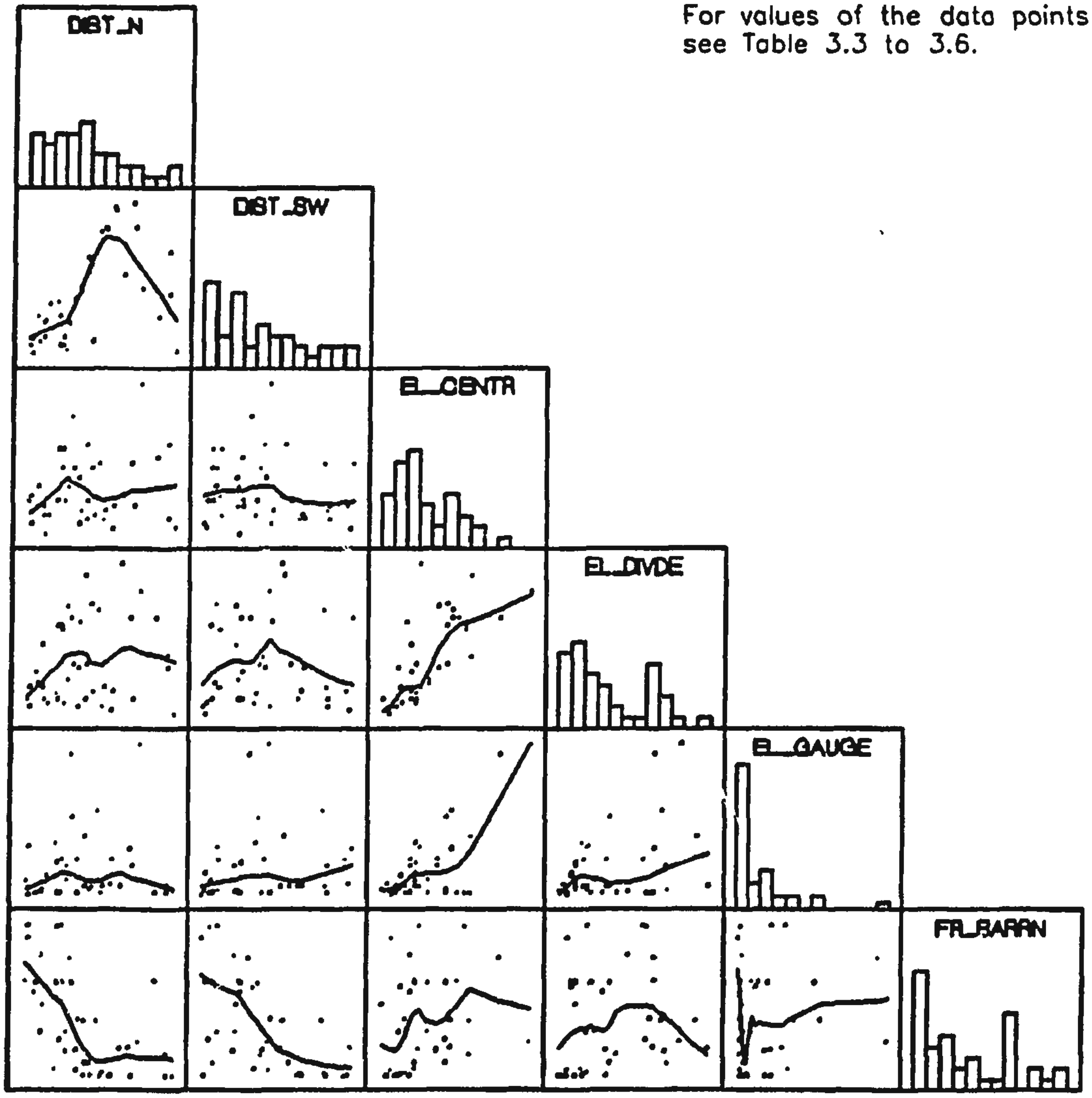


Fig. 4.8 – Scatter Plots of Distance and Elevation Variables

Table 4.1

Pearson Correlation Matrix: Untransformed Variables

	QAvgFld	QDaily	QLow	SpFld	Sp_Low	Sp_Rnoff	Eff−P	Lcv	FD10	FD50	FD90	DA	Dist_N	Dist_SW	Dr_Dens	El_Centr
QAvgFld	1.00															
QDaily	0.92	1.00														
QLow	0.77	0.95	1.00													
SpFld	−0.16	−0.37	−0.46	1.00												
Sp_Low	0.10	0.20	0.32	−0.26	1.00											
Sp_Rnoff	0.01	−0.16	−0.23	0.75	0.32	1.00										
Eff−P	0.02	−0.15	−0.23	0.75	0.31	1.00	1.00									
Lcv	−0.23	−0.32	−0.30	0.51	−0.20	0.28	0.28	1.00								
FD10	0.11	−0.08	−0.25	0.26	−0.67	−0.13	−0.13	0.03	1.00							
FD50	0.14	0.36	0.49	−0.58	0.69	−0.11	−0.10	−0.34	−0.82	1.00						
FD90	0.11	0.29	0.44	−0.52	0.87	−0.06	−0.06	−0.26	−0.74	0.82	1.00					
DA	0.86	0.98	0.96	−0.42	0.17	−0.25	−0.25	−0.33	−0.11	0.37	0.30	1.00				
Dist_N	0.14	0.09	0.03	−0.29	−0.26	−0.52	−0.52	−0.24	0.54	−0.33	−0.16	0.10	1.00			
Dist_SW	0.00	0.06	0.05	−0.44	−0.43	−0.74	−0.74	−0.26	0.37	−0.22	−0.21	0.13	0.51	1.00		
Dr_Dens	−0.28	−0.36	−0.39	0.52	−0.02	0.50	0.49	0.34	−0.05	−0.16	−0.18	−0.39	−0.53	−0.49	1.00	
El_Centr	0.33	0.19	0.10	0.22	0.10	0.36	0.36	−0.04	0.31	−0.29	−0.09	0.10	0.18	−0.12	−0.15	1.00
El_Divde	0.52	0.34	0.19	0.10	0.17	0.22	0.22	−0.07	0.26	−0.16	0.03	0.22	0.21	−0.01	−0.16	0.73
El_Gauge	0.13	0.12	0.09	0.01	0.00	0.05	0.05	−0.26	0.21	−0.17	−0.07	0.07	0.08	0.14	−0.08	0.69
FACLS	−0.13	0.06	0.16	−0.53	0.26	−0.32	−0.32	−0.57	−0.22	0.46	0.32	0.11	0.07	0.17	−0.15	−0.09
Fr_Barn	0.03	−0.03	−0.03	0.45	0.37	0.76	0.76	0.13	−0.32	0.15	0.11	−0.10	−0.59	−0.62	0.34	0.26
Fr_Forst	−0.03	−0.01	−0.03	−0.27	−0.44	−0.63	−0.63	−0.06	0.43	−0.34	−0.24	0.05	0.57	0.58	−0.24	−0.20
Fr_Low	0.01	0.07	0.12	−0.44	0.14	−0.36	−0.35	−0.17	−0.22	0.41	0.29	0.12	0.09	0.14	−0.24	−0.16
Lnc_Sp	−0.55	−0.51	−0.47	0.43	−0.29	0.14	0.14	0.17	−0.10	−0.22	−0.29	−0.48	−0.31	−0.17	0.36	−0.29
Shape	0.21	0.19	0.21	−0.17	0.27	0.03	0.03	0.05	−0.07	0.17	0.23	0.20	−0.10	0.02	−0.30	0.29
Slp1085	−0.29	−0.37	−0.39	0.56	−0.17	0.33	0.33	0.36	0.03	−0.37	−0.27	−0.39	−0.21	−0.23	0.33	0.13

Table 4 .1 Continued

	El_Centr	El_Divde	El_Gauge	FACLS	Fr_Barrn	Fr_Forst	Fr_Low	Lmc_Sp	Shape	Slp1085
QAvgFld										
QDaily										
QLow										
SpFld										
Sp_Low										
Sp_Rnoff										
Eff-P										
Lcv										
FD10										
FD50										
FD90										
DA										
Dist_N										
Dist_SW										
Dr_Dens										
El_Centr	1.00									
El_Divde	0.73	1.00								
El_Gauge	0.69	0.38	1.00							
FACLS	-0.09	-0.23	0.21	1.00						
Fr_Barrn	0.26	0.12	0.14	-0.09	1.00					
Fr_Forst	-0.20	-0.02	-0.13	-0.07	-0.90	1.00				
Fr_Low	-0.16	-0.23	-0.02	0.36	-0.29	-0.15	1.00			
Lmc_Sp	-0.29	-0.39	-0.07	0.08	0.03	0.07	-0.23	1.00		
Shape	0.29	0.37	0.10	-0.09	0.21	-0.18	-0.08	-0.43	1.00	
Slp1085	0.13	0.13	-0.01	-0.24	0.06	0.08	-0.32	0.61	-0.32	1.00

Table 4.2

Pearson Correlation Matrix: Logarithmically Transformed Variables

	QaFld_Ln	Qdly_Ln	QLow_Ln	SpFld_Ln	SpLow_Ln	SpRnf_Ln	Eff−P_Ln	Lcv_Ln	FD10_Ln	FD50_Ln	FD90_Ln	DA_Ln	DstSW_Ln
QaFld_Ln	1.00												
Qdly_Ln	0.96	1.00											
QLow_Ln	0.87	0.96	1.00										
SpFld_Ln	−0.25	−0.48	−0.59	1.00									
SpLow_Ln	0.34	0.46	0.65	−0.28	1.00								
SpRnf_Ln	0.06	−0.06	−0.09	0.71	0.35	1.00							
Eff−P_Ln	0.08	−0.05	−0.08	0.70	0.35	1.00	1.00						
Lcv_Ln	−0.09	−0.24	−0.30	0.45	−0.22	0.22	0.21	1.00					
FD10_Ln	0.11	−0.04	−0.22	0.30	−0.64	−0.16	−0.16	−0.00	1.00				
FD50_Ln	0.14	0.35	0.53	−0.60	0.70	−0.08	−0.06	−0.34	−0.81	1.00			
FD90_Ln	0.19	0.37	0.59	−0.53	0.88	0.02	0.02	−0.24	−0.76	0.82	1.00		
DA_Ln	0.91	0.98	0.95	−0.61	0.37	−0.26	−0.25	−0.27	−0.00	0.35	0.35	1.00	
DstSW_Ln	0.04	0.06	0.05	−0.39	−0.40	−0.72	−0.73	−0.26	0.33	−0.19	−0.20	0.21	1.00
Dst_N_Ln	0.21	0.21	0.14	−0.30	−0.35	−0.59	−0.59	−0.12	0.55	−0.35	−0.27	0.32	0.57
Drd_Ln	−0.26	−0.36	−0.38	0.60	−0.03	0.49	0.48	0.34	−0.01	−0.14	−0.13	−0.45	−0.47
Elc_Ln	0.44	0.35	0.29	0.22	0.24	0.40	0.39	−0.03	0.21	−0.25	−0.00	0.26	−0.03
ElDv_Ln	0.62	0.53	0.48	0.14	0.31	0.27	0.27	−0.05	0.22	−0.14	0.09	0.45	0.10
Elg_Ln	−0.03	−0.05	−0.06	0.05	−0.06	−0.03	−0.04	−0.24	0.11	−0.17	−0.06	−0.05	0.35
FACLS_Ln	−0.21	−0.03	0.10	−0.47	0.22	−0.30	−0.30	−0.52	−0.23	0.50	0.31	0.03	0.17
FrBrn_Ln	0.29	0.22	0.22	0.34	0.48	0.73	0.75	0.06	−0.35	0.24	0.27	0.06	−0.51
FrFst_Ln	−0.18	−0.15	−0.14	−0.28	−0.36	−0.61	−0.61	−0.12	0.29	−0.19	−0.16	−0.02	0.52
FrLws_Ln	0.02	0.18	0.29	−0.59	0.19	−0.44	−0.43	−0.28	−0.24	0.50	0.33	0.27	0.20
Lmc_Ln	−0.88	−0.94	−0.89	0.57	−0.32	0.22	0.21	0.32	−0.01	−0.28	−0.29	−0.95	−0.22
Shp_Ln	0.54	0.53	0.54	−0.18	0.33	0.02	0.02	0.08	−0.06	0.14	0.24	0.51	0.10
Slp_Ln	−0.46	−0.60	−0.59	0.71	−0.12	0.47	0.47	0.35	0.08	−0.41	−0.27	−0.67	−0.20

Table 4.2 Continued

	Dst_N_Ln	Drd_Ln	Elc_Ln	ElDv_Ln	Elg_Ln	FACLS_Ln	FrBm_Ln	FrFst_Ln	FrLws_Ln	Lmc_Ln	Shp_Ln	Slp_Ln
QaFld_Ln												
Qdly_Ln												
QLow_Ln												
SpFld_Ln												
SpLow_Ln												
SpRnf_Ln												
Eff-P_Ln												
Lcv_Ln												
FD10_Ln												
FD50_Ln												
FD90_Ln												
DA_Ln												
DstSW_Ln												
Dst_N_Ln	1.00											
Drd_Ln	−0.49	1.00										
Elc_Ln	0.17	−0.17	1.00									
ElDv_Ln	0.23	−0.17	0.78	1.00								
Elg_Ln	0.17	−0.15	0.47	0.26	1.00							
FACLS_Ln	0.01	−0.17	−0.13	−0.24	0.05	1.00						
FrBm_Ln	−0.59	0.22	0.47	0.42	0.06	−0.09	1.00					
FrFst_Ln	0.57	−0.21	−0.22	−0.06	0.01	0.05	−0.63	1.00				
FrLws_Ln	0.13	−0.22	−0.23	−0.27	−0.00	0.60	−0.29	−0.02	1.00			
Lmc_Ln	−0.32	0.45	0.32	−0.43	−0.06	−0.06	−0.06	0.04	−0.30	1.00		
Shp_Ln	−0.04	−0.30	0.29	0.45	−0.01	−0.10	0.28	−0.18	−0.07	−0.36	1.00	
Slp_Ln	−0.19	0.32	0.16	0.16	−0.04	−0.27	0.24	−0.00	−0.46	0.61	−0.28	1.00

linear for all the relationships, the correlation coefficients for the logarithmic relationships are higher. The exception is Qlow and DA, where the untransformed relationship with DA is stronger. Although the plots are not shown here, the three flow variables are also all related to the unit length of the main channel (LMC-Sp, in m^3/s per km^2), and to shape factor. This results is expected because of the relationships among the basin variables. There are also some weak relationships between these flow variables and El-Dvide. Relationships with other variables are probably masked by the dominant effect of drainage area.

4.1.3 Unit Flow Measures: SpFld, Sp-Low, Sp-Runoff

An examination of the specific flow measures (flows per unit area expressed as m^3/s per km^2) can be helpful in understanding possible relationships with basin variables, because the overwhelming influence of drainage area has been removed. The correlation observed between the natural flow measures and the specific flow measures (e.g., $r = 0.65$ between Qlow-Ln and SpLow-Ln) reflects the fact that there remain some other factors besides drainage area which may explain hydrologic response.

The scatter plots in Figure 4.2 show that the strongest relationship is between Sp-Fld and Eff-P; unit runoff during floods is higher in areas with higher annual average effective precipitation. Smaller basins also appear to have higher unit runoff during floods, as do those with steeper slopes. The relationships appear to be nonlinear rather than linear, although this is not entirely clear due to the considerable scatter in the plots.

The correlation coefficients of the log-transformed variables are slightly higher, also suggesting a nonlinear relationship.

Figure 4.3 shows some additional plots, including Lcv and other basin variables. These plots suggest that basins with higher unit runoff during floods also have greater flood variability (as represented by Lcv). Variation in annual floods could arise for different sets of reasons, and possibly Lcv is one variable which can be better explained with regionalization. Unit runoff during floods also tends to decrease with increasing fractions of lakes and swamps, particularly if the location and size of lakes and swamps means that they control a large fraction of the drainage area.

There are no basin physiographic characteristics associated with Sp-Low; this result is not unexpected because the search for explanatory variables for low flows among basin physiographic characteristics is frequently unsuccessful. Figure 4.4 shows the large scatter and flat slope in the plot of Sp-Low and DA.

Sp-Low is quite closely related to the other low flow measure FD-90, as well as to the availability measure FD-50.

Figure 4.5 shows the scatter plots of Eff-P and some of the distance and basin variables. The strongest relationships are with Dist-SW and Fr-Barrn, with weaker relationships with Fr-Frost and El-Cntrd. Dist-SW is negatively correlated with Fr-Barrn and positively correlated with Fr-Forst. Fr-Barrn and Fr-Forst are naturally related, since by definition the fractions of barren, forest and lakes and swamps must sum to 1.0.

These relationships with Eff-P make physical sense. The weather systems arrive from the southwest, and it is quite evident from a map showing land cover that the areas near the south coast, exposed to these oncoming weather systems, are barren. The more sheltered areas in the lee of the Long Range Mountains and interior uplands are by contrast more forested.

These results are encouraging, because Eff-P appears to be an important explanatory variable throughout the range of flows. An understanding of the relationships between Eff-P and topographic and geographic variables can provide a basis for estimating Eff-P.

4.1.4 Flow Duration Curve Variables:FD-10, FD-50, FD-90

FD-10, FD-50, and FD-90 are the flows which are exceeded 10, 50 and 90 percent of the time. FD-10 is a relatively high flow (about twice the average flow), FD-50 is the median daily flow, and FD-90 is a low flow (about 20 percent of the average flow). They provide a measure of the flashiness of a basin. They are nondimensionalized by dividing by the mean annual flow.

As Figure 4.6 shows, that the strongest relationships the FD variables show are with each other (as well as with Sp-Low, as was shown in Figure 4.4). FD-90, like the other low flow measures, shows no correlation with any basin physiographic characteristics. The plots suggest that FD-50 may be weakly related to FACLS and Fr-Lsw, and this observation is corroborated by the correlation coefficients of the log-transformed variables. FD-10 and FD-50 also show a weak relationship with the distance Dist-N; this observation may suggest a geographic regionalization.

4.1.5 Basin Characteristics

The basin characteristics are generally not correlated with each other, except for those associated with DA or those which are different representations of essentially the same characteristic. Figure 4.7, for example, shows the scatter plots of the variables related to DA. The LOWESS smoothed lines suggest that the relationships are nonlinear, corroborated by the higher correlation coefficients after logarithmic transformation.

The specific length of the main channel LMC-Sp is strongly correlated with DA, i.e., the larger the drainage basin the shorter the relative length of main channel. DA has a negative but weaker relationship with Slp1085; the larger the basin the milder the slope. The related relationship is a moderate positive association between LMC-Sp and Slp1085. Large drainage areas are also weakly associated with a higher shape factor (essentially a more rounded shape).

The relationships among the distance and elevation variables are shown in Figure 4.8; there are no surprises. The two distance variables are obviously somewhat related. The change from positive to negative slope occurs because of the basins on the west coast and Great Northern Peninsula. They are a long distance north, but only a short distance to the sea in a southwesterly direction. There tends to be a decreasing fraction of barren area (and increasing fraction of forest) with both Dist-N and Dist-SW. There is also a weak trend to increased drainage density with Dist-N, perhaps explicable by geology.

The elevation variables are correlated only among each other. The elevation of the centroid has about the same degree of association with the elevation of the divide as with the elevation of the gauge (0.73 and 0.69). The elevation of the divide and the elevation of the gauge are not related.

4.2 Principal Components Analysis

The objectives of a Principal Components Analysis (PCA) are

1) to identify groupings of variables;
2) to identify new meaningful underlying variables:
3) to reduce the dimensionality of the original problem, by allowing the substitution of Principal Components (PC's) or surrogate variables;
4) to identify variables that contribute little to the explanation of the problem.

If variables are combined into components representing the dependent and independent variables for regression, the analysis then becomes a kind of canonical correlation, with the principal components as the canonical variables. PCA clusters variables, and can bring to light previously unnoticed groupings, or generate hypotheses. It transforms the original set of correlated variables to a new set of uncorrelated variables, which are the principal components. These are linear combinations derived in decreasing order of importance. If the first few components account for most of the variance then the true dimensionality of the problem is less than the number of the original variables.

If the original variables are nearly uncorrelated then there is not much point in carrying out a PCA, since the PCA will simply find components close to the original variables in decreasing order of importance. Since principal components analysis is another way of looking at correlations, it is more likely to confirm the results of a careful examination of the scatter plots and the correlation matrix than to produce any new insights.

If the principal components become an important part of the analysis, some attention should be given to the fact that the factors are linear combinations. If some or all of the variables are logarithmically transformed before the components are obtained, there might be problems when the final results are transformed back.

Since the primary purpose of the PCA analysis for this study is exploratory, to identify key variables, the untransformed but standardized data set was used. The analysis was carried out separately on the set of flow variables and on the set of basin characteristics.

4.2.1 Principal Components of Basin Characteristics

Several analyses were done to compare the resulting components for different sets of variables. The set of basin characteristics was analyzed with and without Eff-P. In addition, the results were examined when only one elevation variable was used. The results are described below in more detail, but in general the number of principal components, the "meaning" for each, and the total variance explained did not change much for the different cases.

The plot in Figure 4.9 shows the eigenvalues for the components of basin characteristics. When all variables are included, or when only Eff-P is removed, there is a noticeable break in slope at PC5; when two of the elevations (El-Gauge and El-Divde) are excluded as well as Eff-P, the break occurs at PC4 or 5. The reduction in the first eigenvalue only when Eff-P is removed shows that nearly all the influence of Eff-P is included in the first component.

Tables 4.3, 4.4 and 4.5 show the loadings for the first four and five principal components for the three cases plotted in Figure 4.9. The highest loading for each variable is shaded. For the first two cases (all basin variables, with and without Eff-P), the components are very similar, as expected from the plots of the eigenvalues.

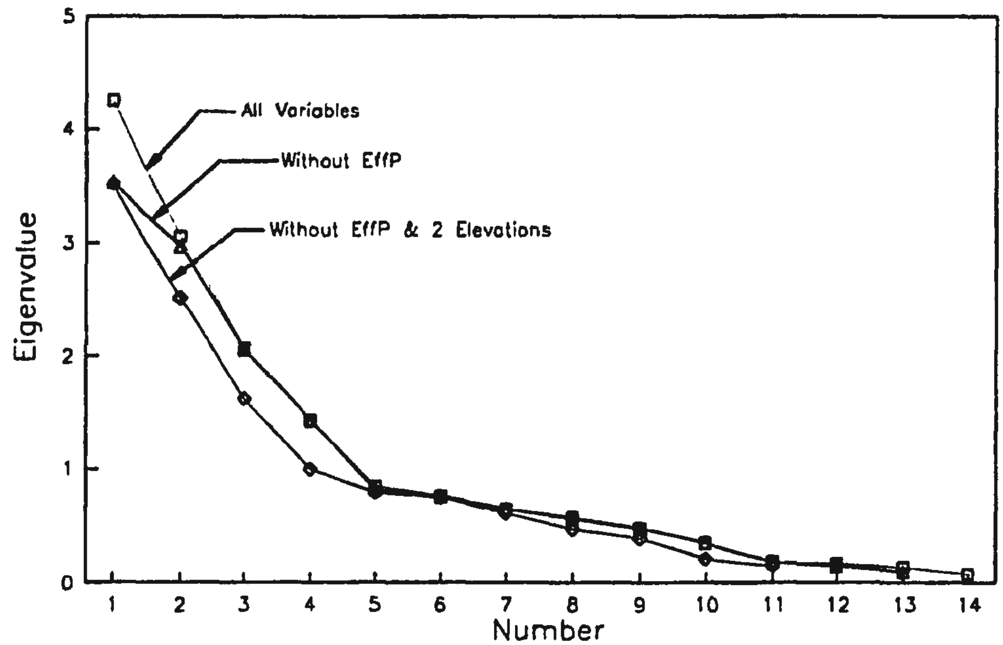


Fig. 4.9 – Eigenvalues: Basin Characteristics

Table 4.3

Rotated Loadings: All Basin Characteristics

a) 5 Principal Components

	1	2	3	4	5	
Fr_Forst	−0.91	−0.10	−0.17	0.21	0.05	
Fr_Barrn	0.90	0.16	−0.01	−0.03	0.24	
Eff−P	0.81	0.25	−0.24	0.33	0.03	
Dist_SW	−0.80	0.01	0.15	−0.27	0.17	
Dist_N	−0.75	0.28	0.20	0.09	−0.24	
Dr_Dens	0.50	−0.17	−0.53	0.17	−0.10	
El_Centrd	0.13	0.95	0.09	0.07	0.04	
El_Gauge	0.01	0.79	−0.06	−0.44	0.08	
El_Divde	−0.01	0.78	0.21	0.35	0.14	
LMC_Sp	0.05	−0.23	−0.86	−0.19	0.09	
Slp1085	0.05	0.20	−0.78	0.27	−0.02	
DA	−0.11	0.09	0.65	−0.00	0.00	
Shape	0.14	0.17	0.55	0.02	0.66	
FACLS	−0.07	−0.02	0.05	−0.85	−0.13	
Fr Lsw	−0.05	−0.15	0.39	−0.42	−0.65	
Percent of Variance Explained	28.4	20.3	13.8	9.6	5.6	Total 77.6 %

b) 4 Principal Components

	1	2	3	4	
Fr_Barrn	0.91	0.19	0.06	0.06	
Fr_Forst	−0.90	−0.13	−0.16	0.25	
Eff−P	0.81	0.22	−0.17	0.36	
Dist_SW	−0.79	0.04	0.13	−0.19	
Dist_N	−0.77	0.23	0.17	0.01	
Dr_Dens	0.51	−0.16	−0.52	0.17	
El_Centrd	0.11	0.92	0.16	0.16	
El_Gauge	0.00	0.85	−0.05	−0.28	
El_Divde	−0.03	0.72	0.31	0.42	
LMC_Sp	0.09	−0.15	−0.86	−0.03	
Slp1085	0.07	0.19	−0.74	0.38	
Shape	0.16	0.18	0.66	0.19	
DA	−0.13	0.05	0.64	−0.09	
FACLS	−0.07	0.09	−0.06	−0.83	
Fr Lsw	−0.10	−0.15	0.22	−0.69	
Percent of Variance Explained	25.7	16.0	17.7	12.7	Total 72.1 %

Table 4.4

Rotated Loadings: All Basin Characteristicsexcept Eff–P

a) 5 Principal Components

	1	2	3	4	5	
Fr_Forst	0.92	−0.13	−0.15	0.18	0.04	
Fr_Barrn	−0.90	0.19	−0.04	0.01	0.24	
Dist_SW	0.78	−0.02	0.15	−0.30	0.18	
Dist_N	0.77	0.24	0.23	0.04	−0.19	
El_Centrd	−0.09	0.95	0.09	0.07	0.06	
El_Gauge	−0.01	0.80	−0.07	−0.42	0.06	
El_Divde	0.04	0.78	0.21	0.36	0.14	
LMC_Sp	−0.05	−0.23	−0.87	−0.17	0.04	
Slp1085	−0.03	0.20	−0.78	0.30	−0.08	
DA	0.09	0.09	0.66	0.01	−0.02	
Dr_Dens	−0.50	−0.14	−0.53	0.21	−0.15	
Shape	−0.15	0.17	0.51	0.02	0.69	
FACLS	0.04	−0.02	0.04	−0.86	−0.13	
Fr_Lsw	0.02	−0.15	0.41	−0.42	−0.63	
Percent of Variance Explained	22.4	17.3	18.9	10.9	7.6	Total 77.1 %

b) 4 Principal Components

	1	2	3	4	
Fr_Barrn	−0.91	0.21	0.04	−0.10	
Fr_Forst	0.90	−0.15	−0.15	−0.22	
Dist_N	0.79	0.20	0.18	0.03	
Dist_SW	0.77	0.03	0.15	0.20	
Dr_Dens	−0.51	−0.15	−0.53	−0.19	
El_Centrd	−0.08	0.92	0.15	−0.15	
El_Gauge	−0.00	0.85	−0.05	0.28	
El_Divde	0.05	0.72	0.30	−0.42	
LMC_Sp	−0.09	−0.15	−0.86	0.03	
Slp1085	−0.05	0.19	−0.74	−0.38	
Shape	−0.18	0.20	0.65	−0.22	
DA	0.12	0.05	0.64	0.09	
FACLS	0.05	0.09	−0.05	0.83	
Fr_Lsw	0.08	−0.16	0.23	0.70	
Percent of Variance Explained	22.8	16.9	18.9	12.8	Total 71.4 %

Table 4.5

Rotated Loadings: All Basin Characteristicsexcept Eff−P El−Gauge, El−Dvide

a) 5 Principal Components

	1	2	3	4	5	
Fr_Forst	0.92	−0.12	0.22	−0.16	−0.04	
Fr_Barrn	−0.89	−0.06	0.04	0.18	0.26	
Dist_SW	0.79	0.13	−0.23	−0.08	0.20	
Dist_N	0.76	0.19	−0.05	0.39	−0.16	
Dr_Dens	−0.52	−0.41	0.31	−0.30	−0.28	
LMC_Sp	−0.05	−0.87	−0.10	−0.28	−0.06	
Slp1085	−0.05	−0.79	0.26	0.20	−0.15	
DA	0.09	0.68	0.00	0.03	0.07	
FACLS	0.04	−0.04	−0.91	−0.05	0.01	
Fr_Lsw	0.02	0.41	−0.59	−0.07	−0.49	
El_Centrd	−0.12	0.04	0.08	0.93	0.14	
Shape	−0.13	0.39	0.03	0.15	0.78	
Percent of Variance Explained	26.2	20.0	12.1	11.0	9.3	Total 78.6 %

b) 4 Principal Components

	1	2	3	4	
Fr_Barrn	−0.91	0.04	0.11	0.24	
Fr_Forst	0.91	−0.13	0.24	−0.18	
Dist_N	0.79	0.12	−0.10	0.33	
Dist_SW	0.77	0.18	−0.14	−0.05	
Dr_Dens	−0.51	−0.48	0.24	−0.33	
LMC_Sp	−0.09	−0.84	0.02	−0.21	
Slp1085	−0.05	−0.78	0.31	0.21	
DA	0.11	0.66	−0.07	−0.02	
Shape	−0.17	0.65	0.23	0.26	
FACLS	0.03	−0.05	−0.83	0.03	
Fr_Lsw	0.07	0.20	−0.77	−0.15	
El_Centrd	−0.10	0.11	0.12	0.93	
Percent of Variance Explained	26.6	20.9	13.4	11.2	Total 72.1 %

About 75 percent of the total variance is explained by the first four or five components for all cases, slightly less for four components, slightly more for five components. The components can be interpreted as follows.

PC1 Geography: Geographic variables related to Eff-P (and Eff-P itself for case 1) load highly. This component can be interpreted as representing Eff-P, or exposure to incoming weather systems. PC1 explains 22 to 28 percent of the variance, depending on whether there are four or five components, and whether or not Eff-P is included.

PC2 Topography: Elevation variables load highly, with El-Cntrd loading highest. It explains 16 to 17 percent of the variance. In Table 4.5, where the other two elevation variables are excluded, the "elevation" component becomes PC4, explaining 11 percent of the variance.

PC3 Relative Size: Variables relating to size load most highly on this component; LMC-Sp loads highest. Drainage density splits almost 50-50 between this component and the first geographic component, usually slightly favouring this "size" component. In case 3 (Table 4.5), when El-Gauge and El-Divde are excluded, Dr-Dens always loads most highly (although still just over 0.5) on PC1.

Because the elevation component is shifted to PC4 in case 3 (Table 4.5), this component essentially becomes PC2 for that case.

PC4 Lakes and Swamps: FACLS loads most highly on the fourth component (and fifth component in Table 4.5). When there are 5 components, it is almost alone on the fourth component, with Fr-Lsw loading much lower; Fr-Lsw and Shape load on PC5. When the fifth component is eliminated, Fr-Lsw loads with FACLS on the fourth component, and Shape joins the "Size" variables on PC2.

For Case 3 (Table 4.5), the loadings are similar, but the component numbers are different, and Fr-Lsw loads slightly differently. Even when there are five components, Fr-Lsw loads most highly on PC2 with FACLS, and Shape is the only variable to have its highest loading on PC5. When the fifth component is eliminated, Fr-Lsw loads more highly on PC2, and Shape loads most highly on the "Size" component, PC2.

This preliminary analysis shows that about 72 percent of the variance can be explained by four principal components, and an additional 6 percent can be explained if a fifth is included. The first four are reasonably meaningful, and each has one highly loaded variable which could possibly be used as a surrogate.

These results indicate that the principal components approach may be useful if subsequent analyses reveal a problem of multicollinearity, or if the dimensionality of the problem must be reduced to obtain significant predictive equations.

These results are not directly comparable with those of Sharp and Moore (1988), because Sharp and Moore used the dimensional data set from the provincial regional flood frequency analysis (Government of Newfoundland, 1984). Variables such as barren area and forest area were highly correlated with drainage area and their first component was thus interpretable as size.

4.2.2 Principal Components of Flow Variables

A brief additional PCA was carried out on the specific flow variables (m^3/s per km^2) together with the dimensionless flow duration variables and Lcv. When the fully dimensional flow values are used, drainage area dominates. Since this result does not contribute much to an understanding of basin response, the flow analysis was done using the variables without the effect of DA.

As with the basin variables, the analysis was carried out on the standardized untransformed data set.

The eigenvalue plot is shown in Figure 4.10. A break in slope occurs around 3 or 4 components, and so the PCA was repeated for both these numbers. The rotated loadings are shown in Table 4.6 for the two cases. The tables show that four components explain almost all the variance (96.6 percent). They can be interpreted as follows.

PC1 Availability, especially at the high flow end: FD-10 and FD-50 load most highly, and FD-90 makes a substantial contribution (0.52);

PC2 General Wetness: Both Sp-Rnoff and Sp-Flood load highly;

PC3 Annual peak flow variability: Lcv alone loads highly;

PC4 Low Flows: Both FD-90 and Sp-Low load highly.

When there are three components, Sp-Low and FD-90 join PC1, making it even more strongly interpretable as an availability component, and increasing its share of the variance explained to 49 percent. The total variance explained by the three components is just over 92 percent. Loadings on the second and third components are almost identical to the four-component case. As with the PCA of basin variables, the components show some potential for use in reducing dimensionality or multicollinearity, should problems with these arise.

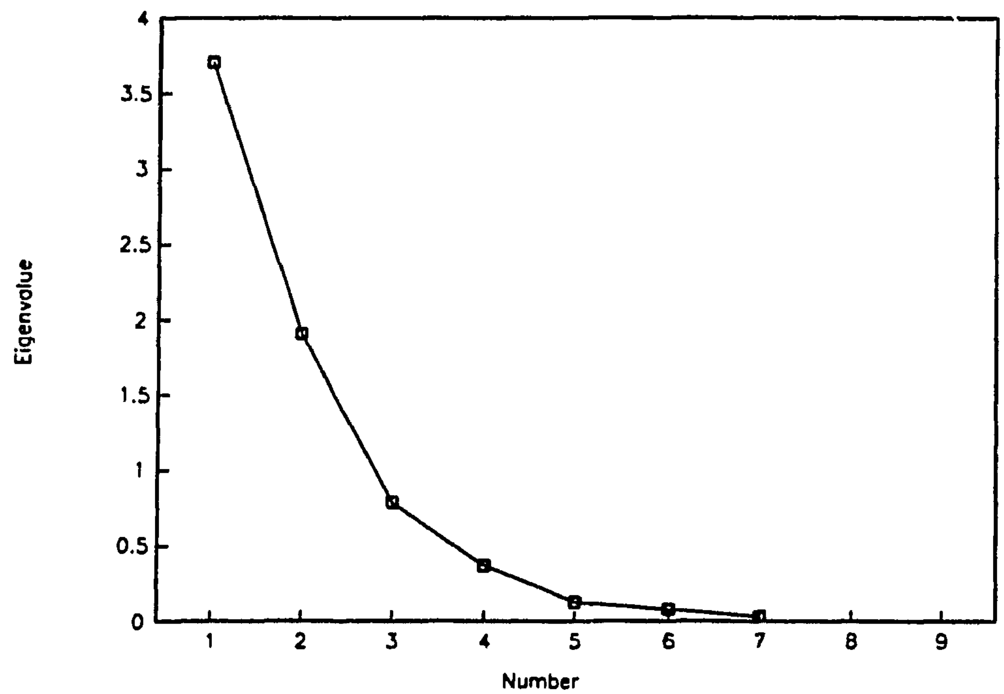


Fig. 4.10 – Eigenvalues: Flow Variables

Table 4.6

Rotated Loadings: Flow Variables

a) 4 Principal Components

	1	2	3	4	
FD-10	-0.92	-0.03	0.06	-0.35	
FD-50	0.81	-0.23	0.23	0.43	
FD-90	0.52	-0.21	0.10	0.80	
Sp-Rnoff	0.04	0.97	-0.08	0.18	
Sp_Fld	-0.23	0.85	-0.30	-0.29	
Lcv	-0.05	0.22	-0.97	-0.11	
Sp-Low	0.39	0.16	0.12	0.89	
Percent of Variance Explained	28.1	26.0	15.9	26.6	Total 96.6 %

b) 3 Principal Components

	1	2	3	
FD-90	0.92	-0.15	0.17	
FD-10	-0.91	0.02	0.15	
FD-50	0.89	-0.25	0.17	
Sp-Low	0.89	0.23	0.22	
Sp-Rnoff	0.14	0.97	-0.08	
Sp_fld	-0.37	0.83	-0.33	
Lcv	-0.11	0.22	-0.95	
Percent of Variance Explained	49.0	26.0	16.4	Total 91.4 %

The remaining sections of this chapter describe the investigations to determine which of the basin variables are most important in explaining Eff-P, high and low flows, and availability. Since Eff-P may act as an input in the equations for the other flow variables, it is examined first. The examination for each of the flow variables begins with a discussion and qualitative assessment, followed by quantitative estimates of relative importance of variables. The principal tool in these detailed investigations was multiple regression.

4.3 Effective Precipitation (Eff-P)

4.3.1 Qualitative Assessment

Regional hydrological studies, especially flood studies, have frequently identified some precipitation measure as an important explanatory variable in hydrologic response. In Newfoundland, however, precipitation is seldom measured at any point which can be considered representative of basin precipitation input, in either gauged or ungauged catchments. Consequently, investigators are left with no variable to represent hydrologic input.

Eff-P can be calculated for every gauged basin. If it can be related to topographic and geographic variables, Eff-P coulu be estimated for any ungauged basin. It could then be used directly to estimate the average flow at ungauged sites and to dimensionalize nondimensional flow duration quantiles, or indirectly to estimate other flow characteristics.

Eff-P varies from over 2000 mm per year on the southwest coast east of Port aux Basques, to about 700mm per year on the northeast coast. Eff-P is plotted at the centroids of the basins in this study in Figure 4.11, and the general trend of decreasing runoff with distance from the incoming weather systems from the southwest is evident.

Eff-P in any basin is clearly not constant over time. This study assumes that the Eff-P for the period of record (minimum 10 years) is representative of the long term MAR. The only exception is Cat Arm, as discussed in Chapter 3.

To get a sense of the extent to which MAR might vary over a typical record length of, say, 10 or 15 years, the 10 and 15 year moving averages were calculated for the basins with the four longest records. The results are presented in Table 4.7. This table shows that the 15 year Eff-P could vary by plus-or-minus about ten percent. These results, while not formal statistical descriptions, provide some background against which to assess estimates and standard errors from multiple regression equations.

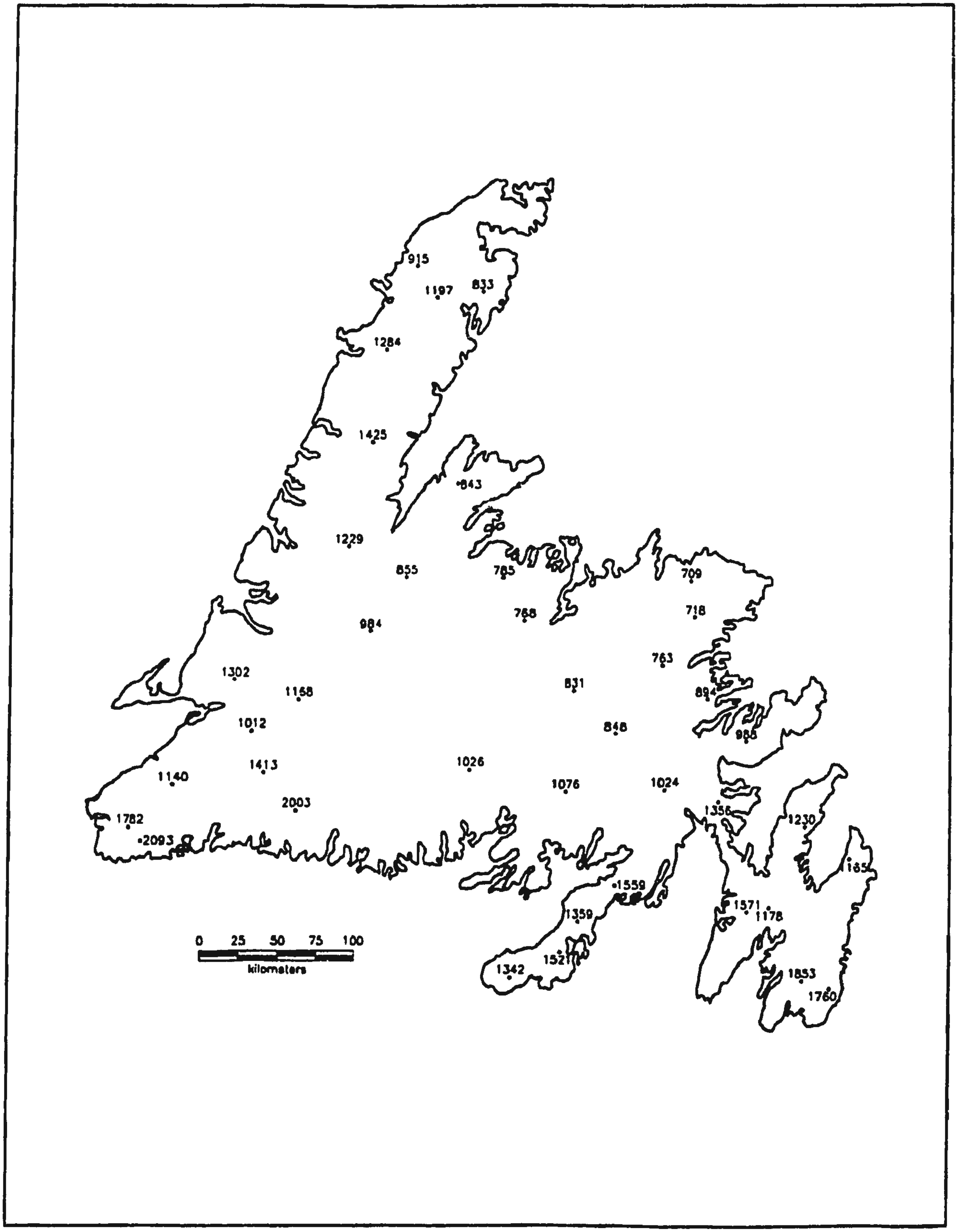


Fig. 4.11 – Effective Precipitation Plotted at Basin Centroids

Table 4.7

Variability of Moving Averages

Basin Name	Upper Humber River	Gander River at Big Chute	Rocky River near Colinet	Bay du Nord River
# of Years of Record	51	43	43	39
Eff–P for Total # of Years	1216	843	1245	1074
10 Year Moving Average	1214	853	1243	1079
Minimum 10 Year Moving Average	1100	774	1080	966
% of Eff–P for Total # of Years	90.46	91.81	86.75	89.94
Maximum 10 Year Moving Average	1367	927	1330	1163
% of Total	112	110	107	108
Eff–P for last 10 Years	1100	824	1228	1061
15 Year Moving Average	1213	860	1247	1086
Minimum 15 Year Moving Average	1132	813	1170	1016
% of Eff–P for Total # of Years	93.09	96.44	93.98	94.60
Maximum 15 Year Moving Average	1287	909	1341	1153
% of Total	106	108	108	107
Eff–P for last 15 Years	1167	839	1276	1080*

* Eff–P for the last 12 years

4.3.2 Eff-P Multiple Regression

The relationships between Eff-P and the basin variables were quantified using multiple regression (MR) techniques. Because MR is an important tool in this research, a brief description and discussion is presented here. The same approach is used throughout this study in all applications of MR.

The steps were as follows.

1. Develop candidate equations using ordinary least squares (OLS) techniques. If nonlinear relationships are expected, use logarithmically transformed variables. Accept variables only if significant at 5 percent level, i.e., p value less than 0.05, and if they do not introduce problems of multicollinearity, i.e., variance inflation factor less than 10, equivalent to tolerance greater than 0.1.
2. Assess results, select final candidate models.
3. Identify outliers using robust regression techniques.
4. Check residual plots.
5. If log-linear equations have been selected, develop the final forms of the equations using nonlinear techniques.

The OLS techniques are very good for preliminary model development and screening because they are well behaved, and provide the best linear unbiased estimators if all assumption are met. Useful results such as standardized coefficients and leverage are relatively easily obtained from most statistical software packages. Alternative models can be easily developed and compared.

There are two potential problems with OLS analysis, however. One is that the analysis may be spoiled by the presence of outliers. Outliers are often hidden in multiple regression, because they do not show up in residual plots or calculations based on residuals, such as studentized or standardized residuals. Alternative statistical techniques have been developed to provide equations which are robust or resistant to a certain amount of contamination in the data. When an equation is fitted using a robust technique, outliers will lie far from the fitted line and can be detected by their large residuals.

The least median of squares (LMS) technique has been found to be very powerful for the detection of outliers, and the robust regression program PROGRESS using LMS was applied in this study for outlier detection (Rousseeuw, 1987). Each candidate equation was tested for outliers; if there was more than one outlier, equations using other independent variables were checked to see whether they might give similarly good results without as many outliers. Equations without outliers were preferred because with the data set used in this study, outliers are likely to be true values, not simply typographical or measurement errors. In practice there may well be ungauged basins similar to an apparent outlier.

The second potential problem with OLS lies not with the technique itself but with the fact that it is often applied to inherently nonlinear models. In order to use OLS techniques in such cases, the variables are logarithmically transformed, as they are in most cases in this study. A fundamental assumption of OLS, however, is that the error terms are additive. This assumption remains true only as long as the analysis remains in the transformed domain. When the estimates are transformed back, the error term becomes multiplicative. Since additive error is assumed, the standard error will be smaller and the coefficients will be biased. Various corrections for bias have been proposed, but these correct only for the bias in the intercept (the constant), not in the slope (the coefficients). This problem is discussed by McCuen et al (1990).

So throughout this study, once candidate models were developed using OLS and robust regression techniques, the final form was defined using nonlinear least squares (NLLS), as provided in SYSTAT. NLLS is based on minimizing the squared deviations of the dependent variable data values from values estimated by the function at the same independent variable data points. Both Quasi-Newton and Simplex estimation methods are available in SYSTAT to seek the minimum of the loss function. Both methods were tried on a number of the candidate models, and they always gave the same results. The coefficients from the NLLS analysis were usually similar in magnitude to those developed using the log-linear methods. It should be noted that although NLLS does not require the additive error assumption, it still assumes constant variable of residual (homoscedasticity).

4.3.3 Eff-P Multiple Regression Analysis

It is not clear from the scatter plots and correlation matrices presented earlier in this chapter whether the relationships between Eff-P and the distance and topographic variables are inherently linear or nonlinear. Both log-transformed and untransformed data sets were therefore used in a multiple regression analysis to determine whether a suitable predictive equation for Eff-P could be obtained. The results are presented in Table 4.8. The linear equation is very straightforward and physically meaningful - the Eff-P is about 1100 mm, reduced by about twice the SW distance to the sea (in km), increased by about 6 times the percent of barren in the basin, and increased further by about 60 percent of the elevation of the centroid (in m). At first glance the importance of Fr-Barrn is somewhat counterintuitive; why should a larger amount of barren area lead to higher average effective pre[illegible]n? One plausible explanation is that evapotranspiration losses are lower in barren basins, although the magnitude of

Table 4.8

Eff–P: Multiple Regression Results

	Linear			Nonlinear			
	OLS	RLS	RLS	NLLS	RNLLS	NLLS	RNLLS
N	40	33	7	40	38	40	37
Constant	1133	1054	1435	1.949	3.000	1.072	1.735
Dist_SW	−2.160	−1.620	−2.030	−0.159	−0.242	−0.194	−0.265
Fr−Barrn (as%	6.070	3.570	3.520	0.081	0.058	0.060	0.038
El−Cntrd	0.644	0.658	1.210			0.153	0.132
r^2	0.74	0.80	0.99	0.97	0.98	0.98	0.98
r^2 adj (OLS)	0.71	0.78	0.98				
r^2 corr (NLLS)				0.67	0.75	0.72	0.80
Outliers	IaM Grdy LtCd SlCv NePl Rtl			SteG SmLm		SteG SmLm Rtl	

Notes:
Log linear constant have been transformed back to natural domain.
Linear results are for Eff−P in mm, nonlinear for Eff−P in m.

OLS Ordinary Least Squares
RLS Reweighted Least Squares
NLLS Nonlinear Least Squares
RNLLS Reweighted Nonlinear Least Squares

the coefficient is then a bit surprising. A more likely reason is that the extent of barren area reflects the degree of exposure to incoming weather systems in Newfoundland. An example may explain this reasoning further.

The Highlands River basin on the west coast, has a relatively short SW distance to the sea (65 km) and a high elevation of its centroid (244 m). Based on those two variables, it would be expected to have a higher Eff-P than Garnish River basin on the Burin Peninsula, with a similar distance to the sea and a lower elevation (152 m). But in fact the Eff-P for the Highlands basin is 1140 mm compared with 1359 mm for Garnish. This difference may be accounted for by the fact that Garnish is in a more exposed location. This exposure is reflected in its greater proportion of barren area (63 percent, compared with 13 percent for Highlands).

The problem with the equation is the large number of outliers (seven), forming a cluster of all basins with Eff-P greater than 1550 mm. A separate MR analysis with and without those basins shows that the same variables are important, but the constant and the coefficients of the equations are different. These results are also included in Table 4.8. The difficulty arises when the equations are used for prediction; how can a potential outlier be assigned to the correct group, so that the correct equation can be used?

As Figures 4.12 and 4.13 show, there are no clearly distinguishing topographic or physiographic characteristics to identify these high runoff basins. Five of these basins are located in very exposed regions, Isle aux Morts, Grandy and Little Codroy on the southwest tip of the island, and Northwest River and Seal Cove on the southeastern tip of the Avalon Peninsula. An ungauged river in these areas might be classed with the high runoff group by judgment. Northeast River at Placentia and Rattle Brook on the Burin Peninsula are slightly less wet, and are somewhat anomalous. They may be affected by the local topography of Placentia Bay in the case of Northeast River, and both Fortune and Placentia Bays in the case of Rattle Brook. Figure 4.14 shows the observed and estimated Eff-P for the basins using the linear equations after reweighting.

The nonlinear equations with either two or three variables are also quite good, and have the advantage of having fewer outliers. Figure 4.15 shows the observed and estimated Eff-P from the nonlinear equations. Both Ste. Genevieve and Salmonier River near Lamaline have lower Eff-P than predicted. In the case of Ste. Genevieve, this result is not unexpected, because the basin is located in a low flat area near the tip of the Great Northern Peninsula, and it is quite conceivable that there is not as much available precipitation in the weather systems north of

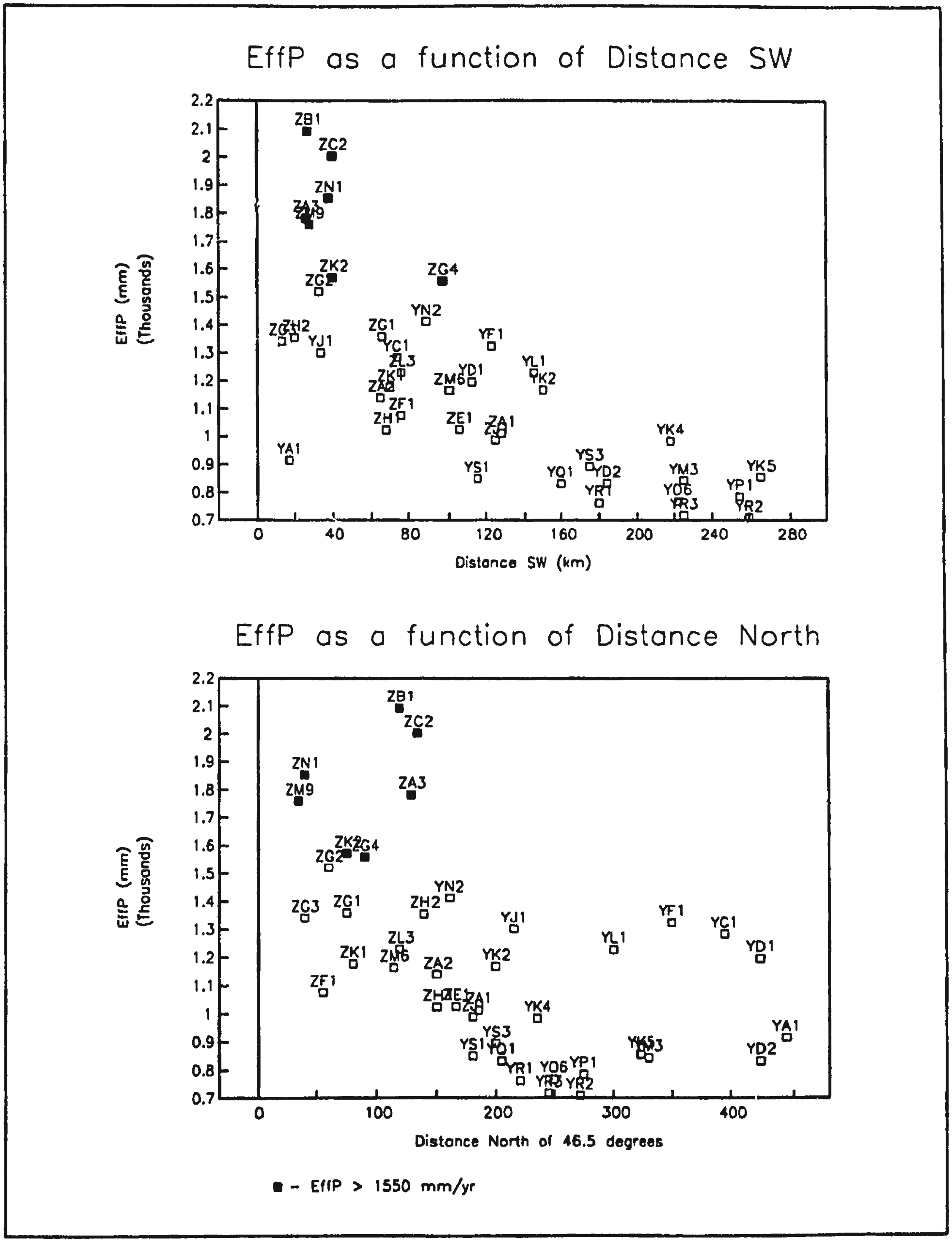


Fig. 4.12 – Effective Precipitation and Distance Variables

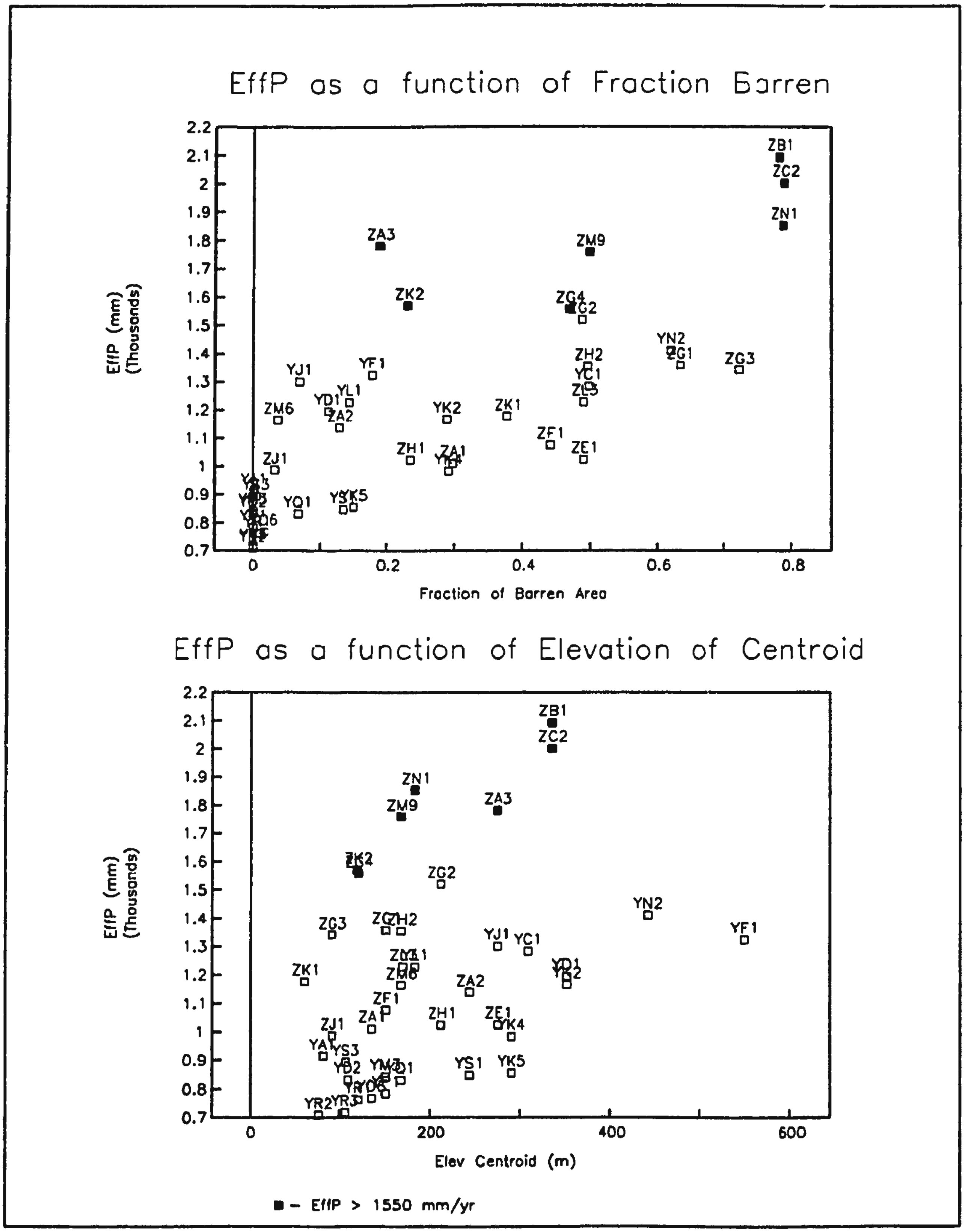


Fig 4.13 – Effective Precipitation and Fraction of Barren and Elevation of Centroid

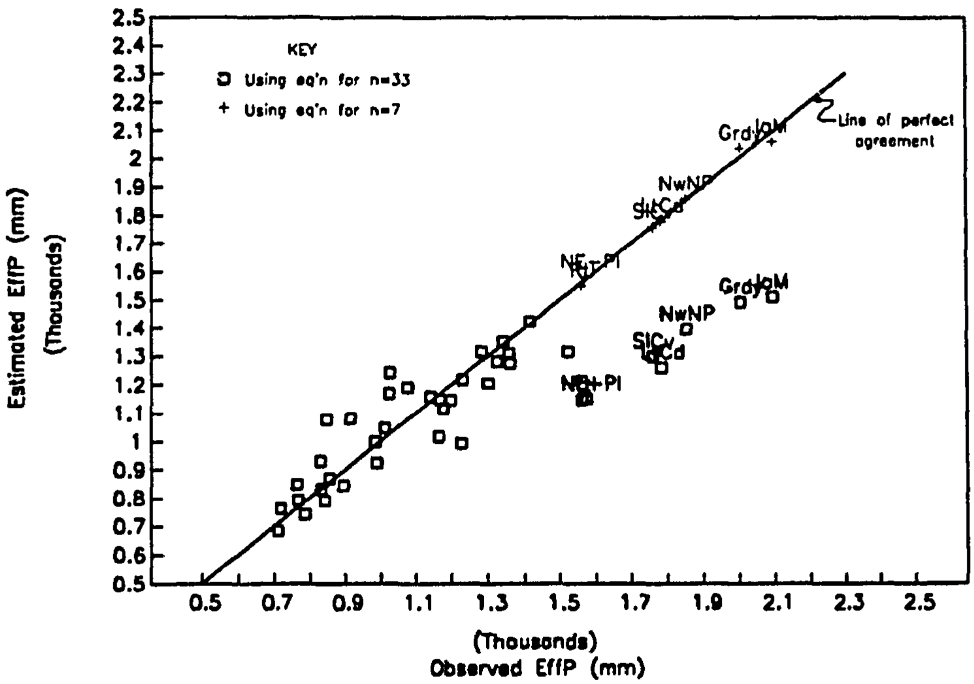
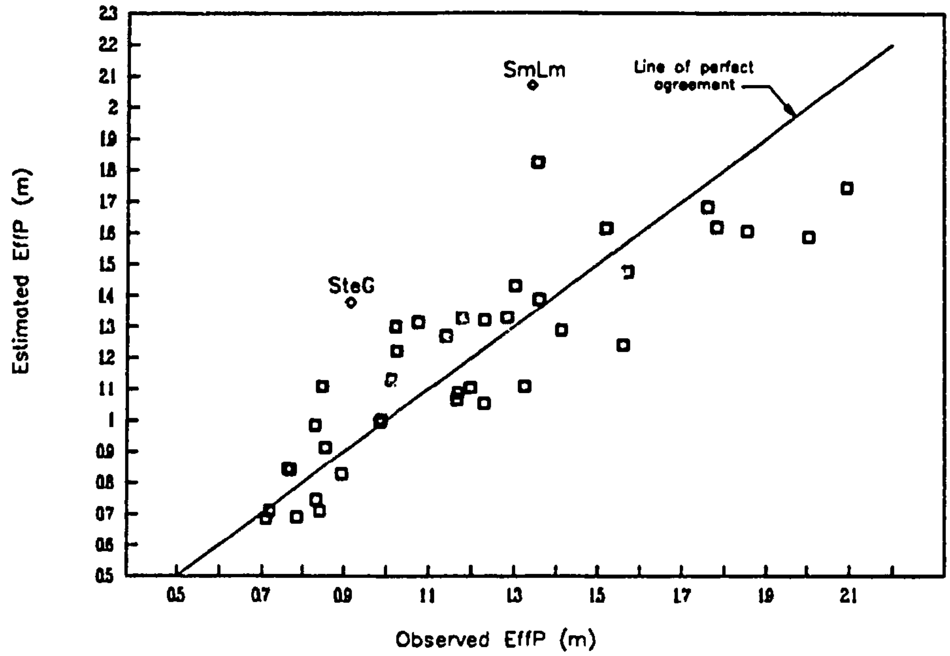
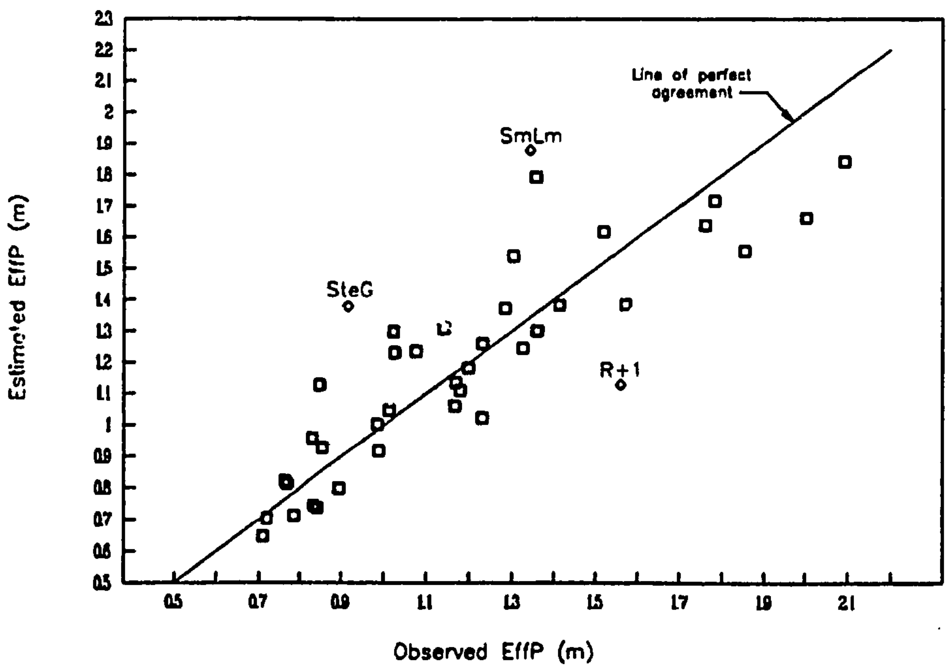


Fig. 4.14 – Observed and Estimated EffP



(a) From 2 Variables



(b) From 3 Variables

Fig. 4.15 – Estimated and Observed EffP: Nonlinear Equation

the Long Range Mountains. Salmonier River is a puzzle; it would be expected to be very wet given its exposed location on the tip of the Burin Peninsula, and the climate station on the coast nearby at St. Lawrence records high rainfall.

In the three variable equation, Rattle Brook (discussed above) is identified as an additional outlier, with higher runoff than predicted.

Additional preliminary attempts were made to develop relationships within geographic regions. These included dividing the island into regions as follows

- along the hydrologic boundary between WSC Y and Z regions;
- according to whether a basin is to windward or to leeward of approaching storms;
- north-south, along the 48th parallel of latitude.

None of these attempts met with any particular success, although the results might be useful in conjunction with clusters based on other flow or basin characteristics. The windward/leeward division is not recommended because it is sometimes difficult to make a clear assignment of an ungauged basin.

In summary, the MR analysis of Eff-P shows that Dist-SW, Fr-Barrn and El-Cntrd are consistently the most important explanatory variables. If an estimate is required of Eff-P at an ungauged basin, all three equations may be tried. In addition, it is important to plot the location of the centroid of the ungauged basin on a map such as Figure 4.11, and to interpolate between known points of Eff-P. The interpolation should take into account the relative Dist-SW, Fr-Barrn, and El-Cntrd of the gauged and ungauged basins. Additional information from climate stations and topographic maps can contribute to the estimate.

4.4 High Flows: Lcv, FD-10, Sp-Fld and Qavgfld

The next set of analyses focused on flood flows, within the range of flows reported in the record, not estimates of flood quantiles. The flow exceeded only 10 percent of the time, FD-10, was also included in this group, although it does not represent a very high flow. For the rivers in the data set, it ranges from 1.9 to 2.7 times the average annual flow. The average annual maximum daily flow, by comparison, might be 5 to 10 times the average annual flow.

The scatter plots in Figure 4.1 to 4.3 suggest that the relationships with the basin variables are nonlinear, so the log transformed data set was used for the initial MR analysis.

4.4.1 Lcv

The analysis began with Lcv, the linear coefficient of variation obtained from L-moments. It represents the variability of annual floods from year to year. Lcv is of particular interest because it has a large influence on the shape of the growth curves used to estimate extreme flood quantiles.

For basins exposed to the same flood-producing hydrologic input, a difference in Lcv would be expected to occur due to differences in basin characteristics. There might also be some geographic or topographic variation, representing the difference in the types of events producing floods, e.g., less intense storms, or snowmelt only compared with mixed rain-on-snow or rain only.

Qualitative Assessment

A brief PCA analysis was undertaken to see whether Lcv might load on a common factor with some basin characteristics. (The PCA analysis described above had separated flow and basin characteristics.) The result, not surprisingly given the low correlation of Lcv with any basin variables, was that Lcv tends to load highly on only one factor, and that no basin characteristics loaded highly on that factor.

The data set was then sorted in order of ascending Lcv to assess qualitatively the difference between basins with high and low Lcv's. The sorted file is presented in Table 4.9. This table shows that the range of Lcv is from 0.11 to 0.314. Some geographic division is immediately suggested, because eight of the ten least variable basins are in WSC hydrologic region Y, and of the 10 most variable basins the reverse is true - 8 are in region Z. Of the ten least variable basins, the two exceptions not in region Y are Seal Cove and Northwest River. They are on the southeastern tip of the Avalon Peninsula, in a very wet and relatively warm area; in addition the areas above their gauges are 100 percent controlled by lakes and swamps.

The remaining ten basins with the lowest Lcv's are generally large (the smallest drainage area of the Y's is 391 km^2), and include the three of the four largest basins in the data set. (The third largest basin ranks #11 in Lcv.) They also tend to have a large fraction of their areas controlled by lakes and swamps - the lowest is 0.75 (Upper Humber), and the second lowest is 0.91. A review of the records of the dates of occurrences of the annual peak daily flows shows that about 90 percent are spring snowmelt events, as expected from their locations and high elevations. Of the ten most variable basins, the two basins not in hydrologic region Z are YM3, Southwest River (Baie Verte), and YO6, Peters River. Southwest River has the second smallest fraction of its area controlled by lakes

Table 4.9

Data Set Sorted by Lcv

YZS	BASINS	LCV	DA	LMC SP	EL GAUGE	EL DIVDE	EL CENTR	SLP1085	DR DENS	SHAPE	FR LSW	FACLS
ZM9	SlCv	0.118	53.6	0.28	35	168	168	0.62	1.130	1.37	0.130	1.000
YL1	UpHm	0.126	2110	0.06	23	701	183	0.40	0.786	1.56	0.115	0.750
YK2	LewB	0.140	470	0.12	137	655	351	0.59	0.627	2.32	0.161	1.000
YK4	Hnds	0.142	529	0.09	198	518	290	0.32	0.637	1.78	0.355	0.950
YF1	Cat	0.144	611	0.05	351	594	549	0.73	0.582	1.86	0.131	1.000
YR3	IndB	0.155	554	0.09	8	137	107	0.22	0.680	1.72	0.310	0.970
YQ1	GdBC	0.156	4444	0.03	23	305	168	0.14	0.452	2.08	0.172	0.910
ZN1	NwNP	0.158	53.3	0.27	114	213	183	0.61	1.089	2.06	0.126	1.000
YS1	TerN	0.159	1365	0.08	84	290	244	0.12	0.726	2.35	0.327	0.920
YK5	Shfd	0.161	391	0.10	110	488	290	1.07	0.191	1.98	0.172	0.940
ZE1	SmLP	0.162	2640	0.04	183	305	274	0.08	0.360	1.75	0.160	1.000
YR1	MdlB	0.165	275	0.18	23	198	122	0.32	0.255	1.93	0.245	0.980
YD2	NeRk	0.167	200	0.19	23	290	110	0.47	0.930	1.65	0.170	0.990
ZG3	SmLm	0.170	115	0.21	2	137	91	0.34	1.550	1.62	0.120	0.920
YS3	SwTN	0.172	36.7	0.30	23	168	107	1.11	0.641	1.43	0.159	1.000
YD1	BvrB	0.173	237	0.17	8	335	351	0.67	0.339	2.23	0.082	0.730
YR2	RedH	0.174	399	0.11	27	122	76	0.21	0.740	1.68	0.320	0.960
YJ1	Hrys	0.177	640	0.09	15	541	274	0.35	1.120	1.81	0.142	0.750
YA1	SteG	0.179	306	0.13	12	107	81	0.23	0.540	1.48	0.354	1.000
ZG4	Rtl	0.189	42.7	0.23	46	152	122	1.10	1.620	1.53	0.180	0.920
ZM6	NBP	0.190	3.63	0.72	122	191	168	2.42	1.038	1.24	0.225	1.000
ZL3	SpCv	0.190	10.8	0.65	76	168	170	1.25	1.090	1.36	0.090	1.000
YC1	Tor	0.195	624	0.08	8	488	309	1.01	0.755	1.45	0.167	0.990
YN2	Lds	0.200	469	0.12	328	488	442	0.22	1.370	2.15	0.160	1.000
ZJ1	SthB	0.205	67.4	0.24	8	137	91	0.50	1.240	1.43	0.150	0.860
ZG1	Grn	0.211	205	0.22	11	381	152	0.60	0.547	2.45	0.101	0.960
ZK1	Rky	0.219	300	0.15	15	168	60	0.50	1.005	2.00	0.113	0.550
ZA1	LtlB	0.220	343	0.19	8	472	137	0.68	1.040	2.45	0.100	0.830
ZF1	BdN	0.223	1170	0.06	8	274	152	0.29	0.612	2.15	0.236	0.960
YP1	ShlA	0.226	63.8	0.24	84	198	152	0.53	0.880	1.62	0.130	0.790
ZH1	PH	0.227	764	0.07	38	244	213	0.35	0.709	1.67	0.639	0.910
ZK2	NE-P	0.231	89.6	0.30	15	213	120	0.55	1.110	1.91	0.300	0.810
ZH2	CbC	0.235	43.3	0.39	53	152	168	0.59	1.110	1.66	0.099	0.920
ZG2	Tds	0.241	166	0.16	8	229	213	1.35	1.350	1.84	0.130	0.920
YM3	SwBV	0.254	93.2	0.20	69	168	152	0.27	0.680	1.67	0.100	0.560
ZB1	IaM	0.262	205	0.16	8	457	335	0.84	0.720	2.09	0.134	0.600
ZC2	Grdy	0.263	230	0.13	84	442	335	1.06	0.960	1.84	0.040	0.340
ZA2	Hlds	0.268	72	0.28	69	533	244	2.19	1.150	1.72	0.050	0.430
YO6	Ptrs	0.269	177	0.24	23	213	137	0.45	0.800	1.93	0.160	0.970
ZA3	LtCd	0.314	139	0.18	8	457	274	1.46	1.460	1.67	0.130	0.730

Table 4.9 Continued

YZS	BASINS	DIST_SW	FR_BARRN	FR_FORST	DIST_N	MAR_MM
ZM9	SlCv	28	0.500	0.370	35	1760
YL1	UpHm	145	0.145	0.740	300	1229
YK2	LewB	150	0.290	0.549	200	1168
YK4	Hnds	218	0.292	0.353	235	984
YF1	Cat	123	0.180	0.689	350	1325
YR3	IndB	225	0.001	0.689	245	718
YQ1	GdBC	160	0.068	0.760	205	831
ZN1	NwNP	38	0.788	0.086	40	1853
YS1	TerN	115	0.136	0.537	180	848
YK5	Shfd	266	0.152	0.676	323	855
ZE1	SmLP	105	0.490	0.350	165	1026
YR1	MdlB	180	0.008	0.747	220	763
YD2	NcRk	184	0.001	0.829	425	833
ZG3	SmLm	13	0.722	0.158	40	1342
YS3	SwTN	175	0.005	0.836	200	894
YD1	BvrB	112	0.112	0.806	425	1197
YR2	RpdH	260	0.001	0.679	272	709
YJ1	Hrys	34	0.069	0.789	215	1302
YA1	SteG	17	0.002	0.644	445	915
ZG4	Rtl	97	0.470	0.350	90	1559
ZM6	NEP	100	0.038	0.737	115	1165
ZL3	SpCv	76	0.490	0.420	120	1230
YC1	Tor	74	0.498	0.335	395	1284
YN2	Lds	89	0.620	0.220	160	1413
ZJ1	SthB	125	0.033	0.317	180	988
ZG1	Grn	66	0.634	0.265	75	1359
ZK1	Rky	70	0.377	0.510	80	1178
ZA1	LtlB	128	0.299	0.601	185	1012
ZF1	BdN	76	0.442	0.322	55	1076
YP1	SblA	255	0.001	0.869	275	785
ZH1	PH	68	0.234	0.107	150	1024
ZK2	NB-P	40	0.230	0.470	75	1571
ZH2	CbC	20	0.496	0.405	140	1356
ZG2	Tds	33	0.488	0.382	60	1521
YM3	SwBV	225	0.001	0.899	330	843
ZB1	IaM	27	0.782	0.084	120	2093
ZC2	Grdy	40	0.790	0.170	135	2003
ZA2	Hlds	65	0.130	0.820	150	1140
YO6	Ptrs	222	0.018	0.822	250	768
ZA3	LtCd	26	0.190	0.680	130	1782

and swamps (0.56) of any basin in the data set, and this, together with its relatively small size and low elevation, may lead to the variability.

Peters River is somewhat larger (177 km^2) but is still relatively small and relatively low, compared to the other basins in region Y. It does have a large fraction of area controlled by lakes and swamps, and its apparent variability may be at least partly due to the fact that during its relatively short period of record it experienced an unusually large flood.

Multiple Regression Analysis

With these preliminary observations, a multiple regression analysis was then carried out

- to identify the most important explanatory variables, assuming one region;
- to identify outliers;
- to assess whether simple geographic subdivisions based on Y-Z regions would identify additional important independent variables.

The results were not especially good, although the division into Y and Z regions led to improved equations as expected. Table 4.10 shows the regression

Table 4.10

Lcv Multiple Regression Results

Regression Coefficients

	All Basins			Hydrologic Region Y					
	OLS Untrans-formed	OLS RLS Lns	OLS RLS Lns	OLS Untrans-formed 1 var	OLS Untrans-formed 4 vars	OLS Lns 1 var	RLS Lns 1 var	OLS Lns 2 vars	RLS Lns 2 vars
N	40	40	40	20	20	20	18	20	19
Constant	0.328	−1.434	no outliers	0.136	0.278	−1.324	−1.523	−1.376	−1.673
DA/1000	−0.015								
Dist−N									
Dist−SW									
El−Divde					−0.00011	Note different sign from Y to Z			
El−Gauge		−0.037							
El−Centrd									
FACLS	−0.143	−0.459			0.107			−0.303	−0.063
Fr−Barrn					−0.120				
LMC−sp		−0.091	sign?	0.306	0.234	0.198	0.123	0.189	0.062
Slope									
r^2	0.39	0.40		0.36	0.59	0.35	0.25	0.41	0.41
r^2 adj	0.36	0.35		0.32	0.49				
Std.Er.	0.04	0.18 lns		0.03	0.03	0.16 lns	0.13	0.16 lns	0.16 lns
Outliers		None				SwBV Peters	Note: r^2 lower wo outliers	Peters	

OLS− Ordinary Least Squares
RLS − Reweighted Least Squares

Table 4.10 Continued

	Hydrologic Region Z						
	OLS Untrans-formed 3 vars	OLS Lns 4 vars	RLS Lns 4 vars	OLS Lns 5 vars	RLS Lns 5 vars	OLS Lns 5 vars	RLS Lns 5 vars
N	20	20	17	20	19	20	17
Constant	0.259	−2.663	−2.450	−2.569	−2.224	−2.082	−1.474
DA		0.095	0.111	0.078	0.073	0.127	0.094
Dist−N		0.194	0.138	0.185	0.110	0.270	0.233
Dist−SW						−0.144	−0.209
El−Divde	0.00014						
El−Gauge	−0.00026	−0.047	−0.048	−0.050	−0.039		
El−Centrd						−0.126	−0.130
FACLS	−0.084			−0.108	−0.110		
Fr−Barrn							
LMC−sp							
Slope		0.235	0.261	0.199	0.192	−0.274	0.226
r^2	0.60	0.73	0.89	0.74	0.76	0.79	0.94
r^2 adj	0.52					0.72	
Std.Er.	0.03	0.13	0.07	0.13	0.11	0.12	0.06
			lns	lns	lns	lns	lns
Outliers		Ltl Bar C by Chn Seal Cove		Seal Cove		BdN SmLm SlCv	

OLS− Ordinary Least Squares
RLS − Reweighted Least Squares

coefficients, coefficients of determination and standard errors for the best results. Because distance and topographic variables had the potential to be important, which may be linearly related, both the untransformed and logarithmically transformed data sets were used.

For the island as a whole, no equation could be obtained that explained even half of the variance. This result suggests that Lcv could be better explained within groupings according to flood cause. The equations were slightly better with the untransformed data set, only two variables being required to give approximately the same r^2 as three variables for the log-transformed set. The most important variables were DA and FACLS for the untransformed data set, and ElG-Ln, FACLS-Ln and LMC-Ln for the log-transformed set. LMC-Ln, the main channel length per unit area, is highly negatively correlated to DA-Ln (0.95) as Table 4.2 showed, so the additional explanatory variable in the log case relates to the elevation of the gauge.

The untransformed data set is slightly better for the case of the Y region as well. The single most important variable is LMC-sp, but only about a third of the variance is explained. Adding three more variables (FACLS, Fr-Barm, and El-Divide) still only explains about half the variance. Further clustering within the Y region, probably separating regions that are known to be dominated by snowmelt floods from those which experience mixed events (or a different clustering other than Y/Z) might help.

Much better results were obtained for the Z region, possibly because the region tends to be subjected to the same type of storm input, so the variability can then be explained primarily on the basis of basin characteristics.

The log-transformed data set led to more significant equations, but the direction of the some of the coefficients are counterintuitive. The most important of these is drainage area. The best equations indicate that variability increases with drainage area, whereas it would usually be expected to decrease with drainage area. The correlation between DA-Ln and Lcv-Ln is low; the partial correlation when there are no variables in the equation is only 0.17. (The scatter plot in Figure 4.3 shows the poor relationship). The partial correlation remains low when Dst-N-Ln, ElC-Ln and FACLS-Ln are the independent variables. As soon as Slp-Ln is added, the partial correlation coefficient rises to 0.49, and when DA-Ln is then included, the overall r^2 rises from 0.66 without DA-Ln to 0.74 with DA-Ln. It is not clear how these variables are interacting hydrologically. Robust regression does not identify the largest drainage basin as an outlier.

Similarly the best regression equations for the Z region shows that variability increases with distance north, whereas we would expect a decrease due to the fact that snowmelt events (which produce less year to year variability) would be more likely to predominate in more northerly basins. The coefficients of Slope, El-Gauge and FACLS (or Dist-Sw where Dist-Sw replaces FACLS) are in the expected directions, however.

It is unlikely that predictive equations would be used for Lcv; the benefit of the MR analysis is to assist in using growth curves for ungauged basins, and in developing regional boundaries for flood frequency analysis.

4.4.2 FD-10

Most high flow analyses focus on extreme events, but for design of hydroelectric projects and water supply systems it can be useful to know what basin characteristics lead to generally high but not extreme flows, e.g., of the order of two to three times the average flow. For this study FD-10 was selected, the flow which is exceeded 10 percent of the time. Like Lcv, it was nondimensionalized, in this case by dividing by the mean annual flow. (Dimensional values lead to clusters of basins based on size and wetness, and it is more difficult to explore the relationships to basin characteristics.)

Qualitative Assessment

The scatter plots in Figure 4.6 and the correlation matrix showed that FD-10 is not highly correlated with any basin characteristics. The highest correlation coefficient was 0.55, between FD10-Ln and Dst-N-Ln. There is no correlation between FD-10 and Lcv or other flood variables; in fact the highest correlations were with other flow duration variables, and Sp-Low (median low flow per unit area). (The correlations must be treated cautiously because of the possible effects of common variables for nondimensionalizing.)

An examination of the data set sorted by FD-10 (Table 4.11) shows that of the basins which have the lowest flows at the 10th percentile (the least flashy, in this definition), 9 are in WSC hydrologic region Z. Of the basins with relatively high FD-10's, only 2 are in region Z. This observation suggests the possibility of some geographic clustering, perhaps based on the mechanism for generating high flows, similar to that for Lcv. If the annual cycle of a basin tended to be dominated by snowmelt runoff, it might have many days with relatively high flows, leading to a high FD-10. The peaks within those long events might be fairly modest, however, and there might be less year-to-year variability. Since the Z region is somewhat warmer with more mixed events than the Y region, this division is quite plausible, especially since the two Z basins which cluster with the Y's have the highest elevations, and therefore could be expected to have more precipitation

Table 4.11

Data Set Sorted By FD−10

YZS	BASINS	FD10	DA	LMC SP	EL GAUGE	EL DIVDE	EL CENTR	SLP1085	DR DENS	SHAPE	FR LSW	FACLS
ZF1	BdN	1.897	1170	0.06	8	274	152	0.29	0.612	2.15	0.236	0.980
ZG2	Tds	1.925	166	0.16	8	229	213	1.35	1.350	1.84	0.130	0.920
ZL3	SpCv	1.971	10.8	0.65	76	168	170	1.25	1.090	1.36	0.090	1.000
ZG1	Grn	2.005	205	0.22	11	381	152	0.60	0.547	2.45	0.101	0.960
ZK2	NB−P	2.027	89.6	0.30	15	213	120	0.55	1.110	1.91	0.300	0.810
ZB1	SmLP	2.028	2640	0.04	183	305	274.32	0.08	0.360	1.75	0.160	1.000
ZN1	NwNP	2.035	53.3	0.27	114	213	183	0.61	1.089	2.06	0.126	1.000
YA1	SteG	2.041	306	0.13	12	107	81	0.23	0.540	1.48	0.354	1.000
ZK1	Rky	2.045	300	0.15	15	168	60	0.50	1.005	2.00	0.113	0.550
ZM9	SlCv	2.140	53.6	0.28	35	168	168	0.62	1.130	1.37	0.130	1.000
YR1	MdlB	2.150	275	0.18	23	198	122	0.32	0.255	1.93	0.245	0.980
YK4	Hnds	2.152	529	0.09	196	518	290	0.32	0.637	1.78	0.355	0.950
ZH1	PH	2.153	764	0.07	38	244	213	0.35	0.709	1.67	0.659	0.910
YJ1	Hrys	2.155	640	0.09	15	541	274	0.35	1.120	1.81	0.142	0.750
YQ1	GdBC	2.171	4444	0.03	23	305	168	0.14	0.452	2.08	0.172	0.910
ZG4	Rtl	2.175	42.7	0.23	46	152	122	1.10	1.620	1.53	0.180	0.920
YK2	LewB	2.190	470	0.12	137	655	351	0.59	0.627	2.32	0.161	1.000
ZH2	CbC	2.226	43.3	0.39	53	152	168	0.59	1.110	1.66	0.099	0.920
YR3	IndB	2.238	554	0.09	8	137	106.68	0.22	0.680	1.72	0.310	0.970
YC1	Tor	2.240	624	0.08	8	488	309	1.01	0.755	1.45	0.167	0.990
YK5	Shld	2.245	391	0.10	110	488	290	1.07	0.191	1.98	0.172	0.940
ZA1	LtlB	2.273	343	0.19	8	472	137	0.68	1.040	2.45	0.100	0.830
ZM6	NBP	2.276	3.63	0.72	122	191	168	2.42	1.038	1.24	0.225	1.000
ZA3	LtCd	2.293	139	0.18	8	457	274	1.46	1.460	1.67	0.130	0.730
YS1	TerN	2.308	1365	0.08	84	290	244	0.12	0.726	2.35	0.327	0.920
ZJ1	SthB	2.346	67.4	0.24	8	137	91	0.50	1.240	1.43	0.150	0.860
ZA2	Hlds	2.358	72	0.28	69	533	244	2.19	1.150	1.72	0.050	0.430
YP1	ShlA	2.363	63.8	0.24	84	198	152	0.53	0.880	1.62	0.130	0.790
ZG3	SmLm	2.372	115	0.21	2	137	91	0.34	1.550	1.62	0.120	0.920
YN2	Lds	2.386	469	0.12	328	488	442	0.22	1.370	2.15	0.160	1.000
YR2	RgdH	2.386	399	0.11	27	122	76	0.21	0.740	1.68	0.320	0.980
YS3	SwTN	2.404	36.7	0.30	23	168	107	1.11	0.641	1.43	0.159	1.000
ZC2	Grdy	2.411	230	0.13	84	442	335	1.06	0.960	1.84	0.040	0.340
ZB1	IaM	2.441	205	0.16	8	457	335	0.84	0.720	2.09	0.134	0.600
YM3	SwBV	2.478	93.2	0.20	69	168	152	0.27	0.680	1.67	0.100	0.560
YL1	UpHm	2.482	2110	0.06	23	701	183	0.40	0.786	1.56	0.115	0.750
YD1	BvrB	2.558	237	0.17	8	335	351	0.67	0.339	2.23	0.082	0.730
YO6	Ptm	2.575	177	0.24	23	213	137	0.45	0.800	1.93	0.160	0.970
YD2	NeRk	2.652	200	0.19	23	290	110	0.47	0.930	1.65	0.170	0.990
YF1	Cat	2.736	611	0.05	351	594	549	0.73	0.582	1.86	0.131	1.000

Table 4.11 Continued

YZS	BASINS	DIST_SW	FR_BARRN	FR_FORST	DIST_N	EFF-P
ZF1	BdN	76	0.442	0.322	55	1076
ZG2	Tds	33	0.488	0.382	60	1521
ZL3	SpCv	76	0.490	0.420	120	1230
ZG1	Grn	66	0.634	0.265	75	1359
ZK2	NB-P	40	0.230	0.470	75	1571
ZB1	SmLP	105	0.490	0.350	165	1026
ZN1	NwNP	38	0.788	[illegible]	40	1853
YA1	SteG	17	0.002	0.644	445	915
ZK1	Rky	70	0.377	0.510	80	1178
ZM9	SlCv	28	0.500	0.370	35	1760
YR1	MdlB	180	0.008	0.747	220	763
YK4	Hods	218	0.292	0.353	235	984
ZH1	PH	68	0.234	0.107	150	1024
YJ1	Hrys	34	0.069	0.789	215	1302
YQ1	GdBC	160	0.068	0.760	205	831
ZG4	Rtl	97	0.470	0.350	90	1559
YK2	LewB	150	0.290	0.549	200	1168
ZH2	CbC	20	0.496	0.405	140	1356
YR3	IndB	225	0.001	0.689	245	718
YC1	Tor	74	0.498	0.335	395	1284
YK5	Shld	266	0.152	0.676	323	855
ZA1	LtlB	128	0.299	0.601	185	1012
ZM6	NEP	100	0.038	0.737	115	1165
ZA3	LtCd	26	0.190	0.680	130	1782
YS1	TerN	115	0.136	0.537	180	848
ZJ1	SthB	125	0.033	0.817	180	988
ZA2	Hlds	65	0.130	0.820	150	1140
YP1	ShlA	255	0.001	0.869	275	785
ZG3	SmLm	13	0.722	0.158	40	1342
YN2	Lds	89	0.620	0.220	160	1413
YR2	RgdH	260	0.001	0.679	272	709
YS3	SwTN	175	0.005	0.836	200	894
ZC2	Grdy	40	0.790	0.170	135	2003
ZB1	IsM	27	0.782	0.084	120	2093
YM3	SwBV	225	0.001	0.899	330	843
YL1	UpHm	145	0.145	0.740	300	1229
YD1	BvrB	112	0.112	0.806	425	1197
YO6	Ptrs	222	0.018	0.822	250	768
YD2	NeRk	184	0.001	0.829	425	833
YF1	Cat	123	0.180	0.689	350	1325

in the form of snow than the other basins in the Z region. The exception at the low end is Ste. Genevieve basin; it may have an unusually low FD-10 for a Y basin because of its low elevation, large degree of control by lakes and swamps (100 percent) and the moderating effect on temperature of its proximity to the sea.

Multiple Regression

A multiple regression analysis was carried out to explore these possibilities. The results, presented in Table 4.12, were not especially encouraging, so no further clustering or analyses was attempted. Generally the observations above were confirmed, but it is a bit surprising that none of the elevation variables were significant.

4.4.3 Sp-Fld and Qavgfld

The most important high flow measure for design of structures, and the one to which most attention has been given in regional hydrological studies, is the peak flow. Peak flow is virtually always dominated by drainage area - the larger the area contributing to the flood, the larger the flood flow.

In order to gain an understanding of the other, less obvious, factors which might suggest clustering of basins based on characteristics other than simply size, the

Table 4.12

FD – 10: Selected Multiple Regression Results

Regression Coefficients				
	All Basins		Region Y	Region Z
All Variables Lns	OLS	RLS	OLS	OLS
N	40	37	20	20
Constant	0.268	0.096	−0.078	0.816
Drd_Ln	0.051	0.055	0.067	
Dist−N	0.093	0.132	0.099	
FrLSw_Ln	−0.044	−0.028	−0.785	
Dist−SW			0.055	
FACLS				0.150
Shp				−0.137
r^2	0.45	0.68	0.62	0.37
r^2 adj	0.41	0.65	0.51	0.29
Std.Er.	0.07	0.05	0.06	0.07
Outliers	St. G Isle aux M Sm Lm			

OLS – Ordinary Least Squares
RLS – Reweighted Least Squares

relationships among the specific flood (average maximum annual daily flow per unit area) and the various basin characteristics were examined.

Qualitative Assessment

The scatter plots in Figures 4.1 to 4.3 and the correlation matrix show that Sp-Fld has the strongest relationships with slope and effective precipitation. Other weaker, but possibly important relationships, are with drainage area, drainage density, specific length of main channel, and fraction of lakes and swamps. All relationships tend to be nonlinear. The relationship with drainage area is negative, because smaller basins tend to have higher peak flows per unit area than larger basins, although in absolute terms of course the larger basins will have higher peaks. The correlation coefficients must be treated with caution where drainage area is a common term because of the possibility of spurious correlation, but the relationships are in general as expected. The principal components analysis does not provide much further information.

The data set was also sorted by specific flood as shown in Table 4.13 in order to assess in a qualitative way the differences between the basins with the highest unit peaks and the ones with the lowest. Not surprisingly, since the Y region tends to have lower runoff in general than the Z region, the ten basins with the lowest unit flood are all Y's, with the exception of the two largest Z basins, which have

Table 4.13

Data Set Sorted By Specific Flood

YZS	BASINS	SPFLD	DA	LMC SP	EL GAUGE	EL DIVDE	EL CENTR	SLP1085	DR DENS	SHAPE	FR LSW	FACLS
YR3	IndB	0.099	554	0.09	8	137	106.68	0.22	0.680	1.72	0.310	0.970
YA1	SteG	0.102	306	0.13	12	107	81	0.23	0.540	1.48	0.354	1.000
ZB1	SmLP	0.106	2640	0.04	183	305	274.32	0.08	0.360	1.75	0.160	1.000
YR1	MdlB	0.106	275	0.18	23	198	122	0.32	0.255	1.93	0.245	0.980
YQ1	GdBC	0.136	4444	0.03	23	305	168	0.14	0.452	2.08	0.172	0.910
YS1	TerN	0.140	1365	0.08	84	290	244	0.12	0.726	2.35	0.327	0.920
ZF1	BdN	0.152	1170	0.06	8	274	152	0.29	0.612	2.15	0.236	0.960
YK4	Hnds	0.171	529	0.09	198	518	290	0.32	0.637	1.78	0.355	0.950
YR2	RgdH	0.177	399	0.11	27	122	76	0.21	0.740	1.68	0.320	0.960
YK5	Shfd	0.200	391	0.10	110	488	290	1.07	0.191	1.98	0.172	0.940
YK2	LewB	0.255	470	0.12	137	655	351	0.59	0.627	2.32	0.161	1.000
ZH1	PH	0.260	764	0.07	38	244	213	0.35	0.709	1.67	0.659	0.910
YL1	UpHm	0.272	2110	0.06	23	701	183	0.40	0.786	1.56	0.115	0.750
ZG1	Grn	0.277	205	0.22	11	381	152	0.60	0.547	2.45	0.101	0.960
YO6	Ptro	0.283	177	0.24	23	213	137	0.45	0.800	1.93	0.160	0.970
ZA1	LtlB	0.289	343	0.19	8	472	137	0.68	1.040	2.45	0.100	0.830
YS3	SwTN	0.290	36.7	0.30	23	168	107	1.11	0.641	1.43	0.159	1.000
ZJ1	StbB	0.298	67.4	0.24	8	137	91	0.50	1.240	1.43	0.150	0.860
ZG2	Tds	0.302	166	0.16	8	229	213	1.35	1.350	1.84	0.130	0.920
YJ1	Hrys	0.309	640	0.09	15	541	274	0.35	1.120	1.81	0.142	0.750
YC1	Tor	0.316	624	0.08	8	488	309	1.01	0.755	1.45	0.167	0.990
YP1	ShlA	0.350	63.8	0.24	84	198	152	0.53	0.880	1.62	0.130	0.790
YN2	Lds	0.372	469	0.12	328	488	442	0.22	1.370	2.15	0.160	1.000
ZK1	Rky	0.374	300	0.15	15	168	60	0.50	1.005	2.00	0.113	0.550
ZM9	SlCv	0.394	53.6	0.28	35	168	168	0.62	1.130	1.37	0.130	1.000
YM3	SwBV	0.399	93.2	0.20	69	168	152	0.27	0.680	1.67	0.100	0.560
YD1	BvrB	0.404	237	0.17	8	335	351	0.67	0.339	2.23	0.082	0.730
ZJ3	SmLm	0.406	115	0.21	2	137	91	0.34	1.550	1.62	0.120	0.920
YF1	Cat	0.417	611	0.05	351	594	549	0.73	0.582	1.86	0.131	1.000
ZK2	NB--P	0.497	89.6	0.30	15	213	120	0.55	1.110	1.91	0.300	0.810
ZH2	CBC	0.528	433	0.39	53	152	168	0.59	1.110	1.66	0.099	0.920
ZN1	NuNP	0.535	53.3	0.27	114	213	183	0.61	1.089	2.06	0.126	1.000
YD2	NeRk	0.536	200	0.19	23	290	110	0.47	0.930	1.65	0.170	0.990
ZA2	Hlds	0.540	72	0.28	69	533	244	2.19	1.150	1.72	0.050	0.430
ZL3	SpCv	0.563	10.8	0.65	76	168	170	1.25	1.090	1.36	0.090	1.000
ZG4	Rtl	0.610	42.7	0.23	46	152	122	1.10	1.620	1.53	0.180	0.920
ZM6	NEP	0.611	3.63	0.72	122	191	168	2.42	1.038	1.24	0.225	1.000
ZA3	LtCd	0.726	139	0.18	8	457	274	1.46	1.460	1.67	0.130	0.730
ZB1	IaM	0.843	205	0.16	8	457	335	0.84	0.720	2.09	0.134	0.600
ZC2	Grdy	0.999	230	0.13	84	442	335	1.06	0.960	1.84	0.040	0.340

Table 4.13 Continued

YZS	BASINS	DIST_SW	FR_BARRN	FR_FORST	DIST_N	EFF-P
YK3	IndB	225	0.001	0.689	245	718
YA1	SteG	17	0.002	0.644	445	915
ZE1	SmLP	105	0.490	0.350	165	1026
YR1	MdlB	180	0.008	0.747	220	763
YQ1	GdBC	160	0.068	0.760	205	831
YS1	TerN	115	0.136	0.537	180	848
ZP1	BdN	76	0.442	0.322	55	1076
YK4	Hnds	218	0.292	0.353	235	984
YR2	RadH	260	0.001	0.679	272	709
YK5	Sbfd	266	0.152	0.676	323	855
YK2	LewB	150	0.290	0.549	200	1168
ZH1	PH	68	0.234	0.107	150	1024
YL1	UpHm	145	0.145	0.740	300	1229
ZG1	Grn	66	0.634	0.265	75	1359
YO6	Ptrw	222	0.018	0.822	250	768
ZA1	LtlB	128	0.299	0.601	185	1012
YS3	SwTN	175	0.005	0.836	200	894
ZJ1	SthB	125	0.033	0.817	180	988
ZG2	Tds	33	0.488	0.382	60	1521
YJ1	Hrys	34	0.069	0.789	215	1302
YC1	Tor	74	0.498	0.335	395	1284
YP1	ShlA	255	0.001	0.869	275	785
YN2	Lds	89	0.620	0.220	160	1413
ZK1	Rky	70	0.377	0.510	80	1178
ZM9	SlCv	28	0.500	0.370	35	1760
YM3	SwBV	225	0.001	0.899	330	843
YD1	BvrB	112	0.112	0.806	425	1197
ZG3	SmLm	13	0.722	0.158	40	1342
YF1	Cat	123	0.180	0.689	350	1325
ZK2	NB-P	40	0.230	0.470	75	1571
ZH2	CBC	20	0.496	0.405	140	1356
ZN1	NwNP	38	0.788	0.086	40	1853
YD2	NeRk	184	0.001	0.829	425	833
ZA2	Hlds	65	0.130	0.820	150	1140
ZL3	SpCv	76	0.490	0.420	120	1230
ZG4	Rd	97	0.470	0.350	90	1559
ZM6	NEP	100	0.038	0.737	115	1165
ZA3	LtCd	26	0.190	0.680	130	1782
ZB1	IaM	27	0.782	0.084	120	2093
ZC2	Grdy	40	0.790	0.170	135	2003

DA's of 1170 and 2640 km^2. Conversely, the ten basins with the highest unit flood flows are all in region Z, with the exception of one Y basin on the Great Northern Peninsula, Northeast Brook near Roddickton (NeRk). It is not clear what makes it respond differently; it is similar in size, mean annual runoff, elevation, slope, fraction of lakes and so on to some other Y basins (e.g., Peters River) yet has a much higher unit flood runoff.

The other Great Northern Peninsula gauges do tend to have relatively high unit flood flows, although not as high as NeRk, so the difference may relate at least in part to a greater amount of snow. As always, there is an exception, in this case Ste. Genevieve (SteG). Although it too is on the Great Northern Peninsula, not far from NeRk, the basin is fairly low, possibly reducing snow accumulation. It is also dominated by a very large lake which could be expected to attenuate flood flows.

Multiple Regression Analysis: Sp-Fld

A brief multiple regression analysis was carried out considering the island as one region using SpFld-Ln as the dependent variable, in order to explore some possible relationships not dominated by drainage area, before moving to a consideration of QavgFld.

Table 4.14 summarizes the coefficients resulting from some of the better equations. DA and Eff-P alone explain about two-thirds to three-quarters of the variance in unit flood flows. Although slope has the highest correlation of any variable with Sp-Fld, there is no other variable which can be combined with it to give a stronger correlation than DA and Eff-P. It is possible that the relationship is spurious, given that Sp-Fld is equal to Qavgfld/DA. At the same time the signs associated with the coefficients are all hydrologically reasonable, and it is certainly accepted that, all other factors being equal, small basins will have higher peaks than large basins. SteG and NeRk, not surprisingly given the discussion above, are identified as outliers.

Adding FACLS-Ln improves the equation, with an adjusted r^2 of 0.86 after removal of outliers. Adding slope and drainage density gives the best 5 variable equation (with Eff-P), but slope is replaced by Dist-SW and Dist-N to give the best 6 variable equation. As noted in Section 4.3.1, all coefficients are significant, with p less than 0.05. No variables were accepted in this study if the associated coefficients were not significant.

Without Eff-P, El-Cntrd and Fr-Forst are required. Fr-Forst is closely related to Fr-Barrn, which was an important explanatory variable along with El-Cntrd in the

Table 4.14

Sp–Fld: Selected Multiple Regression Results

Regression Coefficients

Vars all lns	With Eff–P OLS Lns 2 vars	With Eff–P RLS Lns 2 vars	With Eff–P OLS Lns 3 vars	With Eff–P RLS Lns 3 vars	With Eff–P OLS Lns 5 vars	With Eff–P OLS Lns 6 vars	Without Eff–P OLS Lns 6 vars
N	40	38	40	37	40	40	40
Constant	−0.295	−0.339	−0.361	−0.378	−0.590	−2.330	−1.885
DA	−0.191	−0.185	−0.197	−0.194	−0.111	−0.182	−0.144
DrDens					0.233	0.306	0.402
Dist–SW						0.178	
Dist–N						0.224	
El–Centrd							0.290
FACLS			−0.746	−0.693	−0.631	−0.577	−0.721
Fr–Forst							−0.132
Eff–P	1.190	1.230	0.996	1.112	0.725	1.489	
Slope					0.183		0.198
r^2	0.69	0.77	0.77	0.87	0.81	0.87	0.82
r^2 adj	0.67	0.75	0.76	0.86	0.78	0.84	0.78
Std.Er.	0.34 lns	0.29 lns	0.30	0.22			0.28
Outliers		SteG NeRk		SteG NeRk SlCv			

OLS – Ordinary Least Squares
RLS Reweighted Least Squares

analysis of Eff-P. The resulting 6 variable equation without Eff-P is otherwise very similar to the 5 variable equation with Eff-P.

The results show that both the expected basin characteristics (especially DA and DrDens) and topographic variables are important in explaining flood runoff. The topographic characteristics can be important even when Eff-P is explicitly included in the equation. They may be important for their physiographic effect, e.g. faster/slower runoff, or they may still be representing a hydrologic input specific to storms, in addition to the average annual effective precipitation.

With this information from the specific flood analysis, a multiple regression analysis was carried out using the dimensional average maximum daily flow as the dependent variable.

Multiple Regression: QavgFld

The multiple regression analysis using QavgFld generally confirmed the importance of the explanatory variables obtained in the nondimensional case of Lcv, and in the unit runoff case of Sp-Fld. The analysis started with the one-region case, and because there was some suggestion of geographical differences, a preliminary subdivision was made at the Y-Z boundary. Logarithmically transformed variables were used throughout the analysis because the scatter plots (and previous experience) suggested that the relationships are nonlinear.

Table 4.15 presents selected results of the multiple regression, starting with the island-wide results. As expected from hydrological principles as well as from the correlation matrix, drainage area alone explains over 80 percent of the variance in flood flows from one basin to another. Four outliers were identified. The DA coefficient ranging from 0.75 to 0.82 is not surprising given the results in Table 4.14; since

$$Qavgfld = DA^{1.0}\,(Sp\text{-}fld), \text{ and}$$

$$Sp\text{-}fld = f(DA)^{-0.2}, \text{ then}$$

$$Qavgfld = f(DA)^{0.8}.$$

The next most important variable, again as expected from the analysis of Lcv and Sp-Fld, was the area controlled by lakes and swamps. The r^2 increased to over 90 percent in this case (slightly better if the two outliers were excluded). In both these one and two variable cases, SteG is again an outlier, probably because of the apparently large degree of control provided by a large lake not far upstream of the gauge. Grandy has a very low FACLS. There is no immediately apparent physical explanation for the other outliers.

Table 4.15

QavgFld: Selected Multiple Regression Results

a) All Island
Regression Coefficients

	OLS Lns 1 va.	RLS Lns 1 var	OLS Lns 2 var	RLS Lns 2 var	OLS Lns 2 var	RLS Lns 2 var	OLS Lns 3 vars	RLS Lns 3 vars
N	40	36	40	38	40	39	40	37
Constant	0.168	0.119	−0.059	0.002	−0.362	−0.340	−0.432	−0.432
DA	0.747	0.764	0.753	0.731	0.813	0.814	0.807	0.820
FACLS			−1.161	−1.481			−0.783	−0.705
Eff−P					1.297	1.247	1.095	1.002
r^2	0.837	0.913	0.90	0.91	0.94	0.95	0.96	0.97
r^2 adj								
Std.Er.	0.480	0.360	0.39	0.38	0.31	0.28	0.24	0.20
Outliers		SteG MdlB IndB Grdy		SwBV Hlds		SteG		SteG IndB SmLP

b) Additional Variables

N = 40

	Additional Variables						
# of Vars (all lns)	with 4	without 4	with 5	without 5		with 7	
Constant	−1.266	−1.024	−1.427	−1.625		−0.198	
DA	0.805	0.720	0.825	0.765		0.797	
Eff−P	1.476		1.413			1.259	
FACLS	−0.750	−1.014	−0.730	−0.913		0.695	
Dist−SW	0.181		0.205			0.158	
DrDens			0.177	0.333		0.247	
Dist−N						0.210	
Fr−Barrn		0.085		0.061		0.047	
El−Cntrd		0.264		0.341			
Slp1085							
r^2	0.97	0.95	0.97	0.96		0.98	
r^2 adj	0.96	0.95	0.97	0.96		0.97	
Std.Er.	0.22	0.27	0.21	0.24		0.19	

Table 4.15 Continued

c) Region Y and Z Regression Coefficients
N = 20

	Region Y Without Eff−P = With Eff−P				Region Z with Eff−P	with Eff−P	with Eff−P	without Eff−P
# of Vars	2	3	4	2	3	4	6	5
Constant	−2.600	−2.770	−3.435	0.127	−0.320	−1.362	−1.858	−1.713
DA	0.927	0.890	0.858	0.739	0.778	0.784	0.803	0.742
Fr−LSw	−0.897	−1.003	−0.664					
FACLS				−1.064	−0.814	−0.637	−0.607	−0.781
Eff−P					1.008	1.261	1.007	
Dist−N						0.214	0.285	
Fr−Forst		−0.499					−0.153	−0.283
DrDens			0.316				0.285	0.459
El−Cntrd			0.341					
								0.286
r^2	0.94	0.96	0.98	0.95	0.98	0.99	0.99	0.98
r^2 adj	0.93	0.96	0.98	0.94	0.98	0.98	0.99	0.97
Std Er	0.28	0.22	0.16	0.30	0.19	0.16	0.12	0.21

OLS− Ordinary Least Squares
RLS − Reweighted Least Squares

If Eff-P is allowed as a basin characteristic, it becomes the third explanatory variable, bringing the r^2 up to 0.96, again slightly better if the one outlier, Ste. Genevieve, is included. Additional variables (up to 7 meet the statistical criteria) in approximate order of importance are Dist-SW, Dr-Dens, Dist-N and Fr-Barrn (all lns). As the variables related to Eff-P enter, the tolerance (reflecting multicollinearity) on Eff-P decreases, although it is still acceptable. With three variables, for example, the tolerance is 0.85; when Dist-SW is included as a fourth independent variable, the tolerance decreases to 0.43 (still greater than 0.1, the cutoff point for acceptability).

The distance variables, and possibly Fr-Barrn, are probably representing hydrologic input in some way, in addition to the effect of Eff-P. Dist-SW might represent position along the storm track, for example, and Dist-N might indicate the importance of the snowmelt contribution. If the snowmelt contribution is important, however, one would expect one of the elevation variables to come into play. The distance variable would not otherwise be expected to relate to flood runoff, since they are not true physiographic characteristics.

If Eff-P is not allowed as a basin characteristic, a five variable equation gives results similar to the three variable equation with Eff-P. Fr-Barrn and El-Cntrd appear to be the proxy variables for Eff-P; Dist-SW is not required. Dist-N is not used.

The plots of the residuals did not suggest any particular clustering, except possibly a Y-Z division. The residuals do not group according to the hydrological regions presented in the provincial regional flood frequency analysis (Gov't of Nfld, 1990).

Equations developed for the Y and Z subregions improved the r^2's and reduced the standard error. For a four variable case, for example, the best equations for the Y and Z regions have adjusted r^2s of 0.98 respectively, and standard errors (in the log domain) of 0.16. The best four variable equation for the island has an adjusted r^2 of 0.96 and a standard error of 0.22. The results, presented in Table 4.15(c), also suggest some differences in the two regions, the most important being the relative importance of Eff-P.

For both regions, DA is of course the most important explanatory variable; in the Y region the second most important is Fr-LSw, in the Z region it is the area controlled by lakes and swamps, both similarly suggesting attenuation due to the effects of lakes and swamps. An important point to note with the Y region, however, is that Eff-P is not one of the better explanatory variables, whereas it is the third most important variable in the Z region.

If only one more variable (for a total of three) is allowed in the equation for the Y region, it is Fr-Forst, but if two more are allowed in they are El-Cntrd and Dr-Dens. These results are similar to the four variable case for the island as a whole when Eff-P is excluded. Fr-Barm is more significant for the island than Fr-Forst, but the two are essentially representing the same phenomenon of exposure (or protectedness) from the storm track, as well as possibly some faster runoff (or attenuation in the forested case).

For the Z region, Eff-P and Dist-N are most important, with Dr-Dens and Fr-Forst following, as they do in the Y region. When Eff-P is excluded, Dist-N no longer contributes significantly, and El-Cntrd appears to work with Fr-Forst to represent geographical location.

The idea that Fr-Forst/Fr-Barm are geographical surrogates rather than physiographic parameters could be important when hypotheses are being tested about the effects of forestry. Generalizations cannot be made from one basin to another unless they are similarly situated with respect to incoming weather systems.

The only surprises in this analysis were the importance of the geographic variables even when Eff-P is included as an independent variable, and the lack of importance of slope (which had been important in the Lcv analysis).

4.5 Low Flows: FD-90, Sp-Low, Qlow

In general, attempts to relate basin characteristics to low flows have been less successful than similar attempts for high flows. The explanation usually given is that the factors controlling low flows are difficult to measure or to index. In the provincial low flow frequency analysis for the island of Newfoundland (Gov't of Nfld, 1991), the only physiographic characteristic which could be consistently related to low flows was drainage area, with the fraction of forest being important in one region.

It is possible that fraction of forest is acting as a surrogate for some soil or geological characteristic. This is plausible, since presumably forests grow in areas with more, or at least different, soil than regions without forests. The results of the present study for other flow variables suggest that the fraction of forest may also be acting as a surrogate for hydrologic input - forests seem to be more likely to develop in areas which are sheltered from major weather systems. The two underlying characteristics, one physiographic, the other hydrologic, are not necessarily unrelated; the interactions of weather and soils are complex, and it is plausible that soil is more likely to accumulate to a depth to support forests in a sheltered area.

Compared to other parts of the world where low flow studies have been carried out, the island of Newfoundland should be more amenable to analysis, because there is so little groundwater storage. In other areas with deep soils and aquifers, the difference in subsurface characteristics from one basin to another may cause problems in low flow analysis. The average depth of soil in Newfoundland is less than half a metre, so superficial characteristics, such as area of lakes and swamps, should be relatively more important. In addition, since there is a close and direct link between surface and groundwater, measures of hydrologic input controlling the timing of flows should also be important. If a measure such as depth of snow on the ground in late winter were available, for example, it might turn out to be helpful. Since it is not, one must provide surrogates such as distance north or elevation.

As with the analyses of the other flow variables, the investigation started with an examination of the relationships among the low flow variables themselves, and a qualitative assessment of the sorted data file. It then proceeded to a multiple regression analysis to further investigate the interactions among the variables.

4.5.1 Qualitative Assessment: Relationships among FD-90, SP-Low and Qlow and Basin Characteristics

Sp-Low and Qlow are the same flow measure; Sp-Low is simply the median daily minimum flow per unit area, obtained from the annual series. FD-90 is the flow which is exceeded 90 percent of the time, nondimensionalized by dividing by the average flow. Since the relationships of low flows to Qavg/DA are not strong, it can be compared more or less directly with Sp-Low. Figure 4.14 shows a plot of FD-90 and Sp-Low. On the average the flow will be lower than FD-90 about 36 days per year, whereas it will be lower than Sp-Low only about once every other year.

In the case of the low flows, it appears that the same factors leading to frequent low flows, as represented by low FD-90, are the same as those causing it to have the lowest minimum flows as well.

The scatter plots and the correlation matrix showed that none of the three are correlated with any basin characteristics except Qlow, which is highly correlated with drainage area. Because of the relationship of other variables to DA, it is also correlated to LMC-Sp, Shape, and Slp-1085. The scatter plots and correlation matrix also suggests that the relationships, even if weak, tend to be nonlinear, so the transformed data set was used throughout the low flow analysis.

Table 4.16 shows the data set sorted by FD-90; because of the relationship between the two it is similar when sorted by Sp-Low. The Y-Z split suggested in some of the other sorted tables does not show up here directly. The basins with

Table 4.16

Data Set Sorted by Specific Low Flow

YZS	BASINS	SP LOW	DA	LMC SP	EL GAUGE	EL DIVDE	EL CENTR	SLP1085	DR DENS	SHAPE	FR LSW	FACLS
YM3	SwBV	0.00060	93.2	0.20	69	168	152	0.27	0.680	1.67	0.100	0.560
ZJ1	SthB	0.00110	67.4	0.24	8	137	91	0.50	1.240	1.43	0.150	0.860
YR2	RgdH	0.00145	399	0.11	27	122	76	0.21	0.740	1.68	0.320	0.960
ZM6	NEP	0.00165	3.63	0.72	122	191	168	2.42	1.038	1.24	0.225	1.000
YS3	SwTN	0.00177	36.7	0.30	23	168	107	1.11	0.641	1.43	0.159	1.000
YD2	NeRk	0.00186	200	0.19	23	290	110	0.47	0.930	1.65	0.170	0.990
YO6	Ptrs	0.00186	177	0.24	23	213	137	0.45	0.800	1.93	0.160	0.970
ZH2	CbC	0.00222	43.3	0.39	53	152	168	0.59	1.110	1.66	0.099	0.920
YF1	Cat	0.00226	611	0.05	351	594	549	0.73	0.582	1.86	0.131	1.000
ZG3	SmLm	0.00240	115	0.21	2	137	91	0.34	1.550	1.62	0.120	0.920
YD1	BvrB	0.00240	237	0.17	8	335	351	0.67	0.339	2.23	0.082	0.730
YP1	ShlA	0.00250	63.8	0.24	84	198	152	0.53	0.880	1.62	0.130	0.790
ZC2	Grdy	0.00259	230	0.13	84	442	335	1.06	0.960	1.84	0.040	0.340
ZL3	SpCv	0.00259	10.8	0.65	76	168	170	1.25	1.090	1.36	0.090	1.000
YR1	MdlB	0.00333	275	0.18	23	198	122	0.32	0.255	1.93	0.245	0.980
ZH1	PH	0.00334	764	0.07	38	244	213	0.35	0.709	1.67	0.659	0.910
ZG4	Rtl	0.00342	42.7	0.23	46	152	122	1.10	1.620	1.53	0.180	0.920
ZK1	Rky	0.00350	300	0.15	15	168	60	0.50	1.005	2.00	0.113	0.550
ZA2	Hlde	0.00352	72	0.28	69	533	244	2.19	1.150	1.72	0.050	0.430
YL1	UpHm	0.00363	2110	0.06	23	701	183	0.40	0.786	1.56	0.115	0.750
YK5	Shld	0.00363	391	0.10	110	488	290	1.07	0.191	1.98	0.172	0.940
ZB1	IsM	0.00390	205	0.16	3	457	335	0.84	0.720	2.09	0.134	0.600
YQ1	GdBC	0.00430	4444	0.03	23	305	168	0.14	0.452	2.08	0.172	0.910
ZA1	LtlB	0.00452	343	0.19	8	472	137	0.68	1.040	2.45	0.100	0.830
YS1	TerN	0.00489	1365	0.08	84	290	244	0.12	0.726	2.35	0.327	0.920
YR3	IndB	0.00493	554	0.09	8	137	107	0.22	0.680	1.72	0.310	0.970
YN2	Lds	0.00499	469	0.12	328	488	442	0.22	1.370	2.15	0.160	1.000
ZA3	LtCd	0.00525	139	0.18	8	457	274	1.46	1.460	1.67	0.130	0.730
ZK2	NB–P	0.00554	89.6	0.30	15	213	120	0.55	1.110	1.91	0.300	0.810
YK2	LowB	0.00585	470	0.12	137	655	351	0.59	0.627	2.32	0.161	1.000
YC1	Tor	0.00599	624	0.08	8	488	309	1.01	0.755	1.45	0.167	0.990
YK4	Hnds	0.00616	529	0.09	198	518	290	0.32	0.637	1.78	0.355	0.950
ZG1	Grn	0.00624	205	0.22	11	381	152	0.60	0.547	2.45	0.101	0.960
YJ1	Hrys	0.00628	640	0.09	15	541	274	0.35	1.120	1.81	0.142	0.750
ZE1	SmLP	0.00631	2640	0.04	183	305	274	0.08	0.360	1.75	0.160	1.000
ZG2	Tds	0.00699	166	0.16	8	229	213	1.35	1.350	1.84	0.130	0.920
ZF1	BdN	0.00789	1170	0.06	8	274	152	0.29	0.612	2.15	0.236	0.960
ZM9	SlCv	0.00793	53.6	0.28	35	168	168	0.62	1.130	1.37	0.130	1.000
YA1	SteG	0.00855	306	0.13	12	107	81	0.23	0.540	1.48	0.354	1.000
ZN1	NwNP	0.00893	53.3	0.27	114	213	183	0.61	1.089	2.06	0.126	1.000

Table 4.16 Continued

YZS	BASINS	DIST SW	FR BARRN	FR FORST	DIST N	EFF-P
YM3	SwBV	225	0.001	0.899	330	843
ZJ1	SthB	125	0.033	0.817	180	988
YR2	RgdH	260	0.001	0.679	[illegible]	709
ZM6	NEP	100	0.038	0.737	115	1165
YS3	SwTN	175	0.005	0.836	200	894
YD2	NeRk	184	0.001	0.829	425	833
YO6	Ptm	222	0.018	0.822	250	768
ZH2	CbC	20	0.496	0.405	140	1356
YF1	Cat	123	0.180	0.689	350	1325
ZG3	SmLm	13	0.722	0.158	40	1342
YD1	BvrB	112	0.112	0.806	425	1197
YP1	ShlA	255	0.001	0.869	275	785
ZC2	Grdy	40	0.790	0.170	135	2003
ZL3	SpCv	76	0.490	0.420	120	1230
YR1	MdlB	180	0.008	0.747	220	763
ZH1	PH	68	0.234	0.107	150	1024
ZG4	Rd	97	0.470	0.350	90	1559
ZK1	Rky	70	0.377	0.510	80	1178
ZA2	Hide	65	0.130	0.820	150	1140
YL1	UpHm	145	0.145	0.740	300	1229
YK5	Shld	266	0.152	0.676	323	855
ZB1	IsM	27	0.782	0.084	120	2093
YQ1	GdBC	160	0.068	0.760	205	831
ZA1	LdB	128	0.299	0.601	185	1012
YS1	TerN	115	0.136	0.537	180	848
YR3	IndB	225	0.001	0.689	245	718
YN2	Lds	89	0.620	0.220	160	1413
ZA3	LtCd	26	0.190	0.680	130	1782
ZK2	NE-P	40	0.2[illegible]0	0.470	75	1571
YK2	LewB	150	0.290	0.549	200	1168
YC1	Tor	74	0.498	0.335	395	1284
YK4	Hnds	218	0.292	0.353	235	984
ZG1	Grn	66	0.634	0.265	75	1359
YJ1	Hrys	34	0.069	0.789	215	1302
ZE1	SmLP	105	0.490	0.350	165	1026
ZG2	Tds	33	0.488	0.382	60	1521
ZF1	BdN	76	0.442	0.322	55	1076
ZM9	SlCv	28	0.500	0.370	35	1760
YA1	SteG	17	0.002	0.644	445	915
ZN1	NwNP	38	0.788	0.086	40	1853

the lowest low flows in general seem to be small, without much control by lakes and swamps, but there are several anomalies, especially at the low end.

Cat Arm, for example, is a large basin with 100 percent of its area controlled by lakes and swamps, but it has very low flows. WSC staff report that gauging at low flows was difficult at that location, so there may be a problem with the data. Also the high elevation of the basin may mean that spring runoff is later than in other basins, resulting in very low flows in the late winter. Salmonier River, on the boot of the Burin Peninsula, is of similar size to other basins on the Avalon and Burin peninsulas, and also has a high degree of control by lakes and swamps, but it too has surprisingly low flows (although the Eff-P is unusually low, as discussed in Section 4.3.1). Two basins in the YR region, Ragged Harbour (RgdH) and Indian Bay Brook (IndB), show marked differences. They are very similar on most of the measures which would be expected to control low flows. They both have very low Eff-P, a large degree of control by lakes and swamps (96 and 97 percent), similar drainage areas, and similar elevations. Yet RgdH has the third lowest specific low flow, and IndB has the third highest (of 40).

In order to examine the interactions of geography, topography and basin characteristics, a multiple regression analysis was carried out using the three low flow measures as the dependent variables.

4.5.2 Low Flow Multiple Regression Analysis

A summary of selected results from the multiple regression analysis is presented in Table 4.17. Although the analysis was by no means exhaustive, it does show that the same variables, or variables apparently representing the same phenomena, tend to recur for all three flow measures, and for the all-island and Y-Z subregions.

Of the basin characteristics, DA and FACLS are always important, DA particularly so as expected when the dependent variable is Qlow. The single exception is the equation for FD-90 for Subregion Y. DA is not significant; Dr-Dens may be substituting to some extent. Similarly, Fr-Lsw is more important than FACLS in that equation. (A similar result was found for high flows.)

The other major result is the importance of the distance (geographic) variables. Dist-N or Dist-SW, and often both of them, are always important explanatory variables, even when the Y-Z split, which is basically N-S, has been made. Dist-N may indicate the likelihood of winter precipitation being held as snow, leading to lower low flows at the end of the winter. Dist-SW may indicate a general wet-dry trend; in drought periods, it is reasonable to assume that basins closer to the direction of incoming weather systems are likely to have higher low flows than basins that are more remote.

Table 4.17

Low Flows: Selected Multiple Regression Results

a) FD-90

Regression Coefficients

	One region	Region Y	Region Z
All vars as Lns	Lns	Lns	Lns
N	40	20	20
	3 vars	4 vars	3 vars
Constant	−1.156	4.290	−1.089
DA	0.130		0.106
Dist−N	−0.244	−0.664	−0.449
FACLS	0.468		
Dist−SW		−0.363	0.219 Note sign
Fr−LSw		0.404	
Dr−Dens		−0.254	
r^2	0.37	0.64	0.60
r^2 adj	0.32	0.54	0.53
Std.Er.	0.33	0.31	0.23

b) SpLow

(all lns)	One Region					Region Y	Region Z
	OLS without Eff−P	RLS	with Eff−P	without Eff−P			
# of Vars	4	4	4	4	5	3	4
Constant	−5.033	−3.918	−0.599	−4.127	−4.418	−4.816	−7.069
DA	0.179	0.165	0.234	0.228	0.331	0.230	0.107
Dist−SW	−0.284	−0.486		−0.273	−0.256	−0.428	
FACLS	0.734	0.947	0.811	0.658	0.869	1.575	0.690
Fr−Barrn	0.076	0.113					
Dist−N			−0.252	−0.278	−0.301		−0.732
Fr−Forst							
Fr−LSw							
Shape							
Slp1085					0.290		
Ei−Dvide							0.816
Lnc−Sp							
Eff−P			0.870				
r^2	0.51	0.77	0.53	0.52	0.58	0.65	0.78
r^2 adj	0.46		0.47	0.46	0.52	0.59	0.72
Std.Er.	0.45	0.32	0.44	0.44	0.42	0.42	0.30
Outliers		ShlA IndB SmLm CbC					

Table 4.17 Continued

c) QLow–Ln

	One region	Region Y	Region Z
All vars as Lns	5 vars	3 vars	4 vars
Constant	−5.716	−4.816	−7.069
DA	1.168	1.230	1.107
Dist–SW	−0.273	−0.428	
Dist–N	−0.307		−0.732
El–Dvide	0.373		0.816
FACLS	0.865	1.576	0.690
r^2	0.95	0.94	0.98
r^2 adj	0.95	0.93	0.97
Std.Er.	0.41	0.42	0.30

OLS– Ordinary Least Squares
RLS– Reweighted Least Squares

Unlike the case for the high flows, Eff-P is not required to obtain good low flow relationships. For the one-region four-variable case with Sp-Low as the dependent variable, if Eff-P is used instead of Dist-SW, the results are only marginally better than if Dist-SW is used. The coefficients for the other three variables are very similar in magnitude, suggesting that Dist-SW and Eff-P are acting similarly.

A statistically significant equation for the one-region case with Sp-Low as dependent variable can have as many as 7 independent variables (not shown in the table). The fifth variable is slope; it is difficult to rationalize the importance of slope hydrologically, particularly because the positive sign indicates an increase in low flow with an increase in slope. The other two variables are LMC-sp and El-Gauge (to reach the total of seven, not shown in the table). A long main channel length per unit area is usually associated with small basins, and could suggest flashiness; one would expect it to be related to floods rather than low flows.

El-Gauge has a positive sign, as does El-Divide in the cases in which it enters (Qlow, Table 4-17c). It is not clear how the elevation of the gauge or the divide would affect low flows. Intuitively one would expect that a higher elevation would result in more precipitation stored as snowmelt, and consequently lower, rather than higher low flows. In any event, if that were true, then one would expect the elevation of the centroid to be more important than the gauge or the divide, since it should be most representative of the basin elevation. El-Cntrd, however, is not significant in any equation.

The equation for Sp-Low in the Z region is statistically significant with up to eight variables, with an improvement in adjusted r^2 from 0.79 to 0.86, but it makes no sense to use so many independent variables.

4.6 Availability: FD Variables

Hydrotechnical engineers are frequently interested in the availability of water on a day-to-day basis for such uses as water supply and hydroelectric generation. In these types of applications, it is frequently economic to provide some storage, and so the very lowest flows are not of particular concern. Of more interest is the general so-called "flashiness" of the basin. Flashiness is a concept that presumably arose primarily in connection with flood flows; basins which run off faster than others are considered to be more flashy. More generally a flashy basin can be thought of as one with higher high flows and lower low flows. A flashy river, by this definition, would not have the required flow to satisfy a particular demand rate on as many days as a less flashy river. The less flashy river will thus have greater availability.

The definition itself is not of particular concern here; of interest are

1) whether higher high flows and lower low flows go together;

2) whether there is a suitable dependent variable or "flashiness index" to use to represent availability;

3) what physiographic, topographic or geographic characteristics are associated with flashiness.

This section discusses these three questions.

4.6.1 Qualitative Assessment

The flow duration variables give a good representation of the expected frequency of various flows, and thus are suitable for use in exploring the question of whether higher high flows and lower low flows tend to occur in the same basins. The scatter plots in Figure 4.6 show the relationship among the FD variables and some of the basin variables to which they are related. FD-90 is related positively to the minimum flow variables, but FD-10 did not show a corresponding relationship with the flood flows. These differences are discussed in the relevant high and low flow sections above. For the purposes of analyzing availability, the three flow duration curve variables would seem to offer good prospects.

The PCA with four components showed that FD-10 and FD-50 loaded most highly on the first component. Although FD-90 had a loading of over 0.5 on the first component (rotated), its highest loading was on the fourth component, with Sp-Low. With three components, all the flow duration variables as well as Sp-Low loaded most highly on the first component. (As mentioned in Section 4.1, the PCA analysis was done using the untransformed data set.)

These results suggests that the annual maximum flows may not be part of the same response as the more common range of high flows, the ones that occur 10 percent of the time. These more common ones, however, do seem to associate with the general range of daily flows. Because the three quantiles of flow duration variables are related, the answer to the first question above is positive; higher high flows do tend to occur in the same basins that have lower low flows.

The next question is what is the most suitable index to use for availability.

Flashiness Index

Several possibilities were considered in order to choose a suitable dependent variable representing flashiness. These included

1) a canonical variable composed of a linear combination of the three FD variables;
2) PC1 from the three component analysis;

3) a "flashiness index" similar to a quartile skew measure obtained from the three FD variables as follows

$$I_1 = \frac{FD10 - 2FD50 - FD90}{FD10 - FD90}$$

$$I_2 = \frac{FD10 - 2 - FD90}{FD10 - FD90}$$

Note that in I_2, the middle term in the numerator is actually 2(Qavg/Qavg). In the first equation the equivalent middle terms 2(Qmedian/Qavg).

4) Other measures such as FD10*FD50. Figure 4.16 shows how a low value (e.g., < 0.2) identifies a flashy basin, whereas a higher value (e.g., > 0.4) identifies a non-flashy basin with greater availability.

5) One of the FD variables selected to represent the others.

The first two were rejected at this level of analysis, primarily because they are not especially meaningful and it is consequently more difficult to interpret results. The third and fourth are quite reasonable, but after a detailed comparison of the various flashiness measures for different basins, they did not offer any particular advantage over #5. The FD-50 value was chosen because it correlates well with both the other two, and is very straightforward to calculate and apply. The alternative indices listed above may be useful at a later stage of the analysis, but as a representative measure of availability for the purpose of identifying important

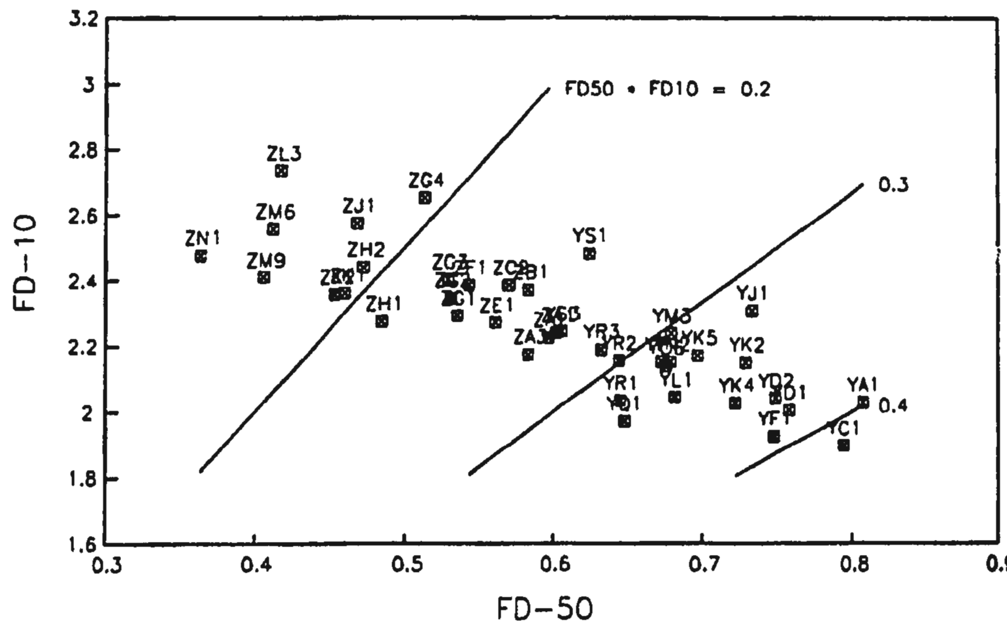


Fig 4.16 – Example of a Flashiness Index

basin characteristics, the FD50 value should be as suitable as any of the others, and it is the simplest.

4.6.2 FD-50

Qualitative Assessment

The data set sorted by FD-50 is presented in Table 4.18. A comparison of the characteristics of the basins with high availability and with low availability offers some indications of the variables which might be significant in a multiple regression analysis. The two basins with highest availability are two large basins on the south coast. Four others in the top ten are on the Burin or Avalon peninsulas. Terra Nova and Gander Rivers (in the eastern part of the Y region) are also in this group; they are both large basins, and have a high degree of control by lakes and swamps.

Ste. Genevieve is anomalous as usual; presumably the extremely high degree of control offered by the large lake just upstream of the gauge has a strong influence on its runoff pattern.

The ten flashiest gauges tend to group geographically into the high basins on the southwest coast, Grandy, Isle aux Morts, and Highlands, and the basins on the

Table 4.18

Data Set Sorted by FD−50

YZS	BASINS	FD50	DA	LMC SP	EL GAUGE	EL DIVDE	EL CENTR	SLP1085	DR DENS	SHAPE	FR LSW	FACLS
YM3	SwBV	0.364	93.2	0.20	69	168	152	0.27	0.680	1.67	0.100	0.560
ZC2	Grdy	0.406	230	0.13	84	442	335	1.06	0.960	1.84	0.040	0.340
YD1	BvrB	0.412	237	0.17	8	335	351	0.67	0.339	2.23	0.082	0.730
YF1	Cet	0.417	611	0.05	351	594	549	0.73	0.582	1.86	0.131	1.000
ZA2	Hlds	0.454	72	0.28	69	533	244	2.19	1.150	1.72	0.050	0.430
YP1	ShlA	0.461	63.8	0.24	84	198	152	0.53	0.880	1.62	0.130	0.790
YO6	Ptrs	0.469	177	0.24	23	213	137	0.45	0.800	1.93	0.160	0.970
ZB1	IaM	0.473	205	0.16	8	457	335	0.84	0.720	2.09	0.134	0.600
ZM6	NEP	0.485	3.63	0.72	122	191	168	2.42	1.038	1.24	0.225	1.000
YD2	NeRk	0.513	200	0.19	23	290	110	0.47	0.930	1.65	0.170	0.990
YS3	SwTN	0.529	36.7	0.30	23	168	107	1.11	0.641	1.43	0.159	1.000
ZJ1	SthB	0.531	67.4	0.24	8	137	91	0.50	1.240	1.43	0.150	0.860
ZA3	LtCd	0.535	139	0.18	8	457	274	1.46	1.460	1.67	0.130	0.730
YN2	Lds	0.543	469	0.12	328	488	442	0.22	1.370	2.15	0.160	1.000
ZA1	LtlB	0.561	343	0.19	8	472	137	0.68	1.040	2.45	0.100	0.830
YR2	RgdH	0.570	399	0.11	27	122	76	0.21	0.740	1.68	0.320	0.960
ZG3	SmLm	0.583	115	0.21	2	137	91	0.34	1.550	1.62	0.120	0.920
ZG4	Rtl	0.583	42.7	0.23	46	152	122	1.10	1.620	1.53	0.180	0.920
ZH2	CbC	0.597	43.3	0.39	53	152	168	0.59	1.110	1.66	0.099	0.920
YC1	Tor	0.602	624	0.08	8	488	309	1.01	0.755	1.45	0.167	0.990
YK5	Shld	0.606	391	0.10	110	488	290	1.07	0.191	1.98	0.172	0.940
YL1	UpHm	0.624	2110	0.06	23	701	183	0.40	0.786	1.56	0.115	0.750
YK2	LewB	0.632	470	0.12	137	655	351	0.59	0.627	2.32	0.161	1.000
YJ1	Hrys	0.644	640	0.09	15	541	274	0.35	1.120	1.81	0.142	0.750
ZN1	NwNP	0.645	53.3	0.27	114	213	183	0.61	1.089	2.06	0.126	1.000
ZL3	SpCv	0.648	10.8	0.65	76	168	170	1.25	1.090	1.36	0.090	1.000
ZH1	PH	0.673	764	0.07	38	244	213	0.35	0.709	1.67	0.659	0.910
ZM9	SlCv	0.676	53.6	0.28	35	168	168	0.62	1.130	1.37	0.130	1.000
YK4	Hnds	0.679	529	0.09	198	518	290	0.32	0.637	1.78	0.355	0.950
YR3	IndB	0.679	554	0.09	8	137	107	0.22	0.680	1.72	0.310	0.970
ZK1	Rky	0.682	300	0.15	15	168	60	0.50	1.005	2.00	0.113	0.550
YQ1	GdBC	0.697	4444	0.03	23	305	168	0.14	0.452	2.08	0.172	0.910
ZK2	NB−P	0.722	89.6	0.30	15	213	120	0.55	1.110	1.91	0.300	0.810
YR1	MdlB	0.729	275	0.18	23	198	122	0.32	0.255	1.93	0.245	0.980
YS1	TerN	0.733	1365	0.08	84	290	244	0.12	0.726	2.35	0.327	0.920
ZG2	Tds	0.748	166	0.16	8	229	213	1.35	1.350	1.84	0.130	0.920
YA1	SteG	0.749	306	0.13	12	107	81	0.23	0.540	1.48	0.354	1.000
ZG1	Grn	0.758	205	0.22	11	381	152	0.60	0.547	2.45	0.101	0.960
ZF1	BdN	0.794	1170	0.06	8	274	152	0.29	0.612	2.15	0.236	0.960
ZB1	SmLP	0.808	2640	0.04	183	305	274	0.08	0.360	1.75	0.160	1.000

Table 4.18 Continued

YZS	BASINS	DIST SW	FR BARRN	FR FORST	DIST N	MAR MM
YM3	SwBV	225	0.001	0.899	330	843
ZC2	Grdy	40	0.790	0.170	135	2003
YD1	BvrB	112	0.112	0.806	425	1197
YF1	Cat	123	0.180	0.689	350	1325
ZA2	Hlds	65	0.130	0.820	150	1140
YP1	ShlA	255	0.001	0.869	275	785
YO6	Ptm	222	0.018	0.822	250	768
ZB1	IaM	27	0.782	0.084	120	2093
ZM6	NEP	100	0.038	0.737	115	1165
YD2	NeRk	184	0.001	0.829	425	833
YS3	SwTN	175	0.005	0.836	200	894
ZJ1	SthB	125	0.033	0.817	180	988
ZA3	LtCd	26	0.190	0.680	130	1782
YN2	Lds	89	0.620	0.220	160	1413
ZA1	LtlB	128	0.299	0.601	185	1012
YR2	RgdH	260	0.001	0.679	272	709
ZG3	SmLm	13	0.722	0.158	40	1342
ZG4	Rtl	97	0.470	0.350	90	1559
ZH2	CbC	20	0.496	0.405	140	1356
YC1	Tor	74	0.498	0.335	395	1284
YK5	Shld	266	0.152	0.676	323	855
YL1	UpHm	145	0.145	0.740	300	1229
YK2	LewB	150	0.290	0.549	200	1168
YJ1	Hrys	34	0.069	0.789	215	1302
ZN1	NwNP	38	0.788	0.086	40	1853
ZL3	SpCv	76	0.490	0.420	120	1230
ZH1	PH	68	0.234	0.107	150	1024
ZM9	SlCv	28	0.500	0.370	35	1760
YK4	Hnds	218	0.292	0.353	235	984
YR3	IndB	225	0.001	0.689	245	718
ZK1	Rky	70	0.377	0.510	80	1178
YQ1	GdBC	160	0.068	0.760	205	831
ZK2	NB-P	40	0.230	0.470	75	1571
YR1	MdlB	180	0.008	0.7[illegible]7	220	763
YS1	TerN	115	0.136	0.5[illegible]	180	848
ZG2	Tds	33	0.[illegible]38	0.382	60	1521
YA1	SteG	17	0.002	0.644	445	915
ZG1	Grn	66	0.634	0.265	75	1359
ZF1	BdN	76	0.442	0.322	55	1076
ZE1	SmLP	105	0.490	0.350	165	1026

leeward side of the Great Northern Peninsula, Northeast Brook and Beaver Brook (both near Roddickton), and Cat Arm. Little Codroy, also on the southwest coast, is just outside the group of ten flashiest, probably because it is a little lower.

The annual hydrographs in the three leeward Great Northern Peninsula basins are very peaky. Winter precipitation is usually stored as snow, leading to dry periods in the late winter, long periods of high flows in the spring, occasionally followed by a second dry period in the summer. At Cat Arm the large degree of control by lakes and swamps and the large drainage area are not sufficient to compensate for this effect, especially given Cat Arm's high elevation (it is the highest of all study basins). Very large storage was provided in the hydroelectric project developed in this basin in order to improve the availability.

Shoal Arm and Peters River are the last two rivers in this group. They are moderately small, in a generally dry, cold area. Neither has a large fraction of lakes and swamps. Although Peters River has a large degree of control by lakes and swamps, according to the measure of FACLS, the actual volume of storage available is relatively low.

These observations suggest that regional differences due to climate will be important, probably represented by the elevation and distance variables. As well, one or both of the variables relating to lakes and swamps (FACLS or Fr-Lsw) will likely explain part of the variance among basins.

Multiple Regression Analysis

A multiple regression analysis was carried out, which basically confirmed most of the above observations. Because the scatter plots and correlation matrix had suggested that the relationships were nonlinear, all variables were logarithmically transformed. As with the other flow variables, because geography/climate does play a part in explaining the differences in basin response, a brief analysis of the two major WSC hydrologic regions, Y and Z, was also carried out. The results are presented in Table 4.19 (a) to (c).

For the all-island case, the four most important variables are DA, Dist-N, FACLS and El-Cntrd. The importance of Dist-N, which can be assumed to be a surrogate climate variable, can be clearly seen in the improvement of the r^2 from 0.44 to 0.66 (0.46 to 0.73 in RLS equations). The relative importance of El-Cntrd decreases substantially when Cat Arm is removed.

The results are also presented for the dimensional case, Q_{FD-50}, as a check on the appropriateness of using the non-dimensional variables. The nondimensional rather than the dimensional, flow duration variables are of most interest for

Table 4.19

FD – 50: Summary of Selected Multiple Regression Results

a) One Region
Regression Coefficients

	FD–50				Qfd–50	
Vars all lns	OLS Lns 3 vars	RLS Lns 3 vars	OLS Lns 4 vars	RLS Lns 4 vars		(m³/s) with Eff–P 5 vars
N	40	38	40	37	40	40
Constant	−0.172	−0.074	0.373	−0.069	−4.237	−3.062
DA	0.058	0.055	0.079	0.073	1.047	1.065
FACLS	0.369	0.323	0.372	0.442		0.347
El–Cntrd	−0.117	−0.132	−0.099	−0.005		
Dist–N			−0.148	0.148		−0.165
Dist–SW						−0.077
Eff–P					1.008	0.732
r^2	0.44	0.46	0.66	0.73	0.98	0.99
r^2 adj					0.98	0.99
Std.Er.	0.16	0.14	0.13	0.11	0.19	0.13
						dimnsl
Outliers	Ptrs SwBV		SteG Cat Rky			

Table 4.19 Continued

b) Hydrologic Region Y

Regression Coefficients					
Vars all lns	OLS Same vars as (a) 4 vars	OLS Best 2 vars	OLS Alt 4 vars	OLS All sig! 9 vars	
N	20	20	20	20	
Constant	0.468	−0.482	0.492	0.826	
DA	0.110	0.078		0.200	*
FACLS	0.474				
El−Cntrd	−0.140				
Dist−N	−0.150				
Fr−Forst			−0.189	−0.334	
El−Gauge			−0.084		
Fr−LSw		0.317	0.423	0.274	
LMC−sp				0.272	*
El−Dvide			0.182	0.173	*
Fr−Barrn				−0.032	*
Dr−Dens				−0.207	
Dist−SW				−0.107	
Shp				−0.538	
r^2	0.640	0.699	0.83	0.95	
r^2 adj	0.640	0.664	0.78	0.91	
Std.Er.	0.110	0.120	0.10	0.06	

* – tol<.2

Table 4.19 Continued

c) Hydrologic Region Z

Regression Coefficients				
Vars all lns	OLS Same vars as (a) 4 vars	OLS Alt 4 vars	OLS 5 vars	OLS 6 vars
N	20	20	20	20
Constant	−0.022	−0.100	−0.106	−0.258
DA	0.038	0.075	0.084	0.180
FACLS	0.393	0.398	0.397	0.414
El−Cntrd	−0.063			
Dist−N	−0.086	−0.132	−0.158	−0.163
Fr−Forst		0.064	0.075	0.068
El−Gauge			0.030	0.044
Fr−LSw				
LMC−sp				0.207
r^2	0.750	0.770	0.81	0.85
r^2 adj	0.680	0.710	0.74	0.78
Std.Er.	0.110	0.100	0.10	0.09

prediction, however, because the results are most likely to be use in selecting a gauge to use as a pattern for daily flows at an ungauged site. MR equations are not likely to be used directly because point FD values are not often required.

Hydrotechnical engineers (or fisheries biologists or others), when presented with a design problem related to the normal range of flows, at an ungauged site, will probably follow a procedure similar to the following.

- Planimeter drainage area.
- Estimate Eff-P, using equations and data from adjacent gauged sites.
- Obtain basin characteristics for ungauged site shown in this present study to be important in explaining availability.
- Examine the flow duration curve and daily flow records for gauged sites with similar characteristics.
- Select a gauged basin for patterning the daily flows.
- Construct a synthetic dimensional flow duration curve or daily flow series by prorating the flow duration curve or daily flow values from the gauged sites by DA and Eff-P.

Table 4-19a suggests that this approach is suitable. DA and Eff-P are the two most important variables in estimating Q_{FD-50}, explaining 98 percent of the variance. The two variable equation is of the form

$$Q_{FD-50} = DA^{1.05}Eff\text{-}P^{1.01}$$

Q_{FD-50} is almost directly proportional to the average annual flow. When other variables are added, the coefficients change, especially for Eff-P. Since Eff-P has been shown to be related to Dist-SW and El-Centrd, it is not surprising that the results are different for these two variables in the dimensional and nondimensional cases. The coefficient for FACLS is very similar in magnitude in the two cases since it is unrelated to either DA or Eff-P. The additional MR analysis was therefore done using nondimensional FD variables.

Some results assuming a regional division along the Y-Z boundary are presented in Table 4-19(b) and (c). The first column in each table shows the results using the same four variables as in the all-island case. The regional subdivision improves the unadjusted r^2 in both regions, from 0.44 to 0.64 in the Y region, and to 0.75 in the Z region.

More improvement can be obtained using other variables, especially in the Y region. There, Fr-Lsw is more important than FACLS (which matches the observations made using the sorted data set). One of the best four-variable equations does not use DA; rather it uses two elevation variables (El-Gauge and El-Dvide), Fr-Forst, and Fr-Lsw. The use of El-Dvide with a positive sign needs further examination before it can be accepted; as discussed in the low flow section above, one would ordinarily expect a high El-Dvide to indicate a greater proportion of snow, and thus less availability, rather than more.

As frequently happens in a MR analysis, apparently very good and statistically significant results can be obtained using many variables. (All coefficients have acceptable p values and tolerances.) An example here is the nine variable equation for the Y region. With the exception of El-Dvide, all the variables are hydrologically reasonable and the coefficients make sense. The low tolerances, while still acceptable, suggest that a more sensible equation could be developed selecting fewer variables.

In the Z region, DA, Dist-N, and FACLS are always important, as they are in the all-island case. Slight improvements can be made with minor changes and by incorporating other variables. If El-Cntrd is dropped and Fr-Forst and El-Gauge are used instead, the adjusted r^2 increases about eight percent. The problem with using El-Gauge is similar to that of using El-Dvide in the Y region - the sign is opposite to that expected.

4.7 Summary of Important Basin Characteristics

Table 4-20 presents the important basin characteristics related to each flow variables, selected from the analysis of each individual flow variable. The associated coefficients (rounded for ease of comparison) are also given, as an illustration for possible weighings in cluster or discriminant analysis. This table shows that DA is always important. For high flows El-Cntrd and FACLS are also very important, as well as other distance/topographic variables (Dist-N, Dist-SW, and/or Fr-Barrn). DrDens seems to be more important in the separate Y-Z equations. For low flows and availability, FACLS and the distance variables are important.

The basic Y-Z geographic division did lead to improved relationships for most flow variables. In Chapter 5, the possibility of further improvements using other methods of regionalization is addressed. The high flow variable Qavgfld is used for this detailed analysis. A similar procedure could be followed for the other flow measures.

Table 4.20

Coefficients of Most Important Variables (without Eff–P)

	One Region			Region Y			Region Z		
	High	Low	Availa–bility	High	Low	Availa–bility	High	Low	Availa–bility
Var (All Ins)	QavgFld	Qlow	Qfd–50	QavgFld	Qlow	Qfd–50	QavgFld	Qlow	Qfd–50
DA	0.7	1.2	1.0	0.9	1.2	1.1	0.7	1.1	1.0
DrDens	x			0.3	x		0.5		
Dist–N	x	–0.3	–0.2	x	x	x	x	–0.7	–0.2
Dist–SW		–0.3	–0.2	x	–0.4	–0.2			
El–Dvide		0.4				x		0.8	
El–Gauge	x					x	x		x
El–Cntrd	0.3			0.3			0.3		
FACLS	–1.0	0.9	0.2	x	1.6		–0.8	0.7	x
Fr–Farm	0.1		0.1			0.1			0.1
Fr–Forst	x			x					x
Fr–LSw	x			x	x	x			
LMC–sp	x			x					
Shape							x		
Slope	x						x		

Based on ordinary least squares analysis with logarithmetically transformed variables
x = variable sig. in other eqn's e.g., for specific runoff or non dim. values

5 Analysis of Regional Subdivisions

The key basin characteristics associated with each of the flow variables (high, low, and available flow) were identified in Chapter 4. The present chapter addresses

- the use of these basin characteristics to identify clusters in basin data-space;
- the advantages, if any, of using these clusters over clusters in geographic space to develop predictive equations for ungauged basins.

This detailed analysis is carried out for the variable Qavgfld, the average of the annual maximum daily flow series. A similar analysis could be carried out using the low flow or available flow data. Qavgfld was chosen because it is frequently used as the index flood in the index flood method of regional flood frequency analysis. Much of the error in this method arises from the uncertainty in estimating the index flood. Clustering has the potential to improve the estimates of the average flood at ungauged basins.

Section 5.1 describes the application of cluster analysis techniques to the basin characteristic data to develop potential clusters in basin dataspace. Section 5.2 then briefly describes the development of equations for each of the candidate regions, and Sections 5.3 and 5.4 provide an assessment of the regionalization methods. An alternative approach is suggested and evaluated in Sections 5.5 and 5.6. Section 5.7 provides comments and a suggested procedure.

5.1 Cluster Analysis

One of the purposes of this study was to compare the grouping of basins by geography with clustering in a dataspace of basin physiographic characteristics. Grouping basins in dataspace assumes that basins separated in geographic space respond similarly to hydrologic events if they have similar physiographic characteristics. In this study, the grouping of basins in dataspace was carried out using the techniques of cluster analysis, as provided in the statistical analysis package SYSTAT (Wilkinson, 1990).

The results of a cluster analysis are very sensitive to the variables selected and their weighting (and indeed, in the case of some cluster techniques, such as Andrews' Fourier plots, to their order), so some thought must be put into their selection. For this study, weighted standardized logarithmically transformed variables were used. The variables selected and their associated weights were taken from the best all-island equation presented in Table 4-20 for Qavgfld. These are

Variable	Coefficient
DA	0.7
FACLS	-1.0
El-Cntrd	0.3
Fr-Barrn	0.1

Cluster analysis results can also be sensitive to the method of clustering, and to the distance criterion and linkage method if hierarchical clustering is used, so several alternatives were tried. The recovery of inherent clusters, if they exist, can be a complex problem, and the present analysis is not intended to provide a detailed comparison of cluster methods. Rather, it attempts to select the methods most likely to be useful and to apply them in a reasonable way, to assess whether cluster analysis shows any promise for improving flow estimates.

The two basic methods of clustering are hierarchical and nonhierarchical. Hierarchical clustering assumes an inherent tree-like structure in the data. Nonhierarchical clustering assumes that the data are spread out in dataspace and can be partitioned. There is no physical reason why basins should group according to a hierarchical or tree-like structure, so clusters obtained by nonhierarchical partitioning might be expected to be more appropriate in this application. Hierarchical clustering has been used successfully in other similar applications, however, and in this study the hierarchical clusters provided a good match to the flow clusters, so both clustering methods were considered.

5.1.1 Nonhierarchical Clustering

The nonhierarchical partitioning proceeded by picking seed cases for the number of clusters specified, spread apart from the centre as much as possible. Cases were reassigned until within-groups sums of squares were minimized. There is no standard objective procedure for selecting the optimum number of clusters, although various stopping rules have been proposed; the recommended procedure is to try several numbers of clusters, and use judgment and knowledge of the problem at hand to select an appropriate number. For this study, reasonable results were obtained when four clusters were specified.

Of the four clusters, two were very small, one with only one member and one with four. These small clusters are distinguished by the small fractions of the basins controlled by lakes and swamps. Similar very small clusters resulted regardless of the number of clusters specified.

Table 5.1 shows the statistical profile of basin characteristics for each nonhierarchical cluster, in terms of the original untransformed variables. Cluster 1, with 24 members, contains most of the large basins; the smallest is 139 km^2. Three other basins larger than 139 km^2 (Isle aux Morts, Grandy and Rocky) are excluded from this group, and are assigned to Clusters 2 or 4, because of their low FACLS. Cluster 2 contains all the basins with FACLS less than 0.73, except Grandy, which is placed alone in Cluster 4. Grandy is unusual because it has the lowest FACLS (0.34) of all the study basins, combined with a relatively high elevation (335 m) and a high Fr-Barrn (0.79). Cluster 3, with 11 members, has

Table 5.1

Statistical Profiles of Nonhierarchical Clusters

Cluster Number	N		DA km²	FACLS fraction	El-Cntrd m	Fr-Barrn fraction
1	24	Min	139	0.73	76	0.00
		Max	4444	1.00	549	0.63
		Mean	801	0.92	229	0.22
		Med	470	0.96	213	0.17
2	4	Min	72	0.43	60	0.00
		Max	300	0.60	335	0.78
		Mean	168	0.54	198	0.32
		Med	149	0.56	198	0.25
3	11	Min	3.63	0.79	91	0.00
		Max	115	1.00	183	0.79
		Mean	52.7	0.93	140	0.34
		Med	53.3	0.92	152	0.47
4	1	Min	230	0.34	335	0.79
		Max	230	0.34	335	0.79
		Mean	230	0.34	335	0.79
		Med	230	0.34	335	0.79

all the basins with drainage areas less than 115 km^2, except Highlands and Southwest River near Baie Verte, which are in Cluster 2 due to their low FACLS.

5.1.2 Hierarchical Clustering

The hierarchical clustering started by combining basins into small clusters, and then progressively agglomerated the smaller clusters into larger ones. (This procedure means that once a basin is assigned to a cluster it can never be reassigned. Undesirable early combinations may persist.)

Four linkage methods were tried, single, complete, average and centroid. As expected, single linkage, in which each member is closer to one other member of its cluster than to any other, produced long stringy clusters. Complete and average linkage methods resulted in similar compact clusters, in which each member is more like every other member. A Euclidean distance measure was used for all linkage methods, consistent with the type of data.

Of the four methods, complete linkage with four clusters produced reasonable groupings, and these were selected for comparison with the nonhierarchical clusters and geographic groups.

Table 5.2 shows the statistical profiles for the resulting clusters in terms of the original untransformed variables. Cluster 1, like Cluster 1 from the nonhierarchical procedure, contains the larger, naturally well regulated basins. It does not include two higher basins in the midsize range (Garnish and Tides Brook), as well as six of the less barren basins in the same midsize range. These eight basins had been included in the otherwise equivalent nonhierarchical Cluster 1, so the total number of basins in Cluster 1 is reduced from 24 to 16. The eight basins are all assigned to hierarchical Cluster 3, resulting in a larger size range in Cluster 3 than was the case for the nonhierarchical Cluster 3. The five basins with the low FACLS which were divided between Clusters 2 and 4 using the nonhierarchical methods are all in Cluster 2. Cluster 4 consists of only the two smallest basins.

Table 5.3 lists all the basins and their cluster assignments by the two methods. The cluster assignment based on the flow characteristic Qavgfld (again standardized log transforms) is also shown for comparison, clustered by partitioning. Just over half (21) of the 40 basins have identical cluster assignments in all three cluster sets. An additional 16 have similar assignments in two of the three cluster sets. Two of the remaining three basins are very small (Northeast Pond and Spout Cove), and the last one, Grandy, has the lowest FACLS. These unusual basins are treated differently in each clustering method.

Table 5.2

Statistical Profiles of Hierarchical Clusters

Cluster Number	N		DA km²	FACLS fraction	El–Cntrd m	Fr–Barrn fraction
1	16	Min	139	0.73	137	0.07
		Max	4444	1.00	549	0.61
		Mean	1059	0.90	281	0.26
		Med	618	0.93	274	0.21
2	5	Min	72	0.34	60	0.00
		Max	300	0.60	335	0.79
		Mean	180	0.50	225	0.42
		Med	205	0.55	244	0.55
3	17	Min	36.7	0.79	76	0.00
		Max	554	1.00	213	0.79
		Mean	167	0.94	129	0.26
		Med	115	0.96	122	0.03
4	2	Min	3.63	1.00	168	0.04
		Max	10.8	1.00	170	0.49
		Mean	7.22	1.00	169	0.26
		Med	7.22	1.00	169	0.26

Table 5.3

Cluster Assignments of Study Basins

Y-Z Region	Basin ID	Nonhier-archical Cluster	Hier-archical Cluster	Flow Cluster
YA1	SteG	1	3	3
YC1	Tor	1	1	1
YD1	BvrB	1	1	1
YD2	NeRk	1	3	3
YF1	Cat	1	1	1
YJ1	Hrys	1	1	1
YK2	LewB	1	1	1
YK4	Hnds	1	1	1
YK5	Shfd	1	1	3
YL1	UpHm	1	1	4
YM3	SwBV	2	2	3
YN2	Lds	1	1	1
YO6	Ptrs	1	3	3
YP1	ShlA	3	3	3
YQ1	GdBC	1	1	4
YR1	MdlB	1	3	3
YR2	RgdH	1	3	3
YR3	IndB	1	3	3
YS1	TerN	1	1	1
YS3	SwTN	3	3	2
ZA1	LtlB	1	1	1
ZA2	Hlds	2	2	3
ZA3	LtCd	1	1	1
ZB1	IaM	2	2	1
ZC2	Grdy	4	2	1
ZE1	SmLP	1	1	1
ZF1	BdN	1	1	1
ZG1	Grn	1	3	3
ZG2	Tds	1	3	3
ZG3	SmLm	3	3	3
ZG4	Rtl	3	3	3
ZH1	PH	1	1	1
ZH2	CbC	3	3	3
ZJ1	SthB	3	3	3
ZK1	Rky	2	2	1
ZK2	NE-P	3	3	3
ZL3	SpCv	3	4	2
ZM6	NEP	3	4	2
ZM9	SlCv	3	3	3
ZN1	NwNP	3	3	3

Note: Shading indicates same cluster assignment.

5.1.3 Andrews Fourier Plots

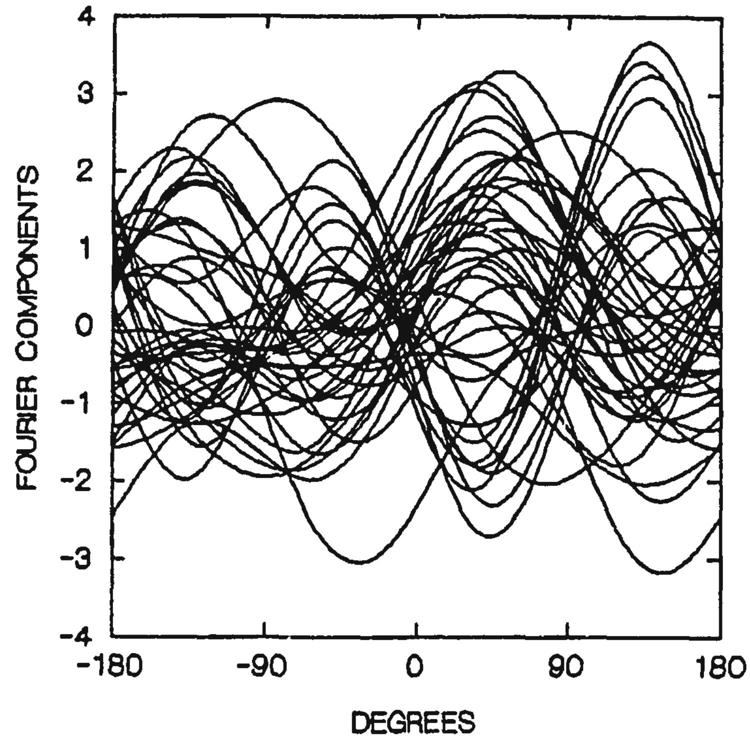
An alternative method of clustering is to use Andrews Fourier plots, which provide a technique for graphical representation of multivariate data (Andrews, 1972). In these plots, each basin characteristic (e.g., DA, FACLS) is represented as a term in a p-dimensional sine-cosine function. When the expression is plotted, the resulting curve becomes a visual representation of the combined basin characteristics. Basins having similar Andrews Fourier plots can be considered as a group. The group can be identified by its typical signature, and ungauged basins can be assigned to the group whose signature matches the plot for the new basin most closely.

Andrews Fourier plots are probably most useful after cluster analysis is completed, perhaps to refine the clusters slightly, and to prepare characteristic cluster signatures in preparation for assigning new basins, so their use in this study is perhaps somewhat premature. Nevertheless, the results of the analysis can be helpful for comparison with other results, and for suggesting future directions.

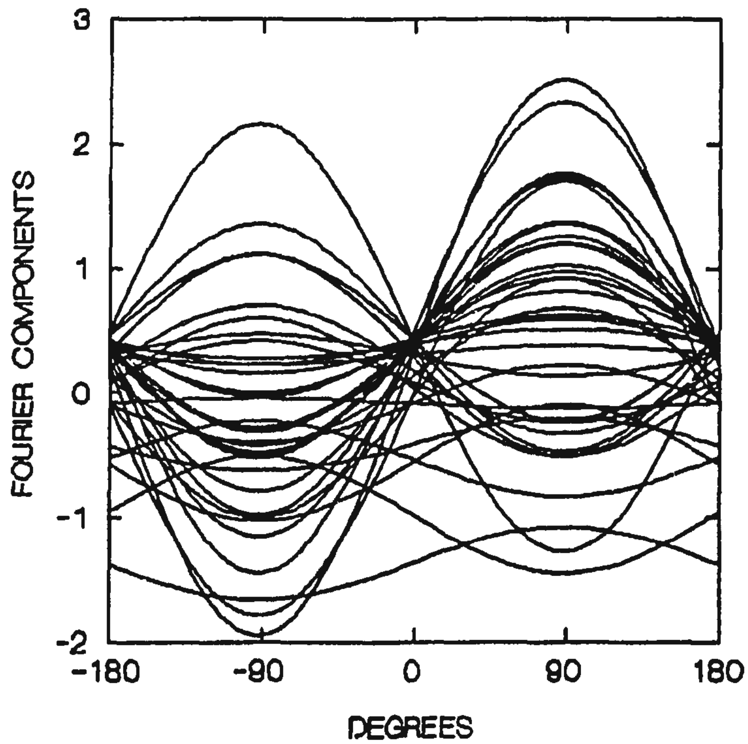
Andrews Fourier plots are sensitive not only to the variables selected but to their order. For this study, the plots were prepared using a standardized data set, and the variables were specified in the order corresponding to the magnitude of the coefficients in the equation in Table 4.20, i.e., FACLS, DA, El-Cntrd, and Fr-Barm. This approach corresponds approximately to the approach taken in developing the clusters using cluster analysis techniques. Since four variables lead to somewhat complex curves, a simpler set of plots using only FACLS and DA was also prepared. Figure 5.1 shows the plots for the combined set of study basins for the two approaches.

In Figure 5.1(a), for the four variable case, groups of basins can be distinguished. The most obvious visual distinction is between the basins with opposite peaks and valleys. The majority have their peaks at approximately -135 degrees and +45 degrees; about eight basins have the reverse pattern. Six of these eight basins are ones that were assigned to Cluster 1 in the nonhierarchical clustering, and to Cluster 3 in the hierarchical clusters. In neither method were they given a cluster of their own. The chief visual distinguishing characteristic between Clusters 1 and 3 in the Andrews Fourier plots is that the Cluster 1 basins have a low first peak and a high second peak. The reverse is true for the Cluster 3 basins. The basins in Clusters 2 and 4 are the odd ones.

Figure 5.1(b) is somewhat simpler, since only two characteristics are represented, FACLS and DA. The Cluster 1 basins in general have their minima at approximately -90 degrees and their maxima at +90 degrees, while the Cluster



(a) Variables FACLS, DA, El–Cntrd, Fr–Barrn



(b) Variables FACLS, DA

Fig. 5.1 – Andrews Fourier Plots

2 and 3 basins have the reverse pattern. The Cluster 2 basins are distinguished from the Cluster 3 basins by being much lower (more negative). As a comparison with the cluster analysis methods, the Andrews Fourier plots based on two variables were grouped visually. The groups corresponded exactly to the nonhierarchical clusters, with the exceptions of the two very small basins, and the two very large basins, which stood out because of the great difference in amplitude of the sine waves. There was also the possibility of further subdivision of Cluster 1.

5.2 Development of Regional Equations

The next step was to develop regression equations for each of the potential regions. The analysis proceeded as follows.

1. Select regions.
2. Develop candidate equations using ordinary least squares multiple regression techniques.
3. Assess results, select final candidate models.
4. Develop final candidate models (regression equations) by
    - identifying outliers using robust regression techniques;
    - refining equations using non-linear least squares;
    - estimating uncertainty and error.

This analysis is described below. The assessment of the regionalization methods follows in Section 5.3.

5.2.1 Candidate Regional Subdivisions

Regions in both geographic and basin characteristic data were considered, as follows.

A. Geographic

- no subdivision (one region);
- WSC hydrologic subregions Y and Z;
- four subregions A to D, as used in the provincial regional flood frequency analysis (boundaries shown in Appendix A).

B. Basin characteristic dataspace

- size (greater than and less than 130 km^2);
- three hierarchical clusters;
- three nonhierarchical clusters (2 and 4 combined equals the second hierarchical cluster);
- four nonhierarchical clusters (2 and 4 separated).

Hierarchical cluster 2 had the identical membership (five members) as nonhierarchical clusters 2 and 4 combined. The hierarchical cluster with the two very small basins was excluded. Because two of the clusters were identical, six clusters were analyzed in total.

5.2.2 Development of Equations for Candidate Regionalization Methods

In Chapter 4, preliminary regression equations were developed for two of the geographical subdivisions, the one-region case and the Y-Z division. Equations with and without the independent variable effective precipitation (Eff-P) were also presented in that chapter. The regression analysis procedures and criteria described in that chapter (Section 4.3) were also applied here.

A similar analysis was also carried out for the other regions obtained using the candidate regionalization methods based on geography and basin characteristics. The most promising results are presented in Table 5.4, a-c. The coefficients associated with all explanatory variables presented in this table are significant ($p < 0.05$) and none introduce problems of multicollinearity (all tolerances are well above the minimum value of 0.1).

Table 5.4

Results of Regional Multiple Analysis

(a) One Region

	OLS Lns 2vars with Eff−P	RLS Lns 2vars	OLS Lns 3 vars with Eff−P	RLS Lns 3 vars	OLS Lns 3 vars no Eff−P	OLS Lns 4 vars	RLS Lns 4 vars	Nonlin LS with Eff−P	Nonlin LS	Nonlin LS
N	40	39	40	37	40	40	38	40	39	40
Constant	−0.362	−0.340	−0.432	−0.432	0.288	−1.024	−1.064	0.489	0.519	1.062
DA	0.813	0.814	0.807	0.820	0.742	0.720	0.738	0.862	0.855	0.748
DrDens										
Dist−SW										
Dst−N										
El−Centrd						0.264	0.260			
El−Divde										
FACLS			−0.783	−0.705	−1.065	−1.014	0.949			−1.339
Fr−Barrn					0.113	0.085	0.083			0.038
Fr−LSW										
LMC										
Shp										
Fr−Forst										
Eff−P	1.297	1.247	1.095	1.002				1.756	1.726	
r^2	0.94	0.95	0.96	0.97	0.94	0.95		0.95	0.95	0.94
r^2 adj/corr			0.96			0.95		0.91	0.91	0.89
Std.Er.	0.31	0.28	0.24	0.20	0.30					
Outliers		SteG		SteG SmLP IndB	None		MdIB SmLP		SteG	None

Table 5.4 Continued

(b) Geographic Regions

	WSC Region				Provincial Region						
	Y		Z		A	B	C	C	D	D	D
	OLS	OLS	OLS	RLS	OLS	OLS	OLS	OLS	OLS	OLS	RLS
	Lns	Lns	Lns	Lns	Lns	Lns	Lns	Lns	Lns	Lns	Lns
		with Eff−P					with Eff−P		with Eff−P		
N	20	20	20	19	12	12	10	10	6	6	5
Constant	−2.770	−0.320	1.067	1.363	−0.097	−0.531	0.673	−3.028	0.491	2.032	0.625
DA	0.890	0.778	0.749	0.752	0.823	0.849	0.640	1.025	0.694		
DrDens						0.608					
Dist−SW			−0.247	−0.319							
Dst−N											
El−Centrd											
El−Divde											
FACLS		−0.814	−1.009	−0.962							
Fr−Barrn											
Fr−LSW	−1.003							−0.821			
LMC										−1.515	−2.387
Shp											
Fr−Forst	−0.499										
Eff−P		1.008					2.295		1.409		
r^2	0.96	0.98	0.96	0.97	0.98	0.96	0.89	0.94	0.97	0.78	0.98
r^2 adj/corr	0.96	0.98	0.96		0.98	0.95	0.86	0.93	0.95		
Std.Er.	0.22	0.19	0.26	0.25	0.18	0.28	0.35	0.25	0.14	0.34	0.11
Outliers				SmLn	None	None		None	None		Harry's

Table 5.4 Continued

(c) Basin Characteristic Regions/Clusters

	Size					Clusters Hierarchical		Non-hierarchical			
	>130 km²		130<DA<1400 km²		<130 km²	H1	H3	NH1	NH3	NH2+4	NH2+4
	OLS Lns	RLS Lns	OLS Lns	RLS Lns	OLS Lns	OLS Lns	OLS Lns	OLS Lns	OLS Lns	OLS Lns	OLS Lns
N	27	26	24	23	13	16	17	24	11	3	5
Constant	−2.664	−2.526	−2.270	−2.888	−0.224	−1.023	0.653	−2.905	−0.[illegible]15	−1.451	−5.539
DA	0.758	0.673	0.660	0.757	0.884	0.696	0.651	0.760	0.736	1.180	1.362
DrDens	0.343	0.317	0.332	0.322			0.406	0.327			
Dist−SW						−0.344					
Dst−N											
El−Centrd	0.510	0.569	0.545	0.523				0.544			0.604
El−Divde						0.532					
FACLS	−1.261	−1.609	−1.152	−1.760				−1.625			
Fr−Barrn					0.563		0.063				
Fr−LSW											
LMC											
Shp									1.309		
Fr−Forst											
Eff−P											
r^2	0.91	0.886	0.874	0.924	0.953	0.85	0.848	0.924	0.956	0.745	0.995
r^2 adj/corr	0.89		0.847		0.944	0.81	0.812	0.908	0.945	0.660	0.990
Std.Er.	0.26	0.241		0.240	0.205	0.27	0.211	0.252	0.208	0.491	0.086
Outliers		Grndy		Grndy	None			None	None	Hlds Rky	

Notes:
1. Hierarchical cluster H2 = Nonhierarchical NH2+NH4
2. Hierarchical cluster H4 consists of the two smallest basins, and is not included.
3. Reweighted least squares not done for NH3 because too few cases.

The number of candidate models which could be used to estimate Qavgfld at an ungauged site is potentially quite large. There are at least 6 possible regions into which a basin can be assigned, and within each region, there are several alternative multiple regression equations to be evaluated. A preliminary screening was therefore carried out to reduce the number of candidate models. The screening had two principal objectives

- to determine whether some of the regions could be eliminated;
- to select only one preferred equation for each region.

After review of the regression results, the only type of regional grouping which was excluded from further analysis at this point was that based on the hierarchical clustering. The regression results were better when based on the nonhierarchical clustering, and for the general purpose of comparing regions in dataspace rather than regions in geographic space, it seemed overrefined at this point to have clusters based on two separate cluster techniques.

The equations were reduced to include only one in each region. The final equations were selected based on

- the fit of the equation (as measured by adjusted r^2 and standard error);
- the overall hydrologic sense of the independent variables, and

- the ratio of the number of cases used to develop the equation to the number of independent variables in the equation. This ratio was kept in the range of 5-10, so that all equations could be compared on roughly the same footing. In the case of provincial region A, drainage area is so important that no additional variable has an acceptable p value. The ratio of the number of cases to the number of variables is therefore slightly higher.

The number of outliers as identified in robust regression was also considered. An equation with no outliers was preferred over one with several outliers, other considerations being approximately equal.

The selection of the final equations was complicated in this study by the fact that one of the most important explanatory variables in some regions is Eff-P. This variable is estimated, not measured like the basin physiographic characteristics. Two approaches were considered for handling Eff-P

1) replacing Eff-P with topographic and distance variables, either directly or using one of the equations developed in Chapter 4 for Eff-P;
2) including Eff-P as an independent variable, and using a first order uncertainty analysis to assess whether the improvement in the

multiple regression results using Eff-P is sufficient to overcome the greater uncertainty in the estimate of Eff-P at ungauged basins.

The one-region case was used for this analysis. Similar results would be expected in other regions since the orders of magnitude of the coefficients and their relative importance tend to be similar.

Substitution for Eff-P

Since effective precipitation can be related to topographic and geographic variables, as described in Section 4.3, the possibility of incorporating these relationships directly into an expression for Qavgfld was explored. The simplest approach is to include the relevant variables directly in the nonlinear equation for Qavgfld, as was done with the log-linear results below, taken for the one region case from Table 4.15, OLS with n=40.

Variable	With Eff-P	Without Eff-P
Constant	-0.432	-1.024
DA	0.807	0.720
El-Cntrd		0.264
FACLS	-0.783	-1.014
Fr-Barm		0.085
Eff-P	1.095	
r^2adj	0.956	0.945
Std. Error	0.24	0.27

Here, Fr-Barrn and El-Centrd are acting as surrogates for Eff-P; in addition, Fr-Barrn may also be representing the characteristic of a basin with a large fraction of barren area to respond with higher flood flows (although Fr-Barrn was never a significant explanatory variable when Eff-P was included). Even with these two variables acting as surrogates for Eff-P, the results are not quite as good.

Various alternative models were developed and tested, incorporating the variables as linear and nonlinear combinations within an overall nonlinear expression for Qavgfld. None of the results were as good as the simple approach presented above, so these results, and similar results for the regions, were used for the remainder of the analysis.

Uncertainty Analysis

Although the results presented above and in Table 5.4 indicate that equations with Eff-P are better, the improvement as indicated by the r^2 and standard error is slight. Unlike basin physiographic characteristics, Eff-P cannot be measured at ungauged basins but must be estimated. In order to assess whether the slight improvement in the regression results using Eff-P is sufficient to overcome the greater uncertainty in its estimate at ungauged basins, an uncertainty analysis was carried out.

The equations presented in Table 5.4(a) for the one region case were used for the analysis. Nonlinear equations were developed for the two variable equation with DA and Eff-P, and the three variable equation with DA, FACLS and Fr-Barm. El-Cntrd was dropped because in the nonlinear analysis the sign of the coefficient was wrong. (The equation without El-Cntrd was also slightly preferable because no outliers were identified in the robust regression.)

The analysis followed the procedures outlined in Chow et al (1989) for first order uncertainty analysis. Briefly, if Y is the dependent variable, and $x_1, x_2, \ldots, x_n$ are the independent variables, with $a_1, a_2, \ldots, a_n$ their associated coefficients, then

$$CV_y^2 = CV_{x1}^2 + CV_{x2}^2 + \ldots + CV_{xn}^2$$

CV_y is then a measure of the uncertainty of the estimated dependent variable due to the uncertainty of measuring or estimating the independent variables.

The CV's for each independent variable were estimated from experience in measuring the physiographic characteristics from mapping and in estimating the Eff-P considering basin specific topography, location, and available data from adjacent gauged basins. The values used and the results are presented in Table 5.5.

This estimate is reasonable compared with the standard errors of prediction based on the log-linear equations. These average just under 10 percent (about 125 to

140 mm) for both the two and three variable equations, estimated for the 40 basins in the data set plus 17 short record basins. (These may be slightly underestimated due to the fact that the regression was carried out in the log domain). They are also consistent with the standard errors of the estimate for the nonlinear regression equations, which are slightly higher (about 180 mm). With an average Eff-P for the study basins of 1200 mm, an average CV of ten percent is equivalent to an average range of 240 mm.

Table 5.5
Uncertainty Analysis

With Eff-P Variable	Est. CV	Coef	Without Eff-P Variable	Est. CV	Coef
DA	0.03	0.862	DA	0.03	0.748
Eff-P	0.10	1.756	FACLS	0.05	-1.339
			Fr-Barm	0.10	0.038
CV_y		0.177			0.071

This table indicates that although the regression results are better for the equation with Eff-P, overall the uncertainty is less using the three variable equation without Eff-P. This results occurs not only because of the uncertainty in estimating Eff-P, but also because the coefficient associated with it is so high. Because the magnitude of the coefficients is similar in the equations for all regions where Eff-P is significant, the results would be similar, and the uncertainty analysis was not repeated.

The equations without Eff-P were therefore chosen in preference to those with Eff-P in other regions. As Table 5.4 showed, however, in several regions, equally good or better equations could be developed without Eff-P, so those results are unaffected.

5.2.3 Development of Equations for Subregions

Based on the results of the preliminary screening, the best equation from the ordinary least squares (OLS) regression analysis without Eff-P was selected for final development. The same regression analysis procedures and criteria described previously in Section 4.3 were followed.

The only exception to the general guideline regarding outliers was the cluster consisting of the four basins assigned to nonhierarchical Cluster 4 plus the single basin assigned to Cluster 2. (These are the same five basins that make up hierarchical Cluster 2.) With only five cases, an equation with only one independent variable was preferred, but no matter which independent variable was tried, there were always two (usually different) outliers, leaving only three cases for a reweighted least squares or nonlinear analysis. For this region, therefore, the nonlinear least squares equation was developed using drainage area alone with all five cases; an alternate equation was also developed with two independent variables (with a much better fit), again using all five cases.

Table 5.6 presents the constants and exponents of the regression equations resulting from the nonlinear least squares analysis, by geographic or basin characteristic region. These are the equations used to evaluate and select the preferred grouping for regionalization. The results of a first order uncertainty analysis for these equations is also presented in the table.

5.3 Assessment of Regionalization Methods

A group of 17 gauged basins with shorter record lengths was used to assess whether it is advantageous to subdivide the island into regions, and if so, to select the preferred method for regionalization. These 17 basins have record lengths of seven to nine years, a period sufficiently long to obtain a reasonable estimate of Qavgfld. They were treated as ungauged basins, and the estimates made from the regression equations were compared with the average of the observed annual maximum daily flow series. Their locations are shown in Figure 5.2.

Table 5.7 lists these basins and their relevant physiographic characteristics. The boxplots in Appendix B compare their characteristics with those of the study basins. In general they are smaller (15 of the 17 are less than 130 km^2), with smaller fractions of the basins controlled by lakes and swamps, and smaller fractions of barren area (conversely more

Table 5.6

Nonlinear Least Squares Regression Results

(a) Geographic Regions

	All Island	WSC Y	Z	Prov A	B	C	D
N	40	20	19	12	12	10	5
Constant	1.082	0.140	12.160	1.103	0.117	0.241	2.529
DA	0.748	0.761	0.756	0.783	1.092	0.841	
DrDens					0.826		
Dist−SW			−0.600				
El−Centrd							
FACLS	−1.339		−0.970				
Fr−Barrn	0.038						
Fr−LSW		−1.079				−0.627	
LMC−Sp							−2.224
Shp							
Fr−Forst		−0.373					
r^2	0.94	0.99	0.98	0.989	0.99	0.98	0.99
r^2 corr	0.89	0.79	0.96	0.979	0.98	0.97	0.96
Sum of Resids	75297	13568	5144	692	9542	8353	823
Error est	45.7	29.1	18.5	8.3	32.6	34.5	16.6
N−p	36	16	15	10	9	7	3
Uncertainty	7.1%	11.6%	6.1%	2.3%	5.3%	6.8%	15.6%
Outliers	None	None	SmLm	None	None	None	Hrys

Table 5.6 Continued

(b) Basin Characteristic Regions/Clusters

	Size (km^2)		Clusters			
	<130	>130	1	3	2+4	2+4 Alt
N	13	26	24	11	5	5
Constant	0.568	0	0.067	0.526	2.227	0.00215
DA	0.793	0.834	0.828	0.810	0.773	1.463
DrDens		0.290	0.303			
Dist-SW						
El-Centrd		0.474	0.415			0.614
FACLS		-2.058	-1.998			
Fr-Barrn						
Fr-LSW						
LMC						
Shp	1.208			1.177		
Fr-Forst						
r^2	0.98	0.967	0.966	0.981	0.850	0.997
r^2 adj/corr	0.92	0.926	0.927	0.924	0.481	0.989
Sum of Resids	185	38023	37403	139	14700	302
Error est	4.3	42.6	44.4	4.2	70.0	12.3
N-p	10	21	19	8	3	2
Uncertainty	8.8%	12.8%	12.1%	8.6%	2.3%	10.2%
Outliers	None	Grndy	None	None	2 incld	

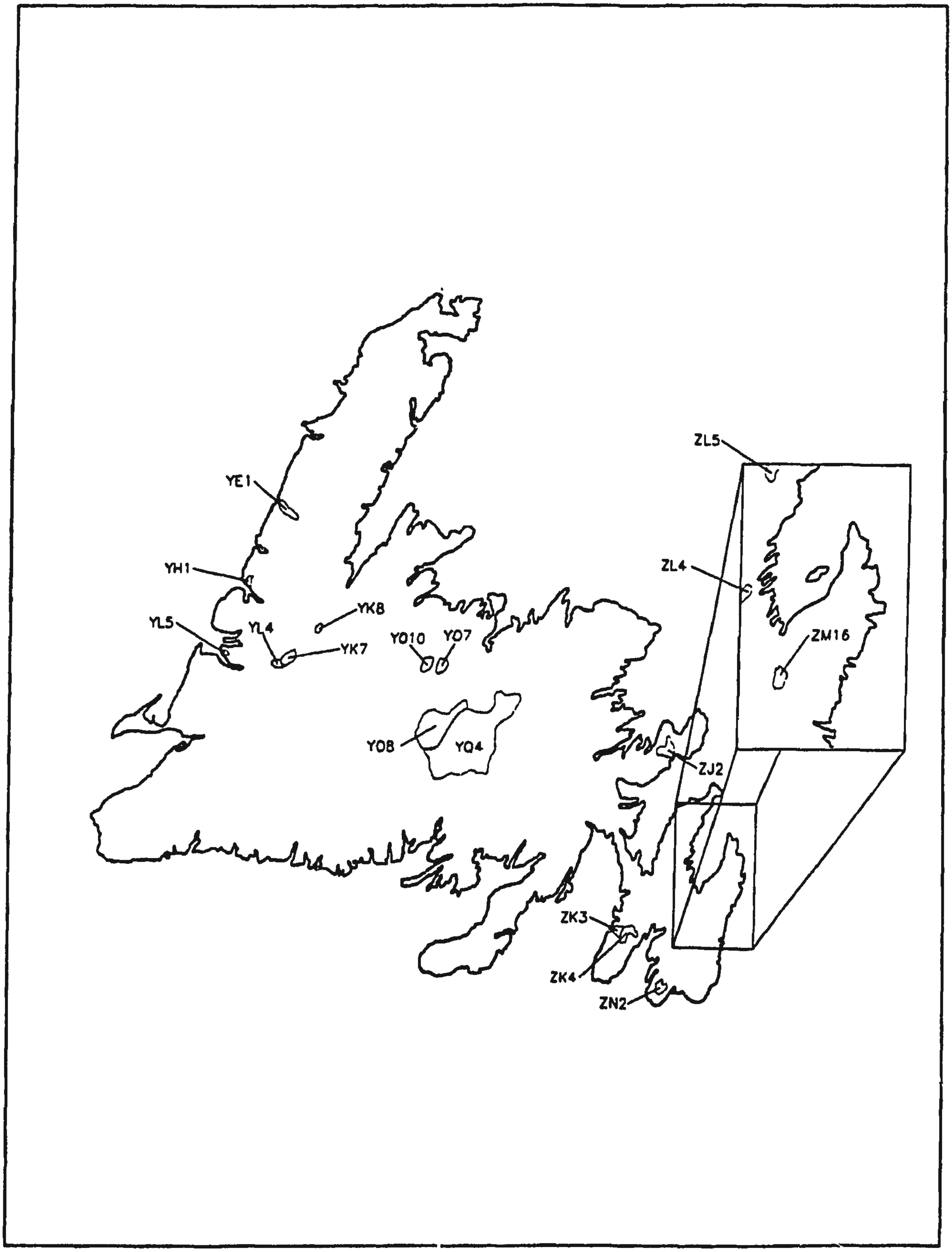


Fig. 5.2 – Locations of Shorter Record Gauges

Table 5.7

Test Basin Data

	Basin ID	Gauge Name	Qavgfld m^3/s	DA	Lmc_Sp	El_Centr	Dr_Dens	Shape
YO10	Jnc	Junction Brk near Badger	11.6	61.6	18.2	213	0.77	1.55
YE1	Grvt	Greavett Brook above Portland Crk Pond	35.7	101.0	24.5	152	0.75	1.64
YH1	BtmC	Bottom Creek near Rocky Hbr	4.9	33.4	14.5	69	1.13	1.69
YK7	Glde	Glide Brook below Glide Lake	21.8	112.0	26.8	30	1.28	1.61
YK8	Boot	Boot Brook at Trans-Canada Hwy	7.4	20.4	10.1	183	1.26	1.47
YL4	SthP	South Brook at Pasadena	27.7	58.5	13.2	46	1.34	1.54
YL5	Rtlr	Rattler Brook near McIvers	11.9	17.0	8.2	[illegible]	1.05	1.10
YO7	Lch	Leech Brook near Grand Falls	23.6	88.3	23.1	351	0.74	1.52
YO8	GtRt	Great Rattling Brk above Tote R. confluence	19.0	779.0	69.0	640	0.69	1.80
YQ4	NWG	Northwest Gander R near Gander Lake	617.6	2150.0	104.2	137	0.45	1.63
ZJ2	SmCh	Salmon Cove River near Champneys	11.9	79.4	18.0	213	1.11	1.33
ZK3	LtBP	Little Barachois R near Placentia	29.8	37.2	14.6	236	1.16	1.48
ZK4	LSNH	Little Salmonier R near North Harbour	73.0	104.0	28.5	152	1.50	1.85
ZL4	Shrs	Shearstown Brook at Shearstown	11.7	28.9	13.4	122	1.14	1.73
ZL5	BigB	Big Brook at Lead Cove	3.8	11.2	6.7	183	1.00	1.52
ZM16	SthH	South River near Holyrood	9.8	17.3	9.7	160	1.01	1.40
ZN2	StSh	St. Shotts R near Trepassey	7.7	15.5	10.3	160	1.03	1.53

	Basin ID	Gauge Name	Fr-LSw	FACLS	Dist-SW km	Fr_Barrn	Fr_Frst	Eff-P
YO10	Jnc	Junction Brk near Badger	0.19	0.81	179	0.00	0.81	0.85
YE1	Grvt	Greavett Brook above Portland Crk Pond	0.12	0.86	36	0.38	0.49	0.95
YH1	BtmC	Bottom Creek near Rocky Hbr	0.13	0.93	41	0.08	0.79	1.10
YK7	Glde	Glide Brook below Glide Lake	0.13	0.96	21	0.00	0.87	1.05
YK8	Boot	Boot Brook at Trans-Canada Hwy	0.24	0.65	200	0.01	0.75	1.05
YL4	SthP	South Brook at Pasadena	0.02	0.08	97	0.05	0.94	0.95
YL5	Rtlr	Rattler Brook near McIvers	0.09	0.46	77	0.00	0.91	1.40
YO7	Lch	Leech Brook near Grand Falls	0.28	0.73	191	0.02	0.70	0.85
YO8	GtRt	Great Rattling Brk above Tote R. confluence	0.24	0.55	142	0.03	0.73	0.85
YQ4	NWG	Northwest Gander R near Gander Lake	0.31	0.44	130	0.03	0.66	0.85
ZJ2	SmCh	Salmon Cove River near Champneys	0.19	0.82	209	0.07	0.74	1.00
ZK3	LtBP	Little Barachois R near Placentia	0.13	0.34	21	0.01	0.86	1.30
ZK4	LSNH	Little Salmonier R near North Harbour	0.46	0.91	46	0.31	0.23	1.60
ZL4	Shrs	Shearstown Brook at Shearstown	0.04	0.39	75	0.27	0.70	1.00
ZL5	BigB	Big Brook at Lead Cove	0.10	1.00	84	0.51	0.39	1.05
ZM16	SthH	South River near Holyrood	0.11	0.90	51	0.68	0.22	1.30
ZN2	StSh	St. Shotts R near Trepassey	0.12	0.82	5	0.00	0.88	1.30

forest). The test basins also tend to be better drained, i.e., they have higher drainage density. The range and distribution of the data for the distance and elevation variables for the two data sets is similar.

Each of the test basins was assigned to the appropriate subregion, for each of the regions in geographic or basin characteristic dataspace. In the case of the geographic regions, the assignment was straightforward, as it was for the case of the basin dataspace region based on drainage area only. For the dataspace region requiring cluster assignment, multiple discriminant analysis (MDA) and visual techniques were considered. With the MDA approach, discriminant scores are calculated from the basin characteristics, and are used to assign each test basin to a cluster.

Considering the relative simplicity of the problem, a visual technique was chosen. The relevant basin characteristics (FACLS, El-Cntrd and Fr-Barrn) for each test basin were plotted on separate plots with drainage area. Particular attention was given to the plot of FACLS and DA, since these were identified as the two most important basin characteristics. The test basins were assigned to the nearest cluster on the plots. The plots used for the assignments are provided in Appendix C. The Andrews Fourier plots for the test basins were also compared with the plots for the cluster to which they had been assigned, and these did not suggest any better assignment (Appendix B).

Qavgfld was then estimated for each test basin for each of the candidate regionalization schemes by selecting the appropriate equation from Table 5.7. Three approaches were used in order to assess the success of each regionalization scheme

- the residuals and the sums of squares of residuals were calculated,
- a pseudo-T value was calculated, and
- a nonparametric rank sum was calculated.

The pseudo-T test was developed as an indicator of the estimated value was within the expected range, given the variance of the model and the variance of the annual maximum daily flow in the basin in question. The test took the form

$$\frac{X_{calc} - X_{obs}}{\sqrt{(Var_{calc} + Var_{obs})^{1/2}}}$$

where X_{calc} is the estimated flow, X_{obs} is the average of the annual series of maximum daily flows, Var_{calc} is the error estimate of the model, and Var_{obs} is the variance of the annual series of maximum daily flows. A t value greater than 2 was somewhat arbitrarily taken to identify an unusually high value; for a standard t test, $t_{(0.025,37)}$ is 2.33.

The nonparametric rank sum test was used to identify the model which tended to give the best estimates in the largest number of cases. In this test, the residuals were ranked from smallest to largest; the regionalization method which resulted in the smallest difference between the observed and estimated flow in a test basin was assigned rank 1 for that basin, and so on. The process was repeated for each basin, and the ranks were summed.

5.4 Results

In general the results were poor. Figure 5.3 shows the scatter around the line of perfect agreement between observed and estimated flows. Figure 5.3(a) shows all the basins, and Figure 5.3(b) enlarges the plot to show the results for the smaller basin more clearly. Table 5.8 lists the observed and estimated Qavgfld for each of the candidate regionalization methods, as well as the residuals, and Table 5.9 provides the results of pseudo-T and rank sum tests. The schemes based on the provincial geographic divisions and on the nonhierarchical clusters appear to give the best results. The one-region assumption gives the worst, suggesting that any regionalization, whether in geographic or basin characteristic dataspace, helps.

From a strictly statistical point of view, the models should be tested using a control group of basins similarly matched to the study basins. There are no test basins in provincial Region D, for example; this fact probably contributes to the success of this method of regionalization, because Region D is the most problematic. In the regionalization based on size, there are only two large basins. These both have low FACLS, and since FACLS is very important in this equation (exponent of -2), the estimates are very poor.

From a practical point of view, however, the test basins are probably a good set to use, because they are quite typical of the basins for which flood estimates may be required. If the equations were tested using a matched group, the results would likely be

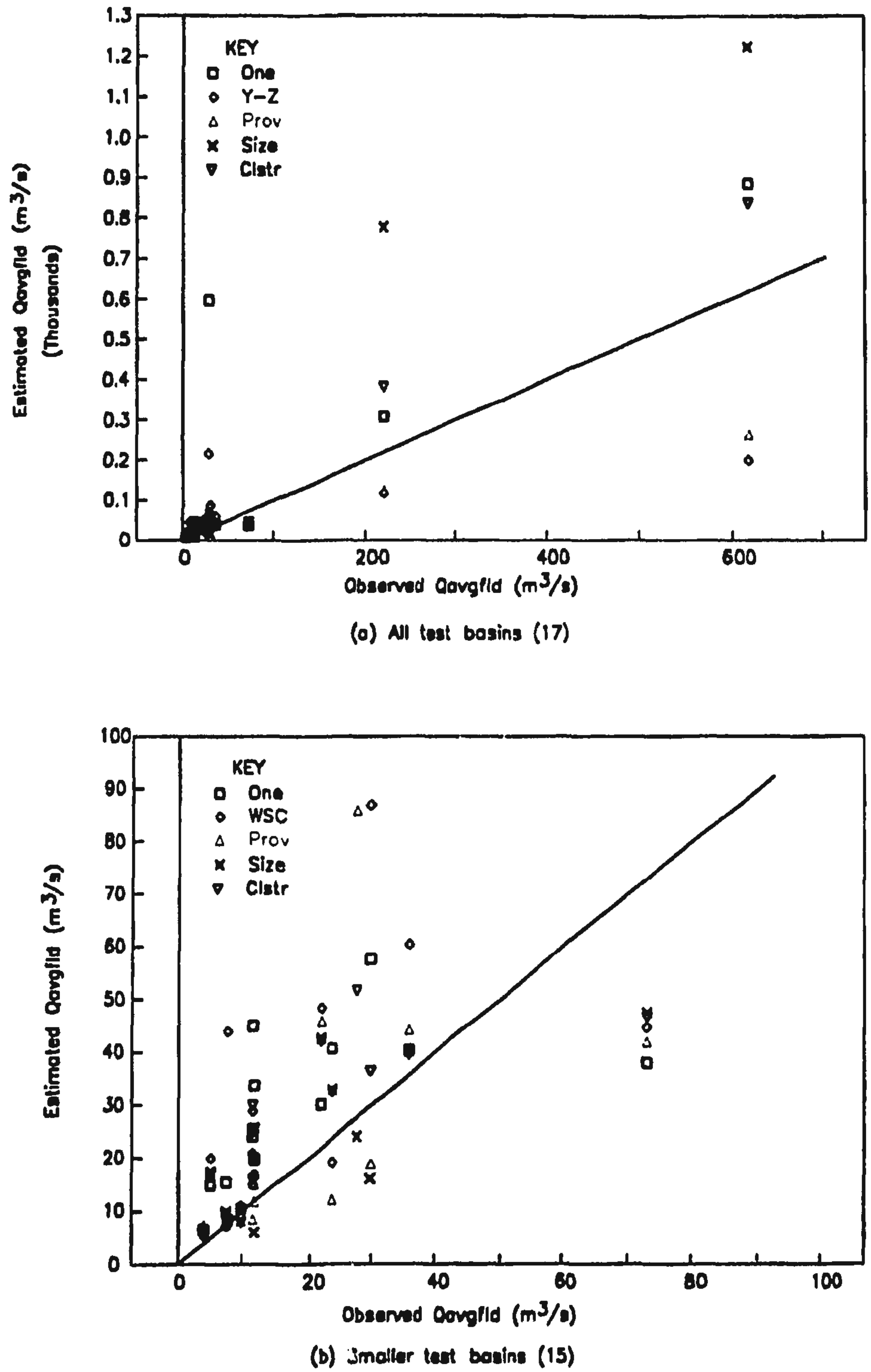


Fig. 5.3 – Results of Regionalization Methods

Table 5.8

Qavgfld Estimates and Residuals for Test Basins

Y_ZS	BASINS	Estimated Qavgfld (m^3/s) One Region	WSC Y–Z	Prov ABCD	Size	Clstr 1.2.3	Qavg Obs	Std Dev
YO10	Jnc	24.1	20.9	8.4	25.3	24.8	11.6	4.3
YE1	Grvt	40.4	60.3	44.2	40.1	39.6	35.7	10.7
YH1	BtmC	15.0	19.9	16.6	17.3	16.7	4.9	1.2
YK7	Glde	30.0	48.3	45.8	42.6	42.1	21.8	7.0
YK8	Boot	15.4	7.2	7.5	9.9	9.5	7.4	4.2
YL4	SthP	596.7	215.7	85.9	24.1	51.7	27.7	15.3
YL5	Rtlr	19.6	16.8	11.8	6.0	19.9	11.9	7.2
YO7	Lch	40.6	19.1	12.1	32.9	32.4	23.6	7.1
YO8	GtRt*	307.5	116.6	123.1	778.7	382.2	219.0	81.2
YQ4	NWG*	886.3	198.9	261.9	1223.7	837.6	617.6	252.1
ZJ2	SmCh	33.7	16.3	15.1	25.7	25.4	11.9	1.8
ZK3	LtBP	57.7	86.9	18.7	16.0	36.4	29.8	12.5
ZK4	LSNH	38.0	44.7	41.9	47.5	46.7	73.0	15.1
ZL4	Shrs	45.0	28.8	15.4	15.9	30.0	11.7	4.9
ZL5	BigB	6.4	5.3	7.3	6.4	6.1	3.8	0.8
ZM16	SthH	10.4	10.9	10.3	8.2	7.9	9.8	3.4
ZN2	StSh	8.4	43.9	9.4	8.3	8.0	7.7	2.4

Residuals

	One Region	WSC Y–Z	Prov ABCD	Size	Clstr 1.2.3
Jnc	12.5	9.3	−3.2	13.7	13.2
Grvt	4.7	24.6	8.5	4.4	3.9
BtmC	10.1	15.0	11.7	12.4	11.8
Glde	8.2	26.5	24.0	20.8	20.3
Boot	8.0	−0.2	0.1	2.5	2.1
SthP	569.0	188.0	58.2	−3.6	24.0
Rtlr	7.7	4.9	−0.1	−5.9	8.0
Lch	17.0	−4.5	−11.5	9.3	8.8
GtRt*	88.5	−102	−95.9	559.7	163.2
NWG*	268.7	−419	−356	606.1	220.0
SmCh	21.8	4.4	3.2	13.8	13.5
LtBP	27.9	57.1	−11.1	−13.8	6.6
LSNH	−35.0	−28.3	−31.1	−25.5	−26.3
Shrs	33.3	17.1	3.7	4.2	18.3
BigB	2.6	1.5	3.5	2.6	2.3
SthH	0.6	1.1	0.5	−1.6	−1.9
StSh	0.7	36.2	1.7	0.6	0.3
Sum of Squared Rs	408114	228482	141172	682631	77756
wo 2 largest	328069	42706	5446	1991	2725

Table 5.9

Test Results for Regionalization Methods

Pseudo-T

One Region	WSC Y-Z	Prov ABCD	Size	Clstr 123
-1.6	-1.3	0.4	-2.9	-2.9
-0.4	-2.1	-0.7	-0.4	-0.4
-1.5	-2.7	-1.9	-5.2	-5.2
-0.8	-3.0	-2.6	-2.8	-2.8
-1.0	0.0	-0.0	-0.5	-0.5
-34.0	-11.6	-3.6	0.2	0.2
-0.8	-0.5	0.0	0.8	0.5
-1.7	0.5	1.3	-1.3	-1.3
-1.1	1.3	1.2	-6.9	-6.9
-1.1	1.7	1.4	-2.4	-2.4
-3.1	-0.9	-0.5	-5.0	-5.1
-2.0	-4.3	0.9	1.1	0.9
2.1	1.8	2.0	1.7	1.7
-4.0	-2.6	-0.6	-0.8	-0.4
-0.4	-0.3	-1.2	-1.2	-1.2
-0.1	-0.2	-0.1	0.4	0.4
-0.1	-7.3	-0.5	-0.2	-0.2

Rank

	One Region	WSC Y-Z	Prov ABCD	Size	Clstr 1.23
	3	2	1	5	4
	3	5	4	2	1
	1	5	2	4	3
	1	5	4	3	2
	5	2	1	4	3
	5	4	3	1	2
	4	2	1	3	5
	5	1	4	3	2
	1	3	2	5	4
	2	4	3	5	1
	5	2	1	4	3
	4	5	2	3	1
	5	3	4	1	2
	5	3	1	2	4
	4	1	5	3	2
	2	3	1	4	5
	3	5	4	2	1
Sum of Ranks	58	55	43	54	45

reasonable, instilling a false sense of confidence in users. With no alternative, the equation would likely be applied in practice with data outside the allowable range. As it is, the poor results serve as a caution. The boxplots in Appendix B compare individual variables in the test data set with the study data set, and these can be used in a preliminary assessment of whether the characteristics of a particular ungauged basin lie within the acceptable range. Because the explanatory variables in multiple regression are multidimensional, however, it may be difficult to recognize whether an extrapolation is being made beyond the range of the original data set. The leverage statistic $h_i = x_o'(X'X)^{-1}x_o$ expresses the distance of a given point x_o from the centre of the sample observations. It may be calculated for an ungauged basin and used as a numerical diagnostic to detect such extrapolation (Helsel and Hirsch, 1992).

The leverage statistics were checked for the one region equation (using the results of the OLS analysis since these statistics are not available for NLLS). Two basins, South Brook at Pasadena (SthP) and Little Barachois near Placentia (LtBP) had leverage statistics beyond the value of 0.4, the highest in the data set used to generate the equation. The leverage statistic for SthP was 2.4, which explains the outlandish estimated flow of 597 m^3/s compared with an observed value of 27.7 m^3/s. The leverage for LtBP was 0.45. The leverage values for three other basins, Rattler at McIvers (Rtlr), Northwest Gander (NWG) and Shearstown Brook (Shrs) were high (>0.3) but within the range of the data of the study set.

So the fact that characteristics of the test data set are somewhat different from those of the study data set partly explains the poor results. An additional explanation is that the hydrologic input is only taken into account indirectly. Most of the equations for the subregions include some variable such as Fr-Barrn, Fr-Forst or Dist-SW which at least in part represents hydrologic input. Because some of these variables are perhaps doing double duty, reflecting both input and basin response, it is not surprising that when the equations are tested they are not very robust.

Hydrologic input can be accounted for by geographical regionalization. Traditional geographic regionalization assigns boundaries based on hydrological homogeneity. At least one part of the meaning of hydrologic homogeneity is homogeneity of input, i.e., similarity of hydrological events. Other types of similarity (such as terrain, geology or vegetation), are usually included in the definition as well, however, thereby clouding the issue. Logically, the two poles of regionalization approaches are

1. to group basins according to similar hydrologic input, and develop equations using basin characteristics;
2. to cluster basins according to basin characteristics, and develop equations using hydrologic input.

All present methods fall somewhere in between. The results of the provincial geographic regionalization for flood frequency analysis are interesting in this respect, because for two or three of the regions, the effect of the regionalization is to reduce the range of

Eff-P within each region, as shown in the boxplots in Appendix A. In those regions, the method accounts for hydrologic input by grouping the basins into areas of similar input. Consequently the equations for those regions do not include the Eff-P variable. The provincial groupings for low flow analysis might show even less variation, since they were specifically grouped by similarity of average annual precipitation and fraction of barren/forest.

5.5 Alternative Approach

How can the results be improved? One alternative is simply to wait for more data. In a few years data from the 17 test basins can be used, and the resulting models should give better results, covering a wider range of basins. The problem of the limited range of data will be reduced, but the question of hydrologic input will still remain.

The alternative approach considered here is to include Eff-P as an independent variable representing hydrologic input. It was initially rejected because of the increased uncertainty of the estimates, but this uncertainty may be compensated for by greater robustness when used with ungauged basins. As a first trial, therefore, a simple equation for the one-region case with DA and Eff-P was developed, and the estimates were compared with those obtained without using Eff-P. The regression equation was

$$Q = 0.489 DA^{0.862} \text{Eff-P}^{1.756}$$

Tables 5.10 and 5.11 present the results together with those previously given in Tables 5.8 and 5.9. They show a marked improvement.

Table 5.10

Qavgfld Estimates and Residuals for Test Basins: Alternate Method for One Region Case

	Basin ID	One Region	WSC Y–Z	Prov ABCD	Size	Clstr 123	One Rgn DA + Eff–P	Qavg Obs	Std Dev
YO10	Jnc	24.1	20.9	8.4	25.3	24.8	12.8	11.6	4.3
YE1	Grvt	40.4	60.3	44.2	40.1	39.6	23.9	35.7	10.7
YH1	BtmC	15.0	19.9	16.6	17.3	16.7	12.0	4.9	1.2
YK7	Glde	30.0	48.3	45.8	42.6	42.1	31.3	21.8	7.0
YK8	Boot	15.4	7.2	7.5	9.9	9.5	7.2	7.4	4.2
YL4	SthP	596.7	215.7	85.9	24.1	51.7	14.9	27.7	15.3
YL5	Rtlr	19.6	16.8	11.8	6.0	19.9	10.2	11.9	7.2
YO7	Lch	40.6	19.1	12.1	32.9	32.4	17.5	23.6	7.1
YO8	GtRt*	307.5	116.6	123.1	778.7	382.2	114.5	219.0	81.2
YQ4	NWG*	886.3	198.9	261.9	1223.7	837.6	274.8	617.6	252.1
ZJ2	SmCh	33.7	16.3	15.1	25.7	25.4	21.3	11.9	1.8
ZK3	LtBP	57.7	86.9	18.7	16.0	36.4	17.6	29.8	12.5
ZK4	LSNH	38.0	44.7	41.9	47.5	46.7	61.9	73.0	15.1
ZL4	Shrs	45.0	28.8	15.4	15.9	30.0	8.9	11.7	4.9
ZL5	BigB	6.4	5.3	7.3	6.4	6.1	4.3	3.8	0.8
ZM16	SthH	10.4	10.9	10.3	8.2	7.9	9.1	9.8	3.4
ZN2	StSh	8.4	43.9	9.4	8.3	8.0	8.3	7.7	2.4

Residuals

	One Region	WSC Y–Z	Prov ABCD	Size	Clstr 123	One Rgn DA + Eff–P
Jnc	12.5	9.3	–3.2	13.7	13.2	1.2
Grvt	4.7	24.[illegible]	8.5	4.4	3.9	–11.8
BtmC	10.1	15.0	11.7	12.4	11.8	7.1
Glde	8.2	26.5	24.0	20.8	20.3	9.5
Boot	8.046	–0.189	0.056	2.483	2.122	–0.209
SthP	569.0	188.0	58.2	–3.6	24.0	–12.8
Rtlr	7.7	4.9	–0.1	–5.9	8.0	–1.7
Lch	17.0	–4.5	–11.5	9.3	8.8	–6.1
GtRt*	88.5	–102	–95.9	559.7	163.2	–104.5
NWG*	268.7	–419	–356	606.1	220.0	–342.8
SmCh	21.8	4.4	3.2	13.8	13.5	9.4
LtBP	27.9	57.1	–11.1	–13.8	6.6	–12.2
LSNH	–35.0	–28.3	–31.1	–25.5	–26.3	–11.1
Shrs	33.3	17.1	3.7	4.2	18.3	–2.8
BigB	2.632	1.491	3.516	2.597	2.294	0.486
SthH	0.567	1.129	0.484	–1.624	–1.933	–0.686
StSh	0.743	36.159	1.736	0.643	0.290	0.590
Squared Rs	408114	228482	141172	682631	77756	129[illegible]78
wo 2 largest	328069	42706	5446	1991	2725	850

Table 5.11

Test Results with Alternate Method

Pseudo–T

One Region	WSC Y–Z	Prov ABCD	Size	Clstr 1.2.3	One Rgn DA + Eff–P
−1.6	−1.3	0.4	−2.9	−2.9	−1.7
−0.4	−2.1	−0.7	−0.4	−0.4	−0.3
−1.5	−2.7	−1.9	−5.2	−5.2	−1.8
−0.8	−3.0	−2.6	−2.8	−2.8	−2.1
−1.0	0.0	−0.0	−0.5	−0.5	−0.3
−34.0	−11.6	−3.6	0.2	0.2	−1.4
−0.8	−0.5	0.0	0.8	0.5	−0.8
−1.7	0.5	1.3	−1.3	−1.3	−0.9
−1.1	1.3	1.2	−6.9	−6.9	−2.0
−1.1	1.7	1.4	−2.4	−2.4	−0.9
−3.1	−0.9	−0.5	−5.0	−5.1	−2.0
−2.0	−4.3	0.9	1.1	0.9	−0.5
2.1	1.8	2.0	1.7	1.7	1.6
−4.0	−2.6	−0.6	−0.8	−0.4	−2.3
−0.4	−0.3	−1.2	−1.2	−1.2	−0.4
−0.1	−0.2	−0.1	0.4	0.4	0.3
−0.1	−7.3	−0.5	−0.2	−0.2	−0.0

Rank Sum Test

	One Region	WSC Y–Z	Prov ABCD	Size	Clstr 1.2.3	One Rgn DA + Eff–P
	4	3	2	6	5	1
	3	6	4	2	1	5
	2	6	3	5	4	1
	1	6	5	4	3	2
	6	2	1	5	4	3
	6	5	4	1	3	2
	5	3	1	4	6	2
	6	1	5	4	3	2
	1	3	2	6	5	4
	2	5	4	6	1	3
	6	2	1	5	4	3
	5	6	2	4	1	3
	6	4	5	2	3	1
	6	4	2	3	5	1
	5	2	6	4	3	1
	2	4	1	5	6	3
	4	6	5	3	1	2
Sum of Ranks	70	68	53	69	58	39

On all measures the simple model with DA and Eff-P is superior to any of the others (although the very large squared residual for NWG, the largest basin, gives the cluster method an overall lower total sum of squared residuals). In addition, all the test basins are within the range of the data as indicated by their leverages.

The plots in Figures 5.4 to 5.6 compare the estimated and observed floods using this one region equation with some of the alternatives derived from the various regionalization methods. Figure 5.4 compares the DA-Eff-P equation with the previous one region model. It shows that while the DA-Eff-P equation underestimates the flow for the two largest basins, due to the fact that it does not take into account the faster response of the large basins with less control by lakes and swamps, the previous one-region equation (with DA, Fr-Barrn, and FACLS) overestimates a similar amount. The previous equation also cannot take into account the extreme wetness of the Little Salmonier North Harbour (LSNH) basin.

Figure 5.4(b) expands the scale to show the results for the smaller basins more clearly. The estimate for SthP by the previous method does not appear on this plot, because it is nearly 600 m^3/s, compared with an observed value of 27.7 m^3/s. The other estimates are generally better with the DA-Eff-P model.

Figure 5.5 shows a similar comparison between the DA-Eff-P (one-region) results and the previous results assuming the provincial regional subdivisions. The DA-Eff-P results

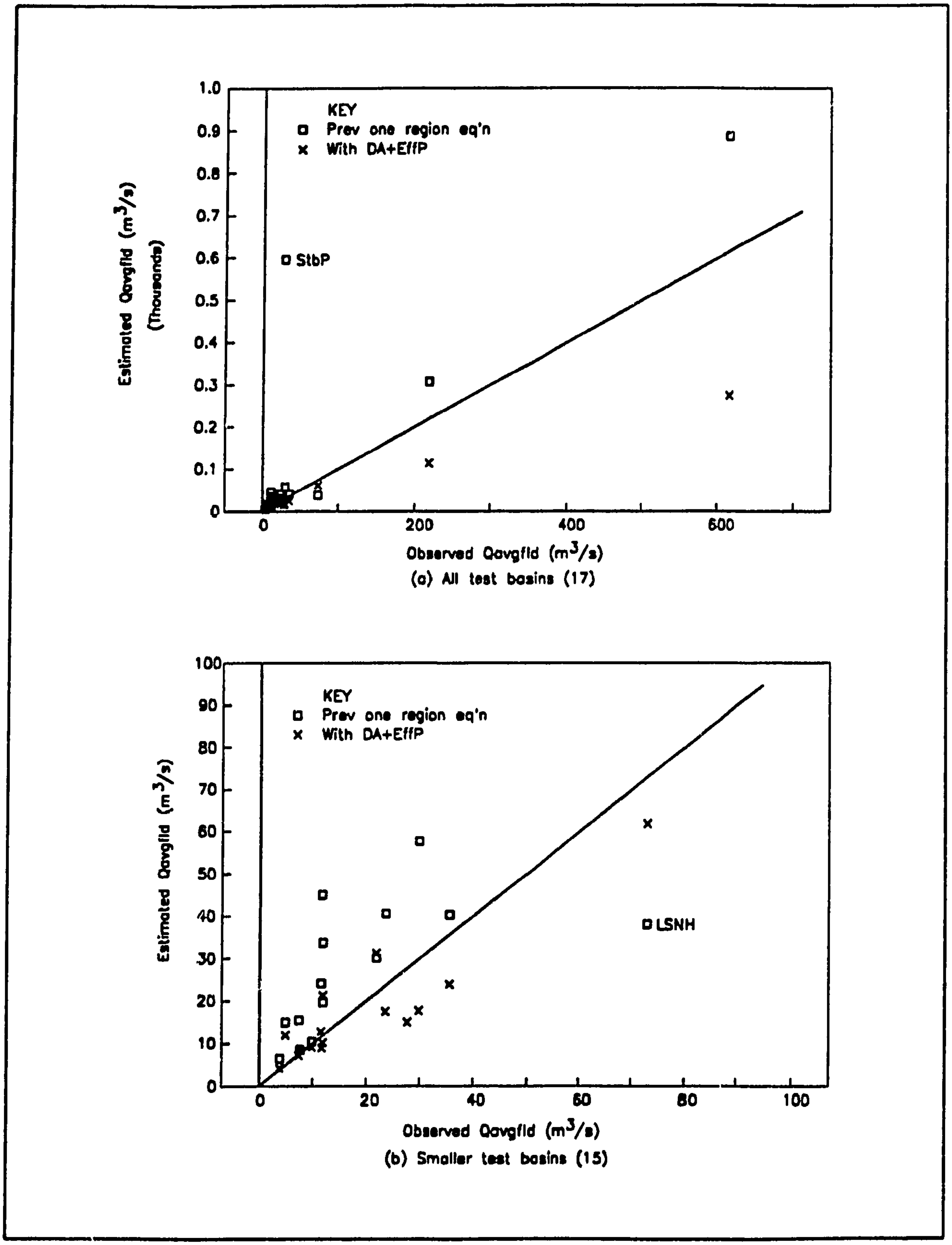
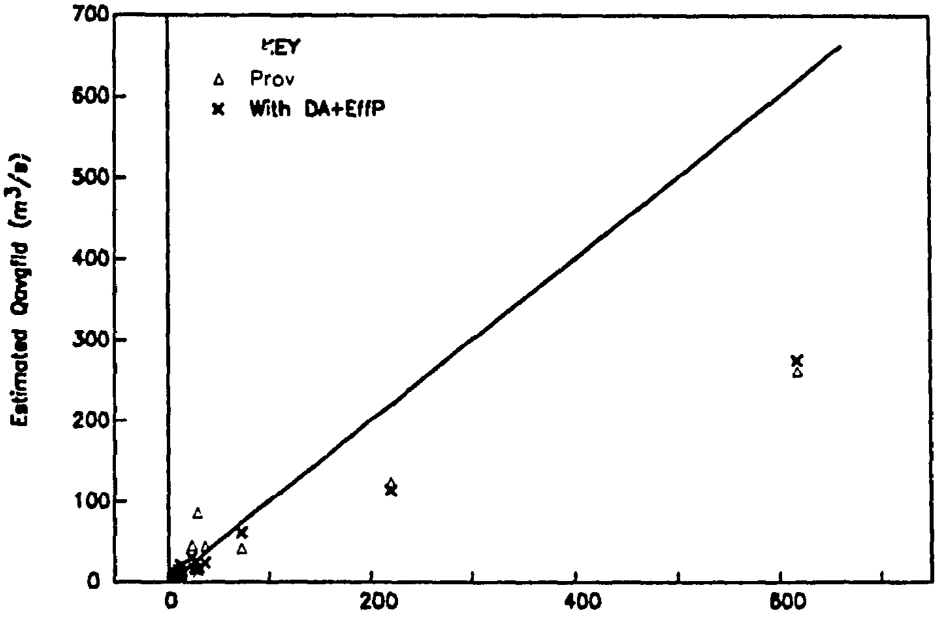
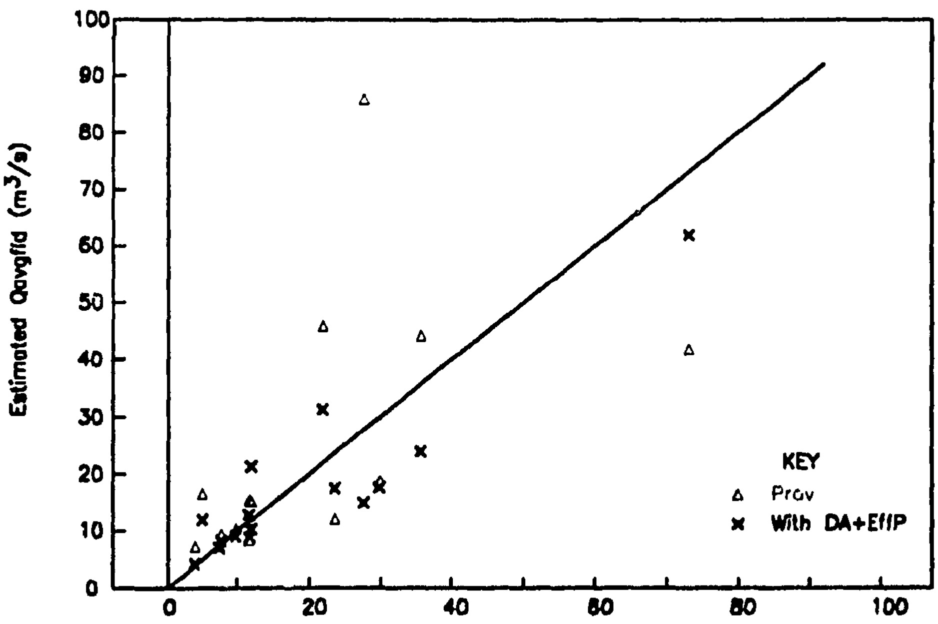


Fig. 5.4 — Comparison of Two One-Region Models



(a) All test basins (17)



(b) Smaller test basins (15)

Fig. 5.5 — Comparison of Provincial Regionalization with DA-EffP Model

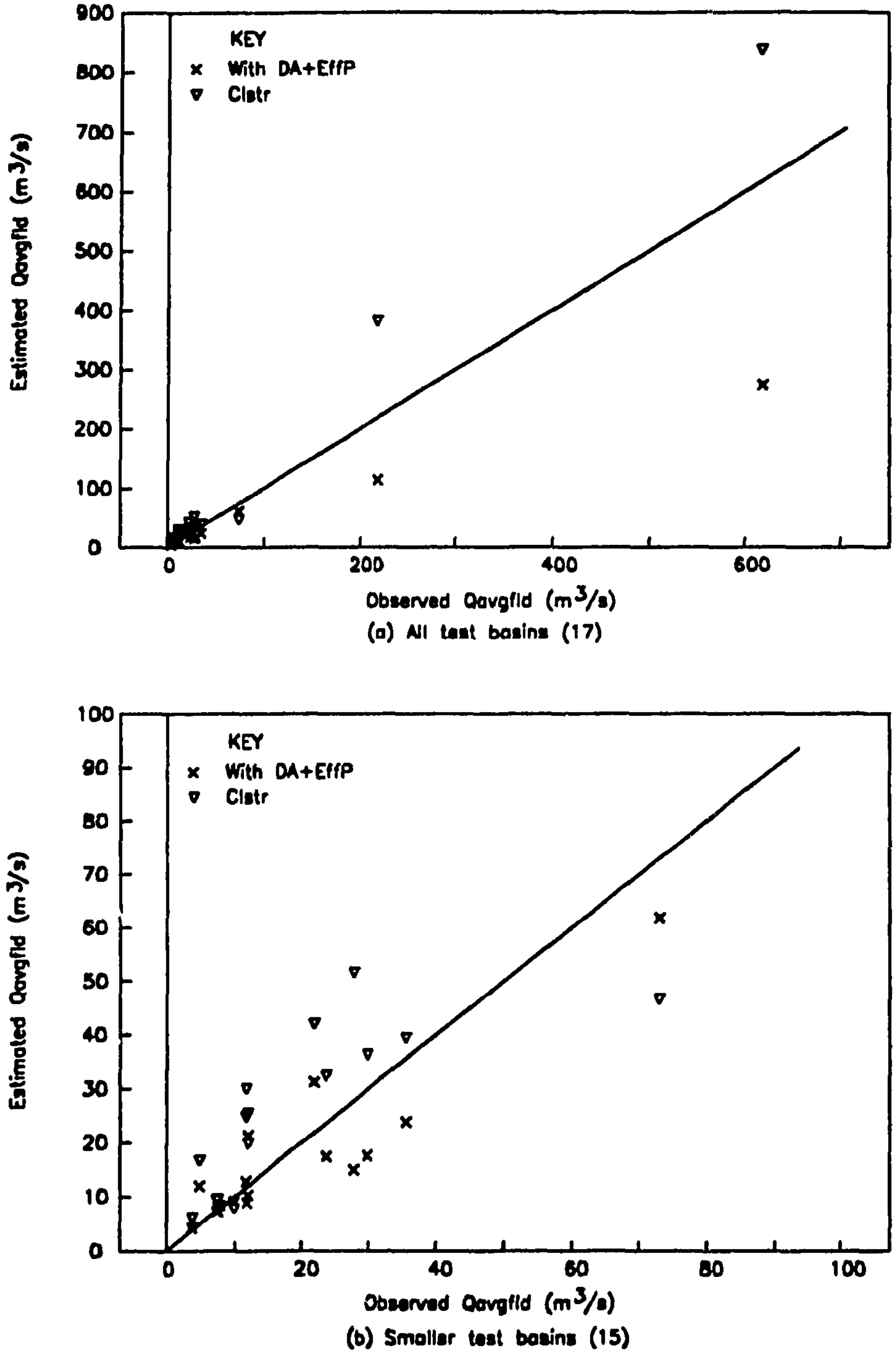


Fig. 5.6 – Comparison of Cluster and DA–EffP Models

are as good or better than the results using the provincial regions. Similarly, Figure 5.6 compares the one-region DA-Eff-P results with the results using the cluster method of regionalization. Again, the results are as good or better.

There are no test basins in the areas of the island where flows are generally considered difficult to estimate, e.g., the south coast or the eastern side of the Great Northern Peninsula, so the performance of the model cannot be checked in those areas.

5.5.1 Estimating Effective Precipitation

One argument against using effective precipitation as an explanatory variable is that it is difficult to estimate. In fact it was because of the additional uncertainty in estimating Eff-P that it was at first rejected. In its favour is that it can be estimated everywhere, and the estimates will virtually always be bounded. The results presented in Chapter 4 of this study identified the important topographic and geographic variables which should be taken into account when estimating Eff-P, even if no equation for Eff-P can be recommended.

For a quick estimate of Eff-P, the isoline maps presented in various water resources studies for different areas of the province may be used (Gov't of Nfld, 1987, 1988, 1989, 1990b, 1992, 1993). This method was used for the test basin estimates. A better estimate can be obtained by using the equations presented in

Chapter 4, together with data from surrounding gauged basins and a consideration of basin elevation and orientation relative to the gauged basins.

5.6 Extension of DA-Eff-P Approach to Regions

The alternative approach using only the two simple variables DA and Eff-P in a one region equation was shown to offer marked improvement over approaches using basin characteristics only. The next question was whether developing new equations within the regions, including Eff-P as an independent variable, might improve the results further. Equations using DA and Eff-P were therefore developed for all subregions in which Eff-P was significant. The coefficients and statistics of these equations are presented in Table 5.12. As expected, the uncertainty is higher in the regions where Eff-P is included as an explanatory variable, and highest where the coefficient associated with Eff-P is highest (e.g., provincial geographic Region C).

As this table shows, Eff-P was not a significant explanatory variable (i.e., $p > 0.05$) in provincial regions A and B. This finding confirms the earlier comment that one effect of geographic regionalization can be to take account of hydrologic input. On the other hand, Eff-P is very important in the other geographic regions, especially region C.

The equations for Region D and for Cluster 2+4 include both DA and Eff-P, although strictly speaking there are too many parameters for the number of cases. The results were very poor with only one independent variable, however, and the coefficients of the

Table 5.12

Alternative Regional Models

(a) Geographic Regions

	All Island	WSC		Prov			
		Y	Z	A	B	C	D
N	40	20	20	12	12	10	6
Constant	0.489	0.334	1.171	1.103	0.117	0.435	1.472
DA	0.862	0.927	0.700	0.783	1.092	0.871	0.700
FACLS							
Dr−Dens					0.826		
Fr−Barrn							
Fr−LSW							
Eff−P	1.756	1.753	1.788			2.4[illegible]	1.569
r^2	0.949	0.993	0.949	0.989	0.990	0.993	0.987
r^2 corr	0.908	0.987	0.890	0.979	0.980	0.987	0.926
ss res	62395	6720	14216	692	9542	3102	1923
n−p	37	17	17	10	9	7	3
error est	41.1	19.9	28.9	8.3	32.6	21.1	25.3
CV(Q)^2	0.031	0.032	0.032	0.001	0.001	0.063	0.025
Uncertainty	0.177	0.178	0.180	0.023	0.033	0.251	0.158

(b) Basin Characterisic Regions/Clusters

	Size (km²)		Clusters		
	<130	>130	1	3	2+4
N	13	27	24	11	5
Constant	0.673	0.497	0.457	0.629	0.972
DA	0.878	0.861	0.871	0.865	0.803
FACLS					
Dr−Dens					
Fr−LSW					
Eff−P	0.380	1.768	1.642	0.620	1.403
r^2	0.976	0.949	0.948	0.983	0.986
r^2 corr	0.889	0.882	0.888	0.933	0.950
ss res	250	60950	57292	123	1416
n−p	10	24	21	8	2
error est	5.0	50.4	52.2	3.9	26.6
CV(Q)^2	0.004	0.033	0.029	0.006	0.022
Uncertainty	0.061	0.183	0.170	0.078	0.149

equations with two variables were quite reasonable. The improvement was such that even with relatively few cases the adjusted r^2's are high. Since there are no test basins in Region D, the equation for that region is never used.

The observed and estimated flows and the residuals are presented in Table 5.13. The plots in Figure 5.7 show the estimated and observed values. The large basins continue to be underestimated, presumably because of their low FACLS. The cluster method takes account of this best. The pseudo-T values and the rankings are given in Table 5.14. The sums of squared residuals are down for all regionalization methods, from about 10 percent to over 200 percent. The number of pseudo-T values with absolute values greater than 2 is reduced from 26 to 8.

The one-region equation continues to perform reasonably well. The Y-Z regions are also quite promising, especially as measured by the sums of squared residuals. If the two largest basins are excluded, the Y-Z regionalization method has the lowest sum of squared residuals, and even with these two it is has the second lowest. Only one pseudo-T value is significant. The Y-Z regionalization is geographic, and may represent some additional information about hydrologic input which is not represented by Eff-P.

The nonparametric rank sums test suggests that clustering basins by their physiographic characteristics is slightly better than assuming only one region, which in turn is slightly

Table 5.13

Qavgfld Estimates and Residuals with DA and Eff-P

		Basin ID	One Rgn	WSC Y-Z	Prov ABCD	Size	Clstr 1-2-3	Qavg Obs	Std Dev
B	YO10	Jnc	12.8	11.5	8.4	23.6	20.1	11.6	4.3
C	YE1	Grvt	23.9	22.0	21.3	38.0	33.0	35.7	10.7
C	YH1	BtmC	11.9	10.2	11.7	15.2	13.9	4.9	1.2
C	YK7	Glde	31.1	28.9	29.9	43.2	38.3	21.8	7.0
C	YK8	Boot	7.2	6.0	6.8	9.7	8.8	7.4	4.2
C	YL4	SthP	14.9	13.3	13.2	23.5	23.8	27.7	15.3
C	YL5	Rtlr	10.2	8.3	11.9	9.2	15.2	11.9	7.2
B	YO7	Lch	17.5	16.0	12.1	32.4	27.4	23.6	7.1
B	YO8	GtRt*	114.5	120.4	123.1	115.0	162.7	219.0	81.2
B	YQ4	NWG*	274.7	308.6	261.9	275.7	367.7	617.6	252.1
B	ZJ2	SmCh	21.3	25.0	15.1	31.4	27.6	11.9	1.8
A	ZK3	LtBP	17.5	23.5	18.7	17.8	25.7	29.8	12.5
A	ZK4	LSNH	61.2	70.0	41.9	47.5	46.7	73.0	15.1
A	ZL4	Shrs	8.9	12.3	15.4	12.5	14.5	11.7	4.9
A	ZL5	BigB	4.3	6.9	7.3	5.7	5.2	3.8	0.8
A	ZM16	SthH	9.1	13.8	10.3	9.1	8.7	9.8	3.4
A	ZN2	StSh	8.2	12.7	9.4	8.3	7.9	7.7	2.4

Residuals

	Basin ID	One Rgn	WSC Y-Z	Prov ABCD	Size Large	Clstr 1.2.3
YO10	Jnc	1.2	-0.1	-3.2	12.0	8.5
YE1	Grvt	-11.8	-13.7	-14.4	2.3	-2.7
YH1	BtmC	7.0	5.3	6.8	10.3	9.0
YK7	Glde	9.3	7.1	8.1	21.4	16.5
YK8	Boot	-0.2	-1.4	-0.6	2.3	1.4
YL4	SthP	-12.8	-14.4	-14.5	-4.2	-3.9
YL5	Rtlr	-1.7	-3.6	-0.0	-2.7	3.3
YO7	Lch	-6.1	-7.6	-11.5	8.8	3.8
YO8	GtRt*	-104.5	-99	-95.9	-104.0	-56.3
YQ4	NWG*	-342.9	-309	-356	-341.9	-249.9
ZJ2	SmCh	9.4	13.1	3.2	19.5	15.7
ZK3	LtBP	-12.3	-6.3	-11.1	-12.0	-4.1
ZK4	LSNH	-11.8	-3.0	-31.1	-25.5	-26.3
ZL4	Shrs	-2.8	0.6	3.7	1.2	2.8
ZL5	BigB	0.5	3.1	3.5	1.9	1.4
ZM16	SthH	-0.7	4.0	0.5	-0.7	-1.1
ZN2	StSh	0.5	5.0	1.7	0.6	0.2
Sum of Sqrd resds		129407	105996	137526	129681	67065
wo 2 largest		867	818	1800	1999	1443

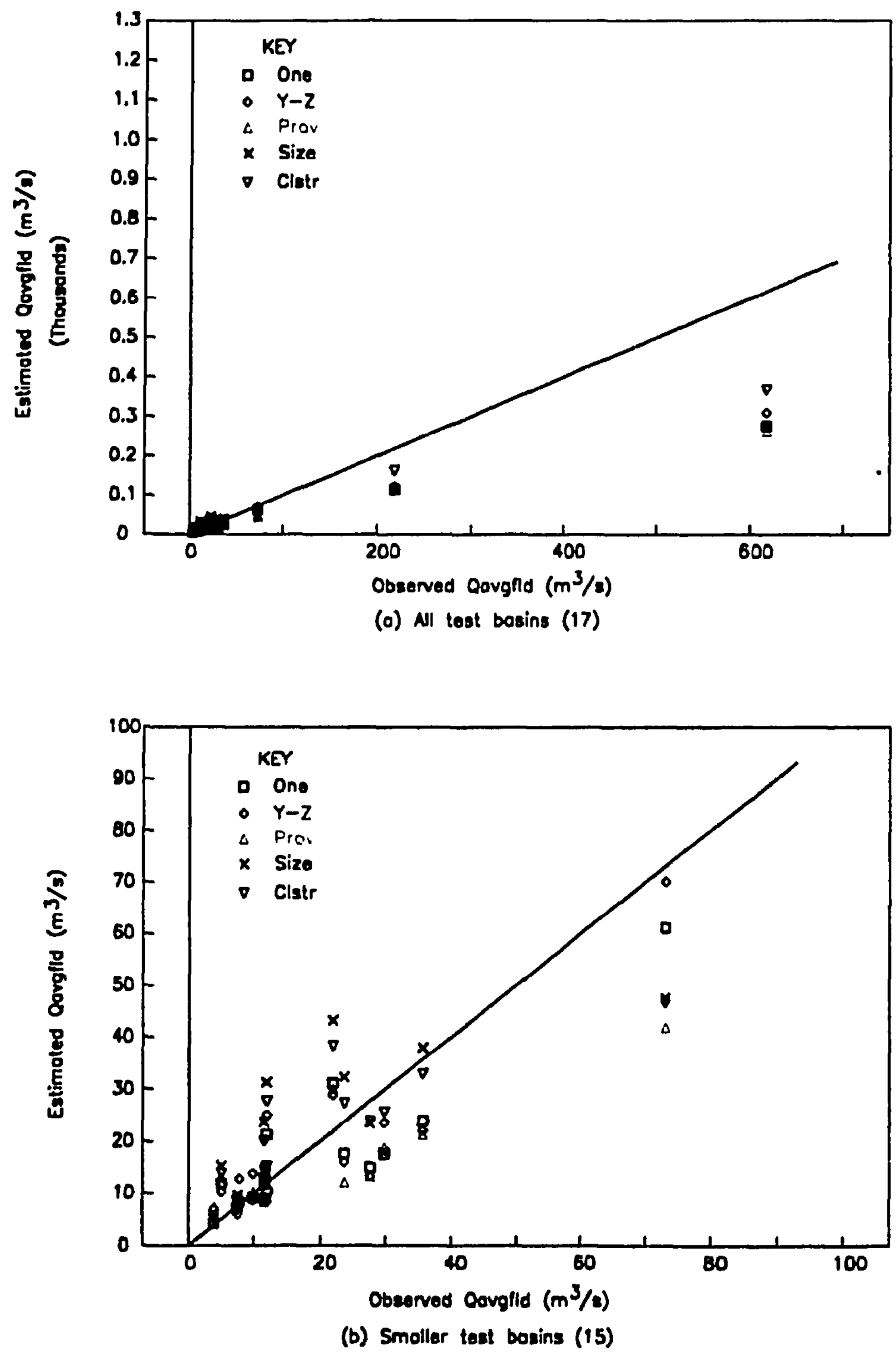


Fig. 5.7 – Results Using DA and EffP, All Regionalization Models

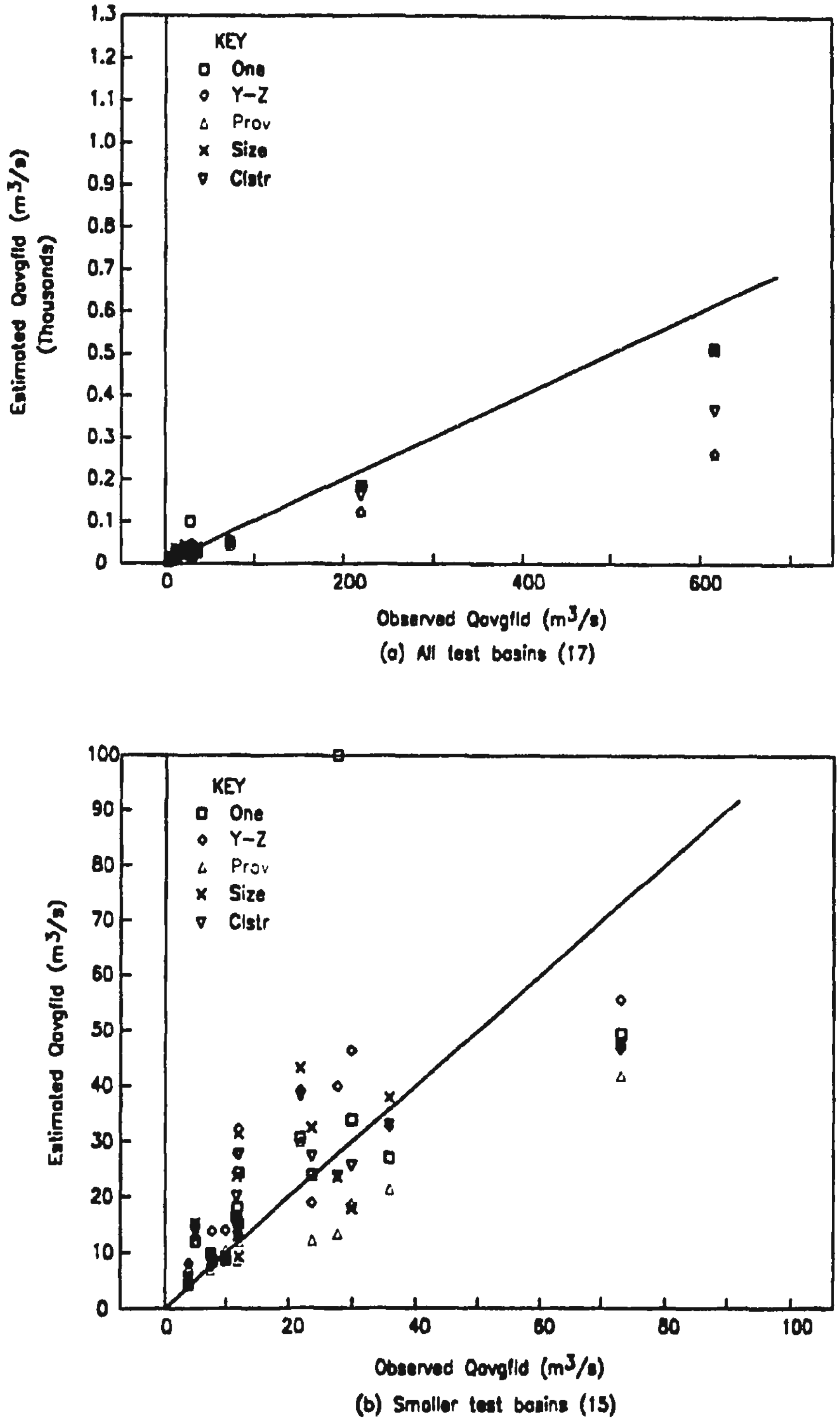


Fig. 5.8 – Results Using Models in Table 5.15

Table 5.14

Test Results with DA and Eff−P

		Basin ID	One Rgn	WSC Y−Z	Prov ABCD	Size Large	Clstr 1−2−3
B	YO10	Jnc	−0.2	0.0	0.4	−2.5	−1.8
C	YE1	Grvt	0.9	1.2	1.2	−0.2	0.3
C	YH1	BtmC	−1.1	−1.1	−1.4	−4.1	−3.9
C	YK7	Glde	−1.0	−0.9	−1.0	−2.9	−2.3
C	YK8	Boot	0.0	0.2	0.1	−0.5	−0.3
C	YL4	SthP	0.8	0.9	0.9	0.3	0.2
C	YL5	Rtlr	0.2	0.4	0.0	0.4	−0.4
B	YO7	Lch	0.6	0.9	1.3	−1.2	−0.5
B	YO8	GtRt*	1.3	1.2	1.2	1.3	0.7
B	YQ4	NWG*	1.4	1.2	1.4	1.4	1.0
B	ZJ2	SmCh	−1.4	−2.3	−0.5	−6.8	−5.9
A	ZK3	LtBP	0.9	0.5	0.9	0.9	0.3
A	ZK4	LSNH	0.7	0.2	2.0	1.7	1.7
A	ZL4	Shrs	0.3	−0.1	−0.6	−0.2	−0.4
A	ZL5	BigB	−0.1	−0.6	−1.2	−0.8	−0.7
A	ZM16	SthH	0.1	−0.6	−0.1	0.2	0.3
A	ZN2	StSh	−0.1	−0.9	−0.5	−0.2	−0.1

Rank Sum Test

	Basin ID	One Rgn	WSC Y−Z	Prov ABCD	Size Large	Clstr 1−2−3
YO10	Jnc	2	1	3	5	4
YE1	Grvt	3	4	5	1	2
YH1	BtmC	3	1	2	5	4
YK7	Glde	3	1	2	5	4
YK8	Boot	1	4	2	5	3
YL4	SthP	3	4	5	2	1
YL5	Rtlr	2	5	1	3	4
YO7	Lch	2	3	5	4	1
YO8	GtRt*	5	3	2	4	1
YQ4	NWG*	4	2	5	3	1
ZJ2	SmCh	2	3	1	5	4
ZK3	LtBP	5	2	3	4	1
ZK4	LSNH	2	1	5	3	4
ZL4	Shrs	4	1	5	2	3
ZL5	BigB	1	4	5	3	2
ZM16	SthH	3	5	1	2	4
ZN2	StSh	2	5	4	3	1
Sum of Ranks		47	49	56	59	44

better than the Y-Z regionalization. (Recall that the clusters were based on similarity of FACLS, El-Cntrd and Fr-Barm.) Because the model variance for the cluster equations is small, however, there are more high pseudo-T values. The large model variance of the one region equation means that none of the pseudo-T values are greater than 2. The clusters based on size alone perform the worst in the rank sums comparison. The provincial regions also do not perform especially well.

The results are not conclusive, but in general they confirm that using DA and Eff-P gives better results for all regionalization methods than using basin characteristics alone. The Y-Z results also suggest that there is room for some further refinement of hydrologic input.

5.6.1 Additional Explanatory Variables

A final comparison was made with models including a third variable in addition to DA and Eff-P, where such a model could reasonably be developed. The coefficients and relevant statistics of the final models are presented in Table 5-15. Additional variables could be included in the equations for one region (all-island), for the Y and Z regions, and for the basins greater than 130 km^2. The third variable was always FACLS except for Region Y where it was Fr-LSw. The estimates and test results are shown in Tables 5-16 and 5-17.

Table 5.15

Alternative Regional Models with a Third Variable

(a) Geographic Regions

	All Island *	WSC Y *	WSC Z *	Prov A	Prov B	Prov C	Prov D
N	40	20	20	12	12	10	6
Constant	0.524	0.333	1.510	1.103	0.117	0.435	1.472
DA	0.843	0.839	0.667	0.783	1.092	0.871	0.700
FACLS	−0.745		−0.710				
Dr−Dens					0.826		
Fr−Barrn							
Fr−LSW		−0.366					
Eff−P	1.194	1.205	0.942			2.495	1.569
r^2	0.964	0.996	0.977	0.989	0.990	0.993	0.987
r^2 corr	0.936	0.992	0.951	0.979	0.980	0.987	0.926
ss res	43338	3956	6347	692	9542	3102	1923
n−p	36	16	16	10	9	7	3
error est	34.7	15.7	19.9	8.3	32.6	21.1	25.3

* − a third variable was used in addition to DA and Eff−P

(b) Basin Characterisic Regions/Clusters

	Size (km²) <130	Size (km²) >130 *	Clusters 1	Clusters 3	Clusters 2+4
N	13	27	24	11	5
Constant	0.673	0.520	0.457	0.629	0.972
DA	0.878	0.844	0.871	0.865	0.803
FACLS		−0.738			
Dr−Dens					
Fr−LSW					
Eff−P	0.380	1.206	1.642	0.620	1.403
r^2	0.976	0.964	0.948	0.983	0.986
r^2 corr	0.889	0.917	0.888	0.933	0.950
ss res	250	43115	57292	123	1416
n−p	10	23	21	8	2
error est	5.0	43.3	52.2	3.9	26.6

Table 5.16

Qavgfld Estimatesand Residuals – Models from Table 5.15

		Basin ID	One Rgn	WSC Y–Z	Prov ABCD	Size	Clstr 1–2–3	Qavg Obs	Std Dev
B	YO10	Jnc	16.3	16.0	8.4	23.6	20.1	11.6	4.3
C	YE1	Grvt	27.0	32.8	21.3	38.0	33.0	35.7	10.7
C	YH1	BtmC	11.9	15.0	11.7	15.2	13.9	4.9	1.2
C	YK7	Glde	30.6	39.2	29.9	43.2	38.3	21.8	7.0
C	YK8	Boot	9.7	7.5	6.8	9.7	8.8	7.4	4.2
C	YL4	SthP	100.0	39.9	13.2	23.5	23.8	27.7	15.3
C	YL5	Rtlr	15.2	13.0	11.9	9.2	15.2	11.9	7.2
B	YO7	Lch	23.9	18.8	12.1	32.4	27.4	23.6	7.1
B	YO8	GtRt*	185.0	123.6	123.1	183.8	162.7	219.0	81.2
B	YQ4	NWG*	514.3	263.8	261.9	510.8	367.7	617.6	252.1
B	ZJ2	SmCh	24.3	32.1	15.1	31.4	27.6	11.9	1.8
A	ZK3	LtBP	33.8	46.4	18.7	17.8	25.7	29.8	12.5
A	ZK4	LSNH	49.5	55.6	41.9	47.5	46.7	73.0	15.1
A	ZL4	Shrs	18.0	27.8	15.4	12.9	14.5	11.7	4.9
A	ZL5	BigB	4.3	7.9	7.3	5.7	5.2	3.8	0.8
A	ZM16	SthH	8.6	13.9	10.3	9.1	8.7	9.8	3.4
A	ZN2	StSh	8.4	13.8	9.4	8.3	7.9	7.7	2.4

Residuals

	Basin ID	One Rgn	WSC Y–Z	Prov ABCD	Size	Clstr 1–2–3
YO10	Jnc	4.7	4.4	−3.2	12.0	8.5
YE1	Grvt	−8.7	−2.9	−14.4	2.3	−2.7
YH1	BtmC	7.0	10.1	6.8	10.3	9.0
YK7	Glde	8.8	17.4	8.1	21.4	16.5
YK8	Boot	2.3	0.1	−0.6	2.3	1.4
YL4	SthP	72.3	12.2	−14.5	−4.2	−3.9
YL5	Rtlr	3.3	1.1	−0.0	−2.7	3.3
YO7	Lch	0.3	−4.8	−11.5	8.8	3.8
YO8	GtRt*	−34.0	−95.4	−95.9	−35.2	−56.3
YQ4	NWG*	−103.3	−353.8	−356	−106.8	−249.9
ZJ2	SmCh	12.4	20.2	3.2	19.5	15.7
ZK3	LtBP	4.0	16.6	−11.1	−12.0	−4.1
ZK4	LSNH	−23.5	−17.4	−31.1	−25.5	−26.3
ZL4	Shrs	6.3	16.1	3.7	1.2	2.8
ZL5	BigB	0.5	4.1	3.5	1.9	1.4
ZM16	SthH	−1.2	4.1	0.5	−0.7	−1.1
ZN2	StSh	0.7	6.1	1.7	0.6	0.2
Sum of Sqrd resds		18066	136204	137526	14647	67065
wo 2 largest		6237	1921	1800	1999	1443

Table 5.17

Test Results with Models from Table 5.15

Pseudo–T

		Basin ID	One Rgn	WSC Y–Z	Prov ABCD	Size	Clstr 1–2–3
B	YO10	Jnc	−0.6	−0.8	0.5	−2.5	−1.8
C	YE1	Grvt	0.7	0.3	1.2	−0.2	0.3
C	YH1	BtmC	−1.2	−2.4	−1.3	−4.1	−3.9
C	YK7	Glde	−1.0	−2.2	−0.9	−2.9	−2.3
C	YK8	Boot	−0.3	−0.0	0.1	−0.5	−0.3
C	YL4	SthP	−4.4	−0.8	0.9	0.3	0.2
C	YL5	Rtlr	−0.4	−0.1	0.0	0.4	−0.4
B	YO7	Lch	−0.0	0.6	1.4	−1.2	−0.5
B	YO8	GtRt*	0.4	1.2	1.2	0.4	0.7
B	YQ4	NWG*	0.4	1.4	1.4	0.4	1.0
B	ZJ2	SmCh	−2.0	−4.2	−0.6	−6.8	−5.9
A	ZK3	LtBP	−0.3	−1.2	0.8	0.9	0.3
A	ZK4	LSNH	1.4	1.1	1.9	1.7	1.7
A	ZL4	Shrs	−0.8	−2.4	−0.5	−0.2	−0.4
A	ZL5	BigB	−0.1	−0.9	−0.6	−0.8	−0.7
A	ZM16	SthH	0.2	−0.7	−0.1	0.2	0.3
A	ZN2	StSh	−0.1	−1.2	−0.3	−0.2	−0.1

Rank Sum Test

	Basin ID	One Rgn	WSC Y–Z	Prov ABCD	Size Large	Clstr 1–2–3
YO10	Jnc	3	2	1	5	4
YE1	Grvt	4	3	5	1	2
YH1	BtmC	2	4	1	5	3
YK7	Glde	2	4	1	5	3
YK8	Boot	2	4	1	5	3
YL4	SthP	5	1	2	4	3
YL5	Rtlr	5	2	1	3	4
YO7	Lch	1	3	5	4	2
YO8	GtRt*	1	.	5	2	3
YQ4	NWG*	1	4	5	2	3
ZJ2	SmCh	2	5	1	4	3
ZK3	LtBP	1	5	3	4	2
ZK4	LSNH	2	1	5	3	4
ZL4	Shrs	4	5	3	1	2
ZL5	BigB	1	5	4	3	2
ZM16	SthH	4	5	1	2	3
ZN2	StSh	3	5	4	2	1
Sum of Ranks		43	62	48	55	47

The estimates for the large basins are improved, but otherwise the results are the same or not as good as without the additional variable. The sums of squared residuals are improved for the one-region method and for the method based on size. This result occurs because including FACLS markedly improves the estimate for the two largest test basins. For most of the test basins, the estimates are no better or even worse, as shown by comparing the sums of squared residuals without the two test basins. There are 13 pseudo-T values greater than 2, compared with 8 in Table 5.14.

Neither the sum of squared residuals nor the rank sum test suggest a clear preference for one regionalization method over another - the one region method has the lowest sum of ranks, but the highest sum of squared residuals when the two largest test basins are excluded. Four of the basins in the one region equation have high leverage (SthP, NWG, LtBP and Shrs) compared with none in the equation with DA and Eff-P only.

5.7 Comments and Suggested Procedure

From the comparisons of the various equations and regionalization methods, we can conclude that dividing the island into regions can improve estimates of average flood flow at ungauged basins, particularly if Eff-P is included to represent hydrologic input. The region/clusters provided here may not be optimal, but they do offer the opportunity to make several estimates before making a choice for design.

If an estimate of an average flood is required at an ungauged basin, the following approach is suggested.

1. Measure the basin characteristics of importance. These include drainage area, area controlled by lakes and swamps, drainage density, fraction of barren and fraction of lakes and swamps; the number of characteristics depends on the region or cluster of the ungauged basin.
2. Estimate Eff-P (as described in Section 5.5.1).
3. Locate the basin in a geographic region using the map in Appendix A and in a cluster using the plots in Appendix A. Calculate standardized values using the basin characteristic data presented in Chapter 3.
4. Use the equations provided in Tables 5.6, 5.12, and 5.15 to obtain estimates of Qavgfld for the appropriate region or cluster. All of the geographic (one-region, Y-Z, ABCD) and cluster options should be considered.
5. Select the most reasonable estimate, taking into account the characteristics of the ungauged basin compared with the basins used in the data set. Use the boxplots in Appendix B as a guide, as well as the tables of characteristics presented in Chapter 3.

6 Conclusions and Recommendations

The principal conclusion of this research is that with careful selection of basin characteristics, good relationships can be obtained between flow measures and basin characteristics on the island of Newfoundland. Stronger relationships may be obtained with subdivision of the island in either geographic or basin characteristic dataspace.

A particular problem in hydrologic analysis in Newfoundland is the lack of data to represent hydrologic input, due to the lack of inland precipitation measurement stations. Because of the nature of the weather patterns affecting the island, topographic and geographic variables can be used to represent hydrologic input, and therefore should be included in the data set. They may be used either as independent variables in a regression equation to estimate a hydrologic input variable, or as surrogates in equations for other flow variables.

The basin characteristics consistently found to be important for a range of flow measures from low to high flows are

- drainage area;
- fraction of area controlled by lakes and swamps, or sometimes alternatively the fraction of the basin area occupied by lakes and swamps;

- fraction of barren area in the basin, or sometimes alternatively the fraction of forest area (inversely related to fraction of barren);
- distance from the sea in a southwesterly direction, and distance north.;
- elevation of the basin.

The last three variables are associated with the location of the basin relative to incoming weather systems and represent hydrologic input.

Other conclusions are as follows.

- **High flows:** For high flows, a more direct representation of hydrological input improves regression relationships. Effective precipitation (total precipitation minus losses) is a suitable variable, which can be estimated using topographic and geographic variables. A procedure was developed to improve the estimate using mapping and data from adjacent gauged basins as well as climate stations where available. Alternatively, the topographic and geographic variables associated with effective precipitation can be incorporated directly in the equations for other flow variables.

  For the average annual maximum daily flow, the important explanatory variables are drainage area, effective precipitation, and the fraction of the drainage area controlled by lakes and swamps. Slope also has some importance. If effective precipitation is not included as a basin characteristic, the elevation of the centroid and some measure of exposure to incoming weather systems (fraction of barren or fraction of forest) are required as surrogates.

- **Flood variability:** The linear coefficient of variation (Lcv) was used as the measure of flood variability. The relationship between Lcv and basin characteristics is not clear; there are obviously several factors interacting which would require a much larger data base to elucidate. In general higher Lcv is associated with higher flood flows, and basins in the southern part of the island (WSC Z region) tend to have higher Lcv's. In this region, Lcv tends to be related to drainage area, distance from the sea, elevation and slope. In the central and northern part of the island (WSC Y region), or when the island is treated as a whole, the important basin characteristic is the area controlled by lakes and swamps.

- **FD-10:** The flow having an exceedance of 10 percent on the flow duration curve was selected as a measure of high (but not flood) flows, in the range of about twice the mean annual flow. As with Lcv, the division into Y and Z geographic subregions resulted in somewhat stronger relationships. In the Y region, about 60 percent of the variance can be explained by drainage density, distance north, and fraction of lakes and swamps. In the Z region, much less of the variance can be explained; the two most important factors appear to be the fraction of the basin controlled by lakes and swamps and the slope.

- **Low Flows**

  The low flow measures selected were the median minimum daily flow from the annual series (both dimensional and as specific low flow), and the 90th exceedance percentile from the flow duration curve (FD-90). The findings for all measures were that the important explanatory variables are drainage area, distance north (and sometimes southwesterly distance from the sea as well), area controlled by lakes and swamps, and fraction of barren. For the low flows, effective precipitation was not especially important.

- **Availability**

  Flow duration measures were used as measures of availability or flashiness. FD-50, the median daily flow, was selected as the most suitable index of availability after consideration of some alternative indices. The findings were that the most important explanatory variables are drainage area, fraction of area controlled by lakes and swamps, and one or more of the distance and elevation variables. As with low flows, effective precipitation is not required to explain availability.

- **Regional Subdivisions - Example with Qavgfld:** The study provides an assessment of regional subdivisions for the purpose of developing regression equations to estimate the mean maximum daily flow (Qavgfld). The conclusions of this part of the study were as follows.

  1. Clustering based on basin characteristics is a promising method of regionalization. Characteristics must be carefully selected and weighted, however.
  2. Some of the geographic regionalization methods are also reasonable. Geographic regions based on similarity of hydrologic input eliminate the need to include a hydrologic input variable in predictive equations. None of the geographic regionalization methods assessed in this study eliminated the need to include a hydrologic input variable at least in some regions.
  3. Effective precipitation is a suitable variable to represent hydrologic input where required. Although there is more uncertainty in estimating effective precipitation than other basin characteristics, which may be obtained by measurement from topographic maps, the improved prediction at ungauged sites compensates for a higher uncertainty.

Recommendations

The recommendations arising from this study are as follows.

1. Estimates of flows at ungauged sites should not be made from regression equations only, unless the ungauged site is very similar in its characteristics to the gauged basins used to develop the regression equations. Nonetheless, the regression equations developed in this work may be used to provide preliminary estimates of flows of interest at ungauged sites, in particular for Qavgfld. These estimates should be modified if the location and characteristics of the ungauged site are not similar to gauged basins in the data set. The results of the regression analysis together with the discussion provided in the present work on the relevant basin characteristics can provide guidance for judgment.

2. Further investigation should be carried out into the possible improvement of regression equations using different geographic regions or clusters in basin dataspace.

3. Similar investigations should be undertaken for other flow measures besides Qavgfld, particularly in about three to five years, when about 20 additional basins will have record lengths of ten years or more and can be added to the data base.

4. The climate network should be expanded to include more climate stations inland at higher elevations.

References

Acreman, M. (1985). "Predicting the mean annual flood from basin characteristics in Scotland," *Hydrological Sciences Journal,* Vol. 30, No. 1, pp. 37-49.

Acreman, M.C. and Sinclair, C.D. (1986). "Classification of Drainage Basins According to their Physical Characteristics: An Application for Flood Frequency Analysis is Scotland," *Journal of Hydrology,* Vol. 84, pp. 365-380.

Andrews, D.F. (1972). "Plots of High-Dimensional Data," *Biometrics,* Vol. 28, pp. 125-136.

Arnell, N. and Beran, M. (1987). "Testing the Suitability of the Two-component Extreme Value Distribution for Regional Flood Estimation," *Regional Flood Frequency Analysis,* ed. V.P. Singh, D. Reidel Publlishing Co., pp. 159-175.

Bhasker, N. R. and O'Connor, C.A. (1989). "Comparison of Method of Residuals and Cluster Analysis for Flood Regionalization" *Journal of Water Resources Planning and Management,* ASCE, Vol 115, No. 6, pp. 793-808.

Blake, G.J., Cook, A.D. and Greenall, D.H,. (1973). "The Use of Principal Component Factor Analysis to Establish the Uniformity of a Hydrological Region in Northland, New Zealand," *Results of Research on Representative and Experimental Basins,* Proceedings of the Wellington Symposium, Dec 1970, Vol 1. IAHS-UNSECO, Paris.

Burn, D.H. (1988). "Delineation of Groups for Regional Flood Frequency Analysis," *Journal of Hydrology,* Vol. 104, pp. 345-361.

Burn, D.H. (1989). "Cluster Analysis as applied to Regional Flood Frequency", *Journal of Water Resources Planning and Management,* ASCE, 115 (5), 576-582.

Burn, D.H. (1990). "An Appraisal of the 'Region of Influence' Approach to Flood Frequency Analysis," *Hydrological Sciences Journal,* Vol. 35, No. 2, pp. 149-165.

Burn, D.H. and Boorman, D.B. (1993). "Estimation of Hydrological Parameters at Ungauged Catchments," *Journal of Hydrology*, Vol. 143, pp. 429-454.

Caissie, D. and El-Jabi, N. (1991). "A Stochastic Study of Floods in Canada: Frequency Analysis and Regionalization," *Canadian Journal of Civil Engineering*, Vol. 18, pp. 225-236.

Cavadias, G.S. (1989). "Regional Flood Estimation by Canonical Correlation," *Canadian Society of Civil Engineering Annual Conference and Ninth Canadian Hydrotechnical Conference*, Vol. XXX, pp. 212 - 231.

Cavadias, G.S. (1990). "The Canonical Correlation Approach to Regional Flood Estimation", *Regionalization in Hydrology: Proceedings of the Ljubljana Symposium, April 1990*, IAHS Publication no. 191, p. 171-178.

Cavadid, L., Cunnane, C., Salas, J.D. and Boes, D.C. (1991). "Flood Frequency Analysis Based on the Box-Cox Transformation," *Water Resources Paper No. 103*, Colorado State University, Fort Collins, Colorado.

Chow, V.T. (Ed), (1964). *Handbook of Hydrology.* McGraw Hill, New York.

Chow, V.T., Maidment, D. R. and Mays, M. (1989). *Applied Hydrology.* McGraw Hill, New York.

Condie, R., Pilon, P.J., Harvey, K.D. and Goertz, H. (1986). "Comparison of Regional Flood Frequency Methods in Southern Ontario using Analysis of Variance Techniques", *Regional Flood Frequency Analysis: Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, ed. V.P. Singh, D. Reidel Publishing Co., pp. 213-222.

Condie, R., (1979). "The Three Parameter Lognormal Distribution Applied to Regional Flood Frequency Analysis by the Index Flood Method", Inland Waters Directorate Technical Workshop in Modelling Activities Related to the Flood Damage Reduction Program, Hull, Quebec.

Cong, S., Le, Y., Vogel, J.L., and Schaake, J.C. (1993). "Identification of the Underlying Distribution Form of Precipitation Using Regional Data," *Water Resources Research*, Vol. 29, No. 4, pp. 1103-1111.

Cunnane, C. (1988). "Methods and Merits of Regional Flood Frequency Analysis," *Journal of Hydrology*, Vol. 100, pp. 269-290.

Dalrymple, T. (Ed), 1960. "Flood Frequency Analysis, Manual of Hydrology, Pt #, Flood Flow Techniques," *U.S. Geological Survey Water Supply Paper 1543-A.*

Everitt, B. (1978). *Graphical Techniques for Multivariate Data.* North-Holland, New York, 117p.

Fennessey, W. and Vogel, R.M. (1990). "Regional Flow Duration Curves for Ungauged Sites in Massachusetts', *Journal of Water Resources Planning and Management*, ASCE, Vol. 116, No. 4, pp 530-549.

Fiorentino, M., Gabriele, S., Rossi, F., and Versace P., (1986). "Hierarchical Approach for Regional Flood Frequency Analysis", *Regional Flood Frequency Analysis: Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, ed. V.P. Singh, D. Reidel Publishing Co., pp. 35-50.

Gabriele, S. and Arnell, N. (1981). "A Hierarchical Approach to Regional Flood Frequency Analysis", *Water Resources Research*, Vol. 27, No. 6, pp. 1281-1289.

Gingras, D. and Adamowski, K. (1993). "Homogeneous Region Delineation Based on Annual Flood Generation Mechanism," *Hydrologic Sciences Journal*, Vol. 38, No. 2, pp. 103.121.

Government of Canada, Department of the Environment, Water Planning and Operations Branch (1971). "Flood Frequency Analysis for Newfoundland Streams", Prepared by Roger Y. Poulin.

Government of Canada, Atlantic Development Board, (1968). *Water Resources Study of the Province of Newfoundland and Labrador*, Prepared by Shawinigan Engineering Company Limited and James F. MacLaren Ltd.

Government of Canada, Department of Energy, Mines and Resources (1970). *Hydrometric Network Plan for the Provinces of Newfoundland, New Brunswick, Nova Scotia and Prince Edward Island*. Prepared by J. Ingledow and Associates Ltd, 2 volumes.

Government of Canada, Inland Waters Directorate (1984). *Hydrologic Design Methodologies for Small Scale Hydro at Ungauged Sites - Phase 1*. Prepared by Acres Consulting Services Ltd.

Government of Canada, Inland Waters Directorate (1985). *Hydrologic Design Methodologies for Small Scale Hydro at Ungauged Sites - Phase II: Feasibility Level Study for Atlantic Provinces*. Prepared by Acres Consulting Services Ltd, 226 pp.

Government of Newfoundland and Labrador, Department of Environment, Water Resources Division and Environment Canada (1984). *Regional Flood Frequency Analysis for the Island of Newfoundland*.

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1987). *Water Resources Study of the Eastern Avalon Peninsula, WRD-SW-1-1*. Prepared by Acres Consulting Services Ltd, St. John's, Nfld.

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1988). *Water Resources Study of the Western Avalon Peninsula, WRD-SW-1-2*. Prepared by Acres Consulting Services Ltd, in association with Colin Karasek Ltd., St. John's Nfld.

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1989). *Water Resources Study of the Bonavista Bay Area, WRD-SW-1-3*. Prepared by ShawMont Newfoundland Ltd, St. John's, Nfld.

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1990a). *Regional Flood Frequency Analysis for the Island of Newfoundland, WRD-HM-90-1.*

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1990b). *Water Resources Study of the Northern Peninsula and Humber Valley, WRD-SW-1-4.* Prepared by Acres International Ltd., St. John's, Nfld.

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1991a). "Characteristics and Estimation of Minimum Streamflows for the Island of Newfoundland", *WRD-HM-91-111.*

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1991b). *Water Resources Study of the Notre Dame Bay Area and Central Newfoundland Region, WRD-SW-1-5.* Prepared by ShawMont Newfoundland Ltd, St. John's, Nfld.

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1992). *Water Resources Atlas of Newfoundland.*

Government of Newfoundland and Labrador, Department of Environment and Lands, Water Resources Division (1993). *Water Resources Study of the Burin Peninsula and Fortune Bay Area, WRD-SW-1-6.* Prepared by Acres International Ltd. in association with Colin Karasek Ltd., St. John's, Nfld.

Greis, N.P. and Wood, E.F. (1981). "Regional Flood Frequency Estimation and Network Design", *Water Resources Research,* Vol. 17, No. 4, pp. 1167-1177.

Greis, N.P. and Wood, E.F. (1983). "Correction to 'Regional Flood Frequency Estimation and Network Design' ", *Water Resources Research,* Vol. 19, No. 2, pp. 589-590.

Guo, X., Salas, J.D. and Boes, D.C. (1990). "Regional Flood Frequency Analysis," *Water Resources Paper No. 102*, Colorado State University, Fort Collins, Colorado, 122 p.

Haines, A.T. and Finlayson, B.L. (1988). "A Global Classification of River Regimes," *Applied Geography*, Vol. 8, pp. 255-272.

Hawley, M.E. and McCuen, R.H. (1982). "Water Yield Estimation in Western United States," *Journal of the Irrigation and Drainage Division*, ASCE, Vol. 108, No. IR1, pp. 25-34.

Helsel, D.R. and Hirsch, R.M. (1992). *Statistical Methods in Water Resources*. Elsevier, New York.

Heo, J., Boes, D.C. and Salas, J.D. (1990). "Regional Flood Frequency Modeling and Estimation," *Water Resources Paper No. 101*, Colorado State University, Fort Collins, Colorado, 132p.

Hosking, J.M., Wallis, J.R., and Wood, E.F. (1985). "An Appraisal of the Flood Frequency Procedure in the UK," *Hydrological Sciences Journal*, Vol. 30, pp. 85-109.

Hosking, J.R.M. (1990). "L-Moments: Analysis and Estimation of Distribtuions Using Linear Combinations of Order Statistics,", Journal of the Royal Statistical Society, B, Vol 52, No 2, pp 105-124.

Hosking, J.R.M. and Wallis, J.R. (1988). "The Effect of Intersite Dependence on Regional Flood Frequency Analysis", *Water Resources Research*, Vol. 24, No. 4, pp. 588-600.

Hosking, J.R.M. and Wallis,J.R. (1992). "Regional Frequency Analysis Using L-Moments," *Research Report*, IBM Research Division, Yorkton Heights, N.Y. 6 p.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). "An Appraisal of the Regional Flood Frequency Analysis in the U.k. Flood Studies Report", *Hydrological Sciences Journal*, Vol. 30, No. 1, pp. 85-109.

Hughes, J.M.R. (1987). "Hydrological Characteristics and Classification of Tasmanian Rivers," *Australian Geographical Series,* Vol. 25, #1, pp. 61-82.

Kuczera, G. (1982). "Combining Site-specific and Regional Information: An Empriical Bayes Approach," *Water Resources Research,* Vol. 18, pp. 306-314.

Kuczera, G. (1983). "A Bayesian Surrogate for Regional Skew in Flood Frequency Analysis." *Water Resources Research,* Vol. 19, pp. 821-832.

Lye, L.M. and Moore, E. (1991). " 'Instantaneous Peak Flow Estimation Procedures for Newfoundland Streams' by U.S. Panu and D. A. Smith': Discussion", *Water Resources Bulletin,* Vol 27, No. 1, p. 125-127.

MacLaren Atlantic Limited (1980). "Regional Flood Frequency Analysis for Mainland Nova Scotia Streams," *Canada-Nova Scotia Flood Damage Reduction Program,* xxx p. xxx

Maidment, D.R. (1993). *Handbook of Hydrology. McGraw Hill, New York.*

Matalas, N.C., Slack, J.R. and Wallis, J.R. (1975). "Regional Skew in Search of a Parent", *Water Resources Research,* Vol. 11, No. 6, pp. 815-826.

McCuen, R.H., Leahy, R.B., and Johnson, P.A. (1990). "Problems with Logarithmic Transfromations in Regression", ASCE *Journal of Hydraulic Engineering,* Vol. 116, No. 3, pp 414 - 428.

Mimikou, M. and Kaemaki, S., (1985). "Regionalization of F-D Characteristics", *Journal of Hydrology,* Vol. 82, pp. 75-91.

Mosley, M.P. (1981). "Delimitation of New Zealand Hydrologic Regions," *Journal of Hydrology,* Vol. 49, pp. 173-192.

Nathan, R.J. and McMahon, T.A. (1990). "Identification of Homogeneous Regions for the Purposes of Regionalisation," *Journal of Hydrology,* Vol. 121, pp. 217-238.

Pearson, C.P. (1991). "New Zealand Regional Flood Frequency Analysis Using L-Moments," *New Zealand Journal of Hydrology*, Vol. 30, No. 2, pp. 53-63.

Pilon, P.F., Alila, Y, and Adamowski, K. (1991). "Assessment of Risk of Flooding Based on Regional Information". *Proceedings of the NATO/ASI Conference on Water Resources and Environmental Engineering*. Porto Carras, Greece.

Pilon, P.J. and Adamowski, K. (1992). "The Value of Regional Information to Flood Frequency Analysis Using the Method of L-Moments," *Canadian Journal of Civil Engineering*, Vol. 19, pp. 139-147.

Pilon, P.J.(1990). "The Weibull Distribution Applied to Low Flow Frequency Analysis", *Regionalization in Hydrology: Proceedings of the Ljubljana Symposium, April 1990,* IAHS Publication no. 191, p. 227-237.

Potter, K.W. (1987). " Research on Flood Frequency Analysis: 1983-1986," *Reviews of Geophysics*, Vol. 25, No. 2, pp. 113-118.

Raines, T.H. and Valdes, J.B. (1993). "Estimation of Flood Frequencies for Ungauged Catchments", *Journal of the Hydraulics Division"*, ASCE, Vol 119, No. 10, pp. 1138-1154.

Riggs, H.G. (1973). "Regional Analyses of Streamflow Characteristics', *Techniques of Water Resources Investigations*, Book 4, Ch B3, USGS, Washington D.C.

Riggs, H.C. (1990). "Regionalization in Hydrology", *Regionalization in Hydrology: Proceedings of the Ljubljana Symposium*, April 1990, IAHS Publication no. 191, p. 159 -169.

Rossi, F., Fiorentino, M., and Versace, P. (1984). "Two-Component Extreme Value Distribution for Flood Frequency Analysis", *Water Resources Research*, Vol. 20, No.7 pp. 847-856.

Rousseeuw, P. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. John Wiley and Sons, N.Y.

Sceviour, E. and Lye, L.M. (1993). "Hydrometric Data Analysis for the Purpose of Regionalisation", Unpublished.

Schaefer, Melvin G. (1983). "Regional Hydrometeorological Analysis for Spillway Design" *International Symposium on Hydrometeorology,* June, 1982, American Water Resources Association, pp. 523-530.

Schaefer, M.G. (1990). "Regional Analyses of Precipitation Annual Maxima in Washington State", *Water Resources Research,* Vol. 26, No. 1, pp. 119-131.

Sharp, J.J. and Moore, E. (1988). "Regional Flood Frequency for the Island of Newfoundland: Analysis Using Principal Components." Faculty of Engineering, Memorial University of Newfoundland, Unpublished, 26 p.

Tasker, G.D. (1982a). "Simplified Testing of Hydrologic Regression Regions," *Journal of the Hydraulic Division,* ASCE, Vol. 108, No. HY10, pp. 1218-1221.

Tasker, G.D. (1982b). "Comparing Methods of Hydrologic Regionalization," *Water Resources Bulletin,* Vol. 18, No. 6, pp. 965-970.

Tasker, G.D. (1986). "Regional Analysis of Flood Frequencies", *Regional Flood Frequency Analysis: Proceedings of the International Symposium on Flood Frequency and Risk Analysis,* ed. V.P. Singh, D. Reidel Publishing Co., pp. 1-10.

Tasker, G.D., Eychaner, J.H. and Stedinger, J.R. (1986). "Application of Generalized Least Squares in Regional Hydrologic Regression Analysis", *USGS Water Supply Paper 2310,* pp. 107-115.

Thomas, D.M. and Benson, M.A. (1970). "Generalization of Streamflow Characteristics from Drainage Basins Characteristics," *U.S. Geological Survey Water Supply Paper No. 1975.* 55 p.

Vogel, R.M. (1992). "Regional Geohydrologic-Geomorphic Relationships for the Estimation of Low Flow Statistics," *Water Resources Research,* Vol. 28, No. 9, pp. 2451-2458.

Vogel, McMahon et al (1993).

Vogel, R.M. and Fennessey, N.M. (1993). "L-Moment Diagrams Should Replace Product Moment Diagrams", *Water Resources Research,* Vol. 29, No. 6, pp 1745-1752.

Wallis, J.R., Matalas, N.C. and Slack, J.R. (1974). "Just a Moment", *Water Resources Research,* Vol 10, No. 2, pp 211-219.

Wallis, J.R., Matalas, N.C. and Slack, J.R. (1977). "Apparent Regional Skew", *Water Resources Research,* Vol. 13, No. 1, pp. 159-182.

Wallis, J. R. and Wood, E.F. (1985). "Relative Accuracy of Log Pearson III Procedures", ASCE *Journal of Hydraulic Engineering,* Vol. 111, No. 7, pp 1043 - 1056.

White, Elizabeth. (1975). "Factor Analysis of Drainage Basin Properties: Classification of Flood Behaviour in Terms of Basin Geomorphology," *Water Resources Bulletin* Vol. 11, No. 4, pp. 676-687.

Wilkinson, Leland (1990). *SYSTAT: The System for Statistics*. Evanston, IL: SYSTAT, Inc.

Wiltshire, S.E. (1986a). "Regional Flood Frequency Analysis 11: Multivariate Classification of Drainage Basins in Britain," *Hydrological Sciences Journal,* Vol. 31, pp. 335-346.

Wiltshire, S.E. (1986b). "Identification of Homogeneous Regions for Flood Frequency Analysis," *Journal of Hydrology,* Vol. 84, pp. 287-302.

Wiltshire, S. and Beran, M. (1987a). "Multivariate Techniques for the Identification of Homogeneous Flood Frequency Regions," *Regional Flood Frequency Analysis: Proceedings of the International Symposium on Flood Frequency and Risk Analysis,* ed. V.P. Singh, D. Reidel Publishing Co., pp. 133-146.

Wiltshire S. and Beran, M. (1987b). "A Significance Test for Homogeneity of Flood Frequency Regions", *Regional Flood Frequency Analysis: Proceedings of the International Symposium on Flood Frequency and Risk Analysis*, ed. V.P. Singh, D. Reidel Publishing Co., pp. 147-158.

Wiltshire, S.E. (1986c). "Regional Flood Frequency Analysis 1: Homogeneity statistics," *Hydrological Sciences Journal*, Vol. 31, pp. 321-333.

Wood, E.F., and Rodriguez-Itrube, I., (1975). "Bayesian Inference and Decision Making for Extreme Hydrologic Events." *Water Resources Research.*, 11(4), pp. 533-542.

Wu, B. (1988). "Bayesian Approaches to Inference of Probabilities of Extreme Hydrologic Events". *M. Sc. thesis*, University College, Galway, Ireland, 99 pages.

Zrinji, Zolt, and Burn, D.H. (1994). "Flood Frequency Analysis for Ungauged Sites Using a Region of Influence Approach", *Journal of Hydrology*, Vol. 153, pp. 1 1-21.