Using Tree Ring Data to Assess Drought

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

Master of Science

By

David Victor Bonin

Department of Civil Engineering
The University of Manitoba

0-612-79932-8

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

USING TREE RING DATA TO ASSESS DROUGHT

BY

DAVID VICTOR BONIN

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree

of

Master of Science

Abstract

The reconstruction of past flow events is of great interest to the water resources engineer. Water resources planning requires the best possible estimates of extreme flow conditions for investment, decision making and design. Unfortunately streamflow records in North America tend to be very short. The multi-year nature of drought events reduces the available sample size making estimates of extreme droughts very difficult.

Tree ring data offer a unique way of addressing this problem. The pattern of a tree's growth rings reflect the environmental conditions experienced during each year. In addition, trees are relatively long lived (up to 500 years) and well distributed in North America. Tree rings are produced annually and can be precisely and reliably linked to climatic variations. This makes them unique and ideal for correlation with annual climatic records.

The purpose of this thesis is to show the utility of using the methods of dendroclimatology, the study of climate through tree rings, to extend streamflow records. These methods use the principle that during drought periods moisture stress proportionally limits tree growth. This limitation is reflected in the width variation of annual growth rings. The climatic information inherently present in ring widths can then be used to extend historical records of low flow back the entire lifetime of the tree.

Three case studies were completed, one in the MacKenzie River Basin and two in the South Saskatchewan River Basin. Two of the case studies verified very well using split sample techniques, one was questionable. The reconstructions extended streamflow records from 59 to 190 years, from 60 to 420 years and from 65 to 352 years.

The results of a comparison between extreme droughts estimated from the gauged data and the reconstructed data showed a decrease in drought severity at all return periods. This was a result of the reconstruction models not being able to reproduce the amount of variance found in the gauged data. The magnitude of streamflow records are smoothed as they are filtered through the tree ring data. The data reconstructed in this study cannot be used in quantitative frequency analysis of extreme drought. Further study is required to determine if it is possible to produce a reconstruction with sufficient explained variance to perform quantitative frequency analysis

Acknowledgements

The support, understanding and patience of many people over the last seven years have made the completion of this thesis possible.

I would like to thank Dr. D. Burn, my thesis advisor, who provided me with his knowledge and guidance and supported me well beyond his obligations. His immense patience and understanding through many difficult times made the completion of this thesis a reality.

My thanks go out to Mr. Bill Girling at Manitoba Hydro for his support and encouragement

Many thanks to Manitoba Hydro who provided the financial assistance that made this project possible.

I would like to thank Acres Manitoba Limited, especially Mr. Joe Groeneveld who allowed me the time off to finish this document and was very supportive in having me finish.

Finally, I would like to thank Gale, my loving wife. Through the long nights and tears you always stuck by me. I am forever grateful.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Purpose

The purpose of this research is to show the utility of using the methods of dendroclimatology, the study of climate through tree rings, to extend streamflow records. These methods use the principle that during drought, or low flow, periods moisture stress proportionally limits tree growth. This limitation is reflected in the width variation of annual growth rings. The climatic information inherently present in ring widths can then be used to extend historical records of low flow back the entire lifetime of the tree.

The objective of this project is to use existing tree ring data, that has been collected from various researchers and deposited in the International Tree Ring Data Bank, to extend streamflow records in and around the Churchill–Nelson River basin. If successful, the techniques used here could help provide better understanding of past drought, assist in the operation of current hydropower projects and assist in the design and planning of future hydropower projects.

1.2 Background

The reconstruction of past flow events is of great interest to the water resources engineer. Water resources planning requires the best possible estimates of extreme flow conditions for investment, decision making and design. For this reason, periods of low flow are of particular interest for hydropower developers and operators.

Low flow estimates govern the estimate of how much firm power can be developed at a hydro site. Poor estimates of firm discharge can lead to over–development, under–development or no development of a potential hydro site.

The problem that each water resources planner in North America has to deal with is that streamflow records are generally very short (usually less than 50 years). Compounding this is the fact that periods of low flow, or drought, usually take place over several years. A single 'event' may take up multiple years of the record (a multiyear event) further reducing the size of the available sample data.

With such small data sets, probability distributions used to perform frequency analysis may provide misleading or erroneous results. This is particularly true if the period of record coincides with a period of anomalous rainfall or runoff.

There are several techniques that have been used to address the problem of short records in drought estimation. Two broad categories of streamflow generation models are used to fill in missing data and extend records: deterministic and stochastic.

Deterministic models are based on the physical characteristics of a drainage basin and hydrologic relationships to translate meteorological records into streamflow. These methods are particularly useful since in most areas the rainfall record is significantly longer than the streamflow record. Some examples of deterministic models in common use are SSARR, HEC-1 and SLURP.

Stochastic models are statistically based and rely on the statistical properties of the available streamflow record or cross-correlation with other, longer, streamflow records. Stochastic techniques include regional regression analysis and synthetic streamflow generation.

One popular technique that falls outside of these two categories is the use of the so-called 'drought of record'. In this technique, project capacity is designed according to the worst drought of recorded history. This does not attempt to address short streamflow records but selects only the low observation from it.

Tree ring data offers a possible method of addressing the problem of short streamflow records. The pattern of a tree's growth rings reflect the environmental conditions experienced during each year of growth. In addition, trees are relatively long lived (up to 500 years) and well distributed in North America. Tree rings are only one source of paleoclimatic data. However, they have the advantage of being annually based, precisely and reliably linked to climatic variations. This makes them a unique source of paleoclimatic data for correlation with annual climatic records. They have been used to augment numerous climatic and hydrologic measurements including temperature, precipitation, Palmer Drought Severity Index (PDSI), streamflow, lake levels and atmospheric circulation (Duvick and Blasing, 1981).

The use of tree rings as a tool for extending climate records is made possible through the 'principle of limiting factors' which governs the tree's annual growth. This principle states that a biological process may not proceed any faster than allowed by the most limiting factor. For tree rings this could be either temperature, moisture, nutrient availability, insect infestation, etc. (Stahle and Cleaveland, 1988).

The amount of growth experienced by an individual tree is affected by many environmental and biological factors. The most significant of these are climatic forcings of temperature and moisture. How strongly trees are affected by these is dependent upon tree species and location. In dry years, where moisture levels limit tree growth, narrower rings are formed. The width of these are proportional to the amount of moisture present during that year. In wet years, where moisture does not limit growth, wider rings are formed, limited in width by some other factor. Hence, the use of tree ring data is most effective for reconstruction of drought related streamflow events because they produce the best correlation with tree growth.

Traditional streamflow record augmentation is often accomplished by exploiting the cross correlation with nearby flow recording gauges. The period of common record between the gauges is used to form a relationship and this is used to extend the shorter record to the length of the longer record. In streamflow reconstruction using tree rings the same general procedure is used except the nearby 'gauges' are tree ring sites (Brockway and Bradley, 1995). The statistical procedures, however, tend to be much more involved than a simple cross correlation analysis.

1.3 Objectives

The scope of this study involves:

- Exploring an approach for using the paleoclimatic information to characterize drought events. Chapters 2 and 3 will discuss the statistical techniques and rationale behind the methods used in reconstructing streamflow using tree rings.
- Determining the availability of tree ring data within and near the study area. Chapter 4 will discuss the available sources of paleoclimatic data.
- Automating the tasks involved in reconstructing streamflow from tree rings. Chapter 5 will discuss several computer applications developed to expedite the reconstruction process.
- Employing the approaches in case studies to demonstrate the feasibility of using tree ring data to reconstruct drought. Chapter 6 will discuss several case studies where the techniques described in the previous chapters are employed.
- Drawing conclusions about the utility of the approaches used. Chapter 7 discusses the outcomes of the case studies and implications for future work.

Chapter 2

Literature Review

This thesis builds upon literature found in the fields of general dendroclimatology, streamflow dendroclimatology and statistical drought analysis. General dendroclimatology presents the basic tools needed to relate climatic parameters to tree rings. It presents the means to choose the tree ring data sets and process them for climate reconstruction. It also presents the multivariate methods that have been successful in climate reconstructions. Streamflow dendroclimatology shows that it is possible to successfully reconstruct streamflow records using tree rings. Statistical drought analysis provides the reason for wanting to reconstruct past streamflow and the means to gauge the success of the reconstructions.

2.1 Tree Ring Data Analysis

Tree ring data have been used for the past 30 years to extend climatic records. Fritts (1971) was a leader in this area and defined the principles by which climate reconstructions could be made. Fritts et al. (1971) were the first to discuss the use of multivariate statistics in tree ring - climate reconstructions. LaMarche (1974) explained some of the inferences that can be made from long tree ring records. Fritts (1976) wrote *Tree Rings and Climate* the definitive book on relating tree ring data to climate. This is one of the most comprehensive books ever published on the subject. More recently Cook and Kairiukstis (1989) wrote the book *Methods of Dendrochronology* that updates and expands on several of the techniques presented by Fritts.

Fritts (1971, 1976) discusses the principles that make tree ring reconstruction of climatic variables possible. The term dendrochronology is defined as the science of dating the annual growth layers in woody plants and the exploitation of the information they contain on the environment. Dendroclimatology is restricted to dendrochonological studies that use climatic information from dated growth layers to study variability in present and past climate. The principles of site selection, sensitivity and cross dating that make tree ring climate relationships possible are discussed below.

The principle of site selection involves using information from a large sample of trees where growth has been limited by the climatic factor in question. The principle of sensitivity is where the person sampling chooses trees that exhibit the most variability in width from one ring to the next. These provide the best indicators of climatic stress.

Cross dating is a procedure that allows the identification of the year in which individual rings are formed. It involves taking a tree of known cutting or coring date and comparing the ring width patterns with those of unknown or known cutting dates to locate them precisely in time or verify the dates of the rings.

Fritts (1976) also discusses that as trees grow the annual rings systematically become thinner with increasing trunk diameter. In order to remove this trend ring widths are standardized into index values. This is accomplished by fitting either an exponential or straight line to the systematic non-climactic effects. Individual ring width values are then divided by the corresponding value of the fitted curve forming what is referred to as a standardized ring width index. Many index series from trees in a localized area are then added together and averaged to form what is called a tree ring chronology.

Fritts et al. (1971) and Fritts (1976) discuss how multivariate techniques provide a way of objectively defining how the ring width growth relates to climatic factors at different periods during the growing season. Multivariate statistics also provide a means of handling and relating data sets consisting of correlated variables. They discuss the use of orthogonal eigenvectors, derived from groups of tree ring and climatic predictors, for dealing with correlated variables and for reducing the number of variables. These concepts are discussed in Chapter 3.

LaMarche (1974) discusses how tree rings are an important source of paleoclimatic data because they are long lived, can be accurately dated and can vary in response to climate changes. He also discusses how cross correlation can be enhanced, when lower frequencies are removed by prewhitening, using the methods presented in Chapter 3. He also reinforces how the best climatic response occurs in trees near climatically determined limits of the distribution, such as near tree lines on mountains and in northern regions.

Cook and Kairiukstis (1989) expand on the concepts presented above and emphasize procedures that can be used in reconstructing climate series. Particular attention is paid to proper cross validation of regression models and the use of verification statistics to gauge the validity of the models.

2.2 Reconstruction of Streamflow Using Tree Rings

Tree ring chronologies are particularly well suited to reconstruct runoff records. They tend to be much longer than instrumental records, they are precisely annual in resolution and they integrate the effects of temperature, precipitation and evapotransporation, the main components that influence streamflow.

Many studies have been done over the last 30 years using tree rings to reconstruct streamflow. The earliest studies by Stockton (1975) were done to gauge the viability of relating streamflow to tree ring chronologies as well as to investigate the various statistical procedures that could be used. Most of the subsequent studies (Phipps, 1983; Cook and Jacoby, 1983; Meko and Graybill, 1995; Meko et al., 2001; Woodhouse, 2001) were done in arid locations to assist in water supply allocation to populated areas. In these locations low flow allocation is a major concern for water resource planners. Longer streamflow records are helpful to better quantify low frequency drought events and make informed policy decisions. Some studies (Smith and Stockton, 1981; Cleaveland and Stahle, 1989) were done in order to assess long term, low probability high flows in addition to low flow events. These were done to better assess low probability flood potential and surplus water allocation. These studies use various statistical approaches to tree ring reconstruction and verification. The following summarizes the approaches used in each study, their purpose, results and conclusions.

Stockton (1975) presented one of the earliest studies. In this study he investigated the use of various statistical procedures in reconstructing streamflow and precipitation as well as the applicability of reconstructing runoff series. Stockton investigated the used of correlation analysis, spectral analysis, analysis of variance, principal components analysis and multiple linear regression. He reconstructed annual streamflow and precipitation on the Bright Angel Creek and Upper Colorado River basins in Arizona and New Mexico from 1564 to 1960.

Phipps (1983) reconstructed monthly summer streamflow on the Occoquan River in Virginia between 1841 and 1975. The purpose of this project was to quantify low flow events to assist in planning of water supplies to populated areas in the region. A monthly streamflow record was available for calibration between 1928 and 1976. Stepwise regression was used to relate lagged tree ring chronologies to monthly streamflow. This study showed R^2 values of between 0.33 and 0.47 with no independent verification. It found that most significant droughts in this area occurred within the gauged record.

Cook and Jacoby (1983) reconstructed streamflow on the Potomac River between 1730 and 1976. The purpose of this project was to provide insight into water supply problems and solutions for Washington, D.C. The goal was to see if the gauged records were representative of long term streamflow in the area. Monthly streamflow records were available between 1907 and 1977. July, August and September streamflow were reconstructed using Stepwise Canonical Regression techniques on prewhitened and lagged streamflow chronologies. The results showed R^2 that varied between 0.28 and 0.48 with a pooled R^2 adjusted for degrees of freedom of 0.36. This reconstruction was verified using independent data and the product moment correlation coefficient and reduction of error verification statistics. This study showed a much better reconstruction from the use of the verification statistics than was implied by the R^2. It showed the danger of using R^2 as the only method of determining calibration reliability. The study identified several periods of persistent low flow prior to the gauged records and showed the potential utility of such studies for planners.

Meko and Graybill (1995) reconstructed streamflow on the Upper Gila River Basin in Arizona and New Mexico between 1663 and 1985. The purpose of this project was to extend the short gauged record to assist in water planning and allocation. The gauged streamflow records extended from 1915-1985. Eight chronologies were prewhitened using a low order autoregressive moving average process and the skewness was removed. These were then lagged forward and back in time and related to streamflow using stepwise regression. The stepwise regression model yielded an R^2 of 0.66. Split sample verification was used to gauge the validity of the model. R^2 values between 0.58 and 0.69 were found for the verification models. The split sample model coefficients were compared for time stability and residuals were analyzed to investigate regression quality. The product moment correlation coefficient and reduction of error statistic were used to validate the split sample models. The conclusion of this study was that the 20th century had an unusually large number of instances of clustered high flow years and high severity multiyear droughts.

Meko et al. (2001) reconstructed streamflow on the Sacramento River in California between 869 and 1999. The purpose of this study was to gain a long term perspective of drought for water allocation planning. The gauged record from 1906-1999 was too short to represent low frequency persistent climate fluctuations. The gauged streamflow record was created by summing the records of four tributaries. Tree ring chronologies were obtained from the International Tree Ring Data Bank. The chronologies were prewhitened using a low order autoregressive process and principal component analysis was used to deal with intercorrelations between tree ring data sets. Stepwise regression was used to relate the principal components of the tree ring chronologies to log streamflow. R^2 values ranged from 0.64 to 0.81. Cross validation was accomplished using the PRESS, RSME and reduction of error statistics. The conclusion to this study was that the use of the 1930s as a design drought is justified. Although there were several more extreme droughts in the past, they were of shorter duration.

Woodhouse (2001) reconstructed streamflow for the Middle Boulder Creek basin in Colorado between 1703 and 1980. The purpose of this study was to gauge the uniqueness of the 20th century low flow events. The gauged record from 1912-1980 was inadequate to assess the low frequency variability in flow fluctuations and allow effective water policy decisions. Tree ring data were prewhitened with a low order autoregressive process but were not orthogonalized using principal component analysis due to the tendency of that procedure to mask some climate signal. Stepwise regression was used to pick a final model with an R^2 value of 0.7.

Split sample verification was used to validate the model using the reduction of error and sign test verification statistics. The conclusion of this study was that the extended record expands the perspective of streamflow variability and provides a tool for water planning and allocation.

Smith and Stockton (1981) reconstructed streamflow on the Salt and Verde Rivers in Arizona between 1580 and 1990. The goal of this project was to confirm long term flood potential and the statistics used in water allocation to populated areas and for hydropower production. Gauged records were available for the Salt and Verde River Basins from 1914 to 1979 and 1895 to 1979 respectively. The tree ring chronologies were prewhitened using ARMA models and tranformed to orthogonal variables using principal components analysis. Stepwise regression against log tranformed streamflow was used to choose a final model. The final model had an R^2 value of 0.73. No discussion of model verification was made. The study found that gauged records contained a large number of high flow events compared to the extended record. Also several extreme and persistent low flow events were found prior to the gauged records.

Cleaveland and Stahle (1989) reconstructed streamflow on the White River in Arkansas between 1700 and 1980. The purpose of this study was to investigate the viability of inter-basin diversion of surplus water. This area is one of the most variable runoff regions in the United States. Gauged data were available from 1931 to 1980. Correlation analysis was used to determine the best season to reconstruct. Annual streamflow was found to yield the highest correlation with tree ring data. The tree ring chronologies were regionalized using a simple averaging. They were prewhitened using a low order autoregressive process. Regression analysis yielded an R^2 value of 0.5. The regression was validated using a standard split sample procedure and the reduction of error verification statistic. In addition, coefficients and moments were compared between the split sample models and the full model to assess time stability. The conclusion of this study showed that tree ring reconstruction could be an important tool in assessing the probability forecasts for the basin.

2.3 Characterizing Drought

Drought, or low flow, events are much more difficult to characterize and analyze than floods. This is due in part to the 'multiyear' nature of drought events where a single event can span many years. The effects of droughts are also much more difficult to quantify. The dramatic effects experienced during floods in the form of property damage and lives lost are not present for droughts (Jarrett, 1991). Drought effects are largely measured in terms of distributed economic loss to an industry or environmental losses, both of which are difficult to quantify.

Much research had been done to properly analyze drought events in a way similar to flood events. Dracup et al. (1980a, 1980b) defined a way in which droughts can be statistically characterized. Joseph (1970) defined a simple method of determining frequency of design drought for water resources projects. Burn and DeWit (1996) expand on this methodology to take into account the multiyear nature of droughts.

Dracup et al. (1980a, 1980b) present a method of statistically characterizing drought events. In their method, four decisions must be made to clearly define a drought event (Dracup et al., 1980b).

The nature of the water deficit must be characterized as either hydrologic (streamflow), meteorological (precipitation) or agricultural (soil moisture) (Dracup et al., 1980a). The parameter of interest is decided by the purpose of the analysis. If the *causes* of drought are of interest, meteorological drought needs to be evaluated. If drought *impacts* are to be quantified, then either hydrologic or agricultural drought is investigated, based on the type of impacts of interest.

The basic time unit must be established as either annual, seasonal or monthly (Dracup et al., 1980a). For hydrologic drought the usual time units are water years. For agricultural drought the basic time unit is the growing season. For meteorological drought the basic time unit can be daily, monthly, seasonal or annual.

The truncation level at which a drought is said to be occurring must be defined (Dracup et al., 1980a). This can be defined as the long term historical mean or some percentage of one standard deviation from the mean for more extreme events. Truncation level can vary greatly depending on the researcher and the effects being researched. In the case of hydropower applications, truncation level can even change temporally based upon the demand for water. This is an area of much contention but for simplicity in this thesis the long term mean is used to define low flow or drought events.

A regionalization or standardization approach must be chosen to allow the drought to be transposed to different areas of the watershed (Dracup et al., 1980a). There are three choices of regionalization:

- Do not regionalize.
- Standardize data and define the region according to similar climate, geomorphology and geography.
- Standardize data and define the regions with similar hydrologic statistics.

Drought events are formulated by first dividing the historical record according to the truncation level (Dracup et al., 1980a). All adjacent time periods which are below the truncation level are then combined into individual drought events.

Drought events are characterized by three attributes (Dracup et al., 1980b). These are:

- Duration, the number of successive time periods the drought persists.
- Severity, the cumulative deficit over the entire drought.
- Magnitude, the average deficit over the drought period.

Magnitude is derived from Severity and Duration, both of which depend on streamflow values as follows:

$$Magnitude(M) = \frac{Severity(S)}{Duration(D)} \tag{2.1}$$

The *impacts* of drought are best measured in streamflow records even though precipitation records often cover a longer period of time and are more complete (Dracup et al., 1980a). Hydrologic (streamflow) drought is characterized as low streamflow lasting an integer number of years (Dracup et al., 1980b). The truncation level is usually selected as the mean annual runoff of the watershed or some percentage of one standard deviation from the mean for more severe drought (Dracup et al., 1980b). Using the mean simplifies the comparison of drought severity because both high and low events have the same scale (Dracup et al., 1980b).

Joseph (1970) discusses the problems associate with hydrologic drought frequency analysis and proposes solutions based on probability theory. The persistent nature of drought is ignored in favor of simplicity. By considering low flows on an annual basis, droughts can be evaluated using traditional probabilistic techniques. For design purposes, since the probability of drought is more important than the cause and effect relationship, this simplification is adequate though not ideal.

A prerequisite to computing drought probabilities is the establishment of a probability distribution to describe the data (Joseph, 1970). The main difficulty is that frequently one or more values in a streamflow sample are zero. This poses a problem with log transformation of the data. Joseph (1970) proposes a two step procedure to solve this problem. The samples are first separated into zero and non-zero drought events and probability density functions are assigned to each. The probability density functions are combined 'a posteriori'. The following equation is used to combine the probabilities of zero and non-zero droughts (Joseph, 1970).

$$F(x) = 1 - (1 - p_0)(1 - p_x) \tag{2.2}$$

Where:

$F(x) =$ probability that a drought will be equal or more severe than magnitude x

$p_0 =$ probability of a zero value drought event

$p_x =$ probability of a non-zero value drought event

The recurrence interval can then be defined as follows (Joseph, 1970):

$$T_x = \frac{1}{F(x)} \tag{2.3}$$

Where:

$T_x =$ recurrence interval (years)

Using binomial theory the probability of nonoccurrence of a drought that is equal or more severe than the T year drought is estimated as follows (Joseph, 1970):

$$P = \left(1 - \frac{1}{T}\right)^T \tag{2.4}$$

From this, assurance is defined as the probability of nonoccurrence of droughts more severe than the design drought of a project during the life of the project (Joseph, 1970).

$$A = \left(1 - \frac{1}{T}\right)^N \tag{2.5}$$

Where:

$A =$ Assurance (probability)

$T =$ Return period of design drought (years)

$N =$ Estimated useful life of the water resources project (years)

This last equation can be used to ascertain the risk involved of drought related impacts during the lifetime of a project.

Burn and DeWit (1996) expand on Joseph's methodology of frequency analysis, taking into account the multiyear nature of drought events. In their methodology, each period of drought, regardless of length, is considered a single event. The implication of this is that the basic time unit varies instead of being static as it was in the other methods. The severity of each drought event is used in a standard frequency analysis as if the time units were equal (Burn and DeWit, 1996). Drought severities are then determined for different probabilities of exceedance. The return period is then determined from the probability of exceedance and the variable time unit is taken into account by multiplying by the average drought duration as follows (Burn and DeWit, 1996):

$$T = \frac{1}{POE} AD \tag{2.6}$$

Where:

$T =$ Return Period (years)

$POE =$ Probability of Exceedance (fraction)

$AD =$ Average duration of all recorded drought events (years)

This allows a return period to be established without ignoring the fact that droughts occur over multiple years and the effects of drought also span many years.

Chapter 3

A Technique of Reconstruction

3.1 Introduction

The goal of this study is to effectively build a statistical regression model between standardized tree ring chronologies and annual streamflow data and to use the resulting reconstruction to analyze low probability drought. In order to accomplish this goal a combination of statistical tests, time series analysis, multivariate statistics and regression analysis are used to process the available data. Most of these procedures were programmed into two applications discussed in Chapter 5. This chapter discusses the methodologies used for this research, the reason they were used, their application and their advantages and disadvantages.

The first step in forming a statistical model relating tree ring data to climate is to presuppose a cause and effect relationship. In the case of hydrological drought the relationship that is assumed is that water stress is limiting to tree growth during drought years. Other possible suppositions are that temperature is limiting to growth or that pollutants are limiting.

The biological process involved in this limitation can be of many forms. The sampling of the trees is especially important for the hypothesis to be born out. Trees taken in areas not limited by moisture will show poor correlation. In most cases, however, tree growth is at least partially limited by a combination of moisture and temperature.

These hypotheses are then investigated by:

1. Preprocessing the data to confirm statistical assumptions.

2. Predetermining possible predictors using physical criteria and correlation analysis.
3. Using orthogonal best subsets regression to determine the best model.
4. Performing final model analysis and regression diagnostics if a satisfactory model is found.
5. Verifying the model using split sample techniques with standard verification statistics.
6. Investigating and removing outliers based on regression diagnostics.
7. Applying the final model to the tree ring data to produce a full reconstruction substituting gauged data where applicable.
8. Analyzing drought using the methods presented in Section 2.3.

All streamflow data in this study are first annualized based on monthly flow records. 12 annualizations are formed for each streamflow data set based on starting month. This is done because there is some uncertainty in how the growing season corresponds to the water year. These are each analyzed as separate streamflow series up until the point at which the best model is chosen.

3.2 Preprocessing

Each standardized tree ring chronology and annualized streamflow data series requires a certain amount of preprocessing before the reconstruction procedure can take place. This is done to avoid problems with data quality, and violations of the fundamental assumptions with multivariate analysis and linear regression to be performed during reconstruction. The three conditions that are of concern are those of non-normality, non-stationarity and autocorrelation. The procedures used to deal with each of these are discussed in the following sections.

3.2.1 Testing and Correcting for Non-Normality

Normally distributed time series and residuals are a requirement for most of the analysis in tree ring reconstructions. Normally distributed error terms are required for both Autoregressive

Moving Average (ARMA) modelling (Box and Jenkins, 1970) and linear regression (Neter et al., 1990). The most fundamental assumption of all multivariate analysis (such as principal component analysis) is normality (Hair et al., 1998). In addition normality is a requirement for t and F statistics (Hair et al., 1998).

The most widely used test for normality is the normal probability plot. This is a graphical method which compares the cumulative distribution of the data values with a cumulative distribution of the normal distribution. A scatter plot of the data pairs is inspected to see if there is any significant deviation from a 45° diagonal line.

A more formal and rigorous test of normality that incorporates this procedure is the probability plot correlation coefficient test presented in Maidment (1993), chapter 13. This test uses the Pearson correlation coefficient between the ordered data and corresponding normal values to satisfy the hypothesis of normality at different confidence limits. The correlation coefficient is calculated as follows:

$$r = \frac{\sum(x_i - \overline{x})(w_i - \overline{w})}{[\sum(x_i - \overline{x})^2 \sum(w_i - \overline{w})^2]^{0.5}} \tag{3.1}$$

Where:

x_i = observation

$\overline{x}$ = average value of all observations

w_i = fitted quantile of the normal distribution

$\overline{w}$ = average of fitted quantiles of the normal distribution

The correlation coefficient is then compared to critical values reproduced in Table 3.1. If the value of r falls below the critical value for the 5% confidence level then a transformation is required to normalize the data.

A widely used normalization method is the Box-Cox transformation (Maidment, 1993, ch. 18). This combines a logarithmic transformation and power transformation into a parameter that can be used in a search algorithm. The equation used for this transformation is as follows:

$$y_t = \begin{cases} \frac{(x_t^\lambda - 1)}{\lambda} & \text{if } \lambda \neq 0 \\ ln(x_t) & \text{if } \lambda = 0 \end{cases} \tag{3.2}$$

Where:

λ = Box-Cox coefficient

x_t = Untransformed data points

y_t = Transformed data points

Using this equation, to transform all of the original data, λ is varied using a search algorithm so that the skewness of the transformed data is minimized. The probability plot correlation test is then reapplied to the transformed data to make sure that the assumption of normality is achieved.

Table 3.1: Lower Critical Values of the Probability Plot Correlation Test Statistic for the Normal Distribution Using $P_i = (i - 3/8)/(n + 1/4)$

	Significance Level		
n	0.10	0.05	0.01
10	0.9347	0.9180	0.8804
15	0.9506	0.9383	0.9110
20	0.9600	0.9503	0.9290
30	0.9707	0.9639	0.9490
40	0.9767	0.9715	0.9597
50	0.9807	0.9764	0.9664
60	0.9835	0.9799	0.9710
75	0.9865	0.9835	0.9757
100	0.9893	0.9870	0.9812
300	0.99602	0.99525	0.99354
1000	0.99854	0.99824	0.99755

Source: Maidment, 1993

3.2.2 Testing and Correcting for Non-Stationarity

A hydrologic time series is stationary if it is free of trends, shifts in the mean or periodicity (Maidment, 1993, ch. 19). Generally speaking an annual streamflow series will be stationary unless some natural or man-made disruption has occurred. Examples of occurrences that will produce non-stationarity are commencement of river regulation, changes in a gauging location or instrumentation, or climate change.

Trends, shifts and periodicity can cause problems in both principal component analysis and linear regression. These occur because of the fundamental assumption that each data series represents a single population. The presence of any non-stationarity indicates that more than one population may be represented. The consequences are poor results in a regression analysis and artificially significant principal components.

No discussion is made, in this thesis, of techniques to remove non-stationarity but tests are presented that identify the presence of trends and shifts in the mean. Periodicity should not be a concern for annual series.

Trends are tested using the standard Mann-Kendall test for trend (Maidment, 1993,ch. 19). This is a non-parametric test for an upward or downward trend in a time series. It is not sensitive to whether the trend is linear or non-linear. For this test a new series is generated by comparing each value in the time series with all the subsequent values. The new series z_k is generated by the following rules:

$$z_k = \begin{cases} 1 & \text{if } y_t > y_{t'} \\ 0 & \text{if } y_t = y_{t'} \\ -1 & \text{if } y_t = y_{t'} \end{cases} \tag{3.3}$$

Where:

z_k = Mann-Kendall series

y_t = Time series value for current time period

$y_{t'}$ = All time series values subsequent to time period t

The Mann-Kendall statistic is then computed by the sum of the points in the z_k series.

$$S = \sum_{t'=1}^{N-1} \sum_{t=t'+1}^{N} z_k \tag{3.4}$$

The test statistic for N > 10 is as follows:

$$u_c = \frac{S+m}{\sqrt{V(S)}} \tag{3.5}$$

$$V(S) = \frac{1}{18}[N(N-1)(2N+5) - \sum_{i=1}^{n} e_i(e_i-1)(2e_i+5)] \tag{3.6}$$

Where:

$m = 1$ if S < 0

$m = -1$ if S > 0

$n =$ number of tied groups

$e_i =$ number of data in the ith tied group

The hypothesis of an upward or downward trend cannot be rejected at an α significance level if $|u_c| > u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution (Maidment, 1993, ch. 19).

The tests for shifts in the mean require the data to be split at the point where the shift is assumed to occur. Although there are rigorous statistical tests for this, it is simplest to observe a plot of the time series and qualitatively determine if a shift has occurred.

Seasonality should not be a problem in annual series, so it is not tested.

3.2.3 Testing and Correcting for Autocorrelation

Autocorrelation is defined as the correlation between successive values in a time series (Fritts, 1976). This occurs when the value of a time series in a given year impacts the values of the following year or years (called lags). The assumption of uncorrelated errors is crucial in linear regression to produce the best possible model. Also, serial correlation reduces the degrees of freedom of a time series effectively reducing sample size. This is especially a problem when dealing with very small sample sizes.

Serial correlation in tree ring time series has been shown to arise primarily from biological factors (e.g. food, storage, crown area and root mass) but some persistence may also be due to climatic forcing (Cleaveland and Stahle, 1989). Usually natural streamflow series do not exhibit significant autocorrelation at an annual scale. For these reasons and the complications caused by autocorrelated series many authors reconstructing streamflow have chosen to remove this persistence and take the risk of losing a minimal amount of climatic signal (Cleaveland and Stahle, 1989; Brinkmann, 1987; Cook and Jacoby, 1983). The method of removing autocorrelation, also called prewhitening, used by these authors is also adopted for this study.

In the case of streamflow series, autocorrelation may be due to storage in lakes and marshes along the stream or some other natural phenomena. For this study significant correlation in streamflow series was not removed. Streamflow records with significant autocorrelation were not reconstructed because the correlation structure between the streamflow and prewhitened tree ring series could not be effectively matched. Although it is possible to remove the autocorrelation structure from the streamflow records to match the prewhitened tree ring chronologies it was decided that too much information could be lost if the streamflow series was too heavily processed. The final streamflow series reconstructed in this way may not resemble the original enough to be useful in planning.

First order autocorrelation is first tested by taking a time series (x_t) and its first order lag (x_{t+1}) and calculating the Pearson correlation coefficient between them.

$$r = \frac{\sum_{t=1}^{n}(x_t - \overline{x}_t)(x_{t+1} - \overline{x}_{t+1})}{(n-1)S_t S_{t+1}} \tag{3.7}$$

Where:

n = the number of data points in the time series minus the lag order (1)
x_t = data points of the unlagged time series
$\overline{x}_t$ = mean of the unlagged time series
x_{t+1} = data points in lag 1 time series
$\overline{x}_{t+1}$ = mean of the lag 1 time series
S_t = Standard deviation of the unlagged time series
S_{t+1} = Standard deviation of the lag 1 time series

The first order autocorrelation coefficient is then tested for significance with a simple t-test.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{3.8}$$

where the null hypothesis is that the two series are independent. The null hypothesis is rejected if $|t| > t_{crit}$ where t_{crit} is from the Student's t distribution with $n-2$ degrees of freedom and exceedence probability of $\alpha/2$ (Maidment, 1993, ch. 19).

If the autocorrelation is found to be significant at the 95% confidence level then an Autoregressive Moving Average (ARMA) model is estimated for the time series and the non-autocorrelated residuals are used in the orthogonal regression analysis. Given the availability of computer programs to fit the ARMA model a full discussion of ARMA modelling is beyond the scope of this thesis. Further explanation can be found in Maidment (1993), chapter 19 and Box and Jenkins (1970).

In order to identify the stochastic process that would best fit the data the autocorrelation function (ACF) and partial autocorrelation function (PACF) are plotted for several lags along with 95% confidence limits. The patterns found here are compared to those found in data of a known stochastic process (Box and Jenkins, 1970). For example if the ACF declines steadily with lag but the PACF becomes essentially 0 after 1 lag then an AR(1) MA(0) process is assumed.

It has been shown that a low order ARMA process (order 1 or 2) is usually sufficient to remove autocorrelation from a tree ring series (Cleaveland and Stahle, 1989). For the tree ring series the fitted model is used to remove the persistence and the serially random residuals are used for further analysis.

3.3 Predetermination of Predictors

One of the problems faced in dendroclimatic studies is that of the tendency to overfit models by adding spuriously significant predictors. When one blindly adds predictors into a model the probability of including chance features in the data that are interpreted as essential features is high (Booy, 1996; Cook et al., 1994). A model built in this fashion will calibrate well on dependent data but will be less useful in predicting independent data. This also reinforces the need for independent verification of the final regression model. The chance of including spuriously correlated variables in the predictor set can be minimized by only including predictors that are likely to be significant, based on physical and statistical characteristics. This has the added benefit of reducing the number of predictors for better computational ease later on.

In this study several criteria were used to choose candidate predictors *'a priori'* from the pool available from the International Tree Ring Data Bank (ITRDB). Judgement also played a critical role in selecting candidate predictors so these criteria were only used as guidelines. The criteria used to select the candidate tree ring chronologies are as follows:

1. They should be less than 500 km from the gauge to be reconstructed.

2. They should be within the gauge sub-basin or an adjacent drainage basin.

3. The length of time overlap between the tree ring set and the gauged data should be greater than 25 years.

4. The statistical quality of the tree ring data set should be high.

5. There should be statistically significant correlations between tree ring series and monthly streamflow.

6. There should be statistically significant correlations between tree ring series and annual streamflow.

Cook (1995) suggests that a limiting distance of 500 km be used as a guideline for climate predictors in order to maximize common signal. This was used as a first screening to form a list of possible candidate predictors.

Predictors that were directly within the gauge sub-basin and adjoining basins were chosen out of this set. Adjoining basins were included because common weather patterns could produce a common climate signal reproduced in each basin. This principle was not upheld in the case of mountain ranges which separate basins. Orographic effects would preclude the existence of a common signal between these basins.

If the tree ring data did not have sufficient length of overlap with the gauged data it was excluded. Any overlap smaller than 25 years causes problems with statistical significance in regression, verification and statistical tests.

All remaining tree ring data sets were inspected for quality, significant non-normality that could not be corrected, non-stationarity and questionable sampling. Any of these conditions would exclude a tree ring series.

A correlation analysis between the tree ring series and monthly streamflow for t and t+1 lags was used to further reduce the number of candidate predictors. This technique has been used in many tree ring climate reconstruction studies to gauge the season to be reconstructed.

In this case, if a tree ring series was not significantly correlated with monthly streamflow or its first order lag, it was discarded.

As a final check, a correlation analysis between the tree ring series and annualized streamflow for t and t+1 lags was used to make sure no significantly correlated data sets were missed by the monthly correlation analysis. This did not exclude any series but if a series with significant correlation to annualized streamflow was found that was excluded by the monthly correlation analysis it was included in the candidate predictor set.

3.4 Orthogonal Regression Analysis

3.4.1 Principal Component Analysis

The most common problems faced in dendroclimatic reconstruction are those of multicollinearity of the predictors and an intractable number of predictors. Both of these can be dealt with effectively using a method in multivariate statistics called principal component analysis (PCA).

Multicollinearity occurs when the independent variables are significantly correlated with each other. This occurs with tree ring data because the trees in an area react to the same macroclimatic signal. This being the case the tree ring records appear very similar with differences occuring due to microclimatic and biological 'noise'.

Multicollinearity of the predictors causes many difficulties with linear regression. It tends to inflate the sampling variability of estimated coefficients (Neter et al., 1990), meaning that the coefficient values cannot be estimated with any degree of certainty and tend to predict poorly. Also the tendency to overfit is increased because the same signal is accounted for multiple times in several predictors. This causes havoc with stepwise regression procedures which hold one coefficient constant while varying the others. A key assumption in stepwise regression is that each predictor is independent. When there is significant correlation between predictors this is no longer true and varying one coefficient while holding constant another one that is related through intercorrelations no longer makes sense.

Dealing with an intractable number of predictors is also a problem. When looking at tree ring data it is not unusual to have upwards of 50 possible tree ring data sets of interest. Taking into account biological carryover effects by including forward and backward lags multiplies this number. Even the best algorithms for stepwise or best subsets regression have difficulty in dealing with this many predictors.

Principal component analysis is a statistical technique where a new set of variables is derived from the original predictors to be orthogonal to each other while preserving as much of the original information as possible. Each of these new variables are a linear combination of the originals but are independent of each other. This is accomplished using a complex procedure called eigenmode analysis the specifics of which are detailed in Draper and Smith (1981) and Press et al. (1988).

Principal components are the eigenvectors of the correlation matrix of the tree ring data sets. The standard matrix representation of this procedure is as follows.

$$\frac{1}{n}F'F\;E = C\;E = EL \tag{3.9}$$

Where:

n = number of years of data

F = predictor matrix

F' = transpose of predictor matrix

E = eigenvector matrix

C = correlation matrix of predictors

L = diagonal matrix of eigenvalues

Each eigenvector has a corresponding eigenvalue which is proportional to the amount of variance represented by the eigenvector. Standard procedure is to present the eigenvectors in order of decreasing variance represented.

Together the set of principal components (eigenvectors) are a more efficient representation of the original data. The most common information or signal is concentrated into the first few components. This means that a small subset of principal components is capable of representing most of the variance found in the original data. Cook (1995) showed how tree ring chronologies that tend to correlate well with climate tend to correlate well amongst themselves and are heavily loaded into the first several principal components. This property can be exploited by making the assumption that important climatic signal is represented in the most important eigenvalues and that smaller eigenvalues represent localized biological 'noise' or microclimatic signal.

Several criteria have been used to choose the Principal Components (eigenvectors) to retain for regression modelling (Cook and Kairiukstis, 1989). The criterion used in this study is the Kaiser-Guttman eigenvalue 1 criterion (Cook et al., 1994). This criterion states that, since an eigenvalue of 1 represents the expected value of an eigenvector in a correlation matrix of random data, an eigenvector must have an eigenvalue of at least 1 to be retained (Cook et al., 1994). Thus the only Principal Components that are kept for regression modelling are those that perform at least as well as random data.

By using this method the problem of multicollinearity in the predictors is averted and the number of predictors is reduced substantially.

3.4.2 Best Subsets Regression

Multiple linear regression is a statistical technique used to analyze the relationship between a single dependent (predictand) variable and several independent (predictor) variables. In this application it is used to form an equation between annualized gauged streamflow and a set of significant principal components derived from standardized preprocessed tree ring chronologies. The basic statistical equation for linear regression is as follows:

$$\hat{y}_t = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m + \epsilon \tag{3.10}$$

Where:

$\hat{y}_t$ = Estimate of predictand (annualized streamflow)

m = Number of predictor variables

$x_1 \ldots x_m$ = Predictor variables (principal components)

$b_1 \ldots b_m$ = Regression coefficients corresponding to each predictor variable

b_0 = Intercept coefficient which scales the regression equation to the mean of the predictand

ϵ = Random error term.

The calculation of regression coefficients is performed using the statistical method of least squares. There are many software packages that perform this analysis, a discussion of the method of least squares is beyond the scope of this thesis. For further information please refer to Neter et al. (1990) or any basic statistics text.

At this stage in the analysis it is known that the significant principal components represent information that is largely common to all of the tree ring chronologies. It is surmised that at least some of this signal is climatic in nature, unfortunately it is not known which combination of principal components represents climate in the form of precipitation, temperature or evapotransporation and their carry-over effects. It is also very likely that some of the principal components represent non-climatic regional information such as pollutant load or insect infestation. For this reason as well as to avoid overfitting the regression model, it is necessary to use statistical techniques to separate out only the important principal components that will help reconstruct streamflow.

Several methods are available to accomplish this goal. Traditionally an automated method known as Stepwise regression is used to choose the best predictors for a model. Stepwise regression is a method of selecting variables for inclusion in the regression model by alternately entering and deleting candidate predictors from the model based on their cumulative predictive power (Hair et al., 1998). This allows the researcher to examine the effect of each predictor on the regression model without having to look at all combinations.

There are two problems with Stepwise procedures. The first is that multicollinearity has a significant negative impact. This has been dealt with by using principal components instead of actual tree ring chronologies. The second is that a threshold significance level must be specified for entering and deleting variables. This means that if a threshold is set too liberally overfitting can occur causing the model to perform poorly on independent data. If the threshold is set too conservatively the best possible model is not found. For this reason the Stepwise regression procedure was discarded for this study. A more robust all-possible-subsets or best-subsets procedure was adopted.

Best subsets regression procedures are methods used to select the smallest possible subset of predictor variables that provide a model with the maximum amount of explained variance. It is different from Stepwise regression in that all combinations and subsets of variables are evaluated. This means that for n possible candidate variables 2^n regression models are evaluated. This is a computationally intense procedure that increases exponentially in computation time and space for each added predictor. This is one reason why reduction of candidate predictors 'a priori' was emphasized in previous sections.

The output of this method is a 'best' model for each number of predictor variables based on the coefficient of determination (R^2), and the coefficient of determiniation adjusted for a reduction in degrees of freedom (R^2_{adj}). Calculation of the coefficient of determination is accomplished as follows:

$$R^2 = \frac{SSR}{SSTO} \tag{3.11}$$

Where:

$R^2 =$ Coefficient of multiple determination
$SSR =$ Regression sum of squares
$SSTO =$ Total sum of squares

R^2 represents the percent of the total variance explained by the regression model. Unfortunately models with different numbers of predictors are not directly comparable using this statistic. R^2 will generally increase with an increase in the number of predictors due to chance alone. It is not possible for this statistic to decrease with an increasing number of predictors because the corresponding decrease in degrees of freedom is not taken into account. A statistic used to choose between models with different numbers of predictors is the R^2_{adj}. This is the same as R^2 except it is adjusted for the loss in degrees of freedom due to the number of predictors. It is calculated as follows:

$$R^2_{adj} = 1 - (1 - R^2)\left[\frac{n-1}{n-m-1}\right] \tag{3.12}$$

Where:

$R^2 =$ Coefficient of multiple determination
$n =$ Number of data points
$m =$ Number of predictors

Finding a model with maximum R^2_{adj} is equivalent to finding a model with a minimum Mean Square Error(MSE) (Minitab Inc., 1996).

3.4.3 Model Determination

At this point there are 12 best subsets analysis performed, one for each annualization of streamflow (January-December to December-November). In addition there is a model with highest explained variance for each number of predictors. This leaves 12 times the number of predictors models to choose from. A formal procedure is required to sort through this large number of competing models to choose the best possible annualization and number of predictors for reconstruction of streamflow. The procedure that was used is as follows:

1. Separate the candidate models into groups with the same number of predictors (12 models for each number of predictors).

2. Choose the model with the highest R^2_{adj} for each number of predictors.

3. Plot the R^2 and R^2_{adj} values on a chart versus the number of predictors.

4. At a certain point on this chart adding additional predictors does not increase adjusted explained variance (R^2_{adj}).

5. The model with the highest R^2_{adj} for the lowest number of predictors is chosen as the model to be used in reconstruction.

3.4.4 Identifying and Removing Outliers

The model that has been identified as the best candidate for reconstruction can now be investigated for influential observations and outliers. Great care must be exercised in removing outliers and influential points as it is very difficult to tell the cause of anomalous points from this type of data. In addition, the nature of the principal of limiting factors makes it likely that at least one high flow year will show up as anomalous, even though it is part of the natural record.

Outliers and influential observations were investigated using Leverage, Studentized Residuals, Cook's Distance and Dfits (Neter et al., 1990). They were evaluated based on guidelines given for these statistics, potential improvement to the regression model and judgement. Only points which showed anomalous results in several of these tests were considered for removal. In addition removal of anomalous observations was only done if their removal significantly improved the regression model.

3.4.5 Final Model Determination

The final model is built using the predictors identified from the best subsets analysis and the data remaining from the investigation of outliers. This model is now ready to be verified using a split sample technique.

3.5 Verification

The most important, yet often ignored, step in building a regression model is verification and validation of the ability of the model to be applied to independent data. This is necessary to confirm that the regression relationship has not been overfit and maintains some degree of universal applicability to independent data.

Verification is usually accomplished using a split sample technique. This is where a portion of the calibration data is withheld from the model building exercise in order to determine how the model performs on independent data. Understandably, many model builders are reluctant to withhold any portion of the available data for this purpose. This is especially true in the case of streamflow since the records are generally very short to begin with.

Techniques have been developed in order to overcome this objection but still provide rigorous verification. One of these consists of splitting the sample into two equal parts and using each as a model building data set. These 'partial' models are then verified using the independent data and statistical tests as well as qualitatively compared to each other in terms of coefficients and statistical properties. Provided that the models validate well and are reasonably similar in form the data sets are combined and the 'full' model is developed using the same form. This 'full' model cannot be validated against independent data but the split sample validation has proven a certain amount of model stability. It has also established the minimum amount of utility that can be expected from the full model, the truth being that a model from the full data set should perform better than the two partial models.

In order to verify the partial data models several standard statistics are used to measure the degree of similarity between the modelled and independent data. There are many statistics used for this purpose by different researchers. Four of these were chosen for this study that have relatively universal applicability in dendrochronology. These are the product moment correlation coefficient test, sign test, product means test and reduction of error statistic.

The product moment correlation coefficient test is the most common test used for statistical verification. It is a basic statistical test that measures the similarity between the shapes of paired time series. It measures the relative variation (or covariance) in common between the two data sets (Cook and Kairiukstis, 1989). This test of significance implies that the variance of the two data sets is linearly related. It does not imply that the values are close to each other or similar in scale. In this respect it is not very robust. The t-test is performed as follows (Maidment, 1993,ch. 17):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{3.13}$$

Where r is the product moment correlation coefficient calculated as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_v)(\hat{x}_i - \bar{\hat{x}}_v)}{\sqrt{\sum_{i=1}^{n}(x_i - \hat{x}_v)^2 \sum_{i=1}^{n}(\hat{x}_i - \bar{\hat{x}}_v)^2}} \tag{3.14}$$

Where:

n = Number of related data pairs (actual and estimated data)

x_i = Actual data point

$\bar{x}_v$ = Mean of actual data

$\hat{x}_i$ = Estimated data point

$\bar{\hat{x}}_v$ = Mean of estimated data

The null hypothesis is that the correlation coefficient(r) is equal to 0. The significance is calculated using Student's t distribution with $1 - \alpha/2$ probability and $n - 2$ degrees of freedom (Maidment, 1993,ch. 17).

The sign test is a non-parametric test of similarity between series based on the number of agreements and disagreements in the sign of the first differences (Cook and Kairiukstis, 1989). If the number of agreements exceeds the number of disagreements by greater than that expected by chance the hypothesis of a relationship existing passes. This test is not sensitive to extremely anomalous data and is simple to apply. It is not very rigorous however as the magnitude of correspondence between variables is not taken into account(Fritts, 1976). Critical values for this test are calculated using the rounded values of a binomial distribution with $\frac{1}{2}$ n degrees of freedom.

The product means test is not a standard statistical test but is used extensively in dendroclimatology. This test attempts to make up for the shortfalls of the sign test by testing the signs and magnitudes of the mean deviations of paired time series. In this test the departures from the mean of each paired data point are multiplied together and gathered into two groups based on sign. The absolute values of the positive and negative data sets are then averaged. The difference between the positive and negative product means is then tested for significance. A positive average which is significantly larger than a negative average indicates that a significant correspondence exists in both direction and magnitude between the two data sets. This test is very rigorous and is a powerful indicator of a relationship when it passes. Unfortunately it has a tendency to underestimate the value of a relationship and fails more often than it should (Cook and Kairiukstis, 1989). When it fails one cannot be positive that no relationship exists because it is very sensitive to anomalous data points. The test statistic is calculated as follows:

$$t = \frac{m_+ - m_-}{\sqrt{\frac{s_+^2}{n_+} + \frac{s_-^2}{n_-}}} \tag{3.15}$$

Where:

m_+ = the mean of the positive products

m_- = the mean of the negative products

s_+^2 = the sample variance of the positive products

n_+ = the number of positive products

s_-^2 = the sample variance of the negative products

n_{-} = the number of negative products

This test statistic is compared to critical values for the Student's t distribution with n-2 degrees of freedom and $1 - \alpha/2$ probability (Cook et al., 1994).

The reduction of error statistic is the most important and powerful verification tool used in dendroclimatology. It is similar but not equivalent to the explained variance statistic (R^2). It is exactly the same as R^2 when applied only to dependent data. It is expressed as follows (Cook and Kairiukstis, 1989):

$$RE = 1 - \frac{SSR}{SSM} \tag{3.16}$$

Where:

SSR = Regression sum of squares, sum of the squared deviations between actual and modelled independent data

SSM = Mean sum of squares, sum of the squared deviations between the actual independent data and the mean of the dependent data

This can be calculated as follows (Cook et al., 1994):

$$RE = 1 - \left[\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x_c})^2}\right] \tag{3.17}$$

Where:

n = the number of independent data points

x_i = the actual independent data

$\hat{x}_i$ = the estimated independent data

$\bar{x_c}$ = is the mean of the actual data in the calibration period

The values of RE can range from +1 to $-\infty$. +1 indicates perfect agreement and a value below 0.0 indicates the regression model does not predict as well as using the calibration mean. Cook and Jacoby (1983) determined that the 95% confidence level for this test is approximately equal to 0.0. This test is very rigorous and sensitive to poor estimates so any positive RE is a good indicator of skill in the model.

3.6 Final Reconstruction

The final reconstruction involves forming a time series made up of the original gauged data extended and completed with the reconstructed data.

A final reconstruction is only made if the final model is determined to have verified well enough to proceed. A positive verification is indicated by a series of positive verification statistics, R^2 values and observed correspondence of the gauged and reconstructed time series. Judgement is used to determine if a final model is of high enough quality for final reconstruction and drought analysis.

Chapter 4

Data Availability

4.1 Streamflow Data

Surface water quantity data have been collected and archived in Canada for over 150 years. Since 1908 streamflow data have been published in a variety of forms. Today, data are collected from a variety of governmental and private agencies and are compiled regionally by Environment Canada and stored by the Meteorological Service of Canada. The Meteorologic Service of Canada stores streamflow, water level, sediment data and gauge data in their HYDEX database. Since 1991 these data have been available in a CD–rom format, called Hydat, which has replaced printed publications (Environment Canada, 2002).

All streamflow data for this study were obtained using the year 2000 version of the Hydat CD which contains data up to 1998 (Environment Canada, 2001).

4.2 Tree Ring Data

Tree Ring data are available from the International Tree Ring Data Bank (ITRDB). This is a central repository administered by the United States National Oceanic and Atmospheric Administration (NOAA). This data bank is part of the World Data Centre for Paleoclimatalogy at the National Geophysical Data Center (NGDC) in Boulder, Colorado, USA. This data centre also houses other types of paleoclimatic data such as ice cores, sediments, pollen, and anecdotal data (Grissino-Mayer, 2002).

The ITRDB provides a permanent location for the storage of well dated high quality dendrochonological data from around the world. It prevents loss of data due to mishandling, laboratory and scientist relocation and demise (Grissino-Mayer, 2002).

Information submitted to the ITRDB is scrutinized for quality and length to ensure high quality error free data. The ITRDB contains more than 6000 data sets representing more than 1500 sites around the world. The data are made freely available to all researchers.

The data are readily available over the World Wide Web at:

http://www.ngdc.noaa.gov/paleo/treering.htm

by downloading the ITRDB display software (National Oceanic and Atmospheric Administration, 2002), a DOS program which allows geographic searches of the data. An example of the display software interface can be observed in Figure 4.1 which shows a picture of available tree ring data in North America.

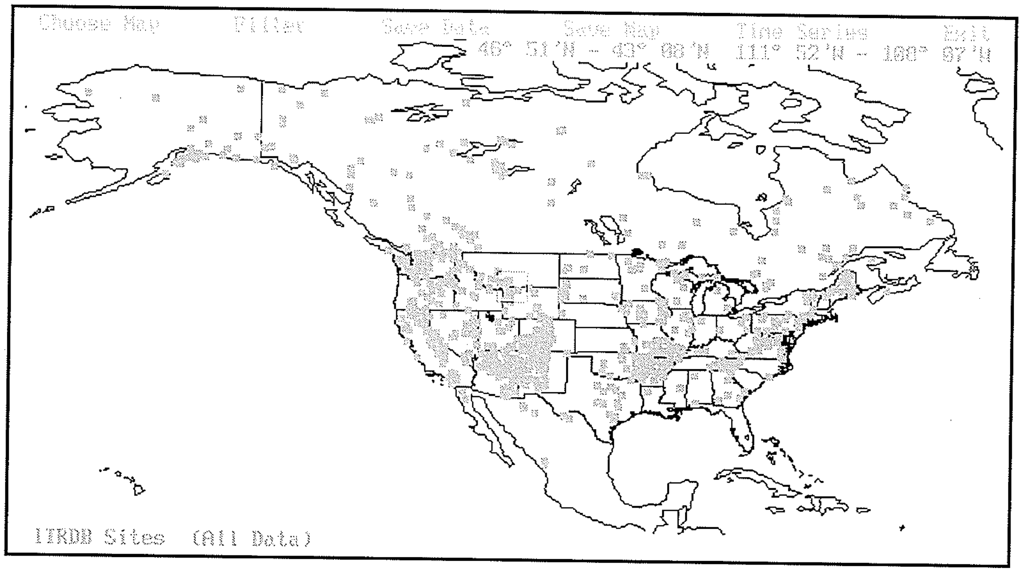


Figure 4.1: Example of ITRDB Display Software User Interface

Chronologies submitted to the ITRDB are processed using standard methods as discussed in Fritts (1976). These methods consist of taking two cores per tree, each tree in an open stand, with a minimum of 10–20 trees per chronology. Biological growth trend is then systematically removed. Biological growth trend is the tendency of the tree rings to become narrower as the tree ages. This is produced because, although on average the same volume of wood is produced every year, the diameter of the tree trunk is becoming larger. This trend decreases as the tree matures and ages because the changes in diameter become less pronounced. Growth trend is removed by fitting a curve form of limited flexibility, to minimize lost climate signal, to the tree ring widths and dividing each ring width by its corresponding value of the fitted line. A curve form of limited flexibility is used for this, usually a reverse exponential or a straight line. This is referred to as standardization of the tree ring data and produces tree ring indices with mean approximately equal to 1 and variance that is constant over the entire tree ring record. The indices for all the trees in the stand are then averaged to reduce the local noise in the signal produced by tree specific factors. The ITRDB has over 3275 chronologies that were processed in this way.

All data for this study were obtained from the ITRDB. Although other data in the study area are known to exist, because several researchers have published papers using new data, it is highly unlikely that they would be willing to pass on this data until they have completed their own projects with it. Collecting of tree ring data is highly labour intensive and costly so a researcher will normally publish all of the work commissioned by the funding agency before submitting the data to the central data bank. This can typically cause a lag of five to ten years between data being sampled and submitted.

Chapter 5

Computer Applications for Tree Ring Reconstruction

Chapter 3 presented the statistical procedures in a tree ring reconstruction of streamflow. Early in this study it was discovered that quality control and consistency would be difficult to preserve with so many techniques and options available. For this reason it was decided to automate the process as much as possible in order to speed and simplify the analysis and make sure each step was rigorously followed.

Three applications were programmed using Visual Basic 5 Professional (Microsoft Corp., 1997). This language offers advantages in that it provides a quality presentation, is easy to program and interfaces well with other applications. The disadvantage is that it is not an efficient platform for performing mathematics, especially matrix operations. Several options were investigated to address this problem from programming in C++ or Fortran to interfacing with an external program. It was decided that there was no reason to reprogram statistical routines that are readily available in commercial statistical software packages. Minitab Release 11 was chosen to serve as the statistical 'engine' for the main Visual Basic programs. Minitab offers the advantage that it is a proven industrial statistical package and it has what was referred to as OLE or ActiveX connectivity. The two terms mean that the program exposes part of it's code to be taken over by a host program. This allows it to exchange information with the host program and allows the host to take control of the 'slave' program's internal routines.

Three programs were written during the course of this thesis. The first is a program to preprocess the raw streamflow data or tree ring chronologies. The preprocessing consists of statistical tests and corrections for non-normality, non-stationariy and autocorrelation. The second program takes the preprocessed streamflow and tree ring data and performs a correlation analysis on it. It identifies tree ring chronologies that are significantly correlated with the chosen streamflow record. The last program uses that preprocessed data to perform orthogonal best subsets regression. This program takes preprocessed tree ring chronologies, extracts the principal components from them, chooses the significant principal components and performs best subsets analysis on these components and all annualizations of a streamflow data set. The final reconstruction and verification is handled with a spreadsheet as there is too much judgement involved for this process to be effectively automated. Drought analysis was done using the Hyfran Software Package (Chair in statistical hydrology, 2002). Program methodologies are presented in the following sections.

5.1 Data Preprocessor

The preprocessor application has three main purposes:

1. To test and correct for non-normality in the data.

2. To test for non-stationarity.

3. To test and correct for significant autocorrelation.

In addition this processor manipulates the data into a form that makes the subsequent analysis easier. It outputs a text file containing the preprocessed data and a Microsoft Word document that shows a description of the data, all of the operations performed and their results. The program is presented as a series of screens which prompt the user with choices based on the information about the data and tests performed.

The first screen is an introduction that allows the user to choose the type of file to be processed. This can be either a tree ring chronology or a streamflow record.

The next screen is a file browser which allows the user to choose the text file to be processed. Data in the text file must be in a certain format. Streamflow data must be in the form of the 1995 Hydat monthly average streamflow print-out. Tree ring chronologies must be in the standard output format from the ITRDB retrieval program.

Once the file is chosen, if it is a streamflow file it is first annualized. This means that the monthly average flows are summed into yearly volumes based on the starting month. For example a March annualization will be made up of summed monthly data from March of the starting year to February of the next year. This is done because the exact correspondence between tree ring chronologies and the streamflow water year is not known at this stage in the analysis. If a tree ring file is chosen it is simply read into memory.

Descriptive statistics are calculated regardless of the data type and tabulated for use later. These statistics are the number of data, mean, variance, standard deviation, standard error, coefficient of variation, skewness and kurtosis. All of this information is printed to the output Word file.

The next screen presents the test and correction for non-normality of the data. Each annualization of streamflow or tree ring chronology is checked for normality using the probability plot correlation coefficient test. This is accomplished using the normal scores calculation available in Minitab. The correlation between the data and the corresponding normal scores is tested. If the data fails this test a Box-Cox transformation is applied to the data as described in Section 3.2.1 . The probability plot correlation coefficient test is then performed a second time to confirm that normality has been achieved. The results of these tests and the coefficients used for Box Cox transformation are recorded in the output file. If normality is not achieved this is also recorded.

The next screen presents the tests for non-stationarity of the data series. First each annualization of streamflow or tree ring chronology is checked for trends using the Mann Kendall test for trend presented in Section 3.2. If a trend is detected in a series no action is taken. The program notifies the user and a note is written into the output log but it is up to the user to address the trend by hand in post processing.

After the Mann Kendall test is completed the data is presented to the user in the form of a time series plot. One plot is shown for a tree ring chronology and 12 for annualizations of streamflow, each on a different screen with prompts to continue. The purpose of this is so the user can determine if there are any shifts in the mean that need to be addressed. Again no action is taken but the graphs are written to the output word file.

The next screen presents the test and corrections for autocorrelation. The test is performed for each annualization of streamflow or tree ring chronology. The program interfaces with Minitab and it is used to extract autocorrelation and partial autocorrelation functions from the data. It then tests for significant first order autocorrelation in the data using the correlation coefficient t-test presented in Section 3.2. If significant first order autocorrelation is found in a streamflow annualization it is noted and the program proceeds to the final screen. No remedial action is taken because the cause of the autocorrelation structure of the gauged data is not known. If significant first order autocorrelation is found in a tree ring series it is noted and the program proceeds to an ARMA modelling screen.

The ARMA modelling screen only appears for tree ring chronologies. It presents graphs of the ACF and PACF and prompts the user to enter a model order for an Autoregressive and/or Moving Average process to be fitted to the data. An understanding of the different possible processes and their effects on the plots of the ACF and PACF is required to make this judgement. Tree ring data usually displays a low order autocorrelation structure. Examples of this type of structure can be found in Box and Jenkins (1970). After the form is chosen, Minitab is again used to perform the ARMA modelling. The prewhitened residuals are extracted and tested again for significant autocorrelation. If there is still significant first order autocorrelation this is noted and the program proceeds to the final screen. If the user wishes to investigate another ARMA the program must be re-initialized.

The final screen simply shows the user that the program has completed its analysis and prompts the user to end the program. On exit all of the processed data are saved in files with the same prefix as the original. Streamflow files they are saved in a *prefix*.str file for the processed annualized data and *prefix*.raw file for the unprocessed monthly data. Processed tree ring data is saved in a *prefix*.trg file. The output Word document showing all the procedures and results is saved in a *prefix*.doc file. All files are saved in the same directory as the original data file. Examples of the output produced by these analyses are presented in Appendix A.

5.2 Correlation Analyzer

The correlation analyzer application performs a correlation analysis between a preprocessed streamflow record and several preprocessed tree ring chronologies. The programmed application is only capable of performing a correlation analysis between uninterrupted monthly streamflow data and uninterrupted tree ring data sets. Annual correlation analysis and monthly correlation analysis on interrupted data sets was accomplished using the same techniques in an Excel spreadsheet. The program is presented as a series of screens which prompt the user with choices based on the information about the data and analysis performed.

The first screen prompts the user whether he/she wants to perform a correlation analysis or orthogonal best subset regression analysis. In this case the user would choose a correlation analysis.

The user is then prompted in a file browser screen to choose from preprocessed tree ring chronology files denoted by a '.trg' file extension. As the files are chosen they are listed on the screen, files chosen accidentally can be removed from this list by a click of the mouse. The number of files that can be chosen or removed is unlimited.

After choosing the tree ring files the user is prompted to choose a monthly streamflow record denoted by a '.raw' file extension. All files are read into memory and the dates that overlap are noted. The user is then prompted to start the correlation analysis.

The correlation analysis proceeds using the methodology presented in Section 3.3 and a plot is presented to the user for each tree ring data set that was chosen. An example of the plot produced can be seen in Figure 5.1. As can be observed the correlation coefficient is presented for each month and t-1 lag of each month along with the 95% significance bands.

Annual correlation analysis is completed using the same methods except it is done on annualized streamflow data in a spreadsheet environment. An example of the output from this analysis is presented in Figure 5.2.

Examples of the output files produced by the correlation analysis software are shown in

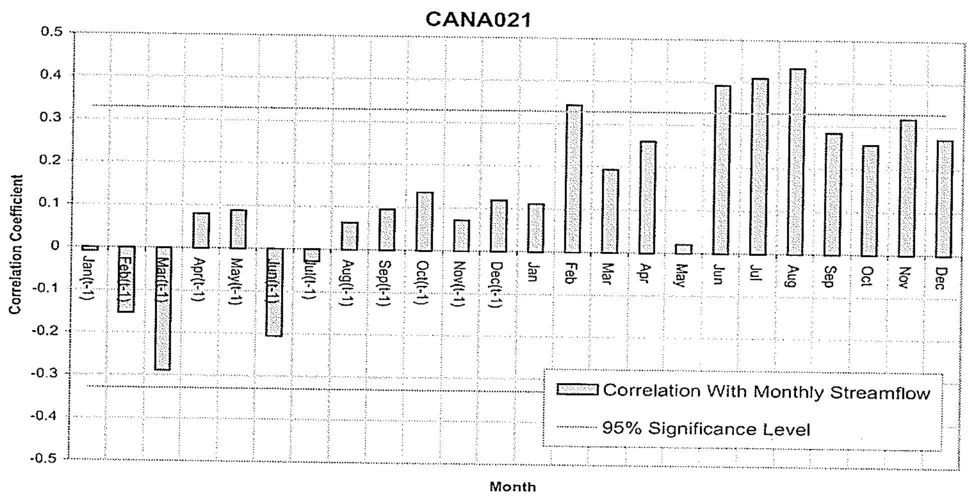


Figure 5.1: Example Correlation Analysis between CANA021 and Monthly Flows of 05AA022

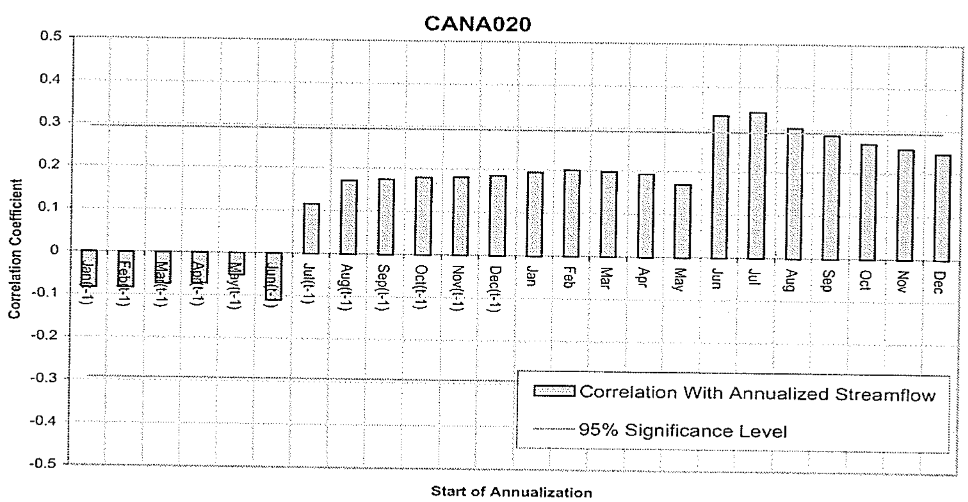


Figure 5.2: Example Correlation Analysis between CANA020 and Annualized Flows of 05AA022

Appendix B.

5.3 Orthogonal Best Subsets Application

The Orthogonal Best Subsets application is invoked after the user has chosen which are the best candidate predictors to be investigated for reconstruction. It automates the following activities:

1. Finding overlapping time periods between preprocessed tree ring chronologies
2. Lagging the preprocessed tree ring chronologies forward and backward in time
3. Extracting principal components from multiple preprocessed tree ring chronologies
4. Identifying the significant principal components using the Kaiser-Guttman eigenvalue 1 criterion.
5. Finding the overlapping time period between principal components and the preprocessed annualized streamflow data set.
6. Performing orthogonal best subsets analysis on the 12 annualizations of the streamflow data sets
7. Presenting the results of these best subsets analysis

The application is only capable of handling uninterrupted data sets. Where there was missing data in a data set the same activities were performed using an Excel spreadsheet and Minitab in tandem. The program is presented as a series of screens which prompt the user with choices based on the information about the data and analysis performed.

The first screen presented prompts the user whether he/she wants to perform a correlation analysis or orthogonal best subset regression analysis. In this case the user would choose an orthogonal best subsets regression analysis.

The user is then prompted in a file browser screen to choose from preprocessed tree ring chronology files denoted by a '.trg' file extension. As the files are chosen they are listed on the screen, files chosen accidentally can be removed from this list by a click of the mouse. The number of files that can be chosen or removed is unlimited.

After choosing the tree ring files the user is prompted into a screen to perform the principal components analysis. This subroutine takes the overlapping period between tree ring data sets and lags them forward and backward one year. This forms a data set where the t-1, t and t+1 lags of each tree ring data set are represented. These are then read into Minitab which processes them extracting the eigenvectors, eigenvalues and transformation coefficients.

The user is then prompted into a screen which determines the significant principal components using the Kaiser-Guttman eigenvalue 1 criterion presented in Section 3.4.1.

After the significant principal components are determined the user is prompted to choose a preprocessed streamflow record denoted by a '.str' file extension. This streamflow data set is overlapped with the principal components.

The user is then prompted to start the best subsets analysis. As the best subsets analysis proceeds graphs are shown for each annualization of streamflow. On this graph each number of predictors is shown on one axis with the corresponding maximum R^2 and R^2_{adj} values shown on the other. These are written to a Word output file as this process continues.

A final plot is made using the output from this analysis consisting of the models with the highest R^2_{adj} for each number of predictors regardless of annualization. An example of this plot is given in Figure 5.3. From this final plot the model that gives the maximum R^2_{adj} is chosen. This model then goes through the process of removing outliers, final model determination, verification and final model reconstruction.

The final drought analysis is accomplished using the methods described in Section 2.3. The statistical frequency analysis is performed using the software package Hyfran (Chair in statistical hydrology, 2002). This package allows many probability distributions to be investigated. The probability distribution that fits the data best is chosen for the final frequency analysis.

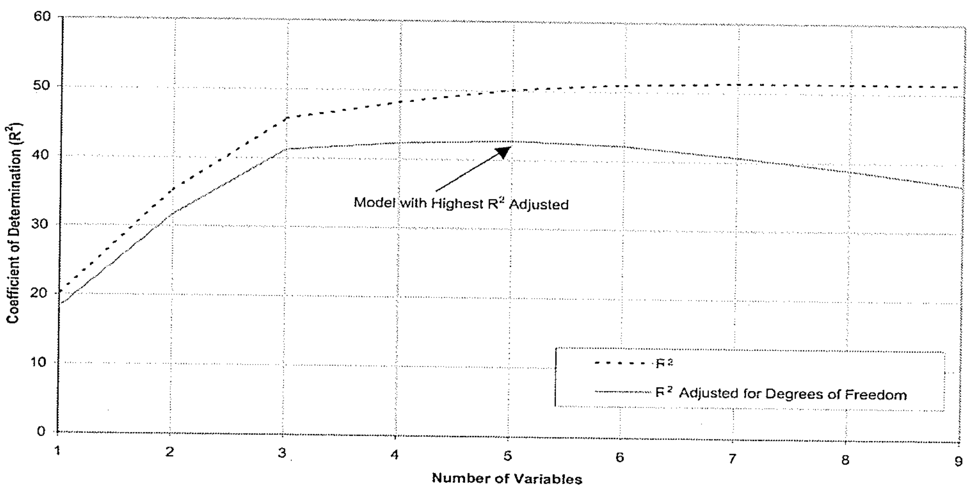


Figure 5.3: Example of Best Subset Analysis Plot

Chapter 6

Case Studies

6.1 Introduction

Three streamflow gauge records were reconstructed using the principles presented in the previous chapters in order to demonstrate their use and potential benefits. The streamflow gauges were chosen based on several factors:

1. Lack of regulation
2. Proximity to available tree ring data sets
3. Length of record overlapping with available tree ring data
4. Proximity to the Nelson-Churchill River Basin

The case study gauges were chosen by first extracting all of the unregulated streamflow gauging stations in Manitoba, Saskatchewan and Alberta from the Hydat archive. These data sets were sorted for length of record and only the ones with greater than 30 years of record were retained. The retained gauge locations were then compared with available tree ring data sets within a 500 km radius. Only those with a greater than 30 year overlap with several tree ring data sets were retained. At this stage only a few gauging stations remained and the three gauging stations with the highest number of tree ring data sets within 500 km were retained. The three gauging station chosen are as follows:

1. 07BE001 Athabasca River at Athabasca
2. 05AA023 Oldman River Near Waldron's Corner
3. 05AA022 Castle River at Beaver Mine Station

Of these data sets one is on the Athabasca River in the Mackenzie River Basin and two are tributaries of the Oldman River in the South Saskatchewan River Basin. The tree ring reconstruction of each of these records are presented in the following Sections.

6.2 Athabasca River at Athabasca

6.2.1 Background Information

The Athabasca River forms the southern most part of the Mackenzie River Basin. This river is shown in Figure 6.1.

The Athabasca River is the longest river in Alberta at 1538 km and runs from Jasper to Lake Athabasca. Originating in the Columbia Ice Field, a 325 km^2 glacier along the continental divide, it flows across three major physiographic regions; the Rocky Mountains, the Interior Plains and the Canadian Shield. The total drainage area is 133 000 km^2.

Today the economic use of this river is mainly tourism and five pulp and paper mills. There are 13 potential hydropower sites along this river but most are of low head and would probably not be economic for development. Grand Rapid is the most noteworthy with a maximum head of approximately 15 m (Denis and Challis, 1916). This site will probably not be developed due to environmental impact.

The gauge that was reconstructed is the Water Survey of Canada gauge 07BE001 described as Athabasca River at Athabasca. This gauge is located at the Town of Athabasca approximately 580 km downstream of the headwaters and 130 km North of Edmonton. The drainage area covered by the gauge is 74 600 km^2. This gauge has been operated since 1914 with discontinuities occurring between 1931 and 1952. Mean annual flow at this location is 429 m^3/s.

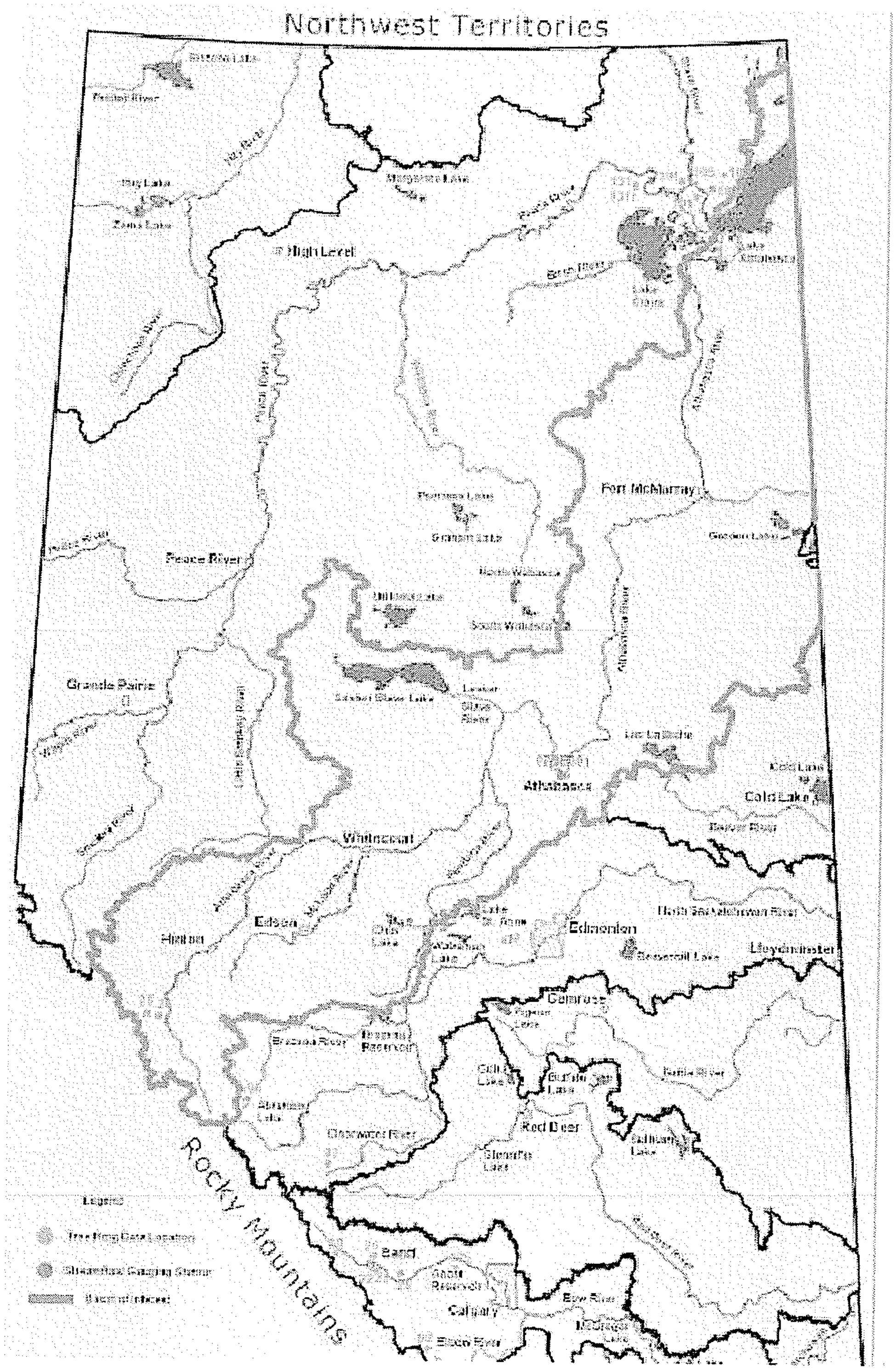


Figure 6.1: Map of Athabasca River Gauge Location

6.2.2 Available Data

Streamflow data for gauge 07BE001 were obtained from the Water Survey of Canada Hydat CD. The monthly average data were converted to 12 series of yearly streamflow volumes based on starting months from January to December. These were then preprocessed by using techniques described in Section 3.2. Eight of the flow series (Jan-Dec, Feb-Jan, Mar-Feb, Apr-Mar, May-Apr, Oct-Sep, Nov-Oct, Dec-Nov) had to be normalized using Box-Cox transformation. No trends or autocorrelation were found.

6.2.3 Predetermination of Predictors

The number of tree ring data sets was narrowed from a possible 140 candidates to 19 based on the criterion suggested by Cook (1995) that they should be less than 500 km from the gauge to be reconstructed. The tree ring data sets to be entered into the best subsets analysis were further selected 'a priori' based on judgement and the other five criteria discussed in Section 3.3.

Only two of the data sets were within the gauging station's sub–basin. These were CANA028 at Pyramid Lake, Alberta and CANA026 at Pyramid and Patricia Lake, Alberta. Of the remaining sets, eight were in the Mackenzie River Basin downstream of the gauge. The final eight were in sub-basins of the Saskatchewan River which abuts the Athabasca River Basin and is of similar geographic characteristics. All of these tree ring series were investigated further.

All tree ring data had more than 25 years of overlap when compared to the two periods of continuous streamflow records from 1914–1929 and 1952–1995. The shortest overlap period was 28 years.

All of the tree ring data sets could be made normal by Box-Cox transformation. They all displayed little or no trend, shifts in the mean or periodicity. Significant autocorrelation was removed with low order ARMA modelling (AR1 or AR1 MA1).

A correlation analysis between tree ring series and **monthly** streamflow yielded 12 tree ring series with significant correlations. These are presented in Table 6.1.

Eight of these are within the Mackenzie River Basin so they could conceivably respond

Table 6.1: Tree Ring Data Significantly Correlated With Monthly Record of 07BE001

Identifier	Description	Minor Basin	Major Basin
CANA021	Tunnel Mountain, Banff, Alberta	Bow River	Saskatchewan River
CANA022	Exshaw, Tunnel and Banff, Alberta	Bow River	Saskatchewan River
CANA026	Pyramid Lake and Patricia Lake, Alberta	Athabasca River	MacKenzie River
CANA028	Pyramid Lake, Alberta	Athabasca River	MacKenzie River
CANA096	Sunwapta Pass, Alberta North	Brazeau River	Saskatchewan River
CANA097	Peyto Lake, Alberta North	Clearwater River	Saskatchewan River
CANA099	Sarrail Glacier, Alberta	Highwood River	Saskatchewan River
CANA102	Revillon Coupe, Alberta	Slave River	MacKenzie River
CANA103	Peace River, Alberta	Slave River	MacKenzie River
CANA104	Peace River, Alberta	Slave River	MacKenzie River
CANA105	Athabasca River, Alberta	Athabasca River	MacKenzie River
CANA135	Towers Ridge, Alberta	Bow River	Saskatchewan River

to the same flow characteristics present in the gauged record. The others are all within the Saskatchewan River Basin so they could be responding to weather patterns common to both basins. Plots of the correlation analysis against monthly flow are shown in Appendix C. In each case the tree ring record is significantly correlated with at least one monthly streamflow record. This indicates at least some useful information within the tree ring series for reconstruction of the streamflow series.

A correlation analysis between tree ring series and **annual** streamflow yielded only four tree ring series with significant correlations. These are presented in Table 6.2.

Table 6.2: Tree Ring Data Significantly Correlated With Annual Record of 07BE001

Identifier	Description	Minor Basin	Major Basin
CANA026	Pyramid Lake and Patricia Lake, Alberta	Athabasca River	MacKenzie River
CANA028	Pyramid Lake, Alberta	Athabasca River	MacKenzie River
CANA105	Athabasca River, Alberta	Athabasca River	MacKenzie River
CANA135	Towers Ridge, Alberta	Bow River	Saskatchewan River

All of these were identified in the monthly correlation analysis. Plots of the correlation analysis against annualized streamflow can be found in Appendix C.

The correlation analysis against monthly streamflow yielded 12 possible data sets to be investigated for model building. The correlation analysis against annual streamflow reaffirmed that four of these data sets have significant common information with the gauged streamflow data. The reduction to 12 data sets based on correlation with monthly streamflow sufficiently simplifies the regression procedure to proceed. The principal components analysis will help further reduce the number of distinct signals and the number of significant predictors available to be entered into a model.

6.2.4 Principal Components Analysis

The overlapping period for all 12 tree ring series was found to be from 1805 to 1965. The tree ring sets were lagged forward and backward one year to account for growth and storage effects forming 36 possible predictors for the reconstruction. This matrix was then orthogonalized and the eigenvectors and eigenvalues tabulated. 12 Eigenvectors had eigenvalues in excess of the Kaiser–Guttman eigenvalue-1 criterion. These components represent 75.6% of the total variance contained in the 36 predictors with the largest single vector representing 9.6% and the smallest 2.9%. These were retained for use in the best subsets model building exercise.

6.2.5 Best Subsets Analysis

The periods of overlap between the tree ring series and the streamflow series from 1914 to 1929 and 1952 to 1965 were used in this analysis. Each monthly annualization was regressed against all possible combinations of the 12 orthogonalized tree ring vectors. For each number (1 to 12) of predictors and each annualization of streamflow the best model was chosen based on R^2 and $R^2_{adjusted}$. The best model for each number of predictors was then separated out based on $R^2_{adjusted}$ and plotted in Figure 6.2. The model that produced the highest $R^2_{adjusted}$ with the least number of predictors was a regression using the 2nd, 4th and 7th highest eigenvalues on the January streamflow annualization. This regression produced an R^2 of 49.5% and $R^2_{adjusted}$ of 43.7%.

6.2.6 Investigation of Outliers

Outliers and influential observations were investigated using Leverage, Studentized Residuals, Cook's Distance and Dfits. They were evaluated based on guidelines given for these statistics,

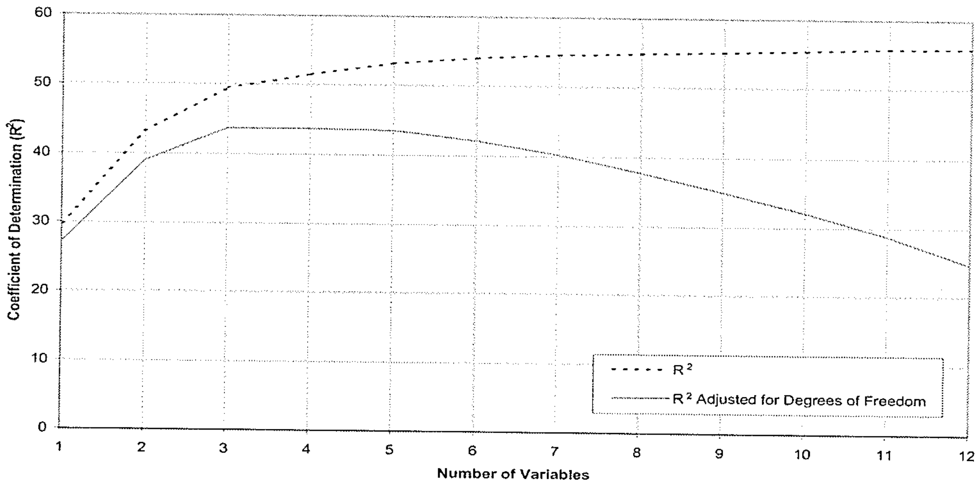


Figure 6.2: Results of Best Subset Analysis of 07BE001

potential improvement to the regression model and judgement. It was found that the data for 1959 had high Studentized Residuals and Dfits. Removal of this point did not significantly improve the regression model. This data point was therefore left in the regression data set.

6.2.7 Model Building and Verification

The final model was built using the regression equation derived from the best subsets analysis. The verification was done using a standard split sample procedure. The data from 1914 to 1929 was first used to build a regression equation (the 'early' model) and this was tested against the independent data from 1952 to 1965. In turn the data from 1952 to 1965 was used to build a regression model (the 'late' model) and verified with the independent data from 1914 to 1929.

The final regression equation takes the following form:

$$streamflow = b_0 + b_1PC_1 + b_2PC_2 + \ldots + b_mPC_m \tag{6.1}$$

Where:

$streamflow$ = Estimate of annualized streamflow

m = Number of predictor variables

$PC_1 \dots PC_m$ = Principal components of tree ring data

$b_1 \dots b_m$ = Regression coefficients corresponding to each principal component

b_0 = Intercept coefficient which scales the regression equation to the mean of the predictand

The regression coefficients can be observed in Table 6.3 as well as the R^2, R^2_{adj} and gauged versus modelled means and standard deviations. Some qualitative observations from the model building portion of this exercise are as follows. The highest R^2_{adj} was 48.2% for the calibration on the 'early' data and the lowest was 39.7% for the calibration on the 'late' data. All three of the R^2_{adj} for the models were within reasonable limits for past streamflow studies in the literature.

Table 6.3: Calibration Statistics

Parameter	**Early Calibration (1914–1929)**	**Late Calibration (1952–1965)**	**Full Calibration (1914–1929 and 1952–1965)**
R^2	0.585	0.536	0.481
R^2_{adj}	0.482	0.397	0.421
B_0	1.40036	1.40038	1.40033
B_1 ($\times 10^{-5}$)	3.757	4.223	4.104
B_2 ($\times 10^{-5}$)	-6.668	-6.507	-7.628
B_3 ($\times 10^{-5}$)	4.874	6.000	3.332
Gauged Mean	14034	13032	13500
Reconstructed Mean	14038	12458	13255
Gauged Standard Deviation	3758	2110	2982
Reconstructed Standard Deviation	2253	1288	1709

The regression coefficients for all models are of the same order and sign. This is a good indicator of model stability for predicting independent data. The means for each of the regression models on the independent data are close to the gauged means. This also indicates a good reconstruction of statistical parameters.

Verification statistics for each of the split models and the full model are shown in Table 6.4.

Table 6.4: Verification Statistics

Parameter	**Calibration Period (1914–1929) Verification Period (1952–1965)**	**Calibration Period (1952–1965) Verification Period (1914–1929)**	**Full Calibration (1914–1929 and 1952–1965)**
Sign Test			
Right	12	11	26
Wrong	2	5	4
Status (Confidence)	Pass(95%)	Fail(90%)	Pass(95%)
Product Means Test			
tvalue	1.600	1.190	1.540
minimum value	1.761	1.746	1.701
Status (Confidence)	Fail(90%)	Fail(90%)	Fail(90%)
Product Moment Correlation Coefficient Test			
tvalue	3.656	2.189	5.095
minimum value	2.977	2.120	2.750
Status (Confidence)	Pass(99%)	Pass(95%)	Pass(99%)
Reduction of Error Test			
tvalue	0.546	0.327	0.460
minimum value	0	0	0
Status	Pass	Pass	Pass

The sign test shows that the 'early' model passes at 95% confidence when applied to independent data as does the 'full' model. The 'late' model however fails at the 90% confidence. This does not conclusively invalidate the model but indicates that the signs of the first differences do not agree between the gauged and modelled data as often as would be expected at random.

None of the product means test results passed at the 90% confidence interval. This test is a very powerful validation tool when the result is positive but is very sensitive to individual deviations, so a negative result cannot invalidate the results.

All models passed the product moment correlation coefficient test at better than 95% confidence. This implies that the variance between the gauged and reconstructed data is linearly related. This is a powerful indicator of association between two variables provided the actual and estimated means are essentially the same.

The reduction of error statistic was passed in all cases. This is the most rigorous and sensitive verification statistic used in tree ring studies. Any positive result is a good indicator that the model is better than using the mean.

Figures 6.3 and 6.4 show how the 'early' regression model performs on independent data. From Figure 6.3 it can be seen qualitatively that the calibrated model does a good job of reconstructing low flow events. As would be expected the high flow events are not as well represented in magnitude. This is true for both the dependent and independent periods. The scatter diagram in Figure 6.4 shows relatively tight correlation between the gauged and modelled data for both the calibration and verification periods. The variance does not appear to change significantly between the calibration and verification data except at higher flow regimes as is expected. This is a good indication that the early model is reconstructing the independent data adequately and provides confidence that a regression model built upon this data will represent past ungauged droughts relatively well provided they are within the realm of the known data.

Figures 6.5 and 6.6 show how the 'late' regression model performs on independent data. Figure 6.5 shows qualitatively that the calibrated model reproduces low flow events well during both the model building period and independent periods. High flows however are not as well represented. The scatter diagram in Figure 6.6 confirms this by showing relatively tight correlation between the gauged and modelled data for low flow events but a greater spread for higher flow events. The variance does not change significantly between the calibration and verification data. This indicates that the late model reconstructs the independent data adequately which was also confirmed by the verification statistics.

Figures 6.7 and 6.8 show the performance of the final reconstruction model using all available data with respect to the gauged data. No independent time period is available for verification in this case. Figure 6.7 shows a good correspondence between gauged low flows and modelled low flows as well as a good correspondence with above average flows although magnitude is not represented well for high flows. Figure 6.8 confirms this by showing relatively tight correlation between the gauged and modelled data especially below 15000 Mm^3.

The split sample verification shows that the 'early' model and 'late' model perform similarly well. The sign and order of the regression coefficients are the same for all models and all passed both the product moment correlation coefficient test and the reduction of error test. This gives confidence in the performance of the full model and indicates a satisfactory result for the reconstruction.

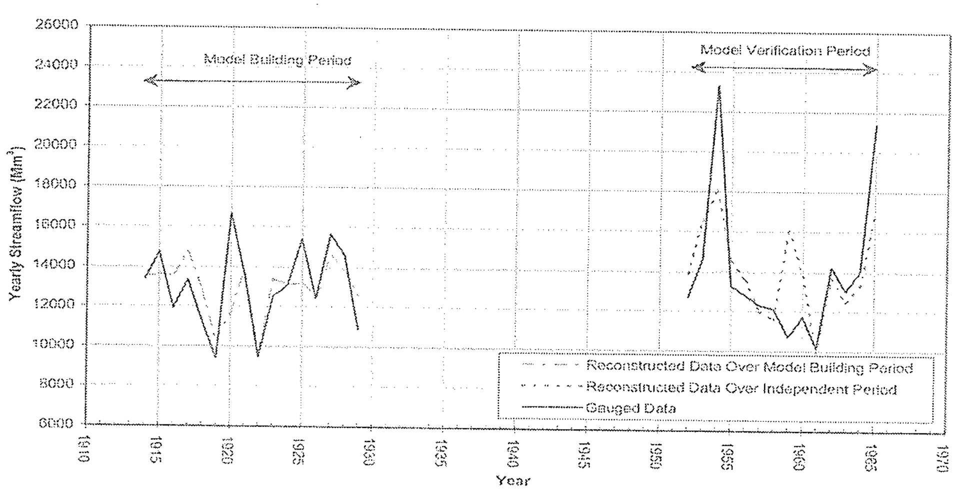


Figure 6.3: Time Series Plot of 07BE001 Using 1914–1929 as the Regression Period and 1952–1965 as the Verification Period

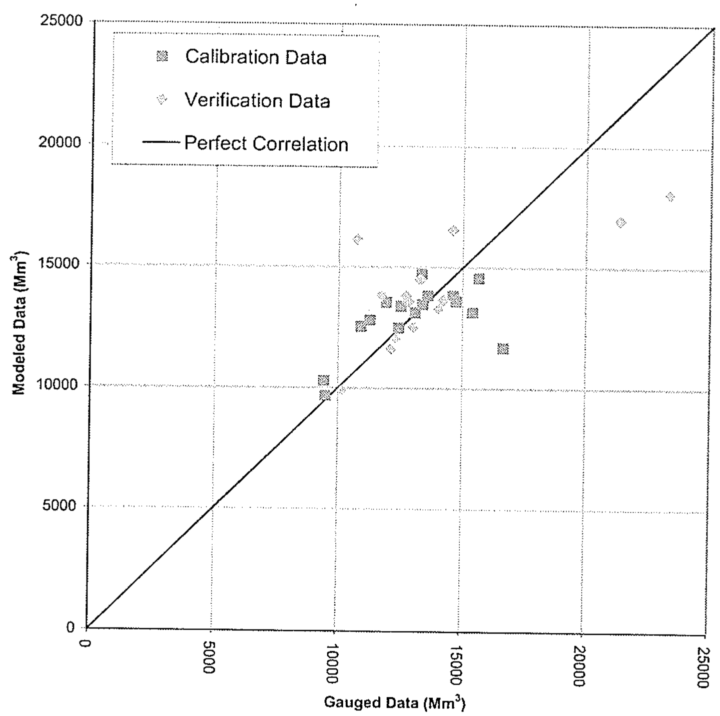


Figure 6.4: Scatter Diagram of Verification Data for 1914–1929 Regression Period and 1952–1965 Verification Period

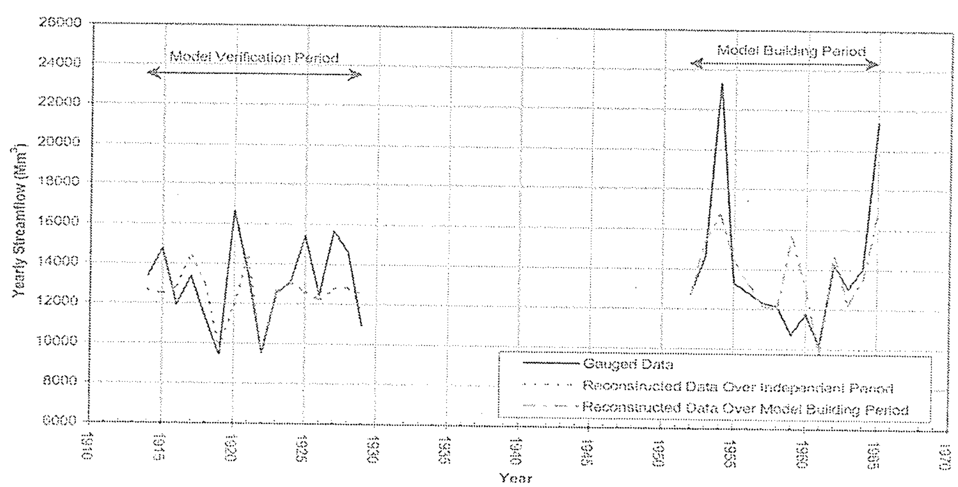


Figure 6.5: Time Series Plot of 07BE001 Using 1952–1965 as the Regression Period and 1914–1929 as the Verification Period

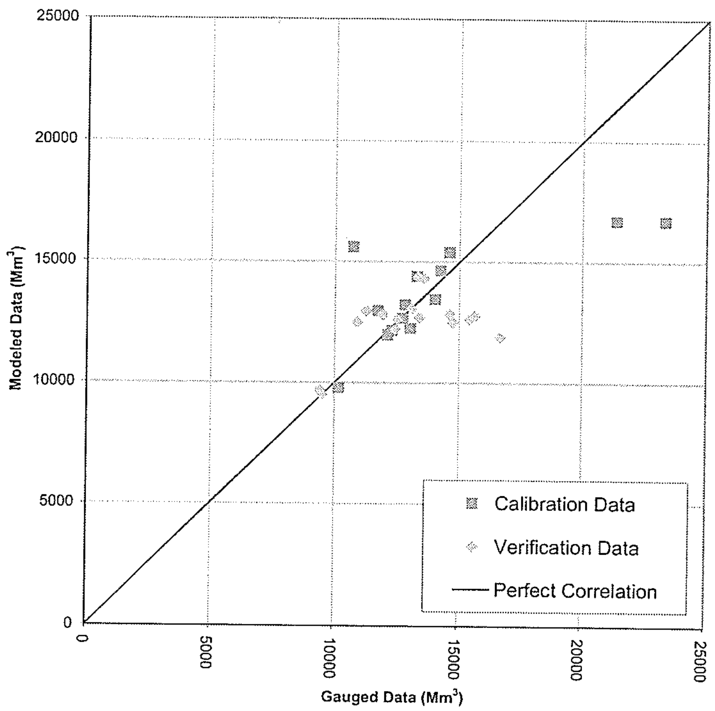


Figure 6.6: Scatter Diagram of Verification Data for 1952–1965 Regression Period and 1914–1929 Verification Period

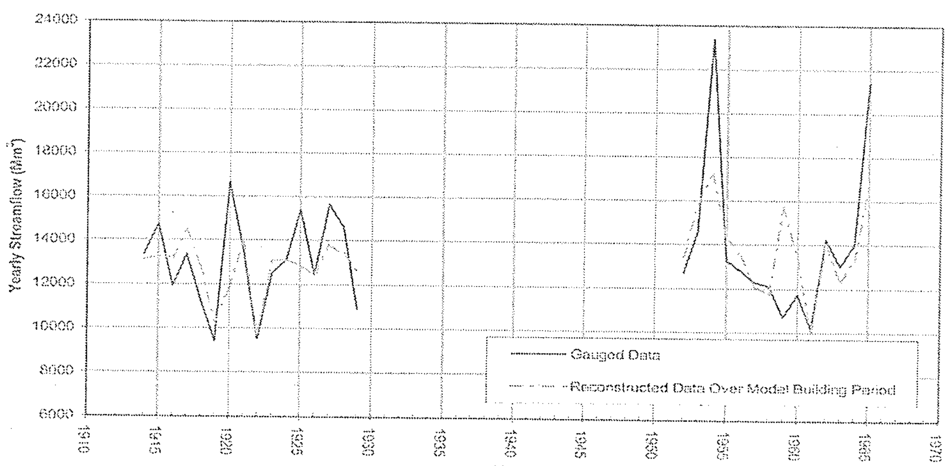


Figure 6.7: Time Series Plot Using Entire Record as the Regression Period with No Verification Period

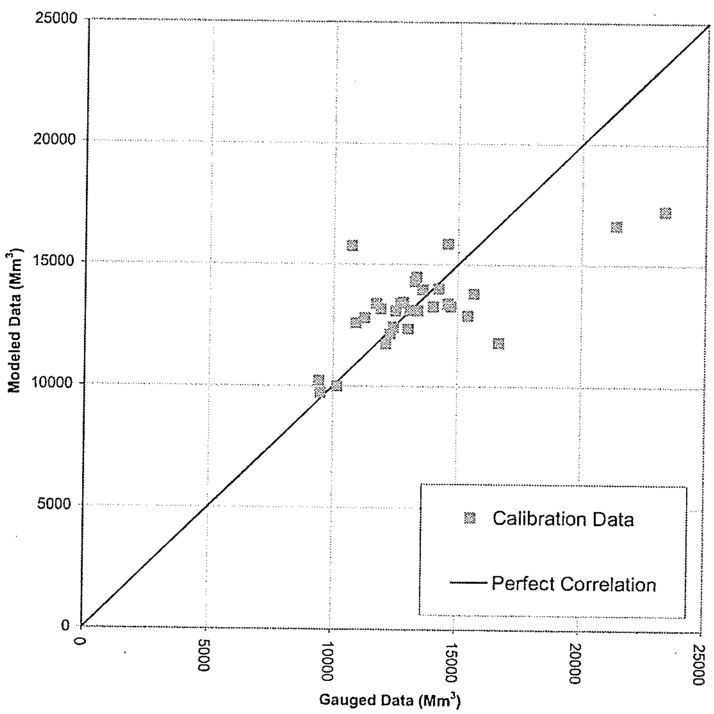


Figure 6.8: Scatter Diagram of Calibration Data for Full Regression Period

6.2.8 Final Reconstruction

Figure 6.9 shows the final reconstruction with gauged data incorporated. Tables of the gauged

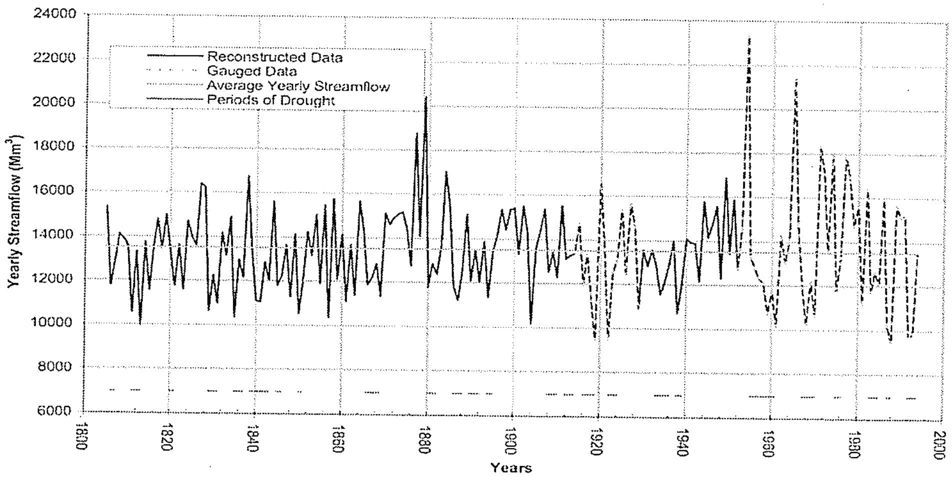


Figure 6.9: Time Series of Reconstructed 07BE001 Flow Record

and reconstructed data may be found in Appendix F.

As can be observed in this plot, there is a distinct reduction in the amount of variance from the gauged to the reconstructed record. In the 59 year gauged record streamflow drops below 10000 $Mm^3/year$ on 4 separate occasions. During the 131 year reconstructed record however streamflow never drops below this level. This indicates that even though the verification results were positive the reconstruction is not reproducing the full variance of the streamflow record. This being said the reconstruction can still be used qualitatively to illustrate the time of occurrence of past severe droughts and long duration drought periods.

6.2.9 Drought Analysis

An analysis was performed on the reconstructed data using the procedure described in Section 2.3. A truncation level of the mean annual runoff volume was used to separate low flow from high flow years. Distinct multiyear droughts were formed by grouping adjacent years of lower than average flow. These are presented in Table 6.5 sorted by Severity (the water deficit over the drought period with respect to mean annual runoff). Several observations can be made about this new drought record:

- The reconstruction identified a record of 49 distinct drought events including 101 low flow years compared to a record of 15 distinct drought events including 33 low flow years from the gauged record.
- Although the highest severity drought identified was from the gauged record, four droughts were identified in the reconstructed record that were at least as severe as the six most severe droughts recorded.
- Three droughts lasting four years or more were identified in the reconstructed record in addition to the three droughts identified in the gauged record

A frequency analysis was performed for illustrative purposes on the gauged and reconstructed data using methods described in Section 2.3. In this analysis the severities for the reconstructed and gauged drought events were fit to multiple distributions using hyfran (Chair in statistical hydrology, 2002). A probability distribution was chosen by qualitatively comparing the different distributions on a single plot. In the case of the reconstructed data shown in Figure 6.10 none of the probability distributions fit the data particularly well. The Weibull was chosen however because it preserves the shape of the data and the values are slightly conservative when compared to the data. With the gauged data shown in Figure 6.11 the Weibull distribution fits the data the best out of the five distributions.

The best fit line was then plotted with 95% confidence limits as well as the original data. The plots for the gauged and reconstructed data can be observed in Figures 6.12 and 6.13. These plots show that the drought analysis using the reconstructed data underpredicts the severity of droughts at all return periods. Table 6.6 also illustrates this quantitatively. This is due to the reduction of variance in the reconstruction and illustrates that the reconstructed record is indeed unsuitable for frequency analysis.

6.2.10 Important Observations

This case study has shown that it is possible to reconstruct low flow data with reasonable accuracy using tree ring data previously collected within the general area. The data analysis

Table 6.5: Historical Drought Periods for 07BE001

Starting Year	Ending Year	Severity (Mm^3)	Magnitude (Mm^3)	Duration (years)	Gauged Data (Y/N)
1955	1961	11387.6	1626.8	7	Yes
1967	1970	8994.8	2248.7	4	Yes
1916	1919	8055.1	2013.8	4	Yes
1992	1994	7485.1	2495.0	3	Yes
1987	1988	7316.3	3658.1	2	Yes
1839	1843	7013.0	1402.6	5	No
1811	1813	6687.9	2229.3	3	No
1829	1831	6664.9	2221.6	3	No
1866	1869	5761.2	1440.3	4	No
1922	1924	5391.1	1797.0	3	Yes
1835	1837	5008.8	1669.6	3	No
1886	1888	4981.3	1660.4	3	No
1933	1936	4577.8	1144.5	4	No
1850	1851	4520.3	2260.2	2	No
1938	1939	4267.6	2133.8	2	No
1983	1985	4101.3	1367.1	3	Yes
1880	1882	3463.5	1154.5	3	No
1904	1904	3315.4	3315.4	1	No
1857	1857	3128.6	3128.6	1	No
1845	1846	2918.1	1459.1	2	No
1890	1892	2861.5	953.8	3	No
1929	1929	2593.1	2593.1	1	Yes
1861	1861	2411.6	2411.6	1	No
1806	1807	2331.5	1165.8	2	No
1981	1981	2314.0	2314.0	1	Yes
1908	1910	2278.6	759.5	3	No
1894	1895	2270.2	1135.1	2	No
1848	1848	2211.0	2211.0	1	No
1863	1863	2105.6	2105.6	1	No
1975	1976	2035.1	1017.6	2	Yes
1815	1815	1939.5	1939.5	1	No
1823	1823	1894.6	1894.6	1	No
1820	1821	1893.0	946.5	2	No
1855	1855	1598.7	1598.7	1	No
1859	1859	1378.9	1378.9	1	No
1943	1943	1332.8	1332.8	1	No
1948	1948	1207.8	1207.8	1	No
1926	1926	1066.5	1066.5	1	Yes
1952	1952	789.4	789.4	1	Yes
1876	1876	753.0	753.0	1	No
1912	1914	727.9	242.6	3	Yes
1931	1931	654.6	654.6	1	No
1963	1963	484.7	484.7	1	Yes
1833	1833	355.3	355.3	1	No
1853	1853	296.5	296.5	1	No
1901	1901	177.7	177.7	1	No
1950	1950	71.7	71.7	1	No
1818	1818	31.2	31.2	1	No
1973	1973	6.3	6.3	1	Yes

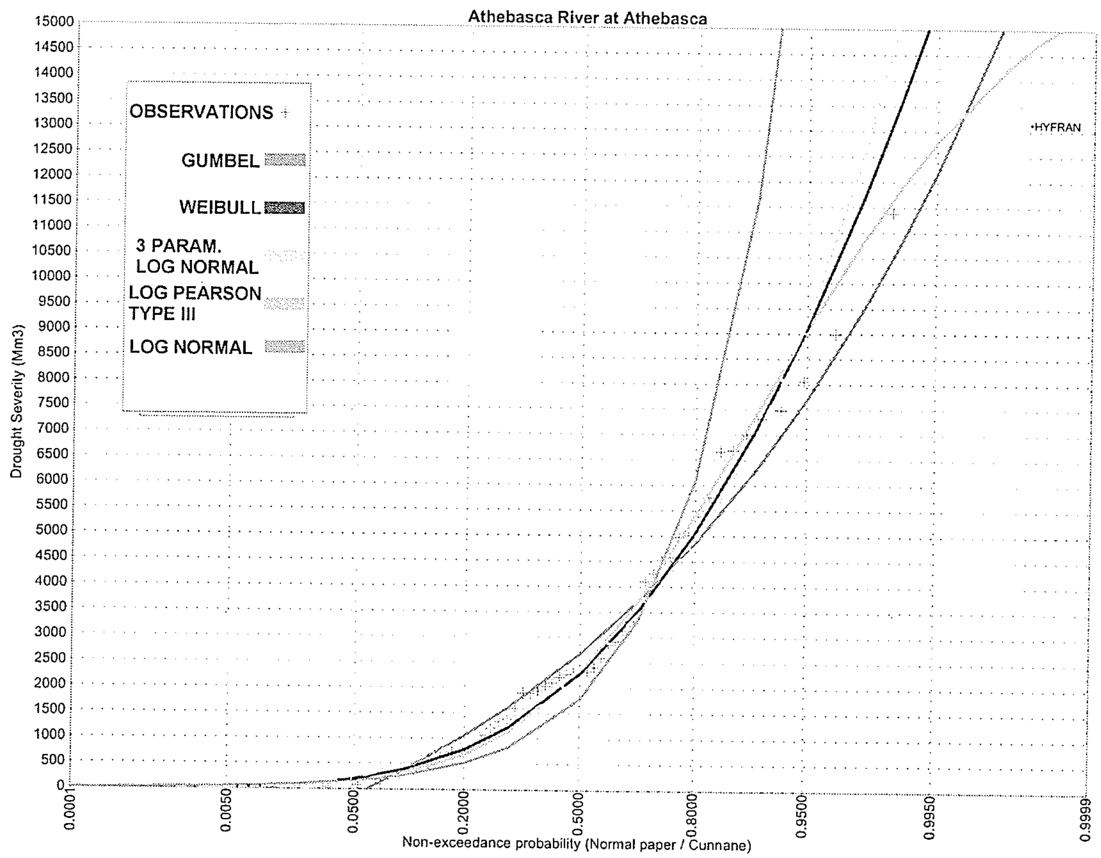


Figure 6.10: Frequency Distributions for Reconstructed Droughts

Table 6.6: Comparison of Drought Frequency Analysis at Various Return Periods for 07BE001

Return Period	Gauged Data	Reconstructed Data
(Years)	Water Deficit (Mm^3/yr)	
2	1.3	0.5
10	6130.8	4947
20	8644.7	6976.4
50	11973.2	9469.6
100	14218.5	11578.3
1000	21042.8	17172.2
10000	28335.5	23474.4

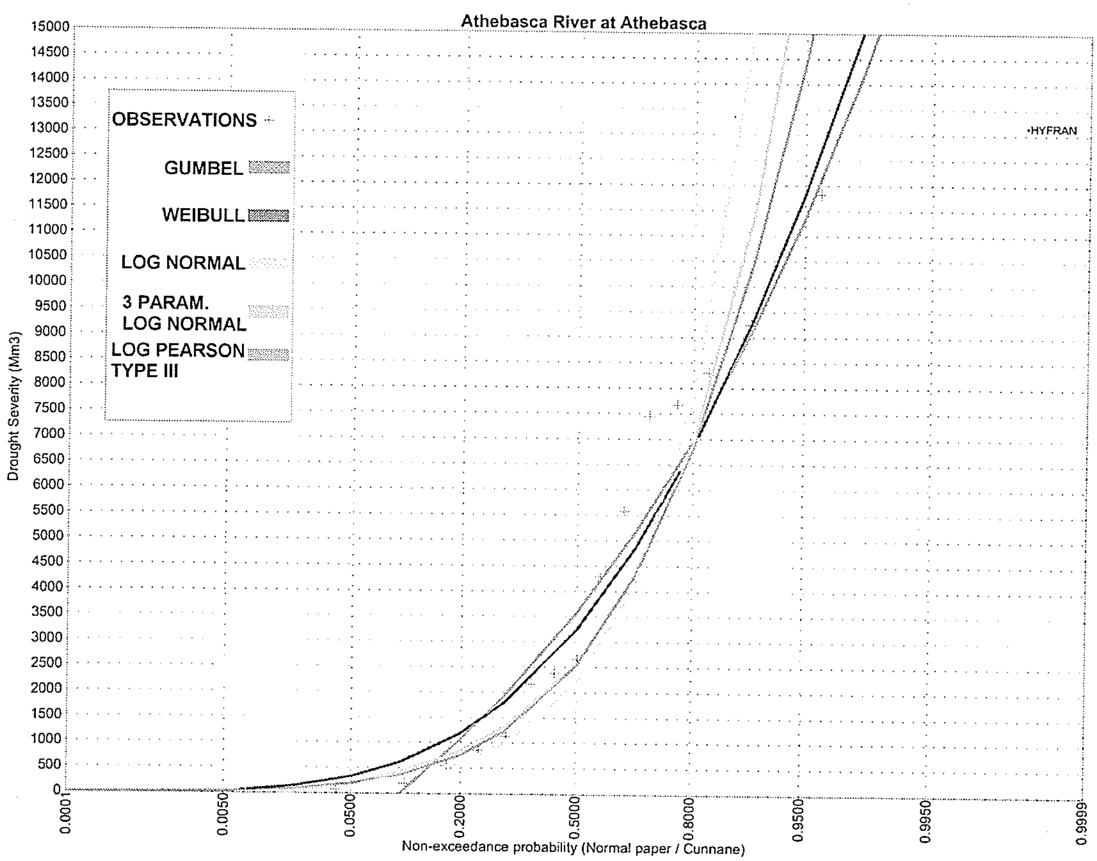


Figure 6.11: Frequency Distributions for Gauged Droughts

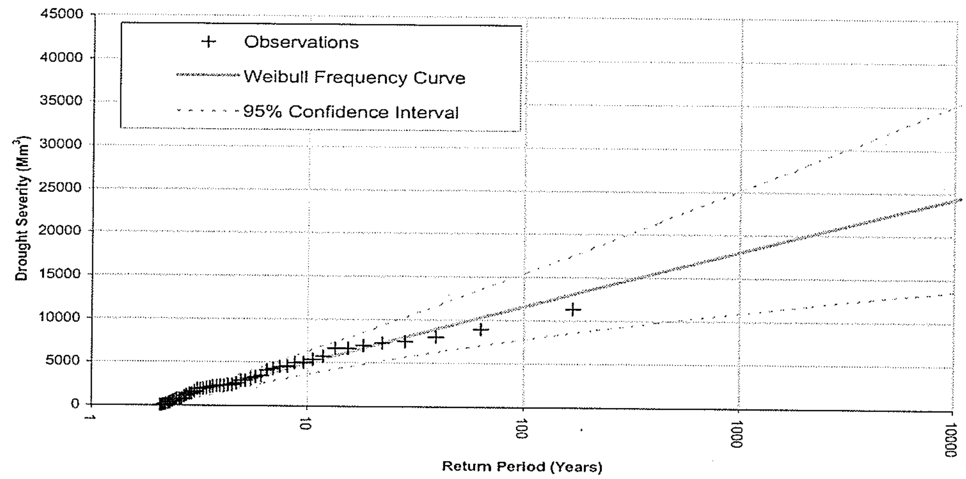


Figure 6.12: Drought Frequency Analysis Based on Reconstructed Data for 07BE001

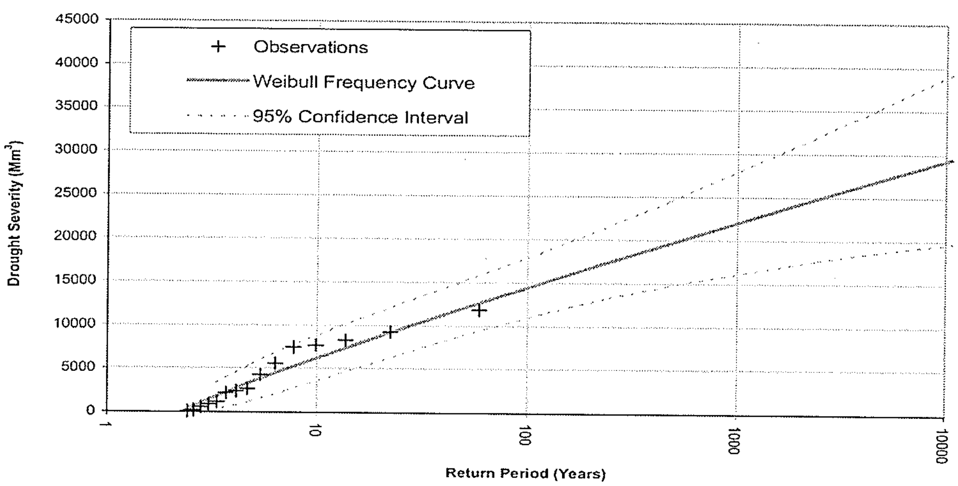


Figure 6.13: Drought Frequency Analysis Based on Gauged Data for 07BE001

presented has shown a tripling of the number of observed multiyear drought events and low flow years in the record. It also illustrates that the reduction in variance that is a result of the low explained variance (R^2) in the reconstruction model makes the data unsuitable for use in a highly quantitative analysis.

6.3 Oldman River Near Waldron's Corner

6.3.1 Background Information

The Oldman River forms the western most part of the South Saskatchewan River Basin. This river is shown in Figure 6.14.

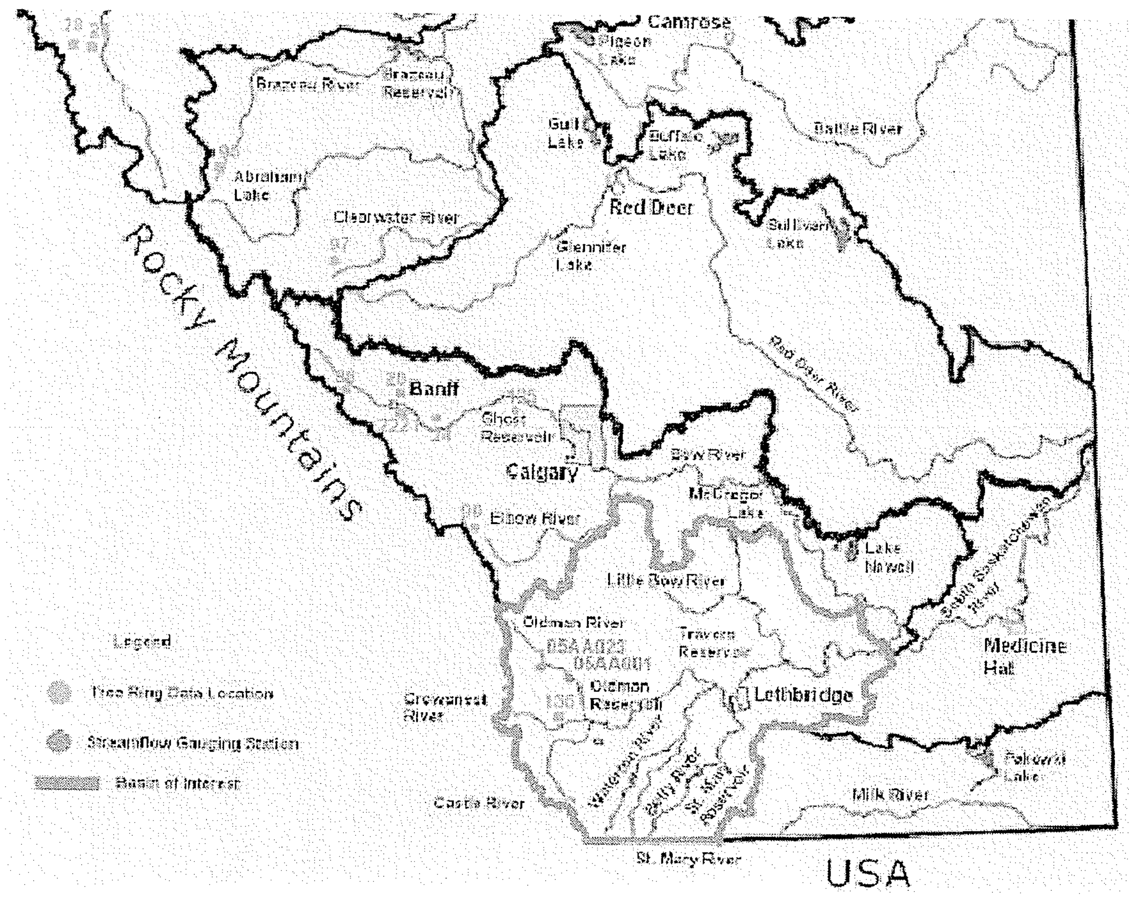


Figure 6.14: Map of Oldman River Gauge Locations

The Oldman River runs from the Rocky Mountains through Lethbridge until it links up with the Bow River to form the South Saskatchewan River. The terrain of the river ranges from mountainous to prairie. The total drainage area is 24 410 km^2.

The Oldman River is very important to Southern Alberta. The Oldman River and the Bow River provide more than 98% of all the irrigation water in Southern Alberta. In 1992, the

Oldman Dam was built to store the water from Oldman River to augment irrigation in low flow years. It was constructed as a multipurpose reservoir for irrigation, industrial consumption, recreation and the potential to install 34 MW of hydro capacity. This hydro installation is currently being approved. It remains the largest reservoir in Southern Alberta.

The gauge reconstructed is the Water Survey of Canada gauge 05AA023 described as Oldman River Near Waldron's Corner.

6.3.2 Available Data

Streamflow data for gauge 05AA023 were obtained from the Water Survey of Canada Hydat CD. Data at this location are available between 1950 and 1990. This did not provide enough overlap with tree ring data sets, in close proximity to 05AA023, for an acceptable regression analysis and verification. It was, therefore, decided to use a gauge downstream on the same river to supplement this record.

The gauge 05AA001, Oldman River Near Cowley, is approximately 20km downstream of 05AA023. Its record extends between 1911 and 1930. There is no overlapping period between the two gauged records. It is surmised that 05AA001 was operated until 'The Dirty Thirties' when funding was cut for most flow monitoring. Operation was then resumed after World War II at a new gauge location with a new gauge number. No major tributaries enter the river between the two gauge locations, therefore a straight pro-ration of monthly data based on the ratio of drainage areas was deemed appropriate. To this end the monthly average data in 05AA001 was multiplied by a factor of $\frac{1400km^2}{1940km^2}$, the drainage area of 05AA023 over the drainage area of 05AA001. These data were then combined with 05AA023 to form a complete record at Waldron's Corner from 1911-1930 and 1950-1990 with a potential minimum overlap of 35 years with tree ring data in the area.

The monthly data were converted to 12 series of yearly streamflow volumes based on starting months between January and December. These were then preprocessed by techniques discussed in Section 3.2. None of the flow data had to be normalized using Box-Cox transformation and no trends or autocorrelation were found.

6.3.3 Predetermination of Predictors

The number of tree ring data sets was narrowed from a possible 140 candidates to 51 based on the criterion suggested by Cook (1995) that they should be within a radius of 500 km from the gauge to be reconstructed. The tree ring data sets to be entered into the best subsets analysis were further reduced 'a priori' based on judgement and the five criteria discussed in Section 3.3 as follows.

Only one of the data sets, CANA136 Crowsnest Pass was within the gauging station's sub-basin. Of the remaining sets, seven were in the Bow River Basin which is also part of the South Saskatchewan River Basin and adjacent to the Oldman River Basin. Four were in sub-basins of the Mackenzie River Basin in areas of similar geography as the Oldman River Basin. The other 39 potential data sets were located on the West side of the Rocky Mountains and, therefore, could not have responded to the same weather patterns as those located East of the Rocky Mountains due to orographic effects. All 12 tree ring data sets identified were investigated further.

All tree ring data had more than 25 years of overlap when compared to the two periods of continuous streamflow records from 1911–1929 and 1950–1995. The shortest overlap was 35 years.

All of the tree ring data sets could be made normal by Box-Cox transformation. They all displayed little or no trend, shifts in the mean or periodicity. Significant autocorrelation was removed with low order ARMA modelling (AR1 or AR1 MA1).

A correlation analysis between tree ring series and **monthly** streamflow yielded eight tree ring series with significant correlations. These are presented in Table 6.7

Seven of these are within the South Saskatchewan River Basin so they could have responded to the same flow characteristics present in the gauged record. One is within the MacKenzie River Basin so it could have responded to weather patterns common to both basins. Plots of the correlation analysis against monthly flows are shown in Appendix D. In each case the tree ring record is significantly correlated with at least one monthly streamflow record. This indicates at least some useful information within the tree ring series for reconstruction of the streamflow series.

Table 6.7: Tree Ring Data Significantly Correlated With Monthly Record of 05AA023

Identifier	Description	Minor Basin	Major Basin
CANA020	Powerhouse, Alberta	Bow River	Saskatchewan River
CANA021	Tunnel Mountain, Banff, Alberta	Bow River	Saskatchewan River
CANA022	Exshaw, Tunnel and Banff, Alberta	Bow River	Saskatchewan River
CANA024	Exshaw, Alberta	Bow River	Saskatchewan River
CANA026	Pyramid Lake and Patricia Lake, Alberta	Athabasca River	MacKenzie River
CANA099	Sarrail Glacier, Alberta	Bow River	Saskatchewan River
CANA135	Towers Ridge, Alberta	Bow River	Saskatchewan River
CANA136	Crowsnest Pass, Alberta	Oldman River	Saskatchewan River

A correlation analysis between tree ring series and **annual** streamflow yielded five tree ring series with significant correlations. These are presented in Table 6.8. All of these were identified in the monthly correlation analysis. Plots of the correlation analysis against annualized streamflow are shown in Appendix D.

Table 6.8: Tree Ring Data Significantly Correlated With Annual Record of 05AA023

Identifier	Description	Minor Basin	Major Basin
CANA020	Powerhouse, Alberta	Bow River	Saskatchewan River
CANA021	Tunnel Mountain, Banff, Alberta	Bow River	Saskatchewan River
CANA022	Exshaw, Tunnel and Banff, Alberta	Bow River	Saskatchewan River
CANA024	Exshaw, Alberta	Bow River	Saskatchewan River
CANA136	Crowsnest Pass, Alberta	Oldman River	Saskatchewan River

The correlation analysis against monthly streamflow data yielded eight possible data sets to be investigated for model building. The correlation analysis against annual streamflow reaffirmed that five of these data sets contain significant common information with the gauged streamflow data. The reduction to eight data sets based on correlation with monthly streamflow simplifies the regression procedure.

6.3.4 Principal Components Analysis

The overlapping period for all eight tree ring series was found to be from 1571 to 1965. The tree ring sets were lagged forward and backward one year to account for growth and storage effects forming 24 possible predictors for the reconstruction. This matrix was then orthogonalized in space and the eigenvectors and eigenvalues tabulated. Seven eigenvectors had eigenvalues in excess of the Kaiser–Guttman eigenvalue–1 criterion. These seven components represent 72.1% of the total variance contained in the 24 predictors with the largest single vector representing 19.5% and the smallest 4.4%. These were retained for use in the best subsets model building exercise.

6.3.5 Best Subsets Analysis

The period of overlap between the tree ring series and the streamflow series from 1911–1929 and 1950–1965 was used in this analysis. Each monthly annualization was regressed against all possible combinations of the seven orthogonalized tree ring vectors. For each number of predictors (1 to 7) and each annualization the best model was chosen based on $R^2_{adjusted}$. The best model for each number of predictors was then separated out based on $R^2_{adjusted}$ and plotted in Figure 6.15. The model that produced the highest $R^2_{adjusted}$ with the least number of predictors was a regression using the 1st to 6th eigenvectors on the December Annualization. This regression produced an R^2 of 46.3% and $R^2_{adjusted}$ of 34.7%.

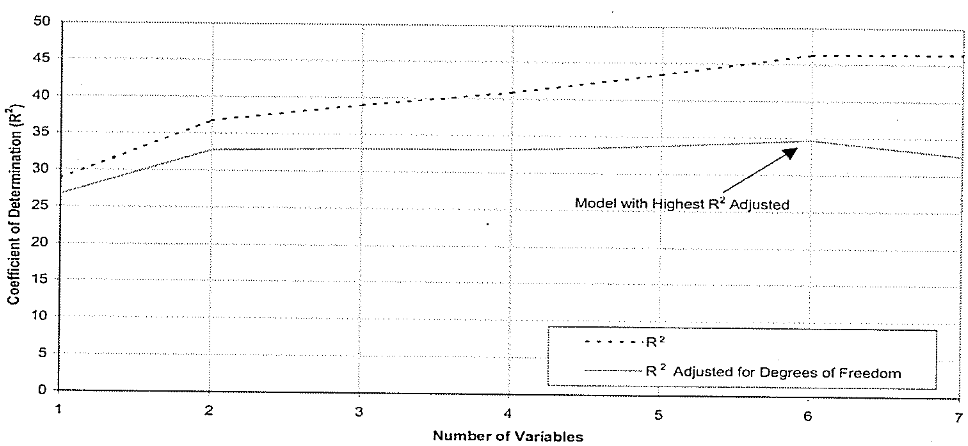


Figure 6.15: Results of Best Subset Analysis of 05AA023

6.3.6 Investigation of Outliers

Outliers and influential observations were investigated using Leverage, Studentized Residuals, Cook's Distance and Dfits. It was found that the data points for 1950 and 1955 had high Studentized Residuals, Cook's Distance and Dfits and that removal of these points significantly improved the regression model. These data points were therefore removed from the regression data set.

6.3.7 Model Building and Verification

The final model was built using the regression equation derived from the best subsets analysis. The verification was handled by a standard split sample procedure. The data from 1911–1929 was first used to build a regression equation (the 'early' model) and this was tested against the independent data from 1950–1965. In turn the data from 1950–1965 was used to build a regression model (the 'late' model) and verified with the independent data from 1911-1929.

The final regression equation takes the following form:

$$streamflow = b_0 + b_1 PC_1 + b_2 PC_2 + \ldots + b_m PC_m \qquad (6.2)$$

Where:

$streamflow$ = Estimate of annualized streamflow

m = Number of predictor variables

$PC_1 \ldots PC_m$ = Principal components of tree ring data

$b_1 \ldots b_m$ = Regression coefficients corresponding to each principal component

b_0 = Intercept coefficient which scales the regression equation to the mean of the predictand

The regression coefficients can be observed in Table 6.9 as well as the R^2, R^2_{adj} and gauged versus modelled means and standard deviations. Some qualitative observations from the model building portion of this exercise are as follows. The highest R^2_{adj} was 54.3% for the regression model built on the 'late' data and the lowest was 50.0% for the regression model on the 'early' data. All three of the R^2_{adj} for the models were within reasonable limits for past streamflow studies in the literature.

Table 6.9: Calibration Statistics

Parameter	**Early Calibration (1911–1929)**	**Late Calibration (1950–1965)**	**Full Calibration (1911–1929 and 1950–1965)**
R^2	0.667	0.754	0.596
R^2_{adj}	0.500	0.543	0.503
B_0	515.19	569.91	521.19
B_1	-27.92	-40.67	-26.161
B_2	49.34	-29.24	42.58
B_3	32.62	44.54	36.058
B_4	-8.76	76.30	-0.21
B_5	12.86	32.41	20.10
B_6	24.38	-64.88	20.25
Gauged Mean	567.137	496.477	526.454
Reconstructed Mean	556.036	564.116	526.455
Gauged Standard Deviation	109.640	143.117	132.901
Reconstructed Standard Deviation	74.446	202.880	102.620

The regression coefficients for the 'full' and 'early' regression models are of the same order and sign. This is a good indicator of model stability for predicting independent data. The 'late' regression model differs from the other two in sign for the B_2, B_4 and B_6 coefficients. This raises some concerns about model stability which will be born out in the verification.

Verification statistics for each of the split models and the full model are shown in Table 6.10. The sign test shows that the 'full' model passes at 95% when applied to the calibration data.

Table 6.10: Verification Statistics

Parameter	**Calibration Period (1911–1929) Verification Period (1950–1965)**	**Calibration Period (1950–1965) Verification Period (1911–1929)**	**Full Calibration (1911–1929 and 1950–1965)**
Sign Test			
Right	11	14	25
Wrong	3	5	5
Status (Confidence)	Pass(90%)	Pass(90%)	Pass(99%)
Product Means Test			
tvalue	0.570	1.510	3.820
minimum value	1.761	1.729	2.730
Status (Confidence)	Fail(90%)	Fail(90%)	Pass(99%)
Product Moment Correlation Coefficient Test			
tvalue	2.17	1.45	6.77
minimum value	2.14	1.73	2.73
Status (Confidence)	Pass(95%)	Fail(90%)	Pass(99%)
Reduction of Error Test			
value	0.482	-0.835	0.596
minimum value	0	0	0
Status	Pass	Fail	Pass

The 'early' and 'late' models, both pass only at the 90% confidence interval. This does not conclusively validate the model but indicates that the signs of the first differences agree between the gauged and modelled data more often than would be expected at random with not more than 90% confidence.

The product means test for both the 'early' and 'late' models failed at the 90% confidence level while the 'full' model passed at greater than 99% on the calibration data. Again one cannot be sure that no relationship exists by a failure of this test as it is very sensitive to large deviations from the mean.

The 'early' model passed the product moment correlation coefficient test at 90% confidence while the 'late' model failed at this level. The 'full' model passed the product moment correlation coefficient test at greater than 95% confidence. A pass implies that the variance between the gauged and reconstructed data is linearly related. It is a powerful indicator of association between two variables provided the actual and estimated means are essentially the same.

The reduction of error statistic was passed for the 'early' and 'full' models. Results for the 'late' model indicate that the regression does not perform as well as the mean of the calibration data. The RE statistic is the same as R^2 for the 'full' model. This is an extremely rigorous and sensitive verification statistic because it has no lower bound (Fritts, 1976). A few bad estimates result in a negative RE statistic. It is used extensively within the literature as the most important indicator of reconstruction reliability.

Figures 6.16 and 6.17 show how the 'early' regression model performs on independent data. From Figure 6.16 it can be seen qualitatively that the calibrated model does a reasonably good job of reconstructing both low and high flow events although during the verification period the low flow events are somewhat better represented. The scatter diagram in Figure 6.17 shows relatively tight correlation between the gauged and modelled data for both the calibration and verification periods. The variance does not appear to change significantly between the calibration and verification data. This is a good indication that the early model is reconstructing the independent data adequately and provides confidence that a regression model built upon this data will represent past ungauged droughts relatively well provided they are within the realm of known data.

Figures 6.18 and 6.19 show how the 'late' regression model performs on independent data. From Figure 6.18 it can be seen qualitatively that the calibrated model does a reasonably good job of reconstructing both low and high flow events during the calibration period but does not reproduce the verification data very well at all. The scatter diagram in Figure 6.19 confirms this by showing relatively tight correlation between the gauged and modelled data for the calibration period but a much greater spread for the verification period. The variance changes significantly between the calibration and verification data. This indicates that the late model does not reconstruct the independent data adequately which was also confirmed by the verification statistics.

Figure 6.20 and Figure 6.21 show the performance of the the final reconstruction model using all available data with respect to the gauged data. Figure 6.20 shows a good correspondence between gauged low flows and modelled low flows as well as a poorer representation of above average flows. Figure 6.21 confirms this by showing relatively tight correlation between the gauged and modelled data especially below 600 Mm^3.

The split sample verification shows that the 'early' model and 'full' model perform similarly well while the 'late' model does not. The sign and order of the regression coefficients are the same for the 'early' and 'full' models and both passed all of the verification statistics. This gives some confidence in the performance of the full model.

The poor performance of the late model could be due to several factors. The combination of the two data sets could have introduced inhomogeneities into the streamflow data. Also the period between 1950 and 1965 had a large number of high flow years which are not well represented by tree rings.

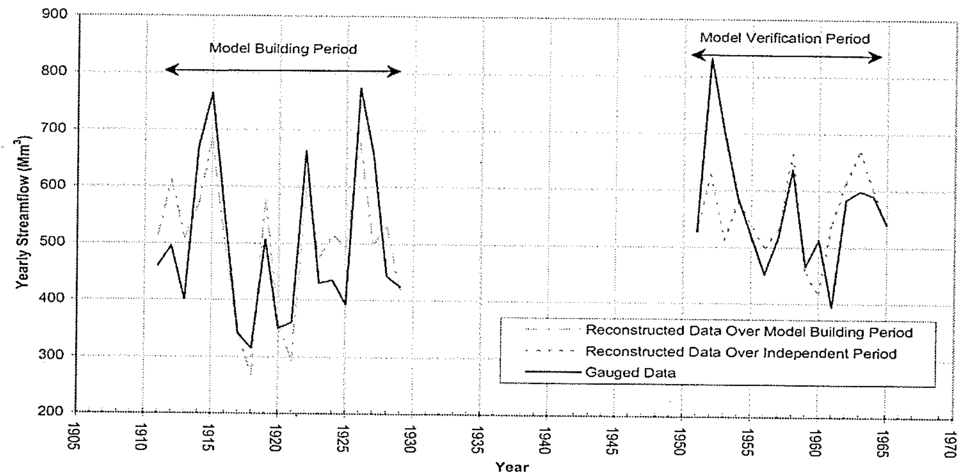


Figure 6.16: Time Series Plot Using 1911-1929 as the Regression Period and 1950-1965 as the Verification Period

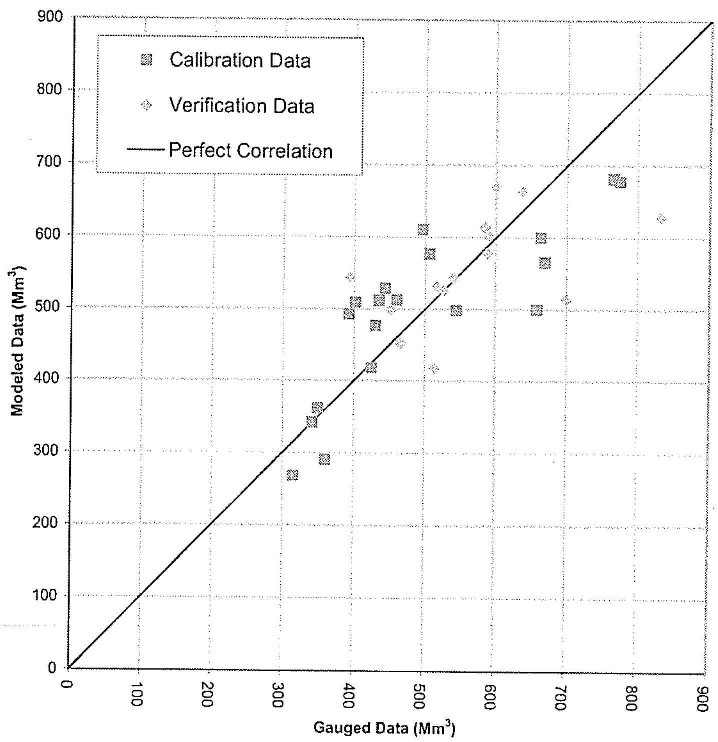


Figure 6.17: Scatter Diagram of Verification Data for 1911-1929 Regression Period

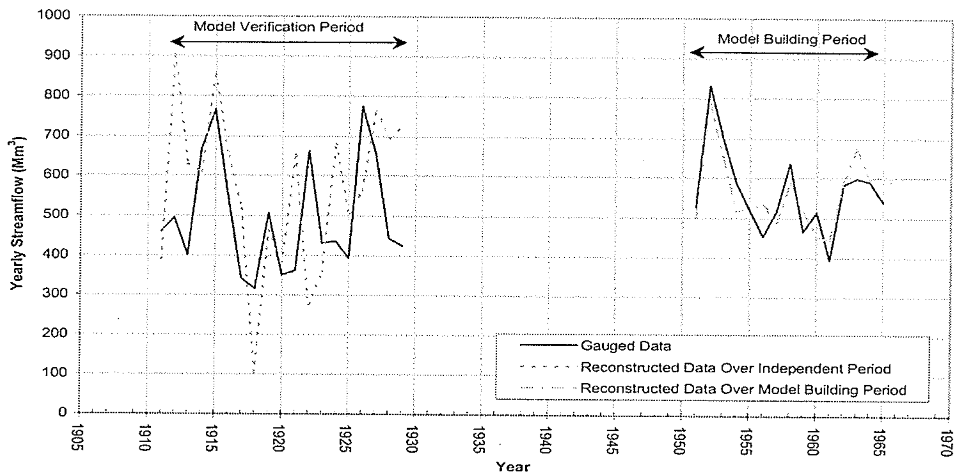


Figure 6.18: Time Series Plot Using 1950–1965 as the Regression Period and 1911–1929 as the Verification Period

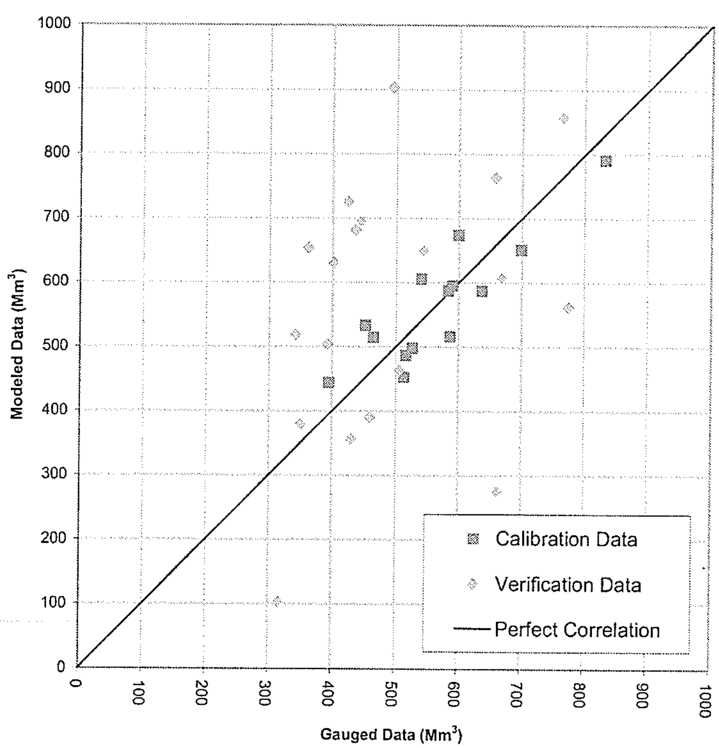


Figure 6.19: Scatter Diagram of Verification Data for 1950-1965 Regression Period

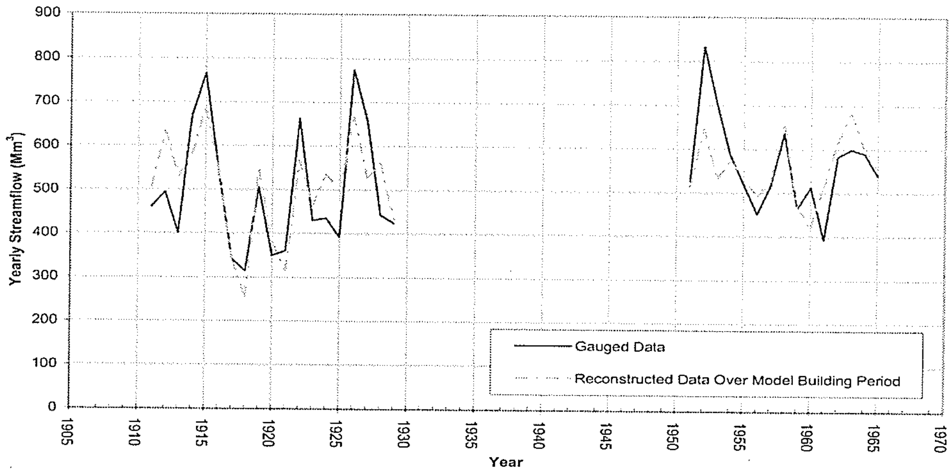


Figure 6.20: Time Series Plot Using Entire Record as the Regression Period with No Verification Period

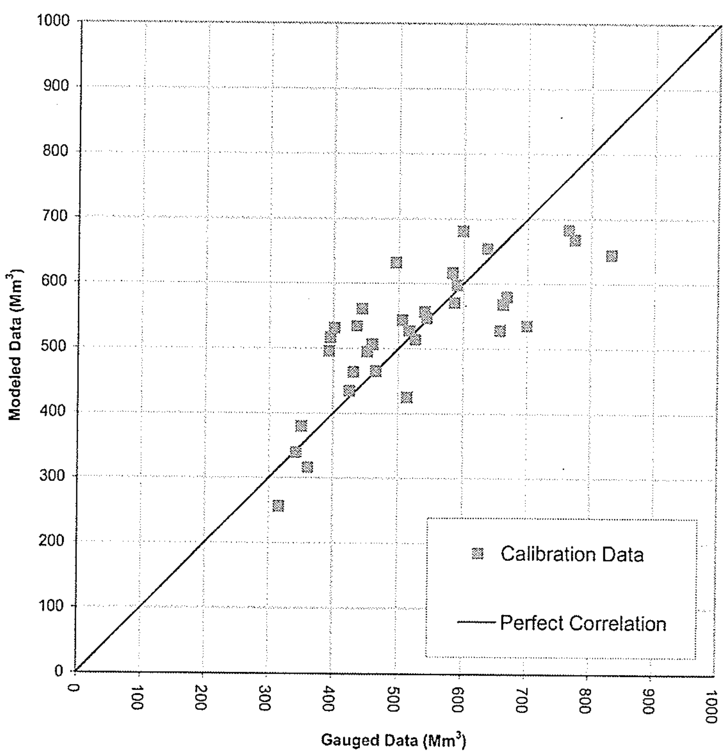


Figure 6.21: Scatter Diagram of Calibration Data for Full Regression Period

6.3.8 Final Reconstruction

Figure 6.22 shows the final reconstruction with gauged data incorporated. Tables of the gauged

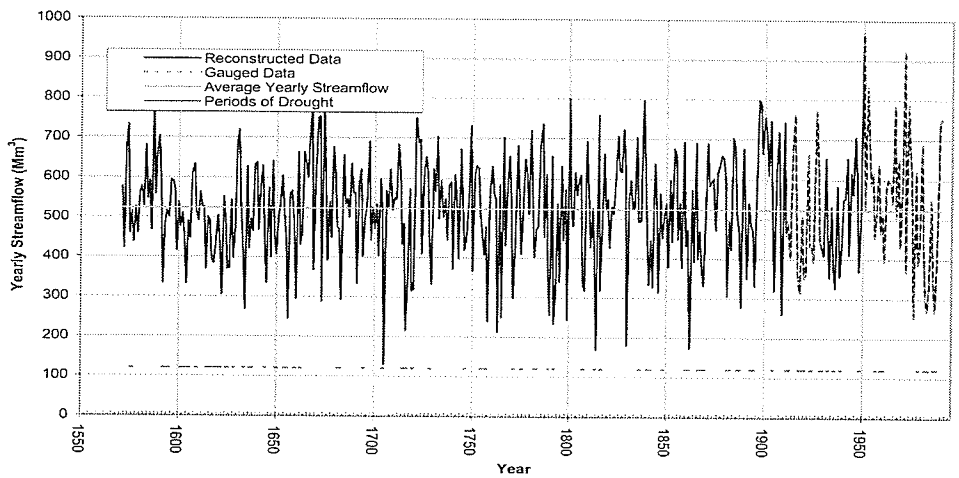


Figure 6.22: Time Series of Reconstructed 05AA023 Flow Record

and reconstructed data may be found in Appendix G.

Unlike the previous case study, the change in variance is not as pronounced. The standard deviation of the reconstruction in this case is 77% of that of the gauged data alone. It was only 47% in the previous case study. This moderate reduction in variance still makes accurate quantitative analysis of drought very difficult as one can never be sure if the quantities reconstructed are indeed representative. The potential for applying this data in a verification role or qualitative analysis still exists, however.

6.3.9 Drought Analysis

An analysis was performed on the reconstructed data based on the procedure described in Section 2.3. A truncation level of the mean annual runoff volume was used to separate low flow from high flow years. Distinct multiyear droughts were then formed by grouping adjacent years of lower than average flow. These are presented in Table 6.11 sorted by Severity. Several observations can be made about this new drought record.

- The reconstruction identified a record of 103 distinct drought events including 199 low flow years compared to a record of 14 distinct drought events including 32 low flow years from the gauged record.

- The highest severity drought identified was from the reconstructed record, three droughts were identified in the reconstructed record that were at least as severe as the three most severe droughts recorded.

- Seven droughts lasting four years or more were identified in the reconstructed record in addition to the two droughts identified in the gauged record

A frequency analysis was performed for illustrative purposes on the gauged and reconstructed data using methods described in Section 2.3. In this analysis the severities for the reconstructed and gauged drought events were fit to multiple distributions using hyfran (Chair in statistical hydrology, 2002). A probability distribution was chosen by qualitatively comparing the different distributions on a single plot. In the case of the reconstructed data shown in Figure 6.23 the Weibull distribution fits the data the best out of the five distributions. With the gauged data shown in Figure 6.24 the Weibull distribution fits the data the best out of the five distributions.

The best fit line was plotted with 95% confidence limits as well as the data. The plots for the gauged and reconstructed data can be observed in Figures 6.25 and 6.26. These plots show that the drought analysis using the reconstructed data underpredicts the severity of droughts at all return periods. Table 6.12 illustrates this quantitativly. This is due to the reduction of variance in the reconstruction and illustrates that even with only a moderate reduction in variance the reconstructed record is still unsuitable for frequency analysis.

6.3.10 Important Observations

The split sample verification showed that a model built on the data between 1950 and 1965 produced poor results. The cause of this could be the combination of two data sets. Even though they are on the same river, the flow provided by the additional drainage area may not have been entirely compensated for by a simple pro-rating of areas. A cross correlation would

Table 6.11: Historical Drought Periods for 05AA023

Starting Year	Ending Year	Severity (Mm^3)	Magnitude (Mm^3)	Duration (years)	Gauged Data (Y/N)	Starting Year	Ending Year	Severity (Mm^3)	Magnitude (Mm^3)	Duration (years)	Gauged Data (Y/N)
1614	1622	901.3	100.1	9	No	1746	1747	182.4	91.2	2	No
1981	1984	758.0	189.5	4	Yes	1628	1629	179.4	89.7	2	No
1917	1921	744.6	148.9	5	Yes	1891	1891	174.1	174.1	1	No
1714	1717	628.8	157.2	4	No	1575	1577	162.7	54.2	3	No
1866	1870	559.7	111.9	5	No	1750	1750	157.9	157.9	1	No
1986	1988	556.0	185.3	3	Yes	1662	1663	157.8	78.9	2	No
1754	1758	527.3	105.5	5	No	1669	1669	156.3	156.3	1	No
1840	1843	514.0	128.5	4	No	1972	1972	156.0	156.0	1	Yes
1792	1793	504.6	252.3	2	No	1740	1740	153.9	153.9	1	No
1860	1862	494.8	164.9	3	No	1694	1695	149.3	74.7	2	No
1813	1814	476.7	238.4	2	No	1851	1851	146.8	146.8	1	No
1789	1790	442.5	221.3	2	No	1858	1858	146.4	146.4	1	No
1601	1606	440.3	73.4	6	No	1676	1676	132.1	132.1	1	No
1719	1720	414.2	207.1	2	No	1864	1864	131.3	131.3	1	No
1704	1705	409.6	204.8	2	No	1743	1743	126.7	126.7	1	No
1807	1808	387.2	193.6	2	No	1647	1647	126.0	126.0	1	No
1633	1634	365.5	182.7	2	No	1702	1702	119.2	119.2	1	No
1654	1656	360.9	120.3	3	No	1775	1776	114.8	57.4	2	No
1881	1883	348.4	116.1	3	No	1724	1724	114.5	114.5	1	No
1935	1936	347.5	173.8	2	No	1943	1943	113.2	113.2	1	No
1830	1830	342.1	342.1	1	No	1599	1599	108.2	108.2	1	No
1771	1772	336.7	168.4	2	No	1572	1572	101.2	101.2	1	No
1624	1626	334.8	111.6	3	No	1821	1821	97.9	97.9	1	No
1681	1683	332.6	110.9	3	No	1811	1811	96.5	96.5	1	No
1886	1888	319.3	106.4	3	No	1969	1969	93.4	93.4	1	Yes
1923	1925	312.3	104.1	3	Yes	1767	1767	93.3	93.3	1	No
1763	1763	311.3	311.3	1	No	1848	1849	85.6	42.8	2	No
1928	1930	294.8	98.3	3	Yes	1853	1853	82.8	82.8	1	No
1893	1895	286.1	95.4	3	No	1698	1698	80.9	80.9	1	No
1799	1799	280.4	280.4	1	No	1797	1797	79.6	79.6	1	No
1728	1729	279.2	139.6	2	No	1737	1737	78.4	78.4	1	No
1765	1765	274.3	274.3	1	No	1956	1957	78.3	39.1	2	Yes
1976	1976	271.5	271.5	1	Yes	1856	1856	75.6	75.6	1	No
1659	1660	268.0	134.0	2	No	1678	1678	74.9	74.9	1	No
1592	1595	264.8	66.2	4	No	1787	1787	71.3	71.3	1	No
1909	1909	263.9	263.9	1	No	1579	1579	64.7	64.7	1	No
1938	1939	261.8	130.9	2	No	1836	1837	62.0	31.0	2	No
1816	1817	254.7	127.4	2	No	1586	1586	57.7	57.7	1	No
1644	1645	253.4	126.7	2	No	1641	1641	56.3	56.3	1	No
1978	1979	240.2	120.1	2	Yes	1875	1875	52.4	52.4	1	No
1673	1673	235.8	235.8	1	No	1700	1700	51.0	51.0	1	No
1782	1784	220.0	73.3	3	No	1802	1802	44.1	44.1	1	No
1932	1933	217.6	108.8	2	No	1609	1610	42.3	21.2	2	No
1911	1913	214.9	71.6	3	Yes	1733	1733	26.7	26.7	1	No
1649	1651	213.4	71.1	3	No	1687	1687	21.5	21.5	1	No
1846	1846	208.6	208.6	1	No	1735	1735	21.1	21.1	1	No
1904	1905	206.6	103.3	2	No	1967	1967	19.2	19.2	1	Yes
1948	1949	200.9	100.5	2	No	1779	1779	13.0	13.0	1	No
1959	1961	198.1	66.0	3	Yes	1823	1823	12.3	12.3	1	No
1636	1638	190.1	63.4	3	No	1709	1709	4.2	4.2	1	No
1691	1691	189.5	189.5	1	No	1819	1819	2.2	2.2	1	No
1795	1795	183.9	183.9	1	No						

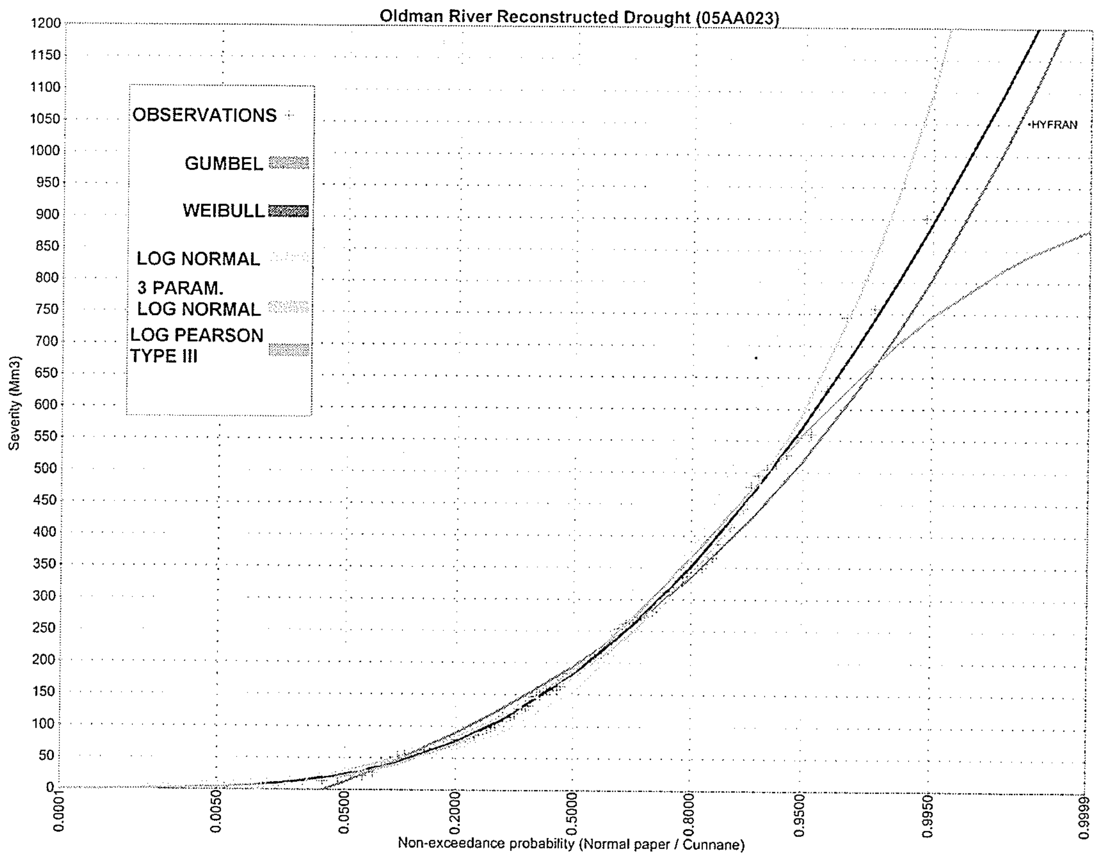


Figure 6.23: Frequency Distributions for Reconstructed Droughts

Table 6.12: Comparison of Drought Frequency Analysis at Various Return Periods for 05AA023

Return Period	Gauged Data	Reconstructed Data
(Years)	Severity (Mm^3)	
10	404	330
20	582	452
50	812	606
100	981	717
1000	1504	1038
10000	1950	1288

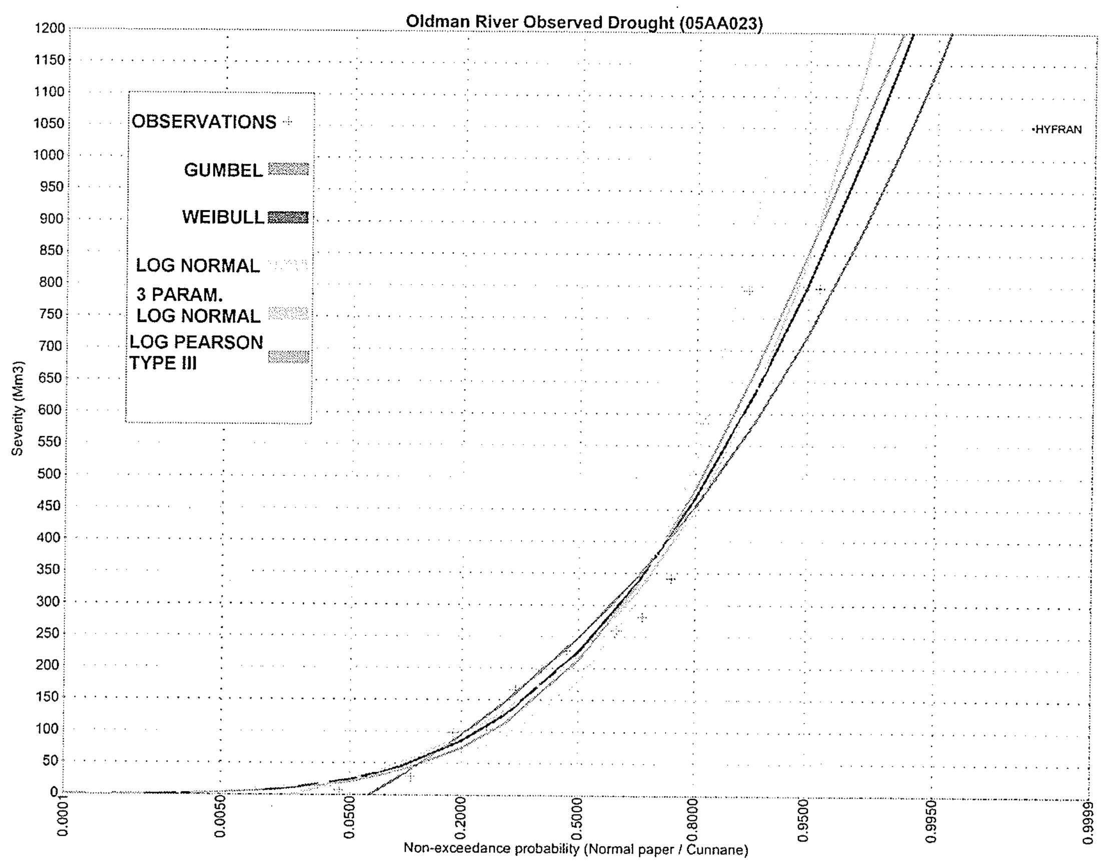


Figure 6.24: Frequency Distributions for Gauged Droughts

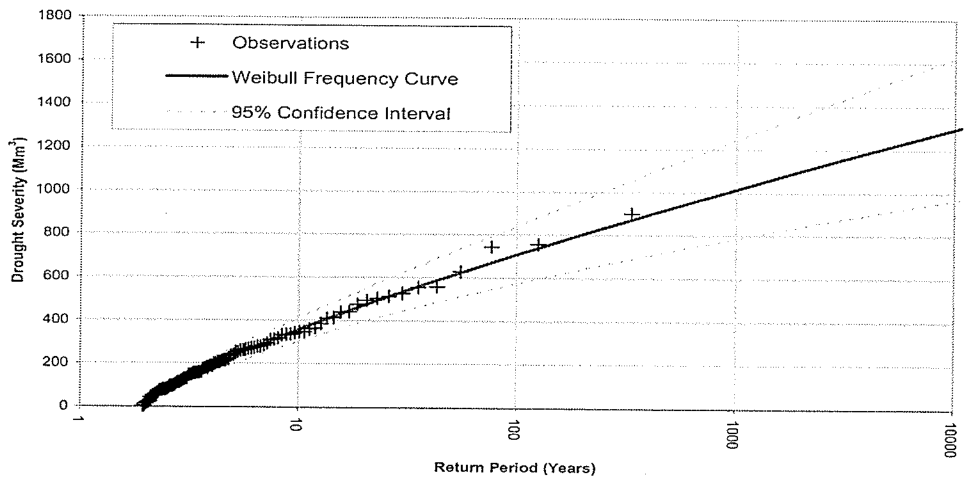


Figure 6.25: Drought Frequency Analysis Based on Reconstructed Data for 05AA023

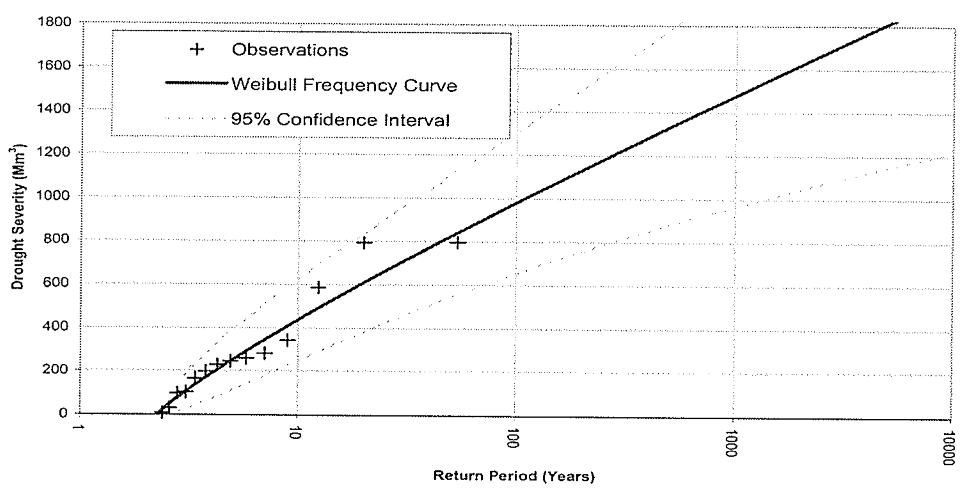


Figure 6.26: Drought Frequency Analysis Based on Gauged Data for 05AA023

have been a more ideal way of relating the two gauging stations. Unfortunately no unregulated overlapping data set was available in the immediate area to accomplish this. In addition the period between 1950 and 1965 had a large number of high flow years which may have biased the model. Even though the 'early' model independent verification and full model verification showed good results the poor performance of the 'late' model verification means that these results must be used with caution. The use of anecdotal data, if available, to further verify the results of the reconstruction may be useful in this case.

The data analysis presented has shown a seven times increase in the number of multiyear droughts and low flow events and a six times increase in the number of low flow years . It also illustrates that even a moderate reduction in variance makes the data unsuitable for accurate quantitative analysis.

6.4 Castle River Near Beaver Mine Station

6.4.1 Background Information

The Castle River is a tributary of the Oldman River Basin that forms the western most part of the South Saskatchewan River Basin. This river is shown in Figure 6.27.

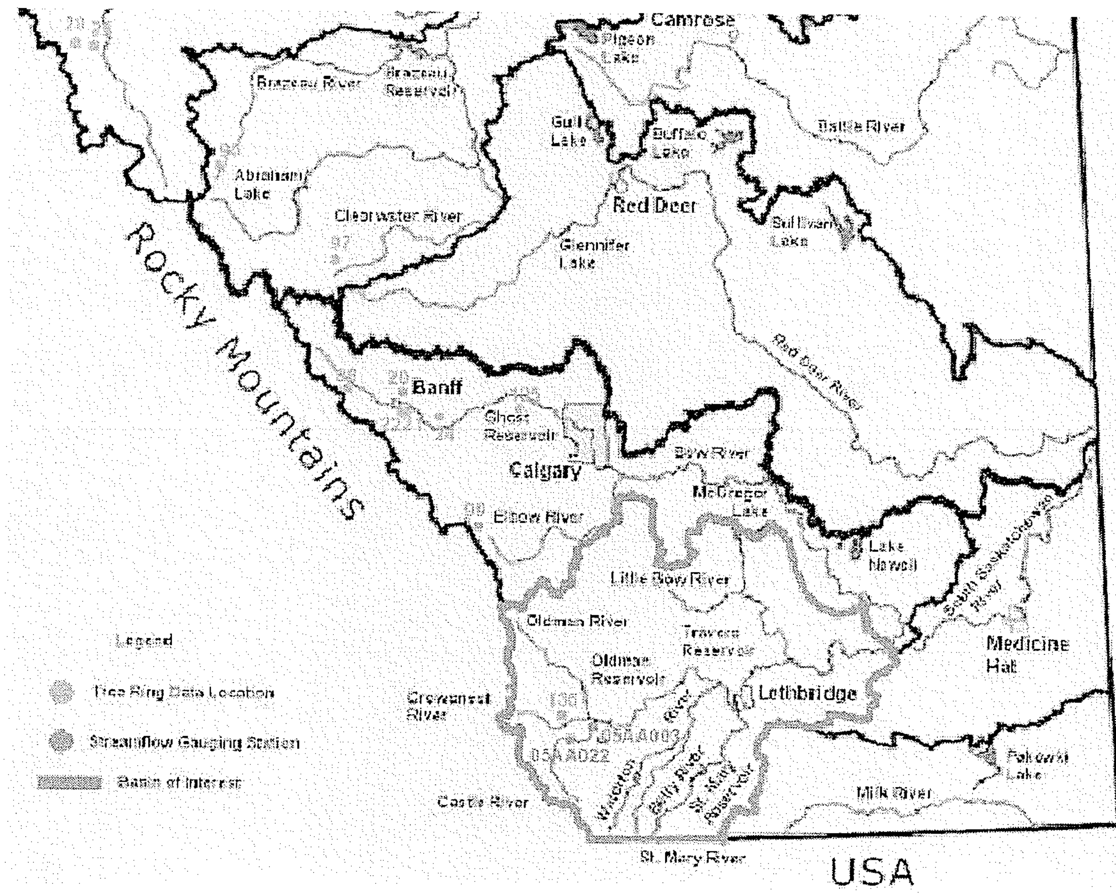


Figure 6.27: Map of Castle River Gauge Locations

The gauge reconstructed is the Water Survey of Canada gauge 05AA022 described as Castle River Near Beaver Mine Station.

6.4.2 Available Data

Streamflow data for gauge 05AA022 were obtained from the Water Survey of Canada Hydat CD. Data at this location are available between 1945 and 1991. This does not provide enough overlap with local tree ring data sets for an acceptable regression analysis and verification. It was, therefore, decided to use a gauge downstream on the same river to supplement this record.

The gauge 05AA003, Castle River Near Cowley, is approximately 20km downstream of 05AA022. Its record extends between 1911 and 1930. There is no overlapping period between the two gauged records. No major tributaries enter the river between the two gauges, therefore a straight pro-ration of monthly data based on the ratio of drainage areas was deemed appropriate. To this end the monthly average data in 05AA003 was multiplied by a factor of $\frac{823km^2}{1120km^2}$, the drainage area of 05AA022 over the drainage area of 05AA003. These data were then combined with 05AA022 to form a complete record at Beaver Mine Station from 1911–1930 and 1945–1990 with a potential minimum overlap of 40 years with tree ring data in the area.

In addition several months of missing data exist in 1949 and 1950. The months of December 1949, and January through March 1950 were filled in by fitting a regression model for each month against the months for which complete records were available (May–Nov). These regression models were calibrated with the complete data between 1945-1991. R^2 for these models was generally low (between 17%-48%) but since these five months collectively only account for less than 14% of the total yearly flow and only one years worth of data is being filled in some additional variance would probably not harm the final reconstructions.

The monthly data were converted to 12 series of yearly streamflow volumes based on starting months between January and December. These were then preprocessed by techniques discussed in Section 3.2. None of the flow data had to be normalized using Box-Cox transformation and no trends or autocorrelation were found.

6.4.3 Predetermination of Predictors

The number of tree ring data sets was narrowed from a possible 140 candidates to 51 based on the criterion suggested by Cook (1995). The tree ring data sets to be entered into the best subsets analysis were further reduced 'a priori' based on judgement and the five criteria discussed in Section 3.3.

Only one of the data sets, CANA136 Crowsnest Pass, is within the gauging station's sub-basin. Of the remaining sets, seven are in the Bow River Basin which is also part of the South Saskatchewan River Basin and adjacent to the Castle River Basin. Four are in sub-basins of the Mackenzie River Basin in areas of similar geography as the Castle River Basin. The other 39 potential data sets are located on the West side of the Rocky Mountains and, therefore, could not have responded to the same weather patterns as those located East of the Rocky Mountains due to orographic effects. All 12 tree ring data sets identified were investigated further.

All tree ring data had more than 25 years of overlap when compared to the periods of continuous streamflow records from 1911–1929 and 1945–1995. The shortest overlap was 40 years.

All of the tree ring data sets could be made normal by Box-Cox transformation. They all displayed little or no trend, shifts in the mean or periodicity. Significant autocorrelation was removed with low order ARMA modelling (AR1 or AR1 MA1).

A correlation analysis between tree ring series and **monthly** streamflow yielded ten tree ring series with significant correlations. These are presented in Table 6.13. Eight of these are within the Saskatchewan River Basin so they could have responded to the same flow characteristics present in the gauged record. Two are within the MacKenzie River Basin so they could have responded to weather patterns common to both basins. Plots of the correlation analysis against monthly flows can be found in Appendix E.

A correlation analysis between tree ring series and **annual** streamflow yielded eight tree ring series with significant correlations. These are presented in Table 6.14. All of these tree ring records were identified in the monthly correlation analysis except CANA020. This record had correlations close to significant but not at the 95% confidence level. The correlation

Table 6.13: Tree Ring Data Significantly Correlated With Monthly Record of 05AA022

Identifier	Description	Minor Basin	Major Basin
CANA021	Tunnel Mountain, Banff, Alberta	Bow River	Saskatchewan River
CANA022	Exshaw, Tunnel and Banff, Alberta	Bow River	Saskatchewan River
CANA024	Exshaw, Alberta	Bow River	Saskatchewan River
CANA026	Pyramid Lake and Patricia Lake, Alberta	Athabasca River	MacKenzie River
CANA028	Pyramid Lake, Alberta	Athabasca River	MacKenzie River
CANA096	Sunwapta Pass, Alberta North	Brazeau River	Saskatchewan River
CANA097	Peyto Lake, Alberta North	Clearwater River	Saskatchewan River
CANA099	Sarrail Glacier, Alberta	Bow River	Saskatchewan River
CANA135	Towers Ridge, Alberta	Bow River	Saskatchewan River
CANA136	Crowsnest Pass, Alberta	Oldman River	Saskatchewan River

Table 6.14: Tree Ring Data Significantly Correlated With Annual Record of 05AA022

Identifier	Description	Minor Basin	Major Basin
CANA020	Powerhouse, Alberta	Bow River	Saskatchewan River
CANA021	Tunnel Mountain, Banff, Alberta	Bow River	Saskatchewan River
CANA022	Exshaw, Tunnel and Banff, Alberta	Bow River	Saskatchewan River
CANA024	Exshaw, Alberta	Bow River	Saskatchewan River
CANA026	Pyramid Lake and Patricia Lake, Alberta	Athabasca River	MacKenzie River
CANA028	Pyramid Lake, Alberta	Athabasca River	MacKenzie River
CANA096	Sunwapta Pass, Alberta North	Brazeau River	Saskatchewan River
CANA136	Crowsnest Pass, Alberta	Oldman River	Saskatchewan River

against annual streamflow shows that it is indeed significantly correlated with streamflow from this station. Plots of the correlation analysis against annualized streamflow can be found in Appendix E.

The correlation analysis against monthly streamflow data yielded ten possible data sets to be investigated for model building. The correlation analysis against annual streamflow reaffirms that seven of these data sets contain significant common information with the gauged streamflow data and also identified one additional potentially significant predictor. The reduction, based on the correlation analysis with monthly streamflow, to 11 data sets simplifies the regression procedure.

6.4.4 Principal Components Analysis

The overlapping period for all 11 tree ring series was found to be from 1639 to 1965. The tree ring sets were lagged forward and backward one year to account for growth and storage effects forming 33 possible predictors for the reconstruction. Nine Eigenvectors had eigenvalues in excess of the Kaiser–Guttman eigenvalue–1 criterion. These nine components represent 79.9% of the total variance contained in the 33 predictors with the largest single vector representing 15.7% and the smallest 4.9%. These were retained for use in the best subsets model building exercise.

6.4.5 Best Subsets Analysis

The period of overlap between the tree ring series and the streamflow series from 1911–1929 and 1945–1965 was used in this analysis. Each monthly annualization was regressed against all possible combinations of the nine orthogonalized tree ring vectors. The best model for each number of predictors was then separated out based on $R^2_{adjusted}$ and plotted in Figure 6.28. The model that produced the highest $R^2_{adjusted}$ with the least number of predictors was a regression using the 1st to 5th eigenvectors on the December Annualization. This regression produced an R^2 of 50.1% and $R^2_{adjusted}$ of 42.7%.

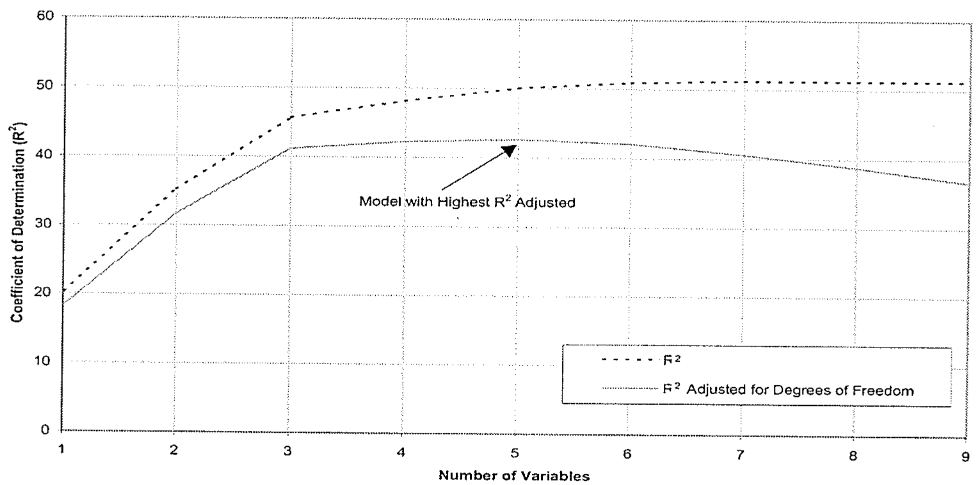


Figure 6.28: Results of Best Subset Analysis of 05AA022

6.4.6 Investigation of Outliers

It was found that one data point, 1927, had high Studentized Residuals and Dfits and that removal of this point significantly improved the regression model. This data point was therefore removed from the regression data set.

6.4.7 Model Building and Verification

The final model was built using the regression equation derived from the best subsets analysis. The data from 1911–1929 was first used to build a regression equation (the 'early' model) and this was tested against the independent data from 1945–1965. In turn the data from 1945–1965 was used to build a regression model (the 'late' model) and verified with the independent data from 1911-1929.

The final regression equation takes the following form:

$$streamflow = b_0 + b_1PC_1 + b_2PC_2 + \ldots + b_mPC_m \tag{6.3}$$

Where:

$streamflow$ = Estimate of annualized streamflow

m = Number of predictor variables

$PC_1 \ldots PC_m$ = Principal components of tree ring data

$b_1 \ldots b_m$ = Regression coefficients corresponding to each principal component

b_0 = Intercept coefficient which scales the regression equation to the mean of the predictand

The regression coefficients can be observed in Table 6.15 as well as the R^2, R^2_{adj} and gauged versus modelled means and standard deviations. Some qualitative observations from the model building portion of this exercise are as follows. The highest R^2_{adj} was 48.1% for the regression model built on the 'early' data and the lowest was 36.7% for the regression model on the 'late' data. All three of the R^2_{adj} for the models were within reasonable limits for past streamflow studies in the literature.

The regression coefficients B_0, B_1, B_2 and B_5 are of the same order and sign for all models. B_3 and B_4 however differ in sign for each model. This raises some concerns about model stability which will be born out in the verification.

Table 6.15: Calibration Statistics

Parameter	**Early Calibration (1911–1929)**	**Late Calibration (1945–1965)**	**Full Calibration (1911–1929 and 1945–1965)**
R^2	0.627	0.525	0.549
R^2_{adj}	0.472	0.367	0.481
B_0	454.26	506.67	484.06
B_1	-25.948	-29.98	-26.760
B_2	27.917	29.21	33.548
B_3	1.12	-36.29	-30.562
B_4	2.75	-3.62	4.608
B_5	28.25	4.67	18.60
Gauged Mean	432.724	543.217	492.220
Reconstructed Mean	432.726	543.222	492.219
Gauged Standard Deviation	95.037	115.367	118.994
Reconstructed Standard Deviation	75.274	83.595	88.189

Verification statistics for each of the split models and the full model are shown in Table 6.16.

The sign test shows that 'full' and 'early' models fail at 90% when applied to the calibration

Table 6.16: Verification Statistics

Parameter	**Calibration Period (1911–1929) Verification Period (1945–1965)**	**Calibration Period (1945–1965) Verification Period (1911–1929)**	**Full Calibration (1911–1929 and 1945–1965)**
	Sign Test		
Right	12	14	24
Wrong	9	4	15
Status (Confidence)	Fail(90%)	Pass(95%)	Fail(90%)
	Product Means Test		
tvalue	0.40	1.70	4.10
minimum value	1.72	1.73	2.71
Status (Confidence)	Fail(90%)	Fail(90%)	Pass(99%)
	Product Moment Correlation Coefficient Test		
tvalue	1.32	3.01	6.71
minimum value	1.72	2.88	2.71
Status (Confidence)	Fail(90%)	Pass(99%)	Pass(99%)
	Reduction of Error Test		
value	0.287	0.656	0.853
minimum value	0	0	0
Status	Pass	Pass	Pass

data. The 'late' model passes at the 95% confidence interval.

The product means test for both the 'early' and 'late' models failed at the 90% confidence level while the 'full' model passed at greater than 99% on the calibration data. Again this test is a very powerful validation tool when a positive result is obtained but it is so sensitive to individual deviations that a negative result does not invalidate the results.

The 'late' and 'full' models passed the product moment correlation coefficient test at 99% confidence while the 'early' model failed at 90%. Again this indicates that the early model does not reproduce variance well while the late model does.

The reduction of error statistic was passed in all cases. However the results for the 'early' data are lower than the others. The positive results are a strong indicator that both the split models have merit.

Figures 6.29 and 6.30 show how the 'early' regression model performs on independent data. From Figure 6.29 it can be seen qualitatively that the calibrated model does a good job of reconstructing both low and high flow events although during the verification period the low flow events are somewhat better represented. The scatter diagram in Figure 6.30 shows relatively tight correlation between the gauged and modelled data for the calibration period. During the verification period for flows above 550 Mm^3 the early model consistently underpredicts the independent data. The variance appears to be a little larger for the verification period than during the calibration period especially for higher flow years. These observations indicate that although the early model does not represent high flows very well it does an adequate job of representing low flows even during independent verification.

Figures 6.31 and 6.32 show how the 'late' regression model performs on independent data. From Figure 6.31 it can be seen qualitatively that the calibrated model does a good job of reconstructing both low and high flow events during the calibration period but does not reproduce the verification data nearly as well. The scatter diagram in Figure 6.32 confirms this by showing relatively tight correlation between the gauged and modelled data for the calibration period but a skewed spread for the verification period. Although the variance only appears slightly greater for the verification period the data are skewed such that the model consistently overpredicts flows over the entire range. These observations indicate that although the 'late' model represents the calibration data well it does not do a very good job of representing independent data.

Figures 6.33 and Figure 6.34 show the performance of the final reconstruction model using all available data with respect to the gauged data. Figure 6.33 shows a good correspondence between gauged low flows and modelled low flows as well as a poorer representation of above average flows. Figure 6.34 confirms this by showing relatively tight correlation between the gauged and modelled data.

The split sample verification shows that even though the 'late' model passed more of the verification statistics the 'early' model qualitatively seems to perform slightly better on independent data. The sign and order of the regression coefficients correspond to those in the full model except for B_3 in the 'early' model and B_4 in the 'late' model. The 'early' model did not pass most of the verification statistics except the most important reduction of error statistic. This gives confidence in the performance of the full model and indicates a satisfactory result for the reconstuction.

The poor performance of the 'early' model in the verification statistics and the 'late' model in the qualitative verification could be due to the choice of how the data was split. The period between 1911 and 1929 had less than 25% high flow years. A calibrated model created from this data could not be expected to accurately reproduce a period such as 1945-1965 which has roughly 60% high flow years. This shows up in verification statistics because they do not discriminate between high and low data points. In the qualitative analysis, however, we are interested in the low flow data much more than the high flow. A model built almost exclusively of low flow data would reproduce similar events very well but would not reproduce the high flows as well.

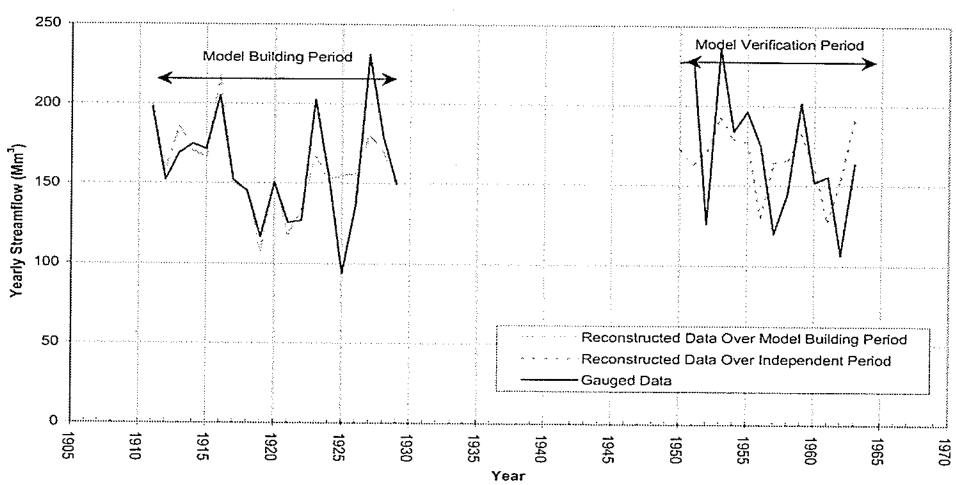


Figure 6.29: Time Series Plot Using 1911-1929 as the Regression Period and 1945-1965 as the Verification Period

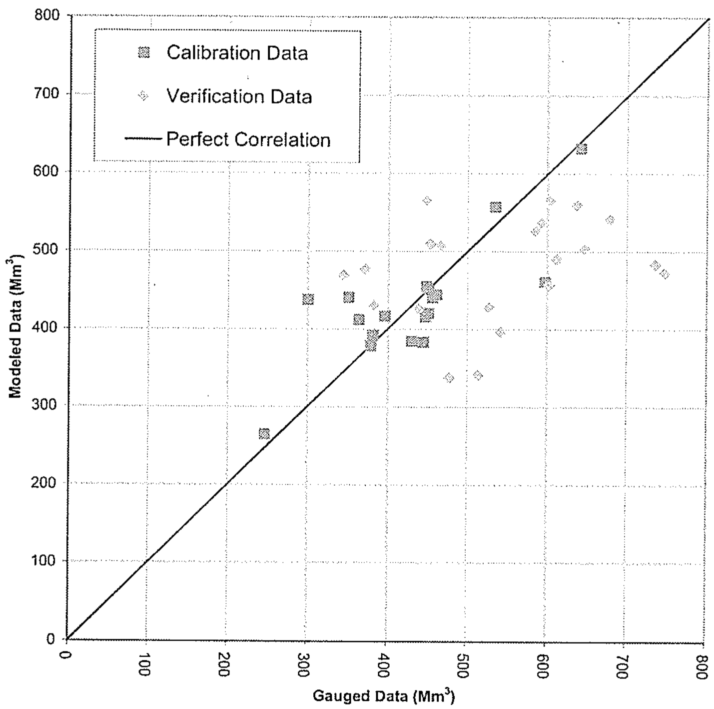


Figure 6.30: Scatter Diagram of Verification Data for 1911-1929 Regression Period

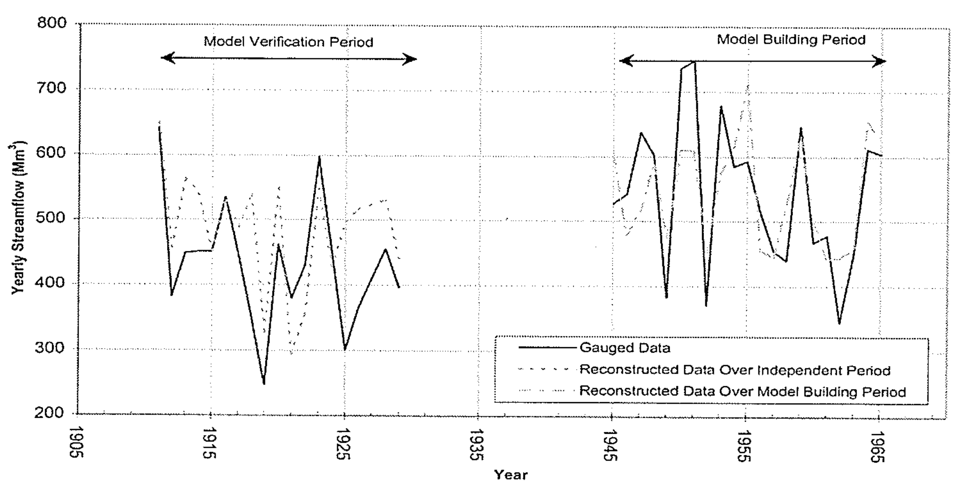


Figure 6.31: Time Series Plot Using 1945–1965 as the Regression Period and 1911–1929 as the Verification Period

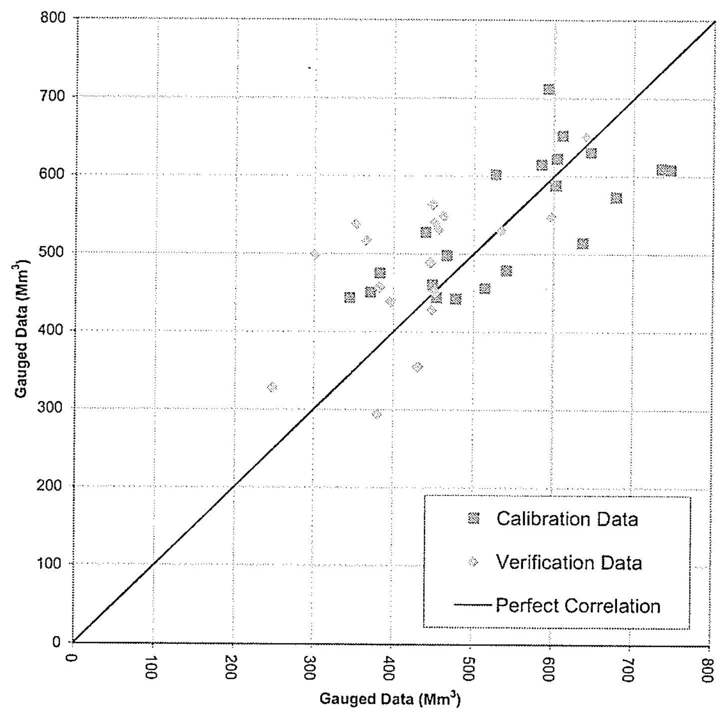


Figure 6.32: Scatter Diagram of Verification Data for 1945-1965 Regression Period

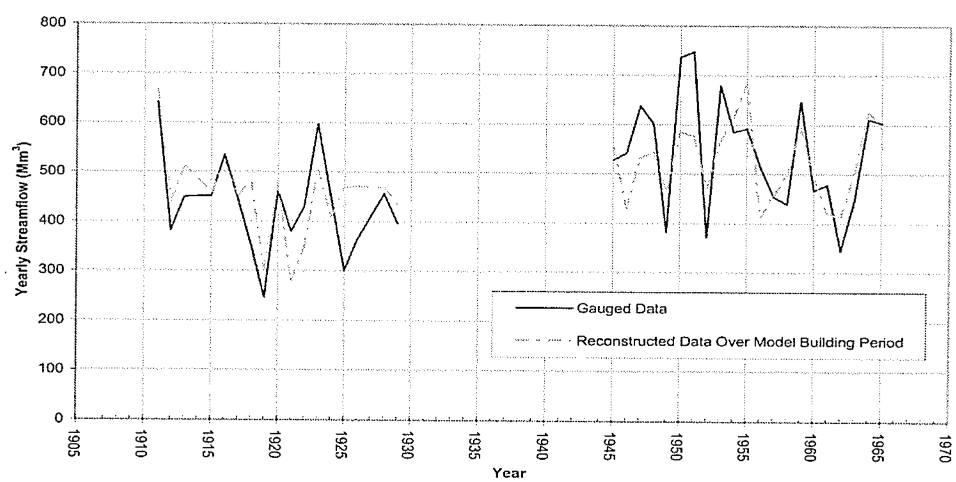


Figure 6.33: Time Series Plot Using Entire Record as the Regression Period with No Verification Period

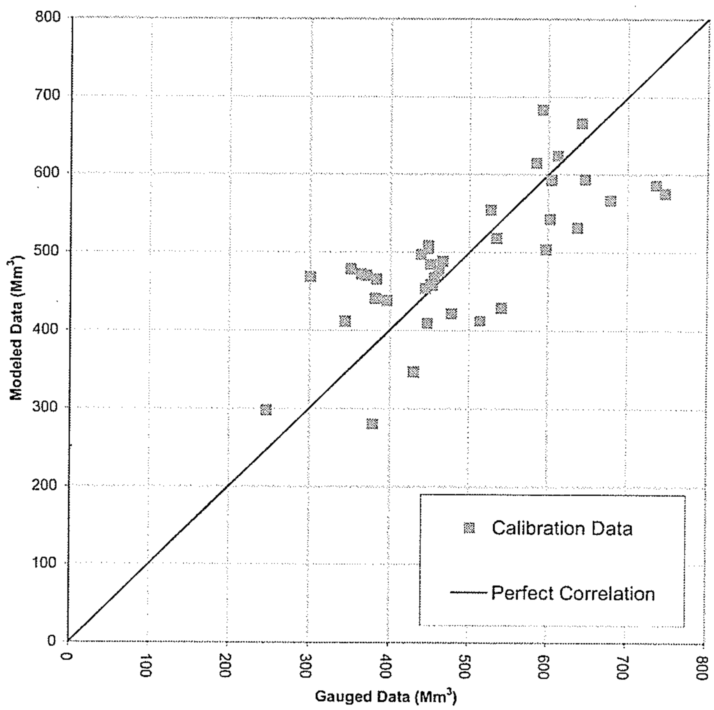


Figure 6.34: Scatter Diagram of Calibration Data for Full Regression Period

6.4.8 Final Reconstruction

Figure 6.35 shows the final reconstruction with gauged data incorporated. Tables of the gauged

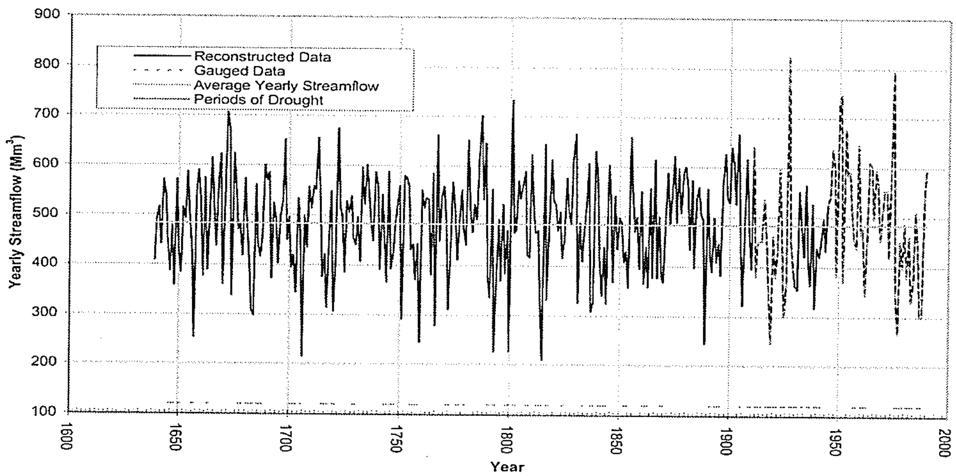


Figure 6.35: Time Series of Reconstructed 05AA022 Flow Record

and reconstructed data may be found in Appendix H.

Again, the reduction in variance in the reconstruction versus the gauge record is not as pronounced as in the first case study. The standard deviation of the reconstruction in this case is 75% of that of the gauged data alone. Again this reduction in variance still makes accurate quantitative analysis of drought unreliable. The potential for applying this data in a verification role or qualitative analysis still exists, however.

6.4.9 Drought Analysis

An analysis was performed on the reconstructed data based on the procedure described in Section 2.3. A truncation level of the mean annual runoff volume was used to separate low flow from high flow years. Distinct multiyear droughts were formed by grouping adjacent years of lower than average flow. These are presented in Table 6.17 sorted by severity. Several observations can be made about this new drought record.

- The reconstruction identified a record of 84 distinct drought events including 172 low flow years compared to a record of 13 distinct drought events including 32 low flow years in the recorded data.

- The highest severity drought identified was from the reconstructed record, five droughts were identified in the reconstructed record that were at least as severe as the three most severe droughts recorded.
- Seven droughts lasting four years or more were identified in the reconstructed record in addition to the two droughts identified in the gauged record

A frequency analysis was performed for illustrative purposes on the gauged and reconstructed data using methods described in Section 2.3. A probability distribution was chosen by qualitatively comparing the different distributions on a single plot. In the case of the reconstructed data shown in Figure 6.36 the Weibull distribution fits the data the best out of the five distributions. With the gauged data shown in Figure 6.37 the Weibull distribution fits the data the best out of the five distributions.

The best fit line was then plotted with 95% confidence limits as well as the data. The plots for the gauged and reconstructed data can be observed in Figures 6.38 and 6.39. These plots show that the drought analysis using the reconstructed data underpredicts the severity of droughts at all return periods. Table 6.18 illustrates this quantitatively. This is due to the reduction of variance in the reconstruction and illustrates that even with only a moderate reduction in variance the reconstructed record is still unsuitable for frequency analysis.

6.4.10 Important Observations

This case study showed how the choice of a split sample can affect the verification. The 'early' model was built almost exclusively of data from drought years. This showed up as poor verification statistics even though the 'early' model clearly reproduced low flow events better than the 'late' split. The verification statistics were largely positive for both models but in hindsight a different type of split sampling that balanced the high and low flow years may have yielded more balanced verification results.

The data analysis presented has shown a thirteen times increase in the number of multiyear droughts and low flow events and a five times increase in the number of low flow years . It also

Table 6.17: Historical Drought Periods for 05AA022

Starting Year	Ending Year	Severity (Mm^3)	Magnitude (Mm^3)	Duration (years)	Gauged Data (Y/N)	Starting Year	Ending Year	Severity (Mm^3)	Magnitude (Mm^3)	Duration (years)	Gauged Data (Y/N)
1917	1922	586.35	97.72	6	Yes	1670	1670	123.22	123.22	1	No
1755	1759	485.15	97.03	5	No	1847	1847	115.02	115.02	1	No
1812	1815	474.94	118.74	4	No	1796	1796	114.27	114.27	1	No
1976	1979	474.54	118.63	4	Yes	1952	1952	112.82	112.82	1	Yes
1681	1684	445.92	111.48	4	No	1692	1692	110.85	110.85	1	No
1986	1988	412.75	137.58	3	Yes	1867	1867	106.68	106.68	1	No
1841	1844	411.05	102.76	4	No	1661	1661	106.54	106.54	1	No
1981	1984	394.88	98.72	4	Yes	1865	1865	106.46	106.46	1	No
1715	1718	370.11	92.53	4	No	1764	1764	102.27	102.27	1	No
1798	1800	366.63	122.21	3	No	1949	1949	100.94	100.94	1	Yes
1928	1931	365.48	91.37	4	Yes	1824	1825	100.02	50.01	2	No
1793	1794	364.12	182.06	2	No	1909	1910	99.91	49.96	2	No
1924	1926	338.08	112.69	3	Yes	1725	1725	98.13	98.13	1	No
1837	1838	331.80	165.90	2	No	1663	1664	95.27	47.63	2	No
1939	1942	311.19	77.80	4	No	1833	1834	92.52	46.26	2	No
1701	1703	291.44	97.15	3	No	1741	1741	92.03	92.03	1	No
1861	1863	288.61	96.20	3	No	1859	1859	88.63	88.63	1	No
1705	1706	288.26	144.13	2	No	1884	1884	86.14	86.14	1	No
1656	1657	285.97	142.98	2	No	1695	1695	80.65	80.65	1	No
1720	1721	273.63	136.81	2	No	1677	1679	78.91	26.30	3	No
1645	1648	263.15	65.79	4	No	1732	1732	76.56	76.56	1	No
1790	1791	260.27	130.14	2	No	1639	1639	75.84	75.84	1	No
1852	1854	254.73	84.91	3	No	1957	1958	74.16	37.08	2	Yes
1889	1889	238.27	238.27	1	No	1776	1777	73.63	36.81	2	No
1771	1773	219.50	73.17	3	No	1729	1730	71.14	35.57	2	No
1869	1870	219.12	109.56	2	No	1933	1934	67.04	33.52	2	No
1894	1896	211.66	70.55	3	No	1973	1973	64.87	64.87	1	Yes
1766	1766	204.79	204.79	1	No	1779	1780	52.72	26.36	2	No
1912	1915	202.05	50.51	4	Yes	1944	1944	50.85	50.85	1	No
1936	1937	200.12	100.06	2	No	1882	1882	49.28	49.28	1	No
1905	1906	197.99	99.00	2	No	1849	1849	48.90	48.90	1	No
1960	1963	197.58	49.40	4	Yes	1708	1708	48.23	48.23	1	No
1751	1751	191.70	191.70	1	No	1667	1667	45.83	45.83	1	No
1817	1818	158.53	79.27	2	No	1642	1642	41.72	41.72	1	No
1891	1892	158.05	79.03	2	No	1738	1738	34.95	34.95	1	No
1831	1831	157.67	157.67	1	No	1768	1768	34.00	34.00	1	No
1746	1748	150.89	50.30	3	No	1699	1699	33.32	33.32	1	No
1650	1651	149.17	74.59	2	No	1969	1969	30.98	30.98	1	Yes
1674	1674	145.31	145.31	1	No	1802	1803	29.83	14.91	2	No
1743	1744	142.54	71.27	2	No	1783	1783	16.76	16.76	1	No
1686	1688	141.95	47.32	3	No	1857	1857	12.85	12.85	1	No
1808	1809	128.99	64.50	2	No	1822	1822	12.26	12.26	1	No

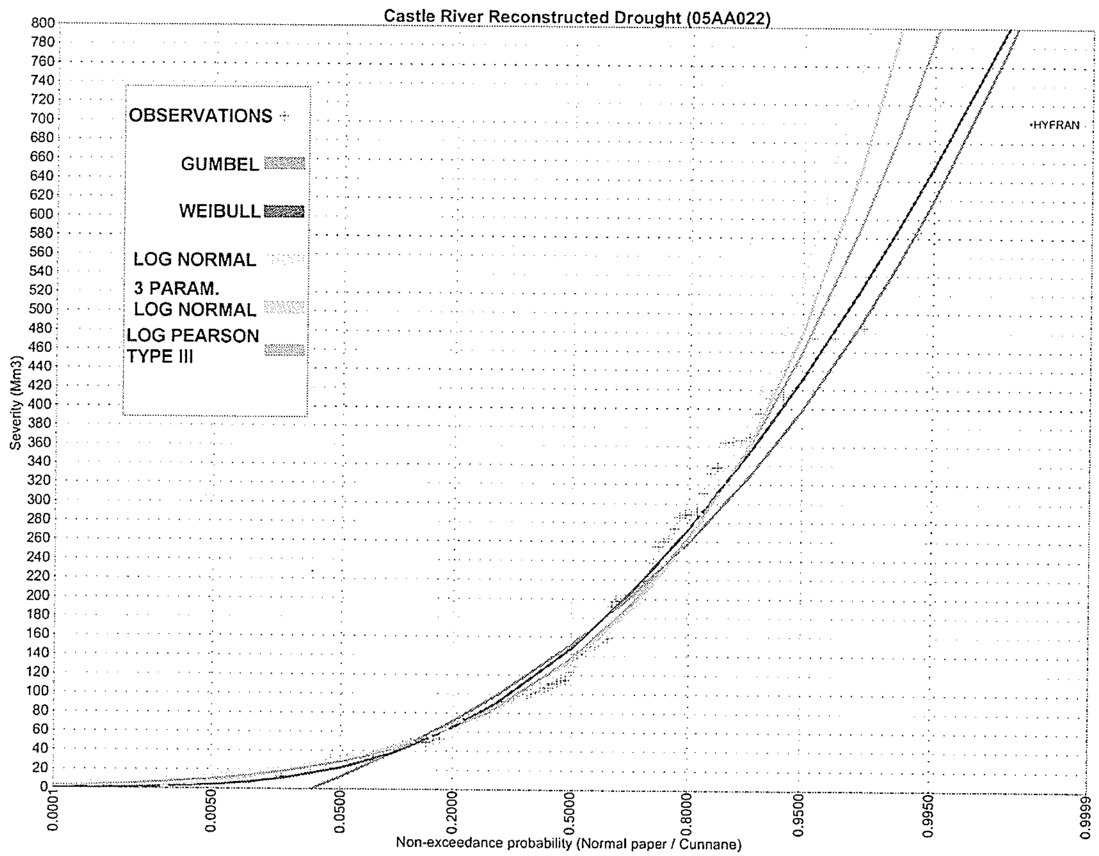


Figure 6.36: Frequency Distributions for Reconstructed Droughts

Table 6.18: Comparison of Drought Frequency Analysis at Various Return Periods for 05AA022

Return Period	Gauged Data	Reconstructed Data
(Years)	Severity (Mm^3)	
10	500	356
20	603	431
50	729	525
100	819	591
1000	1092	795
10000	1339	981

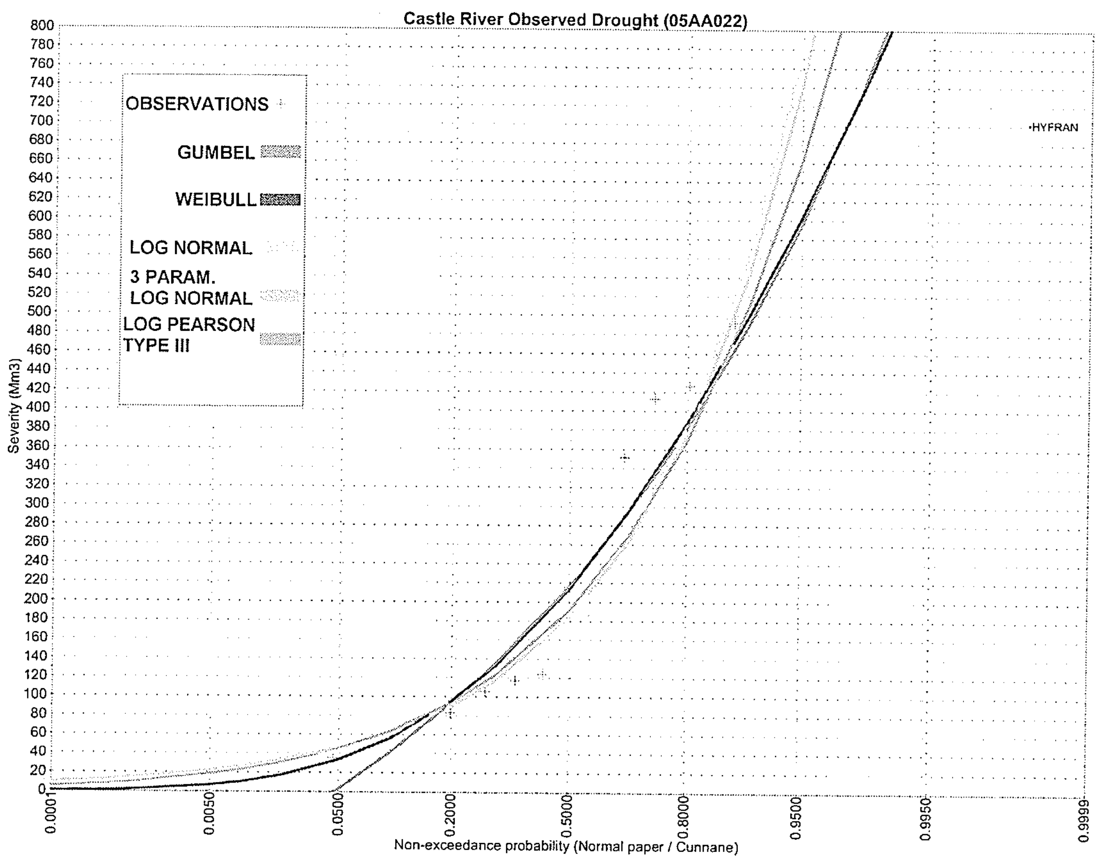


Figure 6.37: Frequency Distributions for Gauged Droughts

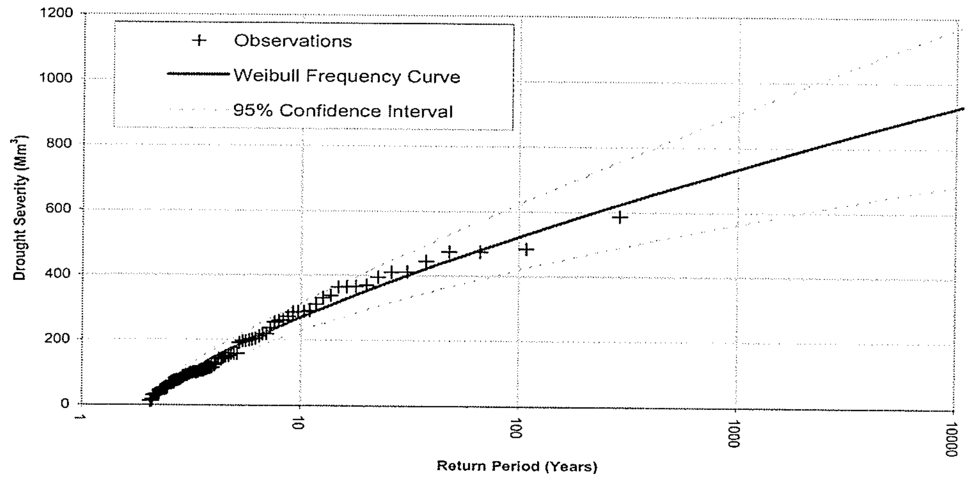


Figure 6.38: Drought Frequency Analysis Based on Reconstructed Data for 05AA022

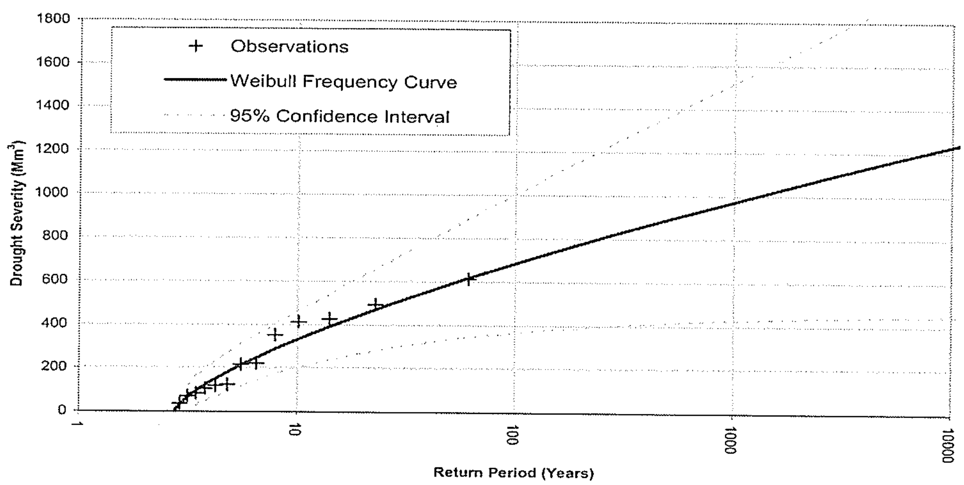


Figure 6.39: Drought Frequency Analysis Based on Gauged Data for 05AA022

illustrates that even a moderate reduction in variance makes the data unsuitable for accurate quantitative analysis.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

The goal of this thesis was to:

1. Explore approaches for using tree ring information to characterize drought events.
2. Determine the availability of tree ring data within and near the Churchill–Nelson River Basin.
3. Employ the approaches in a series of case studies to demonstrate the feasibility of using tree ring data to reconstruct drought.

A literature review of dendroclimatology and its use in reconstructing drought was completed. This review revealed that tree ring data has been used to reconstruct many climatic variables. Further, tree ring data are ideally suited to reconstruct streamflow because both parameters integrate the effects of temperature, precipitation and evapotranspiration. It was also noted that most research to date has been done on reconstructing precipitation and temperature records. Research into reconstruction of streamflow has been very limited. The projects that have been done have mostly concentrated in more arid areas of the United States such as Arizona and New Mexico.

A review of the available streamflow data explains some of the reason for the lack of research into streamflow. It was found that in Canada, the unregulated streamflow records are very short and discontinuous. Very few records were found in the prairie provinces that were long enough to facilitate a statistical reconstruction. All of the streamflow records that were long enough were located in Alberta. Of the five most promising gauging stations found, three of them required the combination of two separate gauging station locations to create a sample large enough for reconstruction. This would explain why temperature and precipitation, whose records tend to be much longer in duration, are more often reconstructed.

A review of available tree ring data revealed that although there are many tree ring data sets available in the United States there are comparably few available in Canada. Tree ring data available in Canada are mostly concentrated in Western Alberta and Quebec. Although this is the case now, research is in progress that should expand this data network appreciably. The work of Eric Nielsen and Scott St. George at Manitoba Energy and Mines as well as Jacques Tardif at the University of Winnipeg should help expand the tree ring network in Manitoba.

Three case studies were completed, one in the MacKenzie River Basin and two in the South Saskatchewan River Basin. Two of the case studies verified very well using split sample techniques, one was questionable. The reconstructions extended streamflow records from 59 to 190 years, from 60 to 420 years and from 65 to 352 years.

The first case study reconstructing the Water Survey of Canada gauge 07BE001, Athabasca River at Athabasca, extended a record that originally spanned between 1914–1929 and 1952–1995 to a record that spans from 1805–1995. The drought record for this set was extended from 15 distinct events to 49 distinct events. An attempt was made to estimate low frequency droughts using standard frequency analysis. It was determined that the reduced variance of the reconstruction model makes the reconstructed data unsuitable for quantitative frequency analysis.

The second case study reconstructing the Water Survey of Canada gauge 05AA023, Oldman River Near Waldron's Corner, extended a record that originally spanned between 1911–1930 and 1950–1995 to a record that spans from 1571–1995. Unfortunately there is some question as to the validity of the reconstruction due to a poor verification. It is suggested that further verification possibly using anecdotal data be done before this reconstruction can be used. The drought record for this set was extended from 14 distinct events to 103 distinct events. Again the reconstruction proved unusable for frequency analysis due to reduced variance compared to the gauged record.

The third case study reconstructing the Water Survey of Canada gauge 05AA022, Castle River Near Beaver Mine Station, extended a record that originally spanned between 1911–1929 and 1945-1990 to a record that spans from 1639–1990. The verification of this data set was successful but could have been improved with a split sampling procedure that equalized the number of high and low flow years in each sample. The drought record for this set was extended from 13 distinct events to 84 distinct events. Also, with this model reconstruction proved unusable for frequency analysis.

The results of a comparison between extreme droughts estimated from the gauged data and the reconstructed data showed a decrease in drought severity at all return periods. This is contrary to what accepted theory would predict. It was found that this is due mainly to the reconstructed model being unable to reproduce the gauged variance. The magnitude of streamflow records are smoothed down as they are filtered through the tree ring data. Although the reconstructed drought magnitudes are unreliable, the reconstructed records could potentially be used in conjunction with other techniques to identify times of severe drought and for verification purposes.

7.2 Recommendations

This thesis has shown that using existing tree ring data to reconstruct streamflow is one viable method of dealing with the problem of short records of drought in Canada.

The research, however, was complicated by the lack of available tree ring data within the Prairie Provinces. In addition, the high quality tree ring data that was available was largely sampled in 1965. This meant that the last 35 years of streamflow record for active hydrometric stations could not be used. This made finding streamflow gauging stations with sufficient overlap with the tree ring data very difficult. Currently there is research being conducted in Manitoba by people such as Eric Nielsen, Scott St. George and Jacques Tardif that will extend the tree ring record in the prairie provinces. It is recommended that as these and other tree ring data sets become available in areas of interest that further research be conducted to help expand our understanding of drought in critical areas of the Prairies.

Difficulty was encountered using the standard split sample verification method where a point in time is used to divide the data set. This occurred because the statistical properties of the two split samples were not essentially equal. It is recommended that in further research of this type a method of splitting that is less affected by statistically anomalous periods in time be adopted. One such method would be the DUPLEX method of sample splitting based on equalized statistical properties presented by Snee (1977)

The results of a comparison between extreme droughts estimated from the gauged data and the reconstructed data showed a decrease in drought severity at all return periods. This was a result of the reconstruction models not being able to reproduce the amount of variance found in the gauged data. The data reconstructed in this study cannot be used in quantitative frequency analysis of extreme drought. Further study is required to determine if it is possible to produce a reconstruction with sufficient explained variance to perform quantitative frequency analysis

The most difficult complication of this study was that the drought reconstructions yielded smaller severity drought events at all return periods. This was shown to be due to the low explained variance of the reconstruction models. Previous tree ring studies identified in the literature have shown that it is possible to produce reconstructions with much higher explained variances in the 70% to 80% range. This however was only accomplished through specific sampling of tree ring data for streamflow reconstruction close to the gauge of interest. It is recommended that future work be conducted using site specific sampling. Statistical methods such as Maintenance of Variance Extension(MOVE) (Hirsch, 1982) should be investigated to help reproduce the missing variance.

References

American Meteorological Society, ed. 1995, March. *13th Annual Conference on Probability and Statistics.* All American Meteorological Society Conferences no. 13. American Meteorological Society, San Fransisco, CA: American Meteorological Society.

Booy, C. 1996. "Regression Analysis." Class notes for statistical methods, course 23.782, University of Manitoba, Winnipeg, Manitoba.

Box, G.E. and G.M. Jenkins. 1970. *Time Series Analysis Forecasting and Control.* San Francisco: Holden-Day.

Brinkmann, W.A.R. April, 1987. "Water Supplies to the Great Lakes–Reconstructed from Tree Rings." *Journal of Climate an Applied Meteorology* 26:530–538.

Brockway, C.G. and A.A. Bradley. September, 1995. "Errors in streamflow drought statistics reconstructed from tree ring data." *Water Resources Research* 31 (9): 2279–2293.

Burn, D.H. and W.J. DeWit. 1996. "Spatial Characterization of Drought Events Using Synthetic Hydrology." *Canadian Journal of Civil Engineering* 23:1231–1240.

Chair in statistical hydrology. 2002. *Software HYFRAN for hydrologic frequency analysis, version 1.1 beta 1.* Quebec: INRS-ETE (University of Quebec). <http://www.inrs-ete.uquebec.ca/activites/groupes/chaire_hydrol/chaire1.html>.

Cleaveland, M.K. and D.W. Stahle. June, 1989. "Tree ring analysis of surplus and deficit runoff in the White River, Arkansas." *Water Resources Research* 25 (6): 1391–1401.

———. 1995, March. "Tree ring reconstruction of climate extremes using discriminant function analysis." In American Meteorological Society 1995, 33–36.

Cook, E.R. 1995, March. "Statistical aspects of drought reconstruction over the coterminous United States from tree rings." In American Meteorological Society 1995, 28–32.

Cook, E.R and G.C. Jacoby. October, 1983. "Potomac River streamflow since 1730 as reconstructed from tree rings." *Journal of Climate and Applied Meteorology* 22 (10): 1659–1672.

Cook, E.R. and G.C. Jacoby Jr. October. 1977. "Tree-ring–drought relationships in the Hudson Valley, New York." *Science* 198:399–401.

Cook, E.R. and L.A. Kairiukstis. 1989. *Methods of Dendrochronology.* Dordrecht: Kluwer Academic Publishers.

Cook, E.R., K.R. Briffa, and P.D. Jones. 1994. "Spatial regression methods in Dedrolimatology: A reviw and comparison of two techniques." *International Journal of Climatology* 14:379–402.

Denis, L.G. and J.B. Challis. 1916. *Water Powers of Manitoba, Saskatchewan and Alberta.* Toronto: Commission of Conservation Canada.

Dracup, J.A., K.S. Lee, and E.G. Paulson Jr. April, 1980a. "On the definition of droughts." *Water Resources Research* 16 (2): 297–302.

——— April, 1980b. "On the statistical characteristics of drought events." *Water Resources Research* 16 (2): 289–296.

Draper, N.R. and H. Smith. 1981. *Applied Regression Analysis.* New York: Wiley.

Duvick, D.N. and T.J. Blasing. August, 1981. "A dendroclimatic reconstruction of annual precipitation amounts in Iowa since 1680." *Water Resources Research* 17 (4): 1183–1189.

Environment Canada. 2001. *Hydat Software Version 2.0.* Concord, Ontario: Greenland International Consulting Inc.

———. 2002, May 30. *Canadian Climate and Water Information.* <http://www.msc-smc.ec.gc.ca/climate/data_archives/water/index_e.cfm>.

Fritts, H.C. 1971. "Dendroclimatology and Dendroecology." *Quaternary Research* 1:419–449.

——— 1976. *Tree Rings and Climate.* New York: Academic Press.

Fritts, H.C., T.J. Blasing, B.P. Hayden, and J.E. Kutzbach. October, 1971. "Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate." *Journal of Applied Meteorology* 10 (5): 845–864.

Grissino-Mayer, H.D. 2002, May 29. *Science of Dendrochronology Web Pages.* <http://web.utc.edu/ grissino/itrdb.htm>.

Hair, J.E., R.E. Anderson, R.L. Tatham, and W.C. Black. 1998. *Multivariate Data Analysis.* New Jersey: Prentice Hall.

Hirsch, R. 1982. "A comparison of four record extension techniques." *Water Resources Research* 18 (4): 1081–1088.

Jarrett, R.D. 1991. "Paleohydrology and its Value in Analyzing Floods and Droughts." *Proceedings of the NHRI Symposium* 8:89–106.

Joseph, E.S. 1970. "Frequency of Design Drought." *Water Resources Research* 6 (4): 1199–1201 (August).

LaMarche, V.C. 1974. "Paleoclimatic Inferences from Long Tree-Ring Records." *Science* 183 (4129): 1043–1048 (March).

Maidment, D.R., ed. 1993. *Handbook of Hydrology.* New York: McGraw-Hill Inc.

Meko, D. and D.A. Graybill. 1995. "Tree-ring reconstruction of upper Gila river discharge." *Water Resources Bulletin* 31:605–616.

Meko, D.M., M.D. Therrel, C.H. Baison, and M.K. Hughes. 2001. "Sacramento River Flow Reconstructed to A.D. 869 from Tree RIngs." *Journal of the American Water Resources Association* 37 (4): 1029–1039.

Microsoft Corp. 1997. *Visual Basic Version 5.* Roselle, IL: Microsoft Corporation.

Minitab Inc. 1996, June. *Minitab Release 11 Reference Manual.* State College, PA: Minitab Inc.

National Oceanic and Atmospheric Administration. 2002, May 29. *ITRDB Display Software.* Boulder, Colorado: National Geophysical Data Centre. <http://www.ngdc.noaa.gov/paleo/treering.html>.

Neter, J., W. Wasserman, and M.H. Kutner. 1990. *Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs.* 3. Boston: Irwin.

Phipps, R.L. 1983. "Streamflow of the Occoquan River in Virginia as reconstructed from tree-ring series." *Water Resources Bulletin* 19 (3): 735–743 (October).

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1988. *Numerical Recipes in C, The Art of Scientific Computing.* Cambridge: Cambridge University Press.

Smith, L.P. and C.W. Stockton. 1981. "Reconstructed stream flow for the Salt and Verde Rivers from tree-ring data." *Water Resources Bulletin* 17 (6): 939–947 (December).

Snee, R.D. 1977. "Validation of regression models: Methods and examples." *Technometerics* 19 (4): 415–428 (November).

Stahle, D.W. and M.K Cleaveland. 1988. "Texas drought history reconstruced and analyzed from 1698 to 1980." *Journal of Climate* 1 (jan): 59–73.

Stockton, C.W. 1975. "Long Term Streamflow Records Reconstructed From Tree Rings." Papers of the laboratory of tree ring research 5, University of Arizona, Tucson, Arizona.

Woodhouse, C.A. 2001. "A Tree-Ring Reconstruction of Streamflow for the Colorado Front Range." *Journal of the American Water Resources Association* 37 (3): 561–569 (June).

Appendix A

Preprocessor Example Output

Streamflow Record Preprocessor Log

Input Filename: 07be001.txt
gsprefix for all files associated with this raw file: 07be001
Station = ATHABASCA RIVER AT ATHABASCA - STATION NO. 07BE001
Latitude = 54:43:20N
Drainage Area = 600 km2
Longitude = 113:17:10W
Flow Type = NATURAL FLOW

January
Number of Data Values = 43
Mean = 13765.828
Variance = 9435880
Standard Deviation = 3071.788
Standard Error = 468.443
Coefficient of Variation = 0.223
Skewness = 1.041
Kurtosis = 1.332
Autocorrellation = 0.132
Mean Sensitivity = 0.215

February
Number of Data Values = 43
Mean = 13767.972
Variance = 9552846
Standard Deviation = 3090.768
Standard Error = 471.338
Coefficient of Variation = 0.224
Skewness = 1.02
Kurtosis = 1.238
Autocorrellation = 0.127
Mean Sensitivity = 0.219

March
Number of Data Values = 43
Mean = 13769.059
Variance = 9614598
Standard Deviation = 3100.742
Standard Error = 472.859
Coefficient of Variation = 0.225
Skewness = 1.001
Kurtosis = 1.175
Autocorrellation = 0.127
Mean Sensitivity = 0.22

April
Number of Data Values = 43
Mean = 13770.259
Variance = 9776012
Standard Deviation = 3126.661
Standard Error = 476.811
Coefficient of Variation = 0.227
Skewness = 0.942
Kurtosis = 1.029
Autocorrellation = 0.12
Mean Sensitivity = 0.226

May
Number of Data Values = 43
Mean = 13750.487
Variance = 9690642
Standard Deviation = 3112.979
Standard Error = 474.725
Coefficient of Variation = 0.226
Skewness = 0.981
Kurtosis = 1.435
Autocorrellation = 0.156
Mean Sensitivity = 0.22

June
Number of Data Values = 43
Mean = 13745.504
Variance = 8829596
Standard Deviation = 2971.464
Standard Error = 453.144
Coefficient of Variation = 0.216
Skewness = 0.716
Kurtosis = 0.38
Autocorrellation = 0.25
Mean Sensitivity = 0.21

July
Number of Data Values = 43
Mean = 13732.001
Variance = 8368002.5
Standard Deviation = 2892.75
Standard Error = 441.14
Coefficient of Variation = 0.211
Skewness = 0.388
Kurtosis = -0.53
Autocorrellation = 0.285
Mean Sensitivity = 0.201

August
Number of Data Values = 43
Mean = 13724.153
Variance = 8697245
Standard Deviation = 2949.109
Standard Error = 449.735
Coefficient of Variation = 0.215
Skewness = 0.567
Kurtosis = -0.367
Autocorrellation = 0.221
Mean Sensitivity = 0.215

September
Number of Data Values = 43
Mean = 13749.504
Variance = 7926587
Standard Deviation = 2815.419
Standard Error = 429.347
Coefficient of Variation = 0.205
Skewness = 0.627
Kurtosis = 0.136
Autocorrellation = 0.258
Mean Sensitivity = 0.206

October
Number of Data Values = 43
Mean = 13749.564
Variance = 8311520
Standard Deviation = 2882.971
Standard Error = 439.649
Coefficient of Variation = 0.21
Skewness = 0.93
Kurtosis = 1.171
Autocorrellation = 0.222
Mean Sensitivity = 0.199

November
Number of Data Values = 43
Mean = 13747.011
Variance = 8770950
Standard Deviation = 2961.579
Standard Error = 451.637
Coefficient of Variation = 0.215
Skewness = 1.002
Kurtosis = 1.274
Autocorrellation = 0.179
Mean Sensitivity = 0.204

December
Number of Data Values = 43
Mean = 13746.046
Variance = 9114963
Standard Deviation = 3019.1
Standard Error = 460.408
Coefficient of Variation = 0.22
Skewness = 1.034
Kurtosis = 1.312
Autocorrellation = 0.156
Mean Sensitivity = 0.208

Annualized Streamflow in: 07be001.str

Checking for normality, January Annualization
Probability Plot Correlation Coefficient = 0.96518
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox transformation
Box-Cox Lamda = -0.944140624999998
Probability Plot Correlation Coefficient = 0.99534
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.059
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.534
Autocorrellation = 0.183
Mean Sensitivity = 0

Checking for normality, February Annualization
Probability Plot Correlation Coefficient = 0.9656
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox transformation
Box-Cox Lamda = -0.947656249999998
Probability Plot Correlation Coefficient = 0.99462
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.055
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.588
Autocorrellation = 0.175
Mean Sensitivity = 0

```
Checking for normality, March Annualization
Probability Plot Correlation Coefficient = 0.96613
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox
transformation
Box-Cox Lamda = -0.916796874999998
Probability Plot Correlation Coefficient = 0.99393
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.091
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.599
Autocorrellation = 0.172
Mean Sensitivity = 0


Checking for normality, April Annualization
Probability Plot Correlation Coefficient = 0.96991
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox
transformation
Box-Cox Lamda = -0.764453124999998
Probability Plot Correlation Coefficient = 0.99537
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.307
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.543
Autocorrellation = 0.159
Mean Sensitivity = 0
```

Checking for normality, May Annualization
Probability Plot Correlation Coefficient = 0.96634
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox transformation
Box-Cox Lamda = -0.692968749999998
Probability Plot Correlation Coefficient = 0.99283
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.441
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.43
Autocorrellation = 0.175
Mean Sensitivity = 0

Checking for normality, June Annualization
Probability Plot Correlation Coefficient = 0.97823
Table Value at 95% Confidence = 0.97311
The data set is sufficiently normal that no transformation is required

Checking for normality, July Annualization
Probability Plot Correlation Coefficient = 0.98603
Table Value at 95% Confidence = 0.97311
The data set is sufficiently normal that no transformation is required

Checking for normality, August Annualization
Probability Plot Correlation Coefficient = 0.98247
Table Value at 95% Confidence = 0.97311
The data set is sufficiently normal that no transformation is required

Checking for normality, September Annualization
Probability Plot Correlation Coefficient = 0.98039
Table Value at 95% Confidence = 0.97311
The data set is sufficiently normal that no transformation is required

```
Checking for normality, October Annualization
Probability Plot Correlation Coefficient = 0.96935
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox
transformation
Box-Cox Lamda = -0.674218749999998
Probability Plot Correlation Coefficient = 0.99362
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.481
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.314
Autocorrellation = 0.254
Mean Sensitivity = 0
```

```
Checking for normality, November Annualization
Probability Plot Correlation Coefficient = 0.9671
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox
transformation
Box-Cox Lamda = -0.867578124999998
Probability Plot Correlation Coefficient = 0.99557
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.152
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.458
Autocorrellation = 0.223
Mean Sensitivity = 0
```

Checking for normality, December Annualization
Probability Plot Correlation Coefficient = 0.96514
Table Value at 95% Confidence = 0.97311
The data set is sufficiently non normal to undergo Box-Cox transformation
Box-Cox Lamda = -0.925781249999998
Probability Plot Correlation Coefficient = 0.99515
Table Value at 95% Confidence = 0.97311
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 43
Mean = 1.08
Variance = 0
Standard Deviation = 0
Standard Error = 0
Coefficient of Variation = 0
Skewness = 0
Kurtosis = -0.476
Autocorrellation = 0.204
Mean Sensitivity = 0

Testing for Stationarity
Mann Kendall Test For Trend

January Annualization Trend analysis
Uc = 0.6697869
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

February Annualization Trend analysis
Uc = 0.6488561
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

March Annualization Trend analysis
Uc = 0.5651327
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

April Annualization Trend analysis
Uc = 0.523271
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

May Annualization Trend analysis
Uc = 0.2721009
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

June Annualization Trend analysis
Uc = 0.6907178
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

July Annualization Trend analysis
Uc = 0.7744411
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

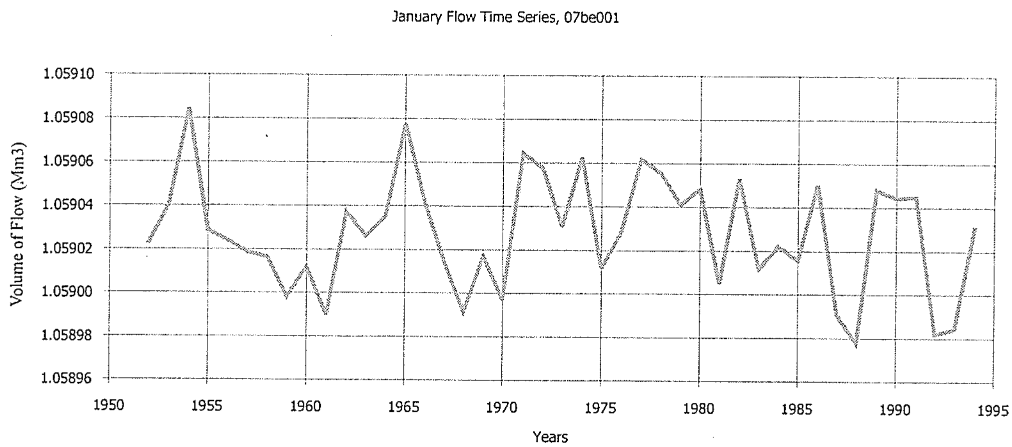
August Annualization Trend analysis
Uc = 0.7325795
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

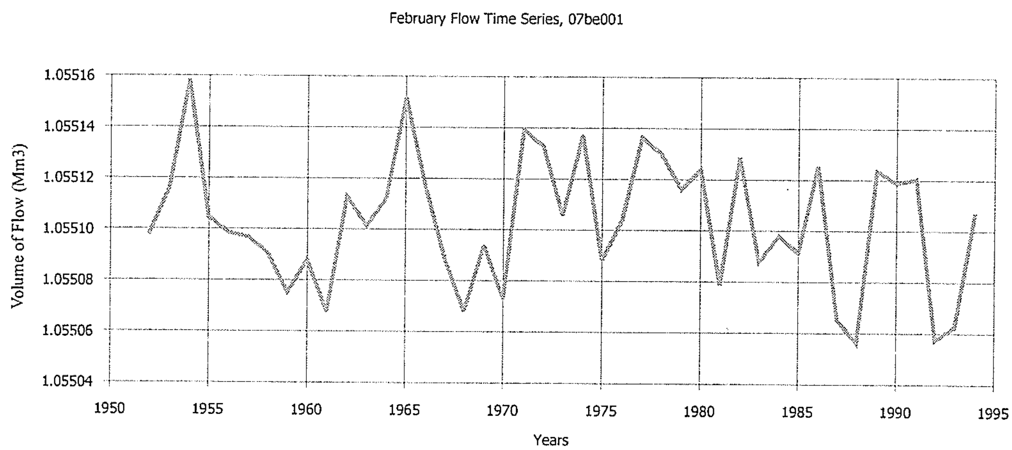
September Annualization Trend analysis
Uc = 0.7535103
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

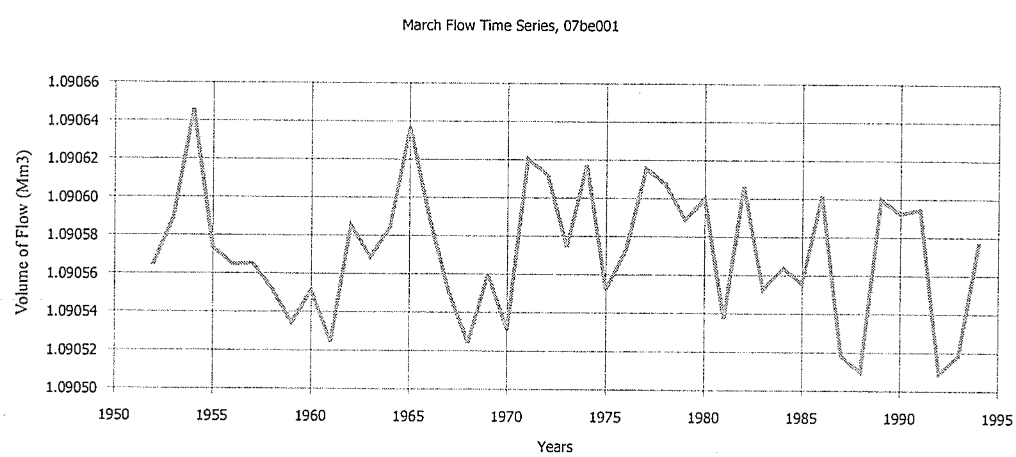
October Annualization Trend analysis
Uc = 0.6697869
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

November Annualization Trend analysis
Uc = 0.6488561
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

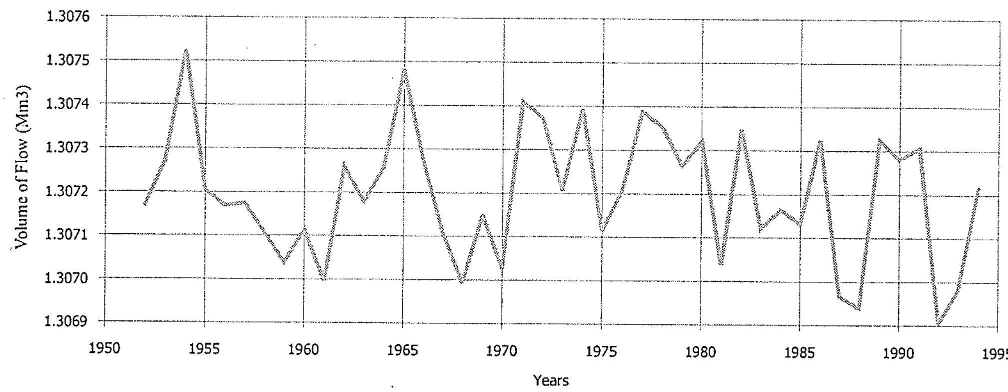
December Annualization Trend analysis
Uc = 0.8163028
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval


January Flow Time Series, 07be001
Volume of Flow (Mm3)
1.05910
1.05908
1.05906
1.05904
1.05902
1.05900
1.05898
1.05896
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years

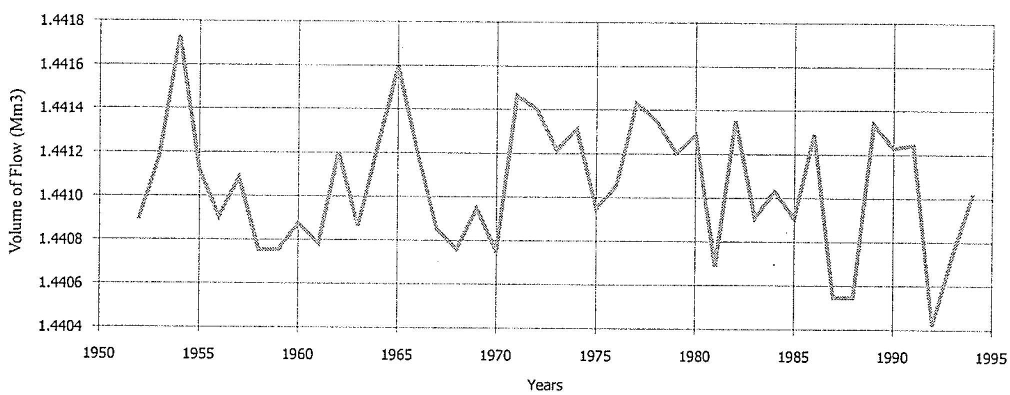

February Flow Time Series, 07be001
Volume of Flow (Mm3)
1.05516
1.05514
1.05512
1.05510
1.05508
1.05506
1.05504
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years


March Flow Time Series, 07be001
Volume of Flow (Mm3)
1.09066
1.09064
1.09062
1.09060
1.09058
1.09056
1.09054
1.09052
1.09050
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years

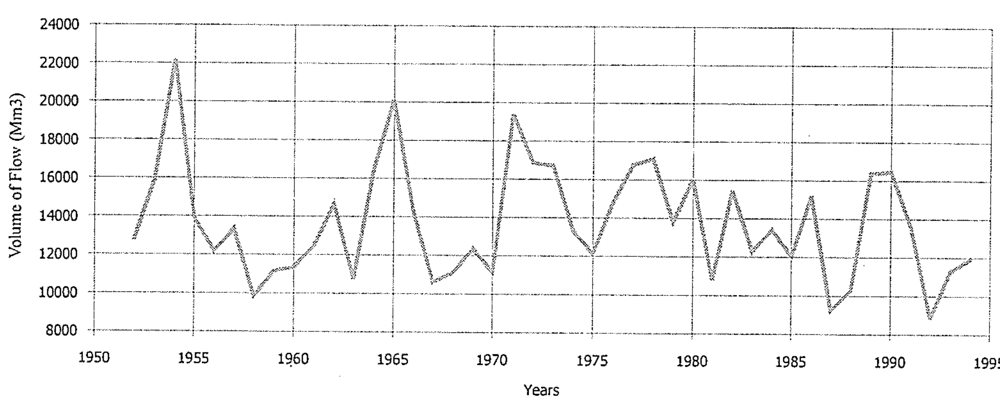
April Flow Time Series, 07be001



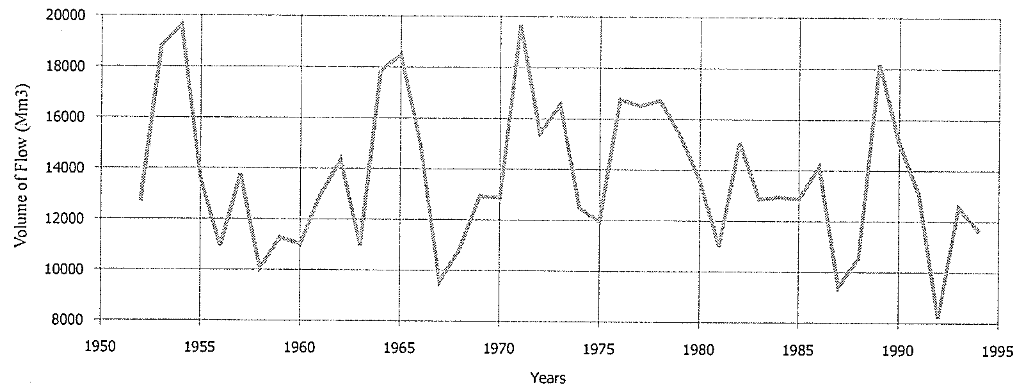
May Flow Time Series, 07be001



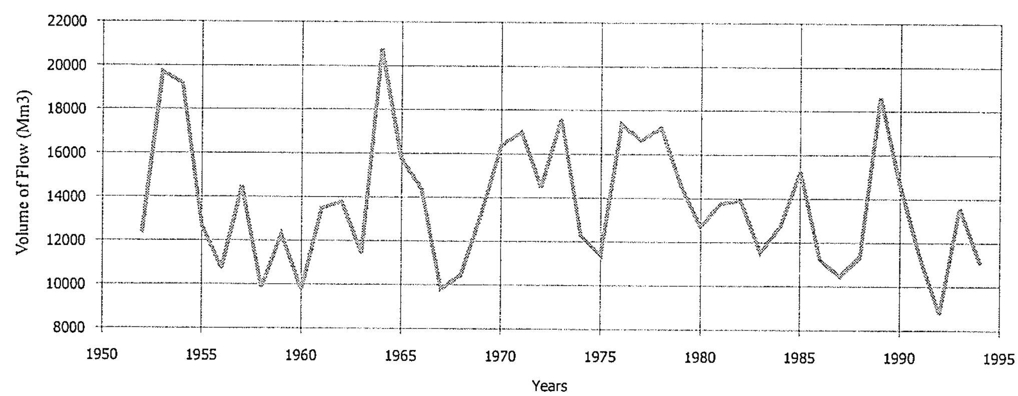
June Flow Time Series, 07be001



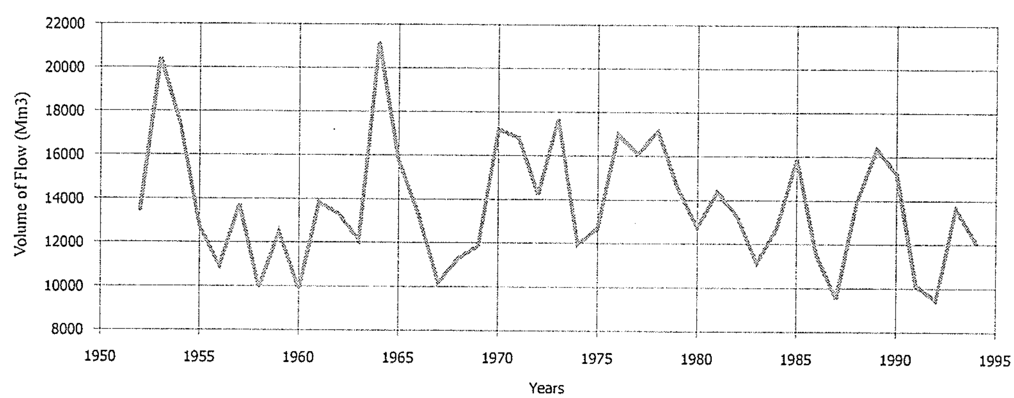
July Flow Time Series, 07be001

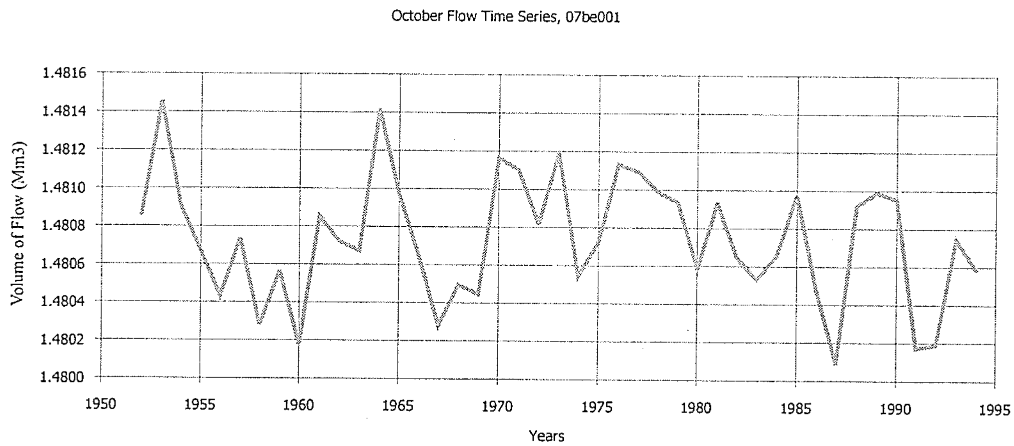


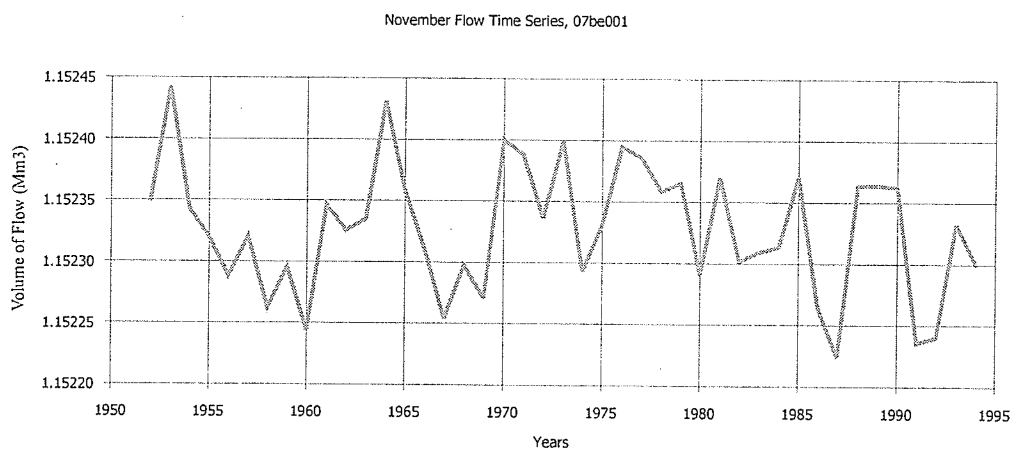
August Flow Time Series, 07be001

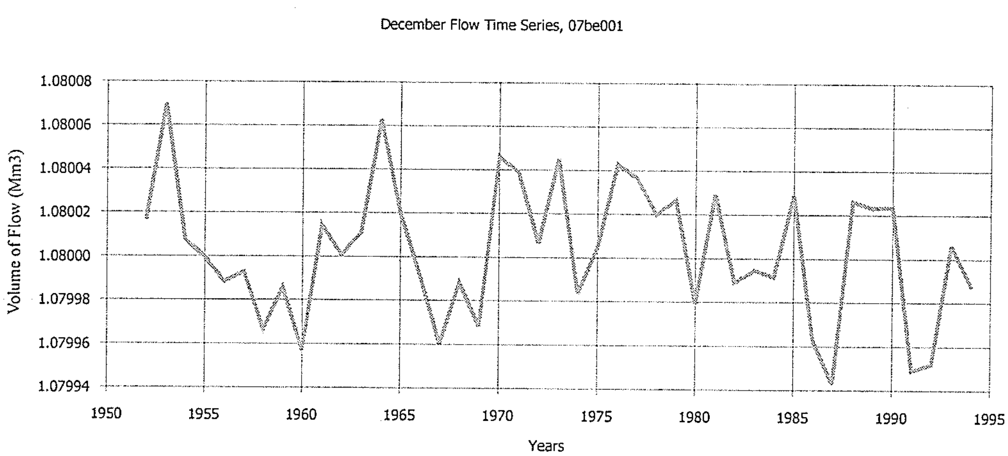


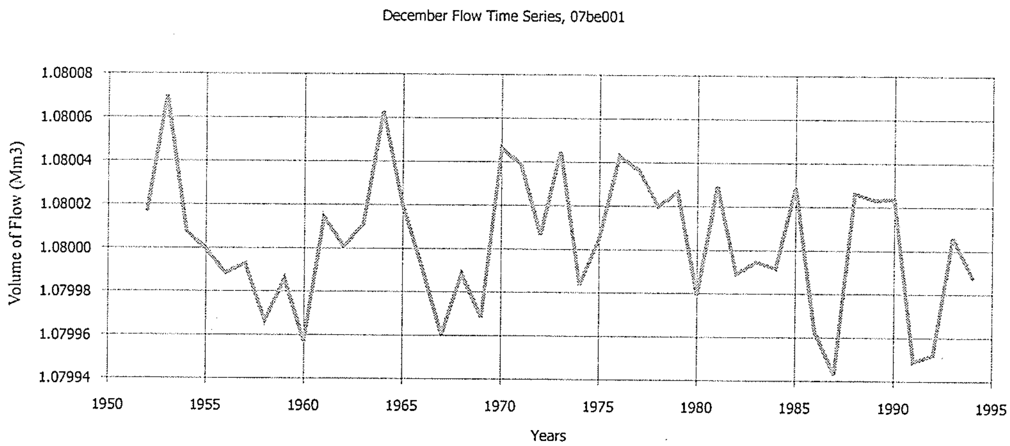
September Flow Time Series, 07be001




October Flow Time Series, 07be001
Volume of Flow (Mm3)
1.4816
1.4814
1.4812
1.4810
1.4808
1.4806
1.4804
1.4802
1.4800
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years


November Flow Time Series, 07be001
Volume of Flow (Mm3)
1.15245
1.15240
1.15235
1.15230
1.15225
1.15220
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years


December Flow Time Series, 07be001
Volume of Flow (Mm3)
1.08008
1.08006
1.08004
1.08002
1.08000
1.07998
1.07996
1.07994
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years


December Flow Time Series, 07be001
1.08008
1.08006
1.08004
1.08002
1.08000
1.07998
1.07996
1.07994
Volume of Flow (Mm3)
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
Years

Check for Autocorrelation

January Annualization

ACF	PACF
0.182873	0.182873
-0.051465	-0.087846
-0.143536	-0.122232
-0.12561	-0.084161
-0.004983	0.018542
0.00571	-0.026424
-0.027619	-0.052632
0.060445	0.0694
0.086826	0.064334
-0.179782	-0.228909
0.055146	0.161174
0.115258	0.104622
-0.077508	-0.182592
-0.180694	-0.17424
-0.111634	0.039418

tvalue for autocorrelation (t*) = 1.1910234389738
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

February Annualization

ACF	PACF
0.174601	0.174601
-0.053243	-0.086361
-0.133631	-0.112704
-0.134451	-0.099137
-0.00838	0.018098
0.012213	-0.017902
-0.028665	-0.057152
0.062574	0.068984
0.083784	0.062611
-0.182093	-0.225692
0.062625	0.162194
0.117451	0.10364
-0.088629	-0.183872
-0.174444	-0.169716
-0.109226	0.031572

tvalue for autocorrelation (t*) = 1.13542630209797
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

March Annualization

ACF	PACF
0.172342	0.172342
-0.056538	-0.08888
-0.125227	-0.103154
-0.139905	-0.10889
-0.012869	0.015844
0.01751	-0.010929
-0.030216	-0.060616
0.064042	0.069256
0.082813	0.060996
-0.183042	-0.22291
0.068839	0.166266
0.117784	0.097751
-0.099055	-0.185447
-0.170271	-0.16313
-0.108694	0.021816

tvalue for autocorrelation (t*) = 1.12027657929955
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

April Annualization

ACF	PACF
0.159364	0.159364
-0.05258	-0.080009
-0.126172	-0.107781
-0.141291	-0.1115
-0.012234	0.014398
0.020614	-0.006803
-0.026337	-0.058656
0.058873	0.060682
0.083407	0.067192
-0.189555	-0.227796
0.085566	0.179054
0.116276	0.0945
-0.096241	-0.179241
-0.169526	-0.166063
-0.109056	0.021032

tvalue for autocorrelation (t*) = 1.03361091237198
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

May Annualization

ACF	PACF
0.174565	0.174565
-0.085879	-0.120008
-0.135932	-0.102758
-0.104155	-0.073929
-0.002954	0.006372
0.026535	-0.003886
0.05925	0.03832
0.024625	0.003393
0.02155	0.029912
-0.172822	-0.178602
0.139023	0.236467
0.096504	-0.001801
-0.100847	-0.136313
-0.18139	-0.139547
-0.159181	-0.083606

tvalue for autocorrelation (t*) = 1.13515794076072
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

June Annualization

ACF	PACF
0.250281	0.250281
-0.100218	-0.173741
-0.210295	-0.151112
-0.146914	-0.074074
-0.096153	-0.093195
0.05765	0.054107
0.118569	0.041815
0.00644	-0.069619
-0.042552	-0.01012
-0.122818	-0.103678
0.198503	0.297238
0.067484	-0.103386
-0.062556	-0.070684
-0.20054	-0.14519
-0.185974	-0.117437

tvalue for autocorrelation (t*) = 1.65525485766618
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

July Annualization

ACF	PACF
0.285335	0.285335
-0.107599	-0.205768
-0.235893	-0.15933
-0.136007	-0.038913
-0.06042	-0.07086
0.016499	-0.008188
0.201093	0.181829
-0.03646	-0.207062
-0.111838	-0.011115
-0.148804	-0.0797
0.185894	0.26788
0.146443	-0.052823
-0.02322	-0.06906
-0.166663	-0.157621
-0.250495	-0.120732

tvalue for autocorrelation (t*) = 1.90624739236432
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

August Annualization

ACF	PACF
0.220921	0.220921
-0.107301	-0.164117
-0.192419	-0.138905
-0.089226	-0.029984
-0.044623	-0.062912
-0.045485	-0.069276
0.116461	0.124974
0.078404	-0.006434
-0.044327	-0.068171
-0.163546	-0.111381
0.097289	0.188112
0.16741	0.074878
-0.048222	-0.131141
-0.187808	-0.125171
-0.222039	-0.162604

tvalue for autocorrelation (t*) = 1.45041513810802
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

September Annualization

ACF	PACF
0.258378	0.258378
-0.110917	-0.190387
-0.142065	-0.06533
-0.137762	-0.109251
-0.049022	-0.013137
-0.042681	-0.076942
0.048308	0.056117
0.06733	0.005839
0.003671	-0.02181
-0.161561	-0.171801
0.124589	0.265154
0.15226	-0.002694
-0.096119	-0.143216
-0.147512	-0.074498
-0.199904	-0.137051

tvalue for autocorrelation (t*) = 1.71252220983566
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

October Annualization

ACF	PACF
0.253609	0.253609
-0.054098	-0.126555
-0.146358	-0.107317
-0.08489	-0.025309
-0.027874	-0.02024
-0.054031	-0.072581
0.012335	0.032087
0.07229	0.051211
0.067222	0.022307
-0.18705	-0.226541
0.075863	0.238665
0.125831	0.036437
-0.080013	-0.199247
-0.186066	-0.09755
-0.183736	-0.066589

tvalue for autocorrelation (t*) = 1.67871075531116
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

November Annualization

ACF	PACF
0.222696	0.222696
-0.030499	-0.084272
-0.149315	-0.130538
-0.107544	-0.049236
-0.016885	0.00678
-0.022494	-0.049033
0.004233	-0.001729
0.061229	0.057052
0.068205	0.036665
-0.176664	-0.2195
0.070641	0.200032
0.118631	0.07997
-0.062651	-0.184453
-0.171602	-0.129407
-0.158577	-0.019328

tvalue for autocorrelation (t*) = 1.46263963668588
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.

December Annualization

ACF	PACF
0.204088	0.204088
-0.027094	-0.071734
-0.129425	-0.114178
-0.12913	-0.084853
-0.012703	0.022881
-0.00694	-0.032171
-0.001253	-0.018861
0.064451	0.062047
0.06976	0.045682
-0.17405	-0.215337
0.063742	0.175236
0.123975	0.104369
-0.076303	-0.185338
-0.16956	-0.154687
-0.1386	0.007967

tvalue for autocorrelation (t*) = 1.33484240511455
Critical t value (tcrit) = 2.0199499130249
The series is not autocorrelated, regression analysis may proceed.
Datafile Printed as 07be001.str

End of analysis.

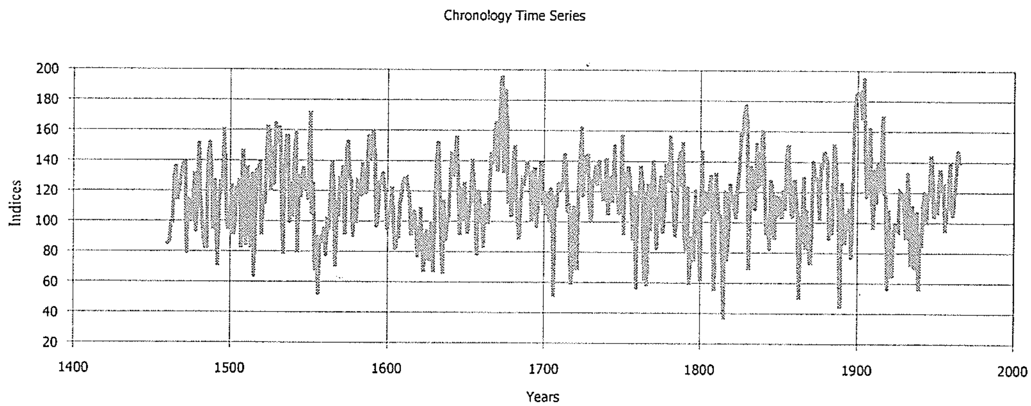
Tree Ring Chronology Preprocessor Log

Input Filename: CANA022.TXT
Prefix for all files associated with this raw file: CANA022
Site Name   : EXSHAW+TUNNEL+BANFF,ALBERTA,
State/Country : CANADA
Location   : 51ø 10'N 115ø 33'W
Year Range   : 1460 - 1965
Elevation   : 1310m
Species Code   : PSME
Common Name   : DOUGLAS-FIR
P. I.   : C. W. FERGUSON AND M. L. PARKER
File Name   : CANA022.CRN
Number of Data Values = 506
Mean = 996.022
Variance = 123159.281
Standard Deviation = 350.941
Standard Error = 15.601
Coefficient of Variation = 0.352
Skewness = 0.463
Kurtosis = 0.59
Autocorrellation = 0.426
Mean Sensitivity = 0.328

Checking for normality
Probability Plot Correlation Coefficient = 0.99276
Table Value at 95% Confidence = 0.996983
The data set is sufficiently non normal to undergo Box-Cox transformation
Box-Cox Lamda = 0.625000000000002
Probability Plot Correlation Coefficient = 0.99806
Table Value at 95% Confidence = 0.996983
The set was transformed to normal using the Box-Cox method.
Number of Data Values = 506
Mean = 116.275
Variance = 709.698
Standard Deviation = 26.64
Standard Error = 1.184
Coefficient of Variation = 0.229
Skewness = 0
Kurtosis = 0.265
Autocorrellation = 0.391
Mean Sensitivity = 0.213

Testing for Stationarity
Mann Kendall Test For Trend

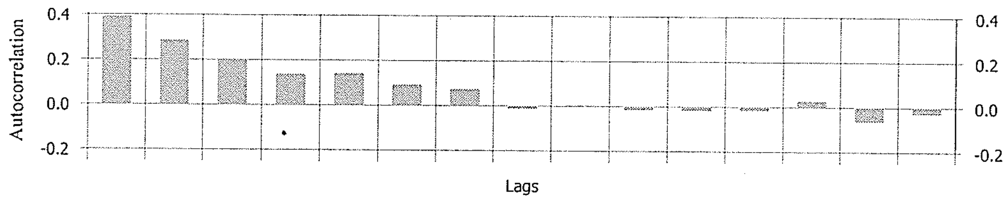
Trend analysis
Uc = 0.1560667
95% Test Stat = 1.960
The hypothesis of a up or downward trend is rejected at the 95% confidence interval

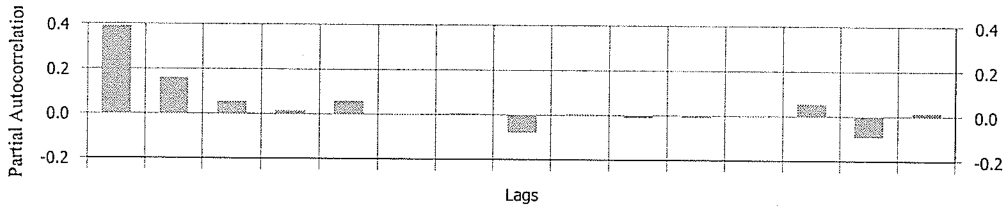


Check for Autocorrelation

ACF	PACF
0.39126	0.39126
0.287953	0.159247
0.202261	0.054797
0.139027	0.014767
0.14041	0.057932
0.091497	-0.002755
0.073788	0.005447
-0.00989	-0.07786
0.002012	0.004207
-0.012237	-0.010902
-0.01591	-0.005436
-0.01256	-0.001856
0.025854	0.053494
-0.060395	-0.090892
-0.027686	0.01312

tvalue for autocorrelation (t*) = 9.54466405012143
Critical t value (tcrit) = 1.97991454601288
The series is autocorrelated, an Arma model is required.





Arma Model Minitab Output

```
ARIMA Model

ARIMA model for C2

Estimates at each iteration
Iteration        SSE      Parameters
    0         474.616    0.100    0.090
    1         438.860    0.250    0.049
    2         427.443    0.385    0.009
    3         427.364    0.392    0.000
    4         427.363    0.393   -0.000
    5         427.363    0.393   -0.000
    6         427.363    0.393   -0.000
Relative change in each estimate less than  0.0010

Final Estimates of Parameters
Type            Coef        StDev          T
AR   1        0.3929       0.0410       9.59
Constant    -0.00019      0.04093      -0.00
Mean        -0.00032      0.06742

Number of observations:  506
Residuals:     SS = 427.190  (backforecasts excluded)
               MS =   0.848  DF = 504

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                  12            24            36            48
Chi-Square    20.6(DF=11)   29.4(DF=23)   47.5(DF=35)   57.5(DF=47)
```

Check for Autocorrelation

ACF	PACF
-0.064232	-0.064232
0.11795	0.114296
0.078049	0.093706
0.03104	0.028971
0.08448	0.070035
0.024705	0.02177
0.063675	0.0456
-0.048676	-0.061813
0.013251	-0.015826
-0.010813	-0.015661
-0.009694	-0.01014
-0.022614	-0.027778
0.069851	0.079258
-0.081157	-0.064884
0.012586	-0.000661

tvalue for autocorrelation (t*) = -1.44426584743814
Critical t value (tcrit) = 1.97991454601288
The series is not autocorrelated, regression analysis may proceed.
Datafile Printed as CANA022.trg

End of analysis.

Appendix B

Correlation Analyzer Example Output

The Following are the Tree Ring Files Used.

CANA041.trg
CANA042.trg
CANA043.trg
Mn002.trg
MN005.trg
MN006.trg
MN008.trg
MN009.trg
MN010.trg
MN013.trg
MN014.trg
MN015.trg
MN016.trg
MN017.trg
MN018.trg
MN025.trg
MN026.trg

Output File For Reconstruction

Correlation Analysis for Investigation of Predictors

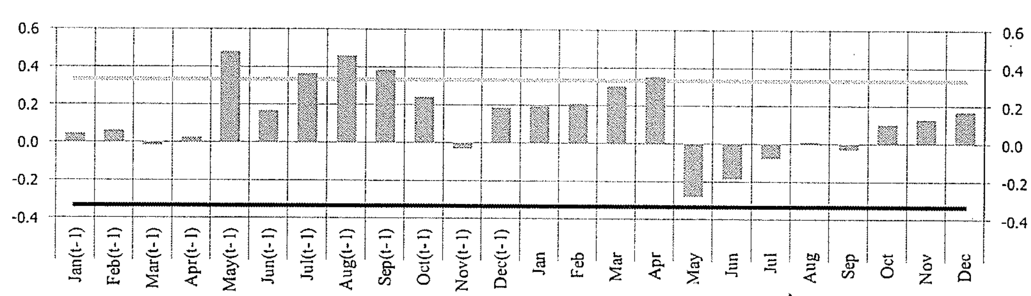
The following is the Monthly Stream Flow and file used.

02AA001.raw

Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	0.044805	0.215029
Feb(t-1)	0.062271	0.30162
Mar(t-1)	-0.016053	-0.0747
Apr(t-1)	0.025498	0.121149
May(t-1)	0.480092	3.123006
Jun(t-1)	0.169339	0.871476
Jul(t-1)	0.366872	2.162622
Aug(t-1)	0.458802	2.925224
Sep(t-1)	0.383411	2.290223
Oct(t-1)	0.24262	1.30762
Nov(t-1)	-0.030301	-0.140021
Dec(t-1)	0.187936	0.978195
Jan	0.195337	1.021381
Feb	0.209959	1.107955
Mar	0.302898	1.701605
Apr	0.353582	2.062745
May	-0.275602	-1.144554
Jun	-0.184285	-0.794281
Jul	-0.075257	-0.340408
Aug	0.006858	0.032277
Sep	-0.029071	-0.134415
Oct	0.103381	0.512094
Nov	0.130534	0.656611
Dec	0.169808	0.874139

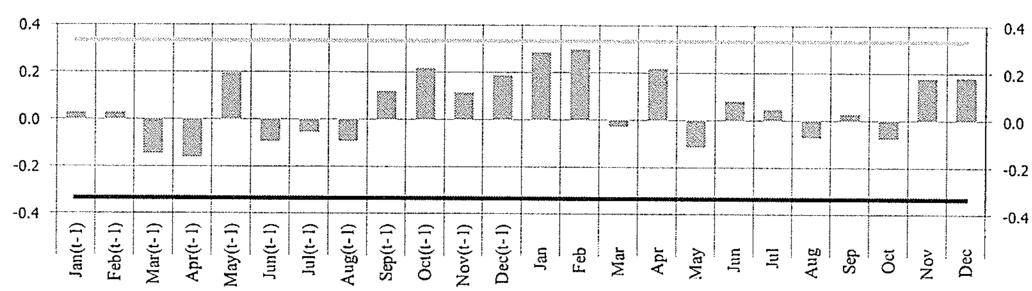
Correlation analysis of CANA041.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	0.028017	0.133289
Feb(t-1)	0.027278	0.129727
Mar(t-1)	-0.146476	-0.641648
Apr(t-1)	-0.160613	-0.699275
May(t-1)	0.196309	1.027087
Jun(t-1)	-0.094361	-0.423083
Jul(t-1)	-0.05432	-0.248134
Aug(t-1)	-0.092367	-0.414518
Sep(t-1)	0.119038	0.594867
Oct(t-1)	0.217563	1.153644
Nov(t-1)	0.114246	0.569371
Dec(t-1)	0.18636	0.969051
Jan	0.284988	1.580815
Feb	0.30045	1.6849
Mar	-0.026737	-0.123763
Apr	0.218089	1.156822
May	-0.111429	-0.495756
Jun	0.081144	0.397048
Jul	0.046022	0.221007
Aug	-0.071736	-0.325016
Sep	0.026304	0.125033
Oct	-0.076629	-0.346396
Nov	0.178984	0.926507
Dec	0.181467	0.940787

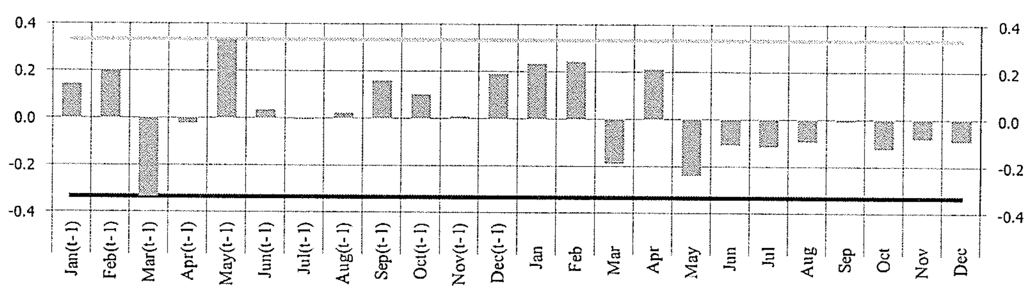
Correlation analysis of CANA042.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	0.14272	0.722995
Feb(t-1)	0.201948	1.060313
Mar(t-1)	-0.334173	-1.356988
Apr(t-1)	-0.022261	-0.103272
May(t-1)	0.333128	1.913383
Jun(t-1)	0.034887	0.166568
Jul(t-1)	-0.004876	-0.022816
Aug(t-1)	0.021087	0.099969
Sep(t-1)	0.160036	0.81903
Oct(t-1)	0.100134	0.49511
Nov(t-1)	0.008001	0.037677
Dec(t-1)	0.190399	0.992523
Jan	0.235036	1.26045
Feb	0.242466	1.306658
Mar	-0.188793	-0.812164
Apr	0.210779	1.112857
May	-0.238133	-1.003801
Jun	-0.107069	-0.477296
Jul	-0.11619	-0.515835
Aug	-0.094487	-0.423622
Sep	-0.006725	-0.031438
Oct	-0.12417	-0.549303
Nov	-0.080649	-0.363886
Dec	-0.089365	-0.401597

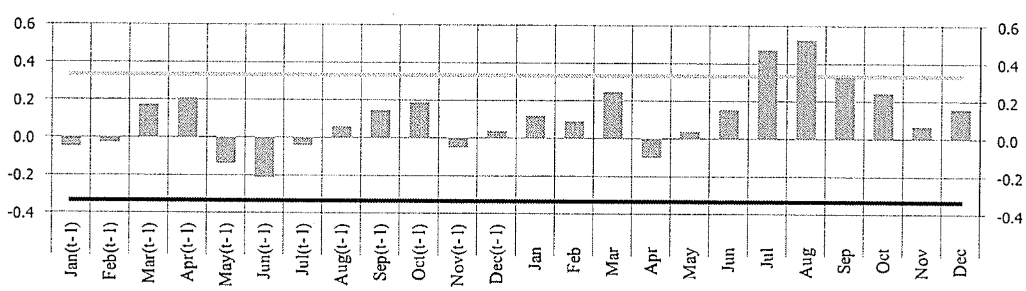
Correlation analysis of CANA043.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.049558	-0.226895
Feb(t-1)	-0.025108	-0.116316
Mar(t-1)	0.175073	0.904113
Apr(t-1)	0.207905	1.09569
May(t-1)	-0.137844	-0.60612
Jun(t-1)	-0.212816	-0.906397
Jul(t-1)	-0.038565	-0.177494
Aug(t-1)	0.060135	0.290943
Sep(t-1)	0.145933	0.740661
Oct(t-1)	0.187495	0.975634
Nov(t-1)	-0.048464	-0.222
Dec(t-1)	0.039144	0.187302
Jan	0.119899	0.599463
Feb	0.089904	0.442023
Mar	0.247203	1.336368
Apr	-0.098667	-0.441519
May	0.037527	0.179415
Jun	0.155827	0.795499
Jul	0.473108	3.057105
Aug	0.523672	3.558919
Sep	0.331288	1.900193
Oct	0.244588	1.319943
Nov	0.068389	0.332337
Dec	0.158215	0.808832

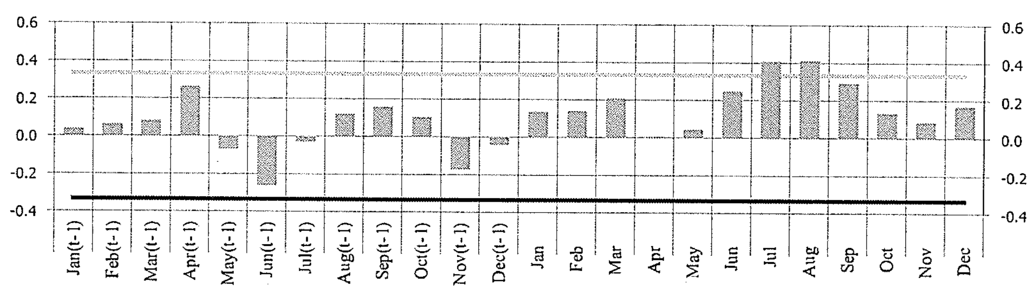
Correlation analysis of Mn002.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	0.037086	0.177267
Feb(t-1)	0.061368	0.297101
Mar(t-1)	0.082127	0.402073
Apr(t-1)	0.263968	1.443162
May(t-1)	-0.066979	-0.304137
Jun(t-1)	-0.262586	-1.096107
Jul(t-1)	-0.02742	-0.126884
Aug(t-1)	0.121072	0.605728
Sep(t-1)	0.159555	0.816331
Oct(t-1)	0.104015	0.515416
Nov(t-1)	-0.171313	-0.742446
Dec(t-1)	-0.037408	-0.172268
Jan	0.136748	0.690341
Feb	0.141293	0.715169
Mar	0.209468	1.105016
Apr	0.006121	0.0288
May	0.043044	0.206386
Jun	0.248133	1.342228
Jul	0.407657	2.484392
Aug	0.411662	2.51732
Sep	0.292756	1.632799
Oct	0.1346	0.678653
Nov	0.084885	0.416204
Dec	0.170385	0.877416

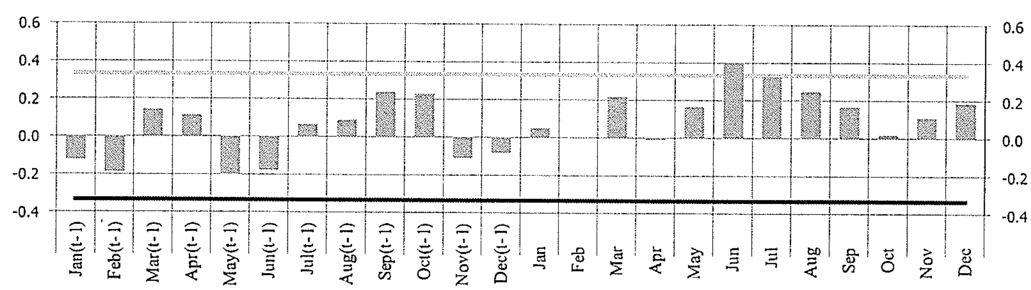
Correlation analysis of MN005.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.119368	-0.529191
Feb(t-1)	-0.186906	-0.804687
Mar(t-1)	0.143426	0.726869
Apr(t-1)	0.114559	0.571034
May(t-1)	-0.19976	-0.855405
Jun(t-1)	-0.178813	-0.772482
Jul(t-1)	0.064677	0.313677
Aug(t-1)	0.089336	0.439095
Sep(t-1)	0.238147	1.279735
Oct(t-1)	0.227862	1.216286
Nov(t-1)	-0.108388	-0.482889
Dec(t-1)	-0.079648	-0.359538
Jan	0.04936	0.237451
Feb	0.000218	0.001021
Mar	0.217677	1.154332
Apr	-0.004214	-0.019725
May	0.164968	0.84676
Jun	0.401635	2.435344
Jul	0.326754	1.867863
Aug	0.246419	1.331437
Sep	0.167052	0.858529
Oct	0.018901	0.089502
Nov	0.109891	0.546324
Dec	0.185946	0.966656

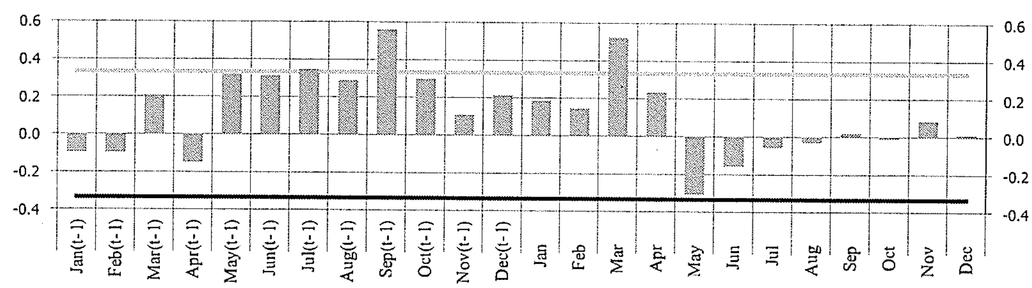
Correlation analysis of MN006.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.096074	-0.430424
Feb(t-1)	-0.09685	-0.433747
Mar(t-1)	0.206055	1.084674
Apr(t-1)	-0.146873	-0.643273
May(t-1)	0.32055	1.824017
Jun(t-1)	0.313578	1.775256
Jul(t-1)	0.347385	2.016944
Aug(t-1)	0.290491	1.61758
Sep(t-1)	0.559729	3.956661
Oct(t-1)	0.299674	1.67962
Nov(t-1)	0.109982	0.546805
Dec(t-1)	0.213262	1.127742
Jan	0.18601	0.967023
Feb	0.145608	0.738867
Mar	0.522249	3.543955
Apr	0.233171	1.248924
May	-0.304757	-1.251413
Jun	-0.154792	-0.675628
Jul	-0.054802	-0.250277
Aug	-0.028379	-0.131261
Sep	0.019765	0.093634
Oct	-0.007412	-0.034638
Nov	0.083986	0.41159
Dec	0.01323	0.06247

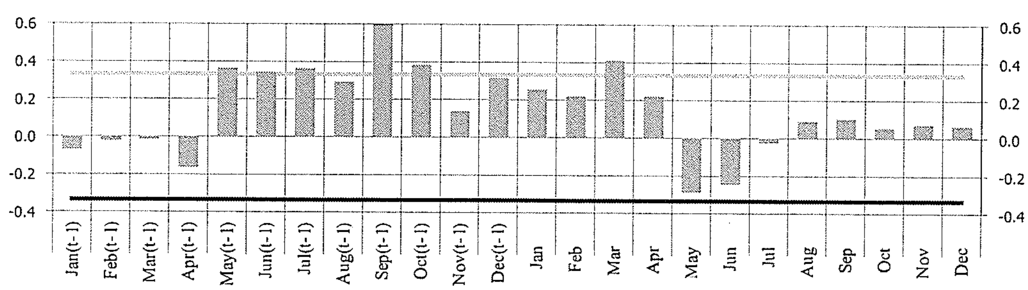
Correlation analysis of MN008.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.067298	-0.305544
Feb(t-1)	-0.021357	-0.099122
Mar(t-1)	-0.013117	-0.061124
Apr(t-1)	-0.162306	-0.706132
May(t-1)	0.36644	2.159338
Jun(t-1)	0.344121	1.993013
Jul(t-1)	0.363326	2.135741
Aug(t-1)	0.292284	1.629625
Sep(t-1)	0.598233	4.426849
Oct(t-1)	0.383522	2.291095
Nov(t-1)	0.139047	0.702883
Dec(t-1)	0.317657	1.80372
Jan	0.254474	1.382367
Feb	0.219071	1.16276
Mar	0.408632	2.492381
Apr	0.220674	1.172473
May	-0.285725	-1.181914
Jun	-0.243446	-1.024001
Jul	-0.024116	-0.111776
Aug	0.0907	0.446132
Sep	0.10333	0.511826
Oct	0.055719	0.268944
Nov	0.074361	0.362524
Dec	0.065309	0.316849

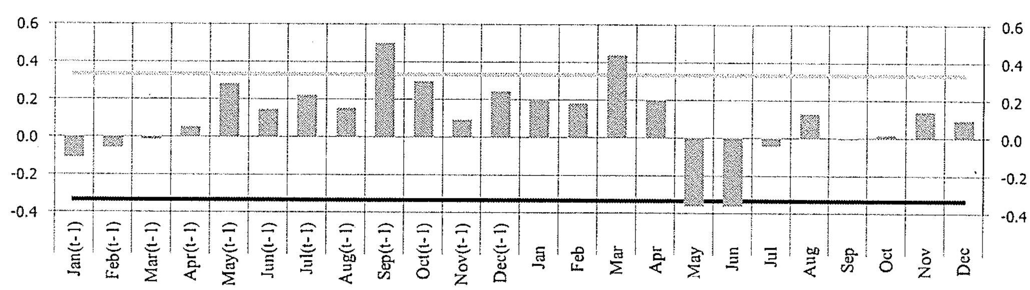
Correlation analysis of MN009.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.107671	-0.479847
Feb(t-1)	-0.05692	-0.259691
Mar(t-1)	-0.011401	-0.053174
Apr(t-1)	0.054144	0.261126
May(t-1)	0.283131	1.568481
Jun(t-1)	0.145142	0.736303
Jul(t-1)	0.221074	1.174902
Aug(t-1)	0.153502	0.782552
Sep(t-1)	0.498585	3.302572
Oct(t-1)	0.297487	1.664765
Nov(t-1)	0.0924	0.45492
Dec(t-1)	0.244564	1.31979
Jan	0.19653	1.028383
Feb	0.181874	0.943132
Mar	0.43833	2.743292
Apr	0.202545	1.063851
May	-0.358528	-1.442779
Jun	-0.359161	-1.444992
Jul	-0.037198	-0.171316
Aug	0.130262	0.65514
Sep	-0.003183	-0.014906
Oct	0.016483	0.077955
Nov	0.141607	0.716893
Dec	0.096077	0.473984

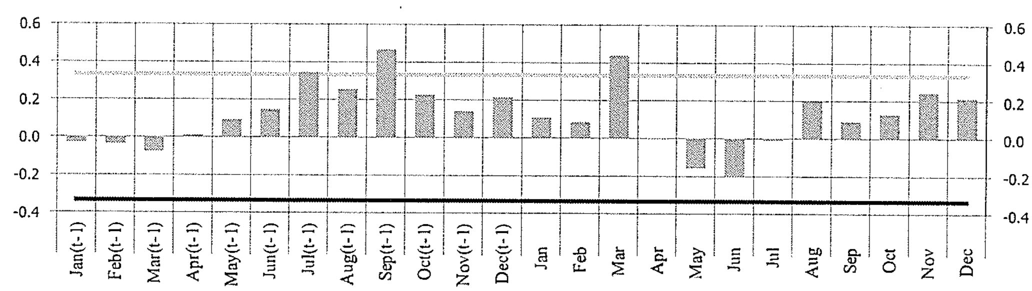
Correlation analysis of MN010.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.02771	-0.128209
Feb(t-1)	-0.035708	-0.164573
Mar(t-1)	-0.074277	-0.336131
Apr(t-1)	0.009824	0.046305
May(t-1)	0.095309	0.469995
Jun(t-1)	0.147598	0.74984
Jul(t-1)	0.34554	2.003399
Aug(t-1)	0.257352	1.400705
Sep(t-1)	0.467045	3.000717
Oct(t-1)	0.224732	1.197158
Nov(t-1)	0.141056	0.71387
Dec(t-1)	0.214252	1.133688
Jan	0.108774	0.540436
Feb	0.084736	0.415438
Mar	0.437147	2.733012
Apr	-0.002451	-0.011481
May	-0.154853	-0.675875
Jun	-0.194553	-0.834925
Jul	-0.005655	-0.026451
Aug	0.199151	1.043803
Sep	0.090113	0.443105
Oct	0.129635	0.651751
Nov	0.244731	1.32084
Dec	0.21502	1.138309

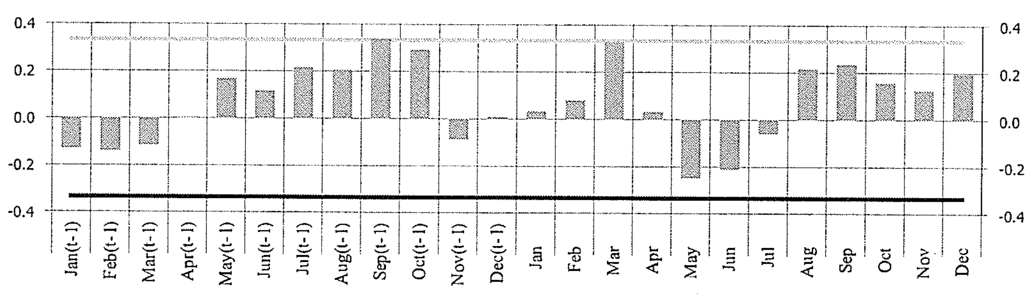
Correlation analysis of MN013.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.13053	-0.57581
Feb(t-1)	-0.139489	-0.612909
Mar(t-1)	-0.115026	-0.510935
Apr(t-1)	0.000736	0.003452
May(t-1)	0.167753	0.862491
Jun(t-1)	0.115613	0.576629
Jul(t-1)	0.215042	1.138444
Aug(t-1)	0.204922	1.077944
Sep(t-1)	0.33547	1.930225
Oct(t-1)	0.291469	1.624143
Nov(t-1)	-0.086332	-0.388511
Dec(t-1)	0.00818	0.038525
Jan	0.031266	0.148997
Feb	0.078738	0.384775
Mar	0.329422	1.88686
Apr	0.032021	0.152658
May	-0.246904	-1.037104
Jun	-0.210036	-0.895582
Jul	-0.059431	-0.270824
Aug	0.216489	1.147161
Sep	0.235445	1.26298
Oct	0.158705	0.811574
Nov	0.125609	0.630054
Dec	0.19699	1.031084

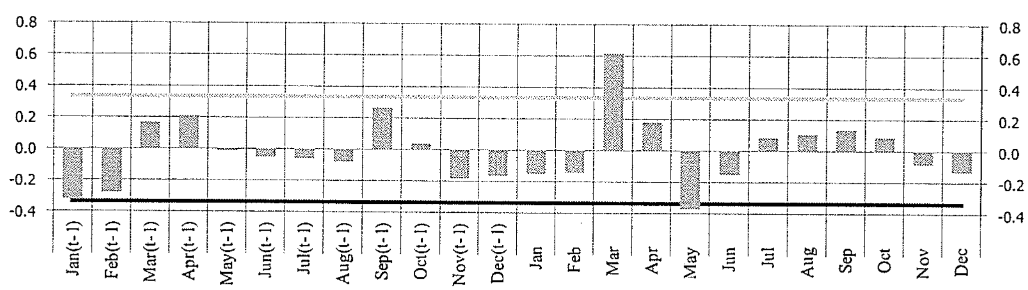
Correlation analysis of MN014.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.319366	-1.304118
Feb(t-1)	-0.27808	-1.153726
Mar(t-1)	0.164697	0.845228
Apr(t-1)	0.209486	1.105123
May(t-1)	-0.009973	-0.046548
Jun(t-1)	-0.048187	-0.220759
Jul(t-1)	-0.059754	-0.272256
Aug(t-1)	-0.079422	-0.358555
Sep(t-1)	0.264759	1.448262
Oct(t-1)	0.038803	0.185641
Nov(t-1)	-0.182712	-0.788025
Dec(t-1)	-0.16057	-0.6991
Jan	-0.149634	-0.654577
Feb	-0.14024	-0.616008
Mar	0.618035	4.69043
Apr	0.178605	0.924333
May	-0.367988	-1.47572
Jun	-0.149917	-0.655735
Jul	0.081979	0.401318
Aug	0.105757	0.524559
Sep	0.136962	0.691509
Oct	0.087574	0.430018
Nov	-0.083511	-0.376303
Dec	-0.127725	-0.564137

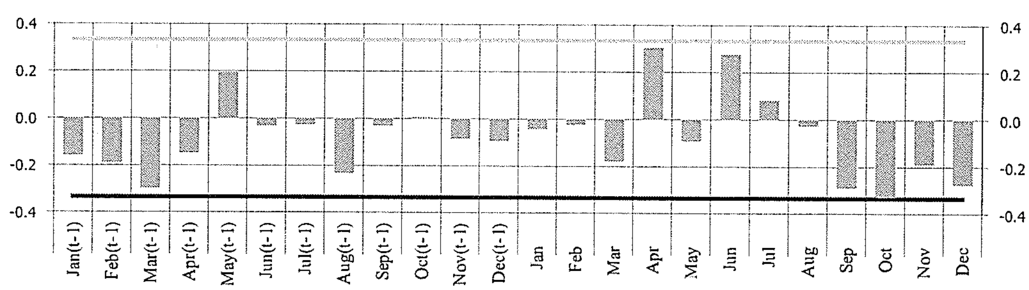
Correlation analysis of MN015.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.159217	-0.693614
Feb(t-1)	-0.188939	-0.812743
Mar(t-1)	-0.297483	-1.22496
Apr(t-1)	-0.147285	-0.644964
May(t-1)	0.194385	1.015806
Jun(t-1)	-0.031787	-0.146778
Jul(t-1)	-0.026617	-0.123214
Aug(t-1)	-0.23378	-0.987188
Sep(t-1)	-0.030067	-0.138951
Oct(t-1)	0.004904	0.023057
Nov(t-1)	-0.084218	-0.379364
Dec(t-1)	-0.094334	-0.422963
Jan	-0.040648	-0.186894
Feb	-0.022441	-0.104097
Mar	-0.178625	-0.771732
Apr	0.30578	1.721357
May	-0.093082	-0.41759
Jun	0.275005	1.514905
Jul	0.080986	0.396241
Aug	-0.026588	-0.123082
Sep	-0.288976	-1.193853
Oct	-0.326053	-1.328063
Nov	-0.188776	-0.812096
Dec	-0.276292	-1.147108

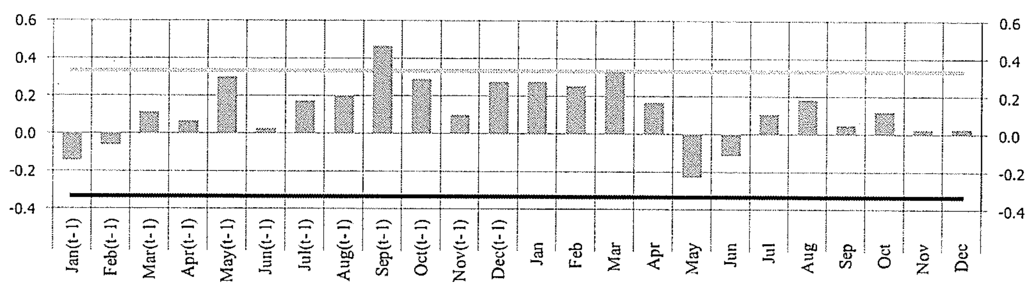
Correlation analysis of MN016.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.14278	-0.626464
Feb(t-1)	-0.061544	-0.280176
Mar(t-1)	0.112346	0.559303
Apr(t-1)	0.065831	0.319471
May(t-1)	0.300313	1.683969
Jun(t-1)	0.026323	0.125123
Jul(t-1)	0.171409	0.883234
Aug(t-1)	0.200058	1.049151
Sep(t-1)	0.462652	2.960316
Oct(t-1)	0.290494	1.617596
Nov(t-1)	0.098922	0.488793
Dec(t-1)	0.275967	1.52121
Jan	0.274923	1.514367
Feb	0.25287	1.372181
Mar	0.329806	1.8896
Apr	0.16838	0.866046
May	-0.226637	-0.959806
Jun	-0.112015	-0.498234
Jul	0.105798	0.524774
Aug	0.183742	0.953905
Sep	0.048961	0.235484
Oct	0.121343	0.607177
Nov	0.027085	0.128794
Dec	0.027231	0.1295

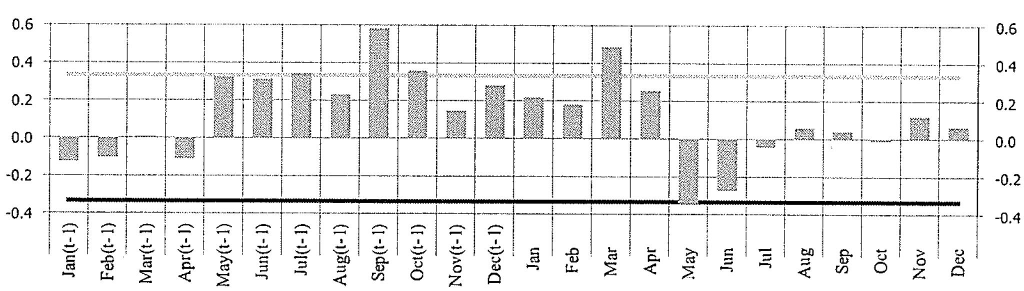
Correlation analysis of MN017.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.125252	-0.553823
Feb(t-1)	-0.103552	-0.462352
Mar(t-1)	0.007049	0.033182
Apr(t-1)	-0.110214	-0.490619
May(t-1)	0.322902	1.840584
Jun(t-1)	0.314365	1.780734
Jul(t-1)	0.341983	1.97741
Aug(t-1)	0.232049	1.242007
Sep(t-1)	0.583114	4.236003
Oct(t-1)	0.358123	2.096608
Nov(t-1)	0.145975	0.740889
Dec(t-1)	0.279864	1.546862
Jan	0.217936	1.155899
Feb	0.181669	0.941951
Mar	0.486058	3.180111
Apr	0.252139	1.36754
May	-0.341453	-1.382784
Jun	-0.272613	-1.13347
Jul	-0.042265	-0.19418
Aug	0.059581	0.288174
Sep	0.041532	0.198977
Oct	-0.007	-0.032719
Nov	0.118751	0.593332
Dec	0.064381	0.31219

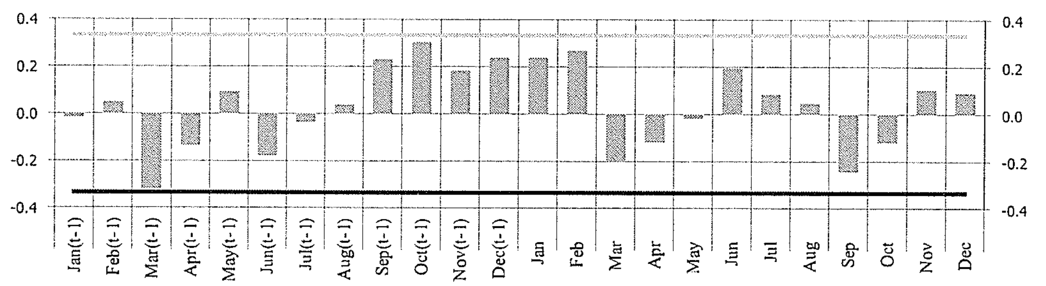
Correlation analysis of MN018.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	-0.015885	-0.073923
Feb(t-1)	0.049151	0.236421
Mar(t-1)	-0.319656	-1.305162
Apr(t-1)	-0.133397	-0.587714
May(t-1)	0.093036	0.458215
Jun(t-1)	-0.17674	-0.764199
Jul(t-1)	-0.034957	-0.161171
Aug(t-1)	0.038815	0.185699
Sep(t-1)	0.229338	1.225338
Oct(t-1)	0.303586	1.706317
Nov(t-1)	0.182321	0.945704
Dec(t-1)	0.239004	1.285068
Jan	0.239585	1.288681
Feb	0.268074	1.46971
Mar	-0.196187	-0.841361
Apr	-0.119783	-0.530932
May	-0.01807	-0.084001
Jun	0.194616	1.017161
Jul	0.08348	0.408997
Aug	0.045256	0.217242
Sep	-0.245355	-1.03124
Oct	-0.119647	-0.530362
Nov	0.103896	0.514792
Dec	0.090822	0.446765

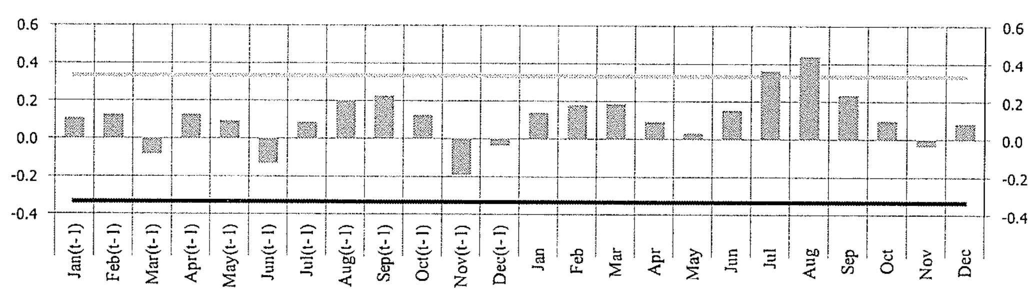
Correlation analysis of MN025.trg



Correlation Graphs of Tree Ring Series Vs Monthly Streamflow
T Critical Value for Correlation Test = 2.03570008277893
Data set sample size35

Month	Corr	Tval
Jan(t-1)	0.107195	0.532117
Feb(t-1)	0.124732	0.625345
Mar(t-1)	-0.082243	-0.370807
Apr(t-1)	0.128423	0.645211
May(t-1)	0.091936	0.452523
Jun(t-1)	-0.131035	-0.57791
Jul(t-1)	0.086842	0.426251
Aug(t-1)	0.203877	1.071737
Sep(t-1)	0.225504	1.201865
Oct(t-1)	0.125152	0.627602
Nov(t-1)	-0.190599	-0.819311
Dec(t-1)	-0.032114	-0.148264
Jan	0.138195	0.69823
Feb	0.180072	0.932758
Mar	0.185816	0.965901
Apr	0.092577	0.455835
May	0.032629	0.155604
Jun	0.154764	0.789575
Jul	0.360619	2.115342
Aug	0.440316	2.760602
Sep	0.237032	1.272813
Oct	0.09711	0.479357
Nov	-0.031899	-0.14729
Dec	0.083555	0.409383

Correlation analysis of MN026.trg



Appendix C

Correlation Analysis for 07BE001

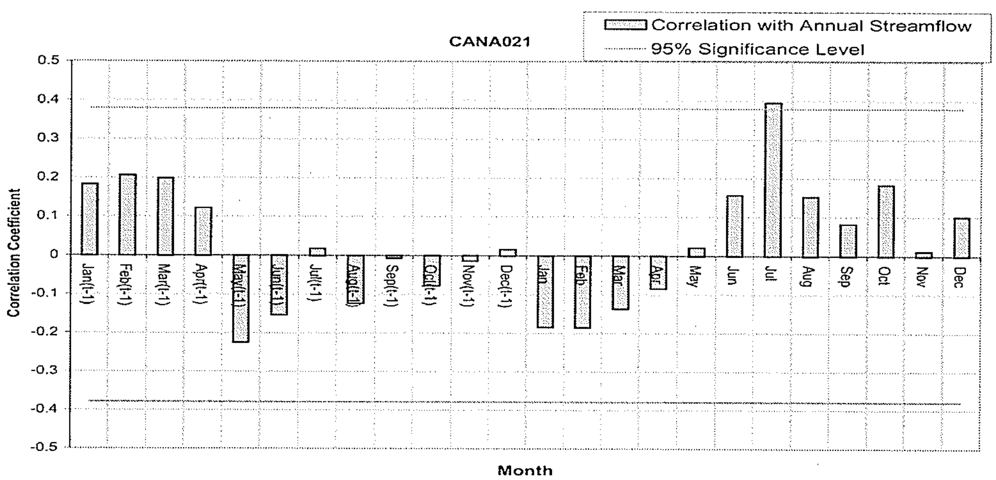


Figure C.1: Correlation between CANA021 and Monthly Flows of 07EB001

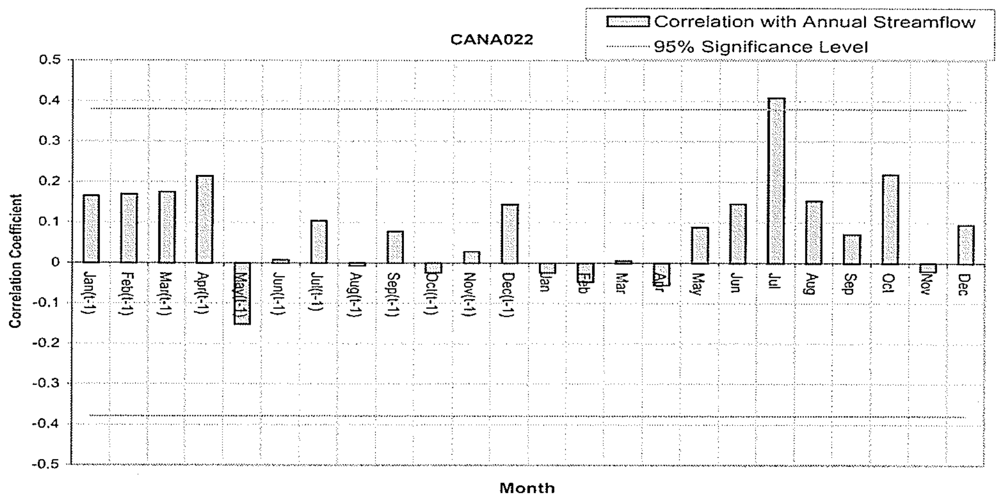


Figure C.2: Correlation between CANA022 and Monthly Flows of 07EB001

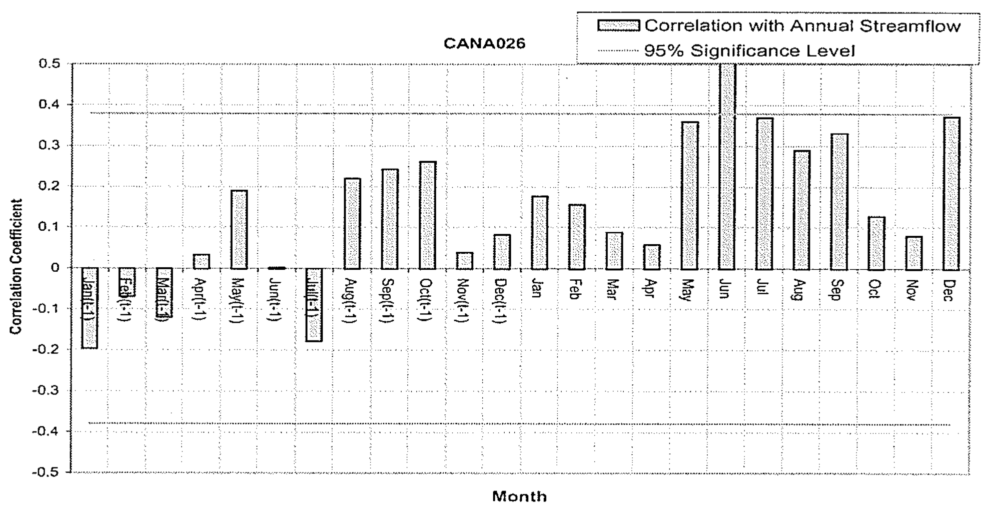


Figure C.3: Correlation between CANA026 and Monthly Flows of 07EB001

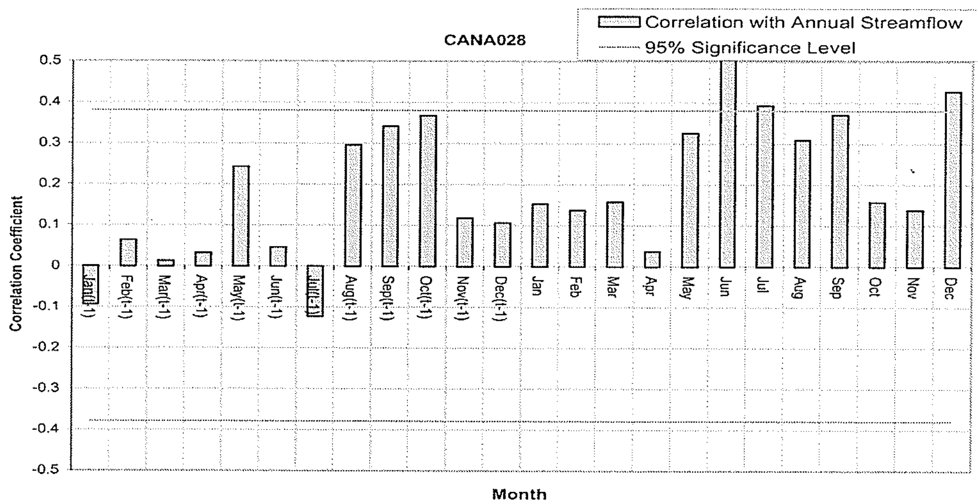


Figure C.4: Correlation between CANA028 and Monthly Flows of 07EB001

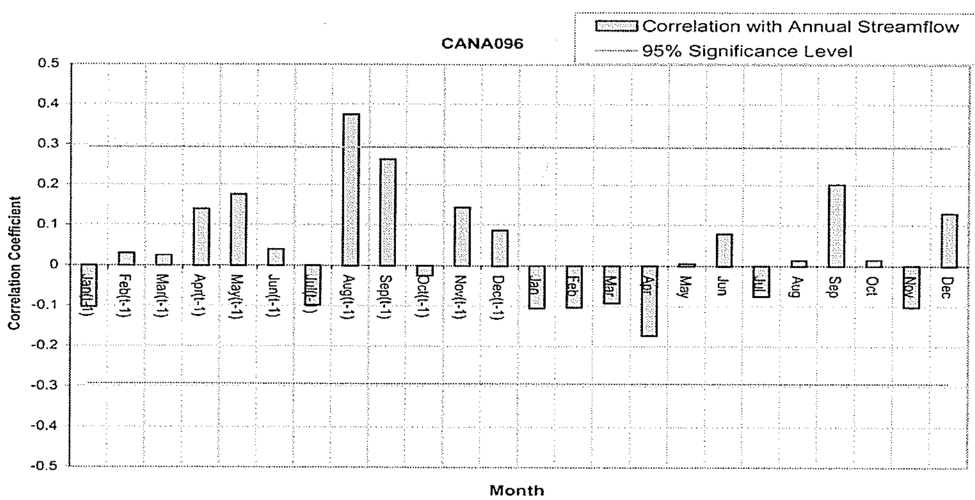


Figure C.5: Correlation between CANA096 and Monthly Flows of 07EB001

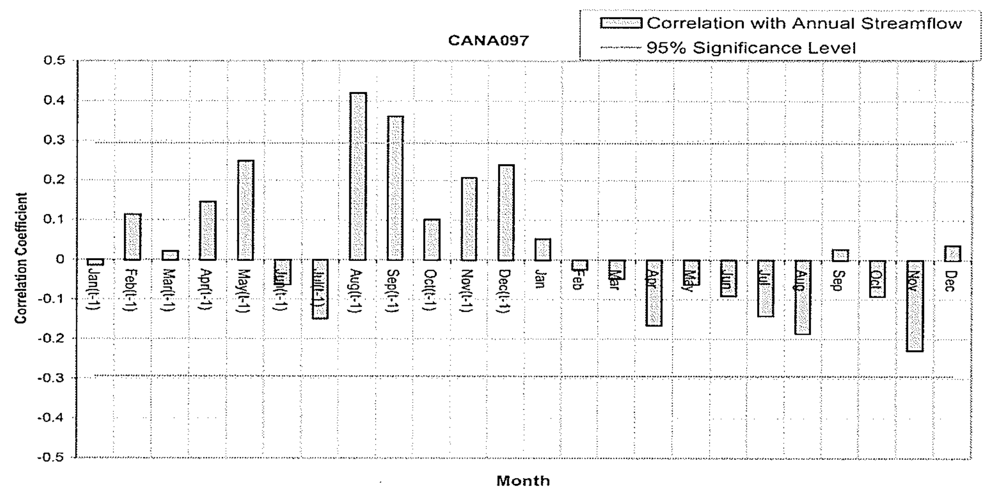


Figure C.6: Correlation between CANA097 and Monthly Flows of 07EB001

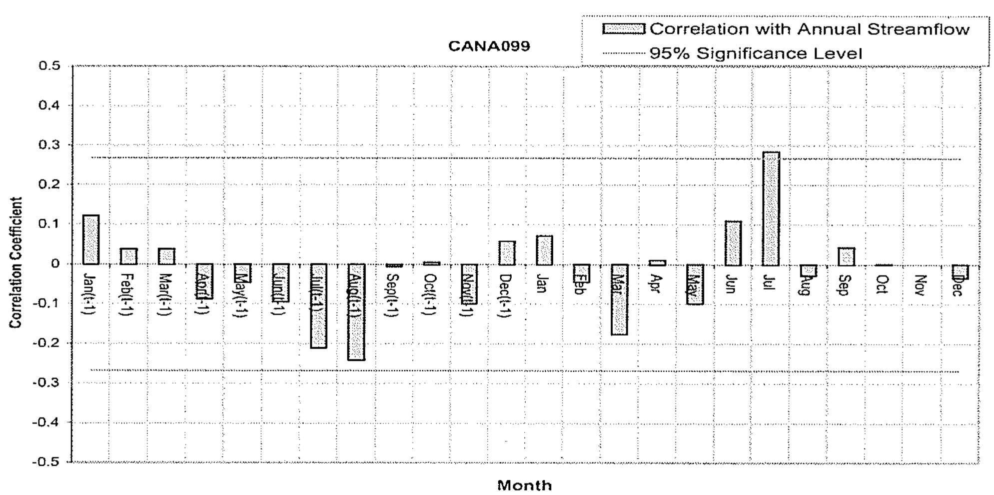


Figure C.7: Correlation between CANA099 and Monthly Flows of 07EB001

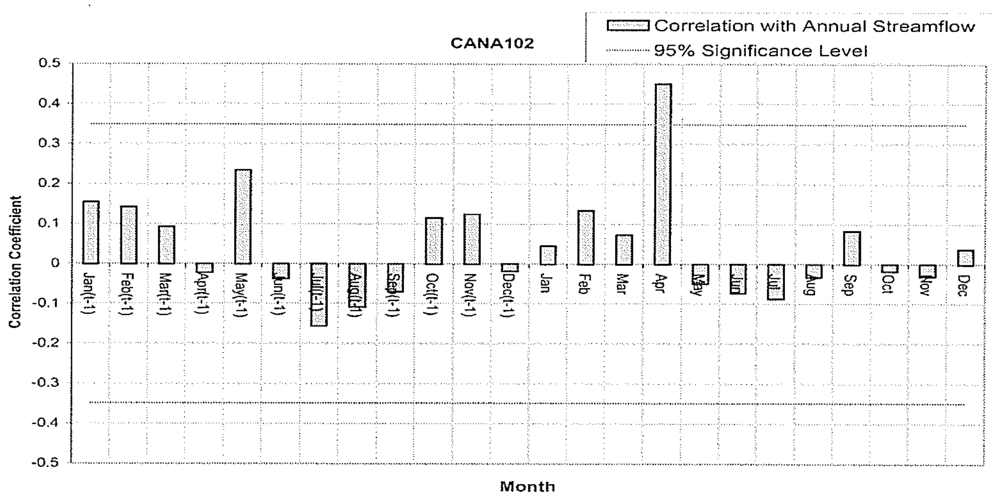


Figure C.8: Correlation between CANA102 and Monthly Flows of 07EB001

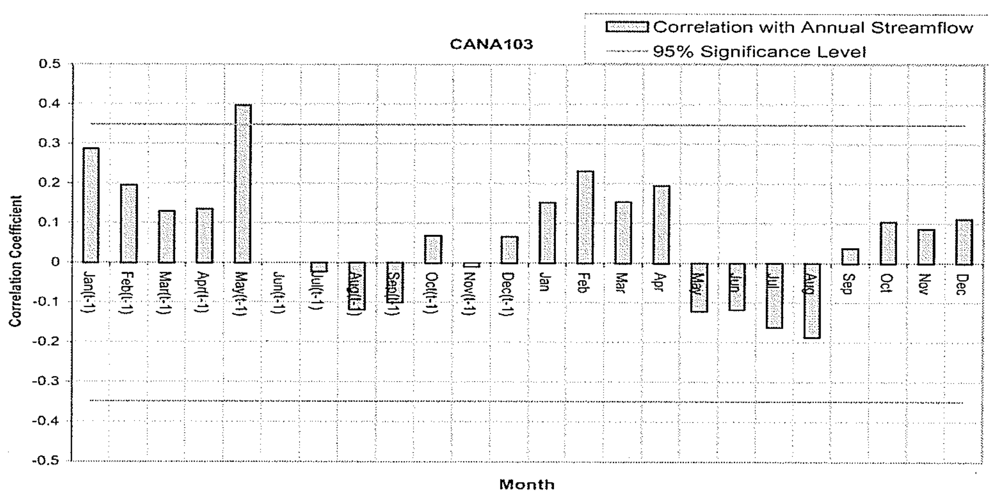


Figure C.9: Correlation between CANA103 and Monthly Flows of 07EB001

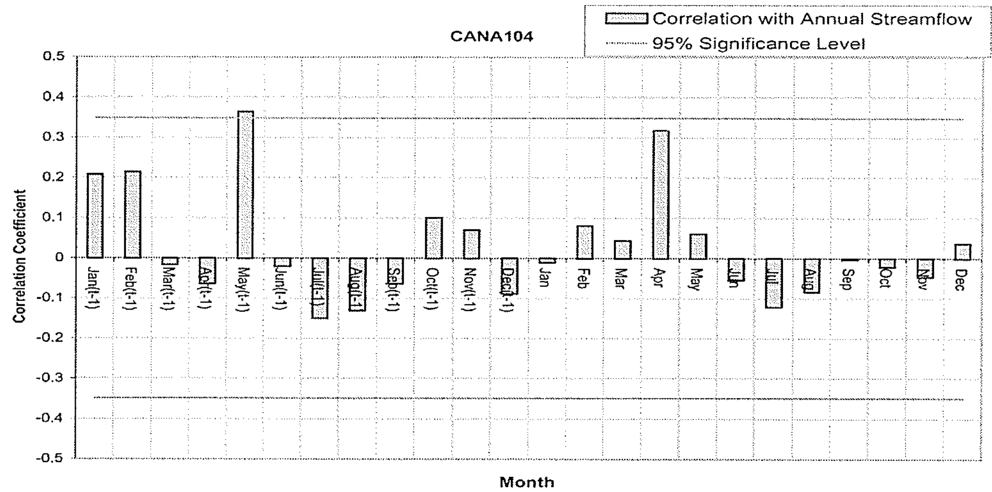


Figure C.10: Correlation between CANA104 and Monthly Flows of 07EB001

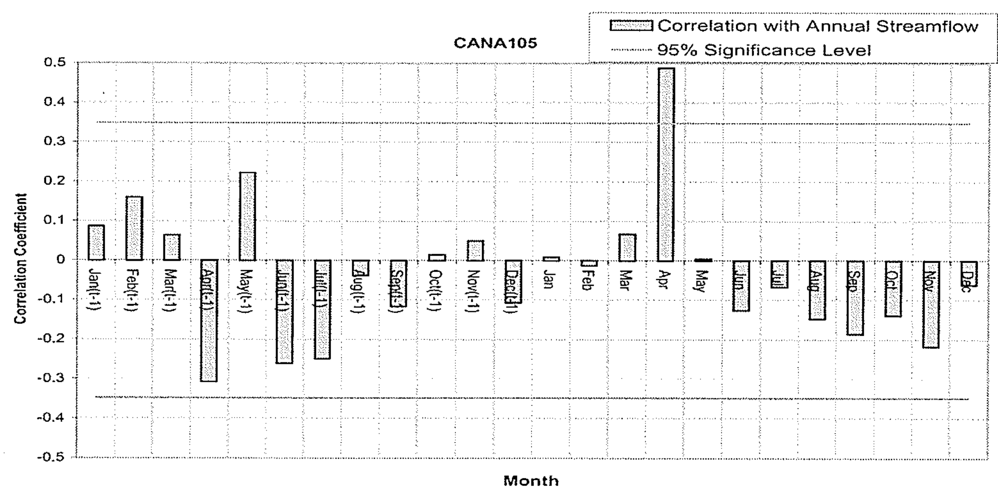


Figure C.11: Correlation between CANA105 and Monthly Flows of 07EB001

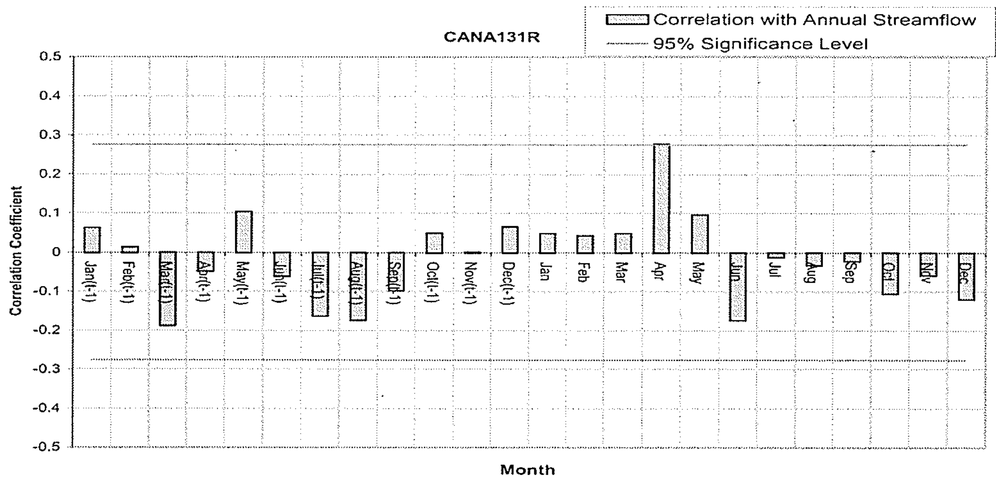


Figure C.12: Correlation between CANA131R and Monthly Flows of 07EB001

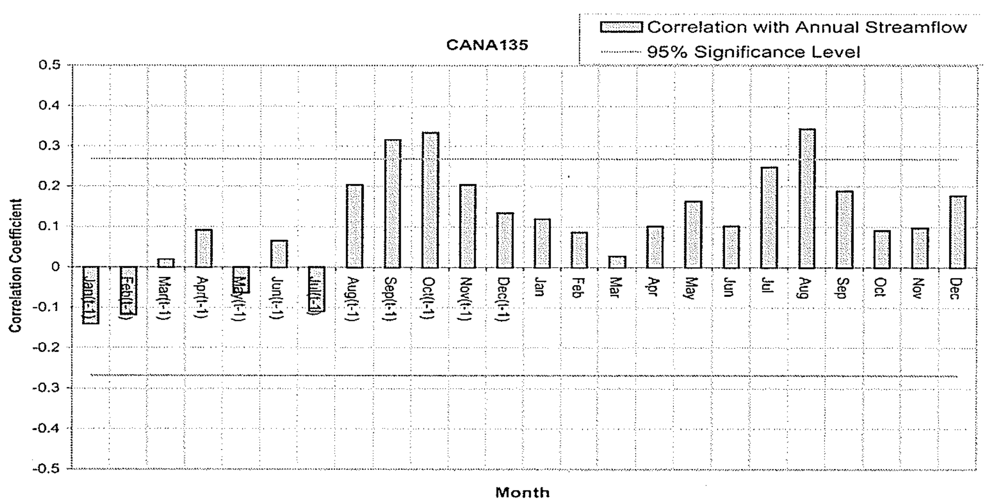


Figure C.13: Correlation between CANA135 and Monthly Flows of 07EB001

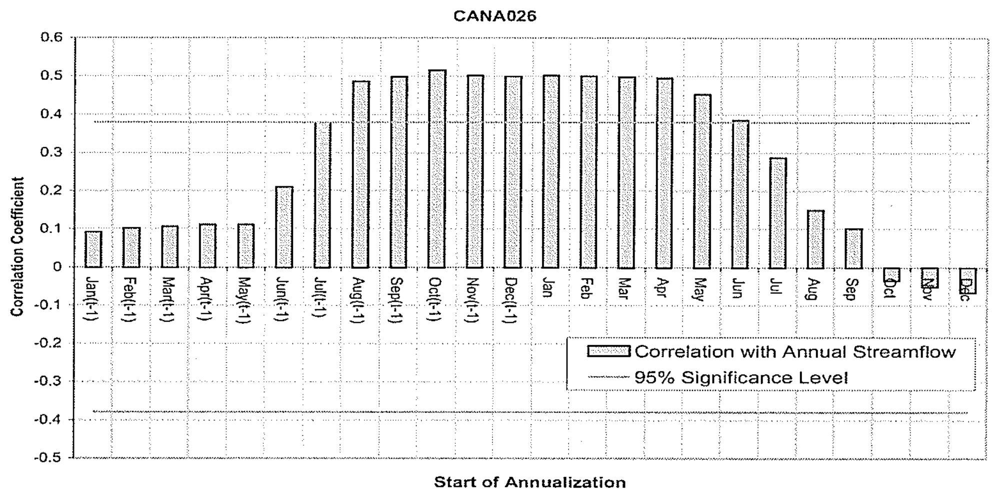


Figure C.14: Correlation between CANA026 and Annualized Flows of 07EB001

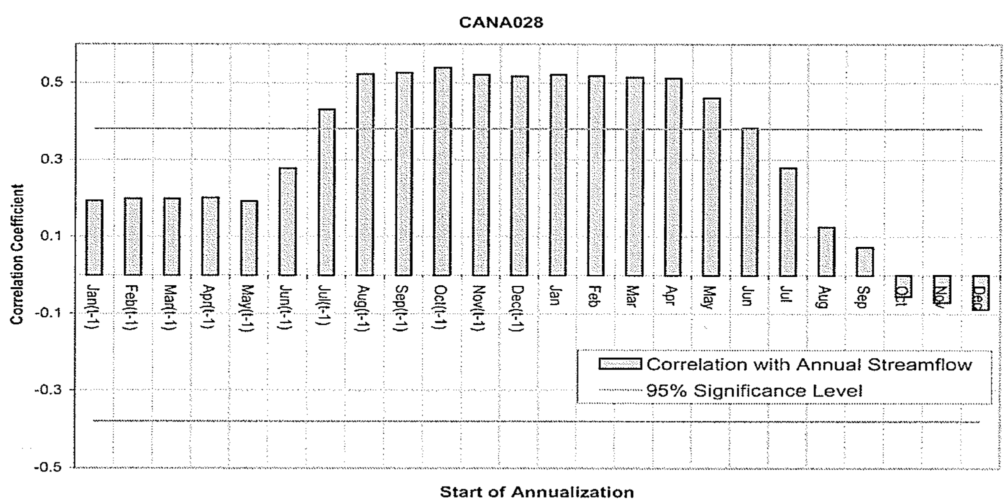


Figure C.15: Correlation between CANA028 and Annualized Flows of 07EB001

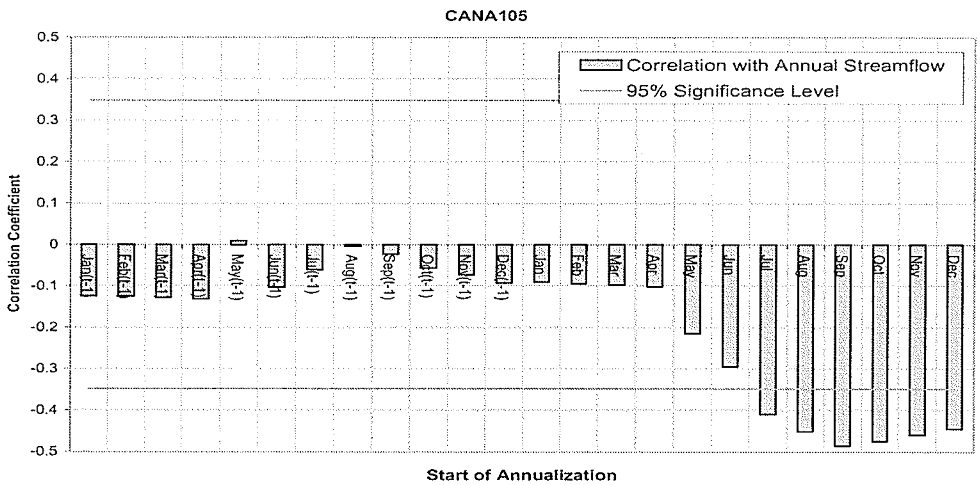


Figure C.16: Correlation between CANA105 and Annualized Flows of 07EB001

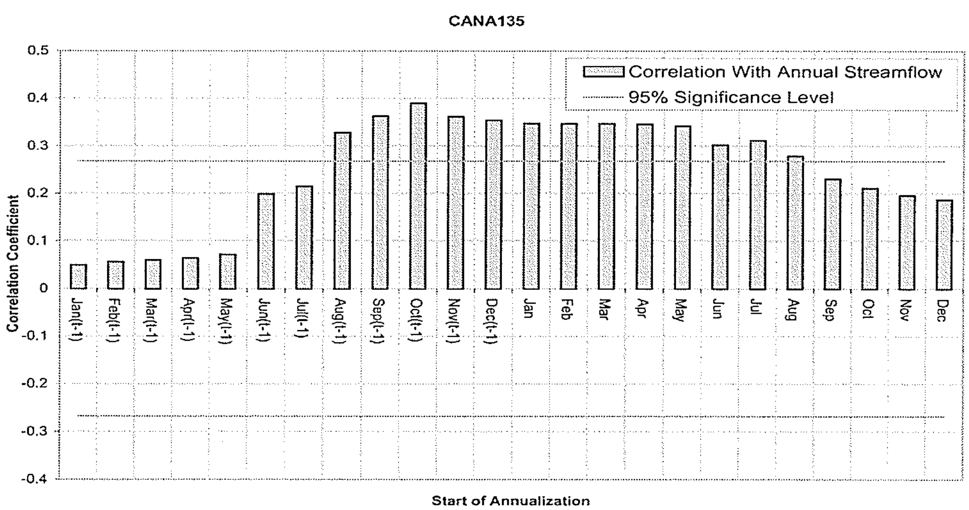


Figure C.17: Correlation between CANA135 and Annualized Flows of 07EB001

Appendix D

Correlation Analysis for 05AA023

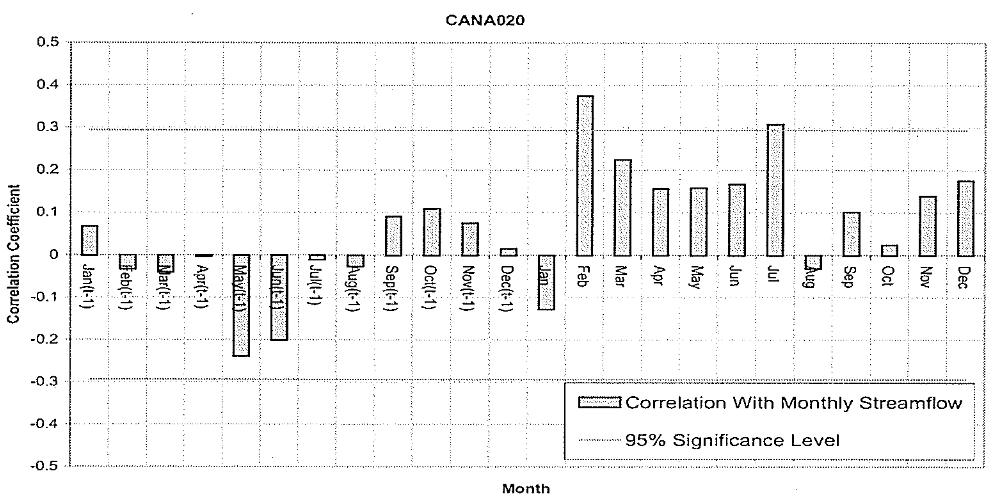


Figure D.1: Correlation between CANA020 and Monthly Flows of 05AA023

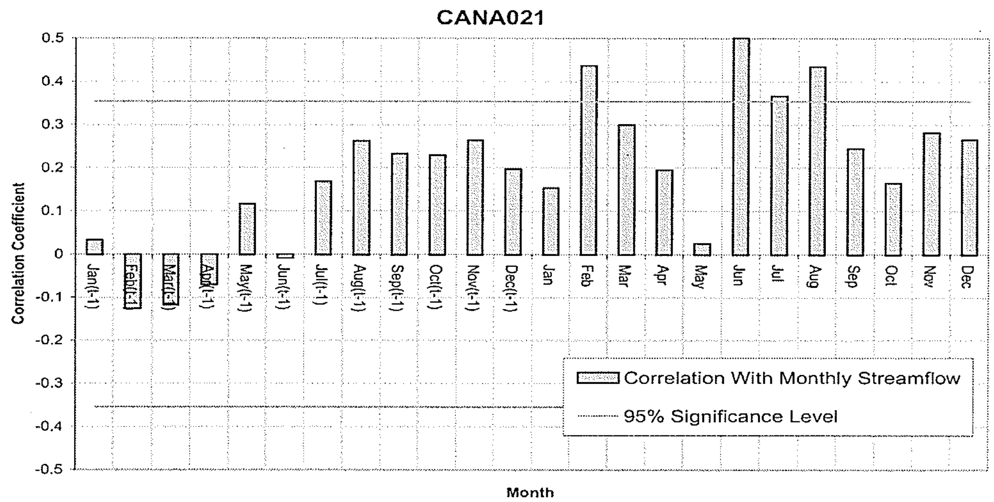


Figure D.2: Correlation between CANA021 and Monthly Flows of 05AA023

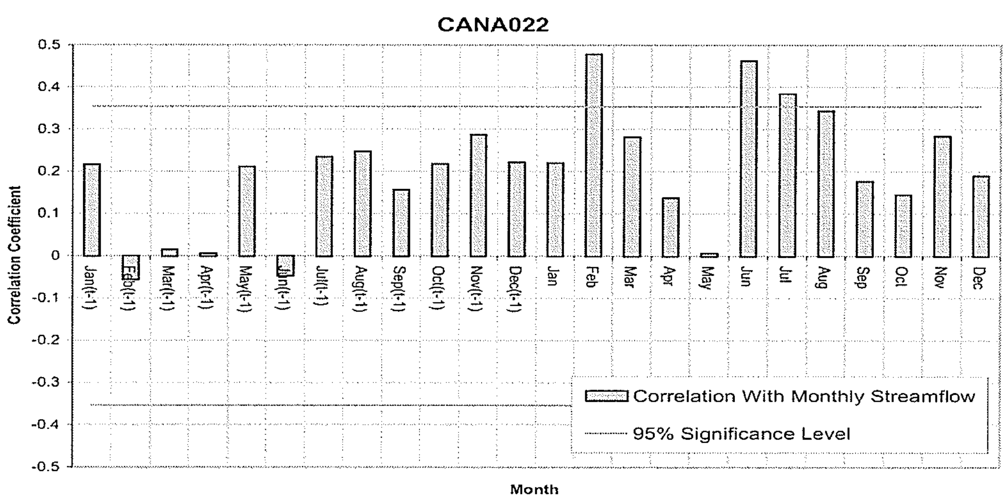


Figure D.3: Correlation between CANA022 and Monthly Flows of 05AA023

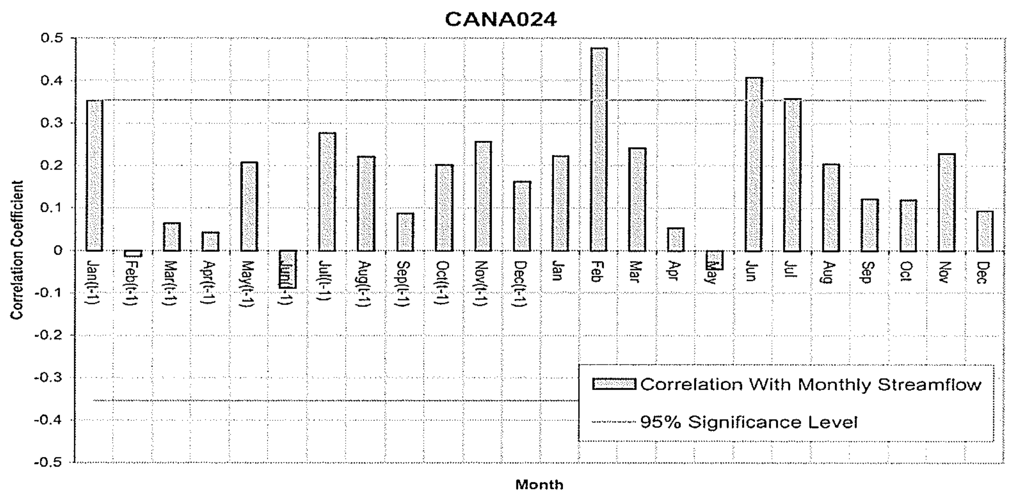


Figure D.4: Correlation between CANA024 and Monthly Flows of 05AA023

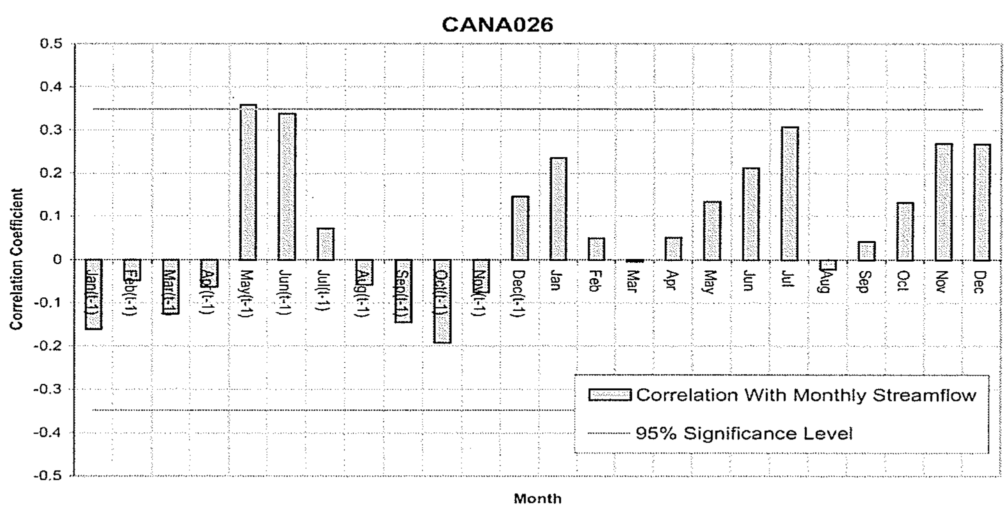


Figure D.5: Correlation between CANA026 and Monthly Flows of 05AA023

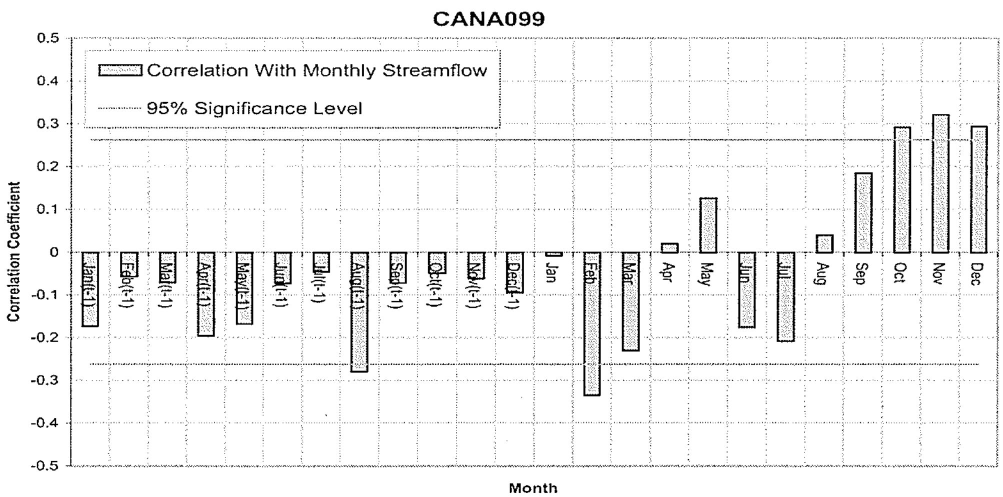


Figure D.6: Correlation between CANA099 and Monthly Flows of 05AA023

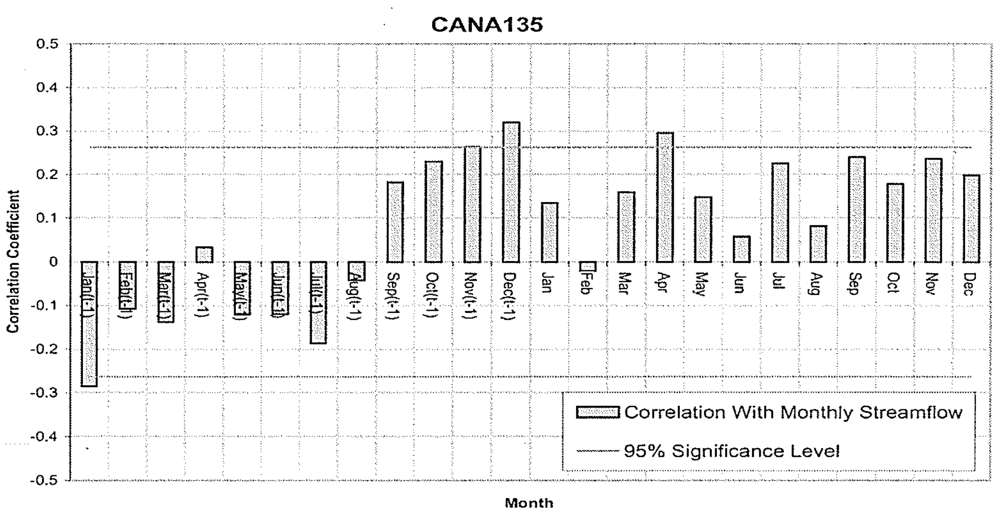


Figure D.7: Correlation between CANA135 and Monthly Flows of 05AA023

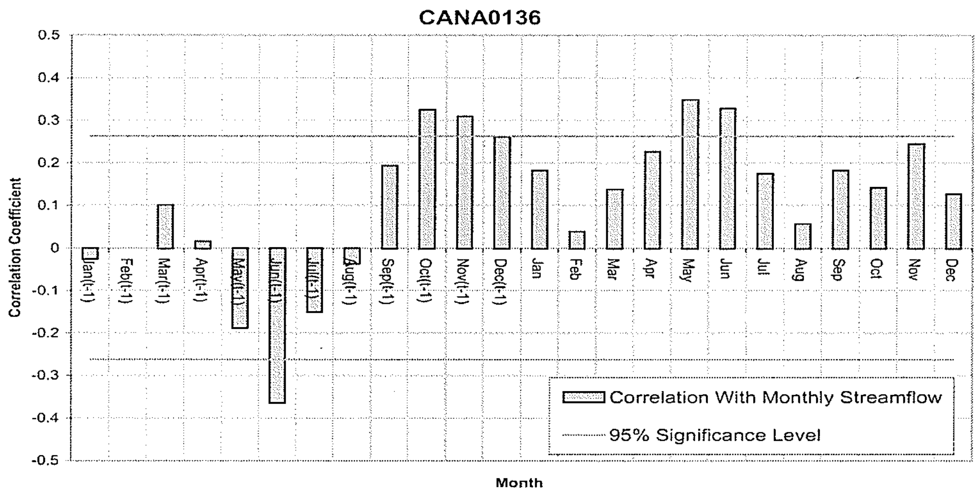


Figure D.8: Correlation between CANA136 and Monthly Flows of 05AA023

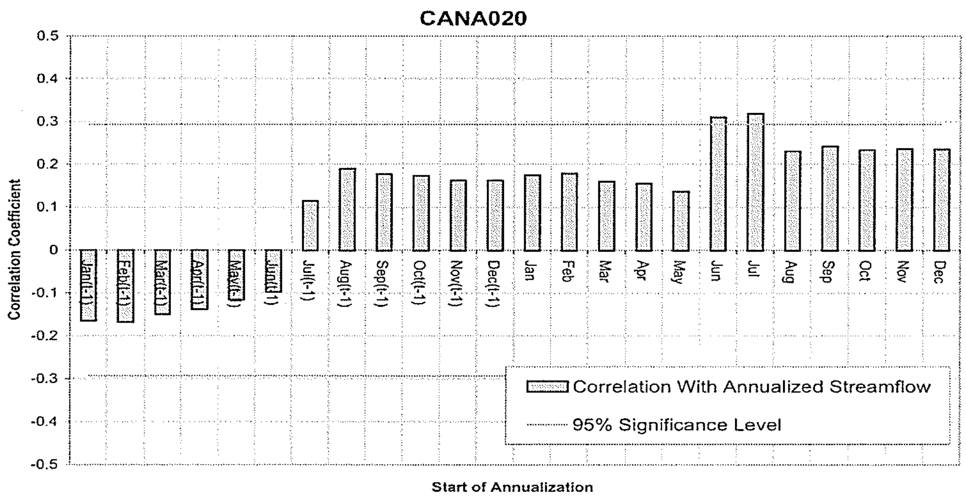


Figure D.9: Correlation between CANA020 and Annualized Flows of 05AA023

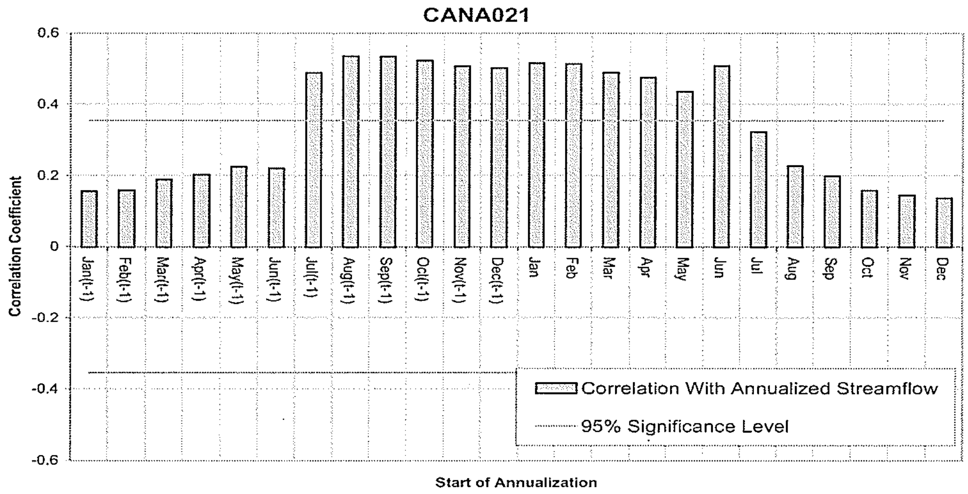


Figure D.10: Correlation between CANA021 and Annualized Flows of 05AA023

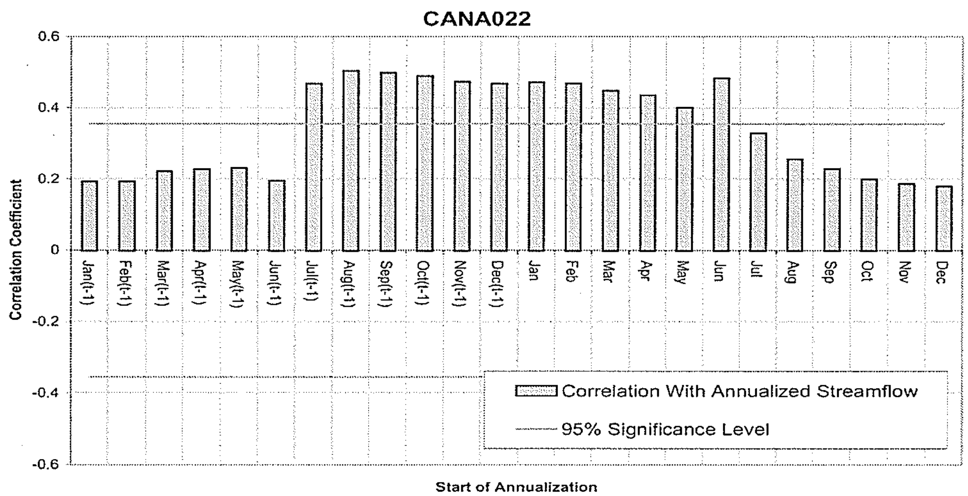


Figure D.11: Correlation between CANA022 and Annualized Flows of 05AA023

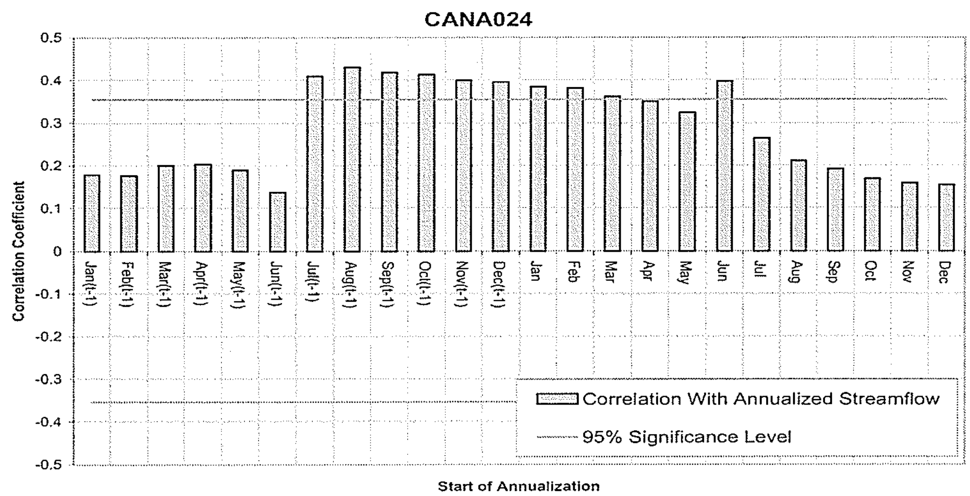


Figure D.12: Correlation between CANA024 and Annualized Flows of 05AA023

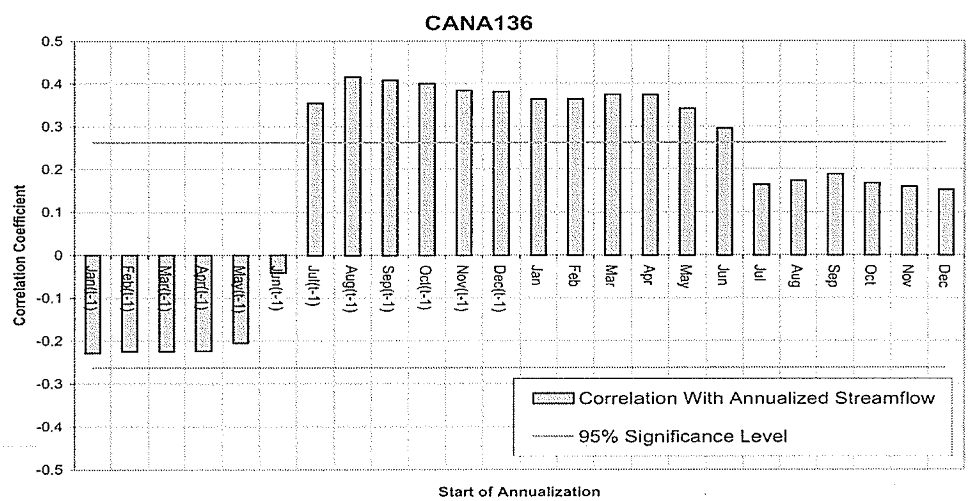


Figure D.13: Correlation between CANA136 and Flows of Annualized 05AA023

Appendix E

Correlation Analysis for 05AA022

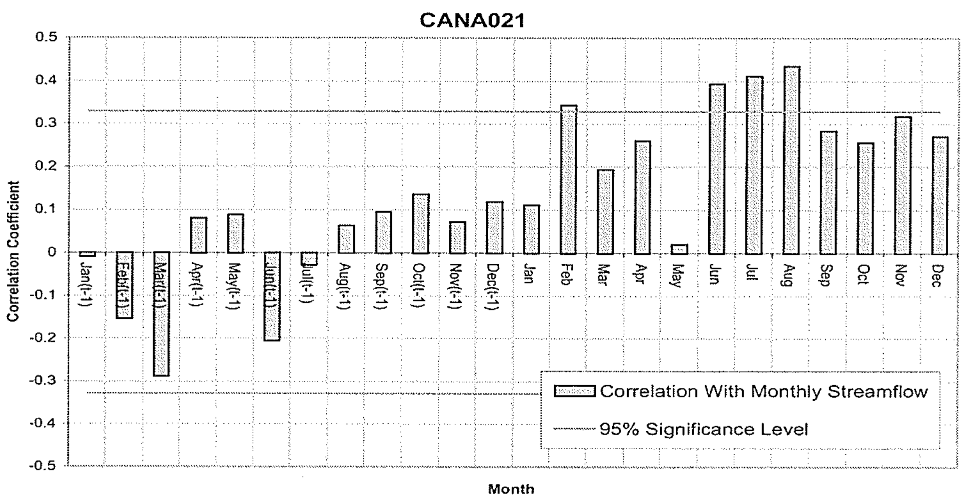


Figure E.1: Correlation between CANA021 and Monthly Flows of 05AA022

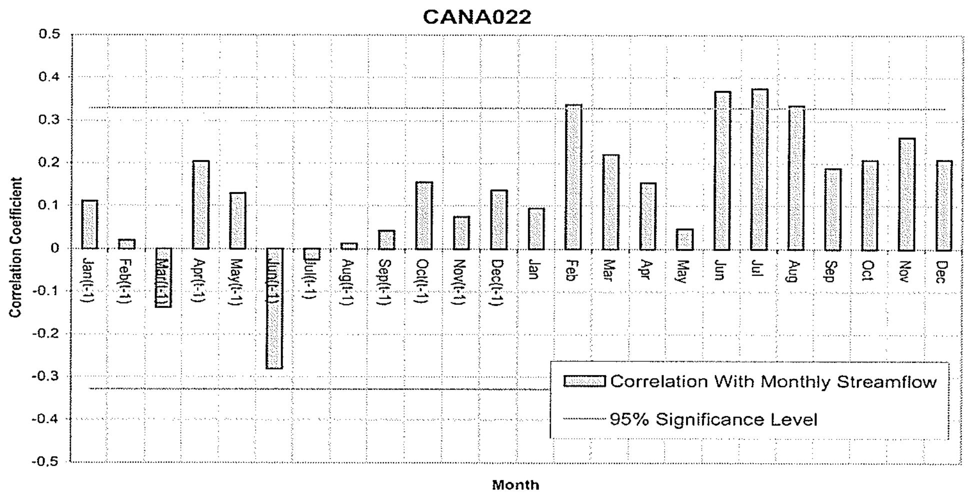


Figure E.2: Correlation between CANA022 and Monthly Flows of 05AA022

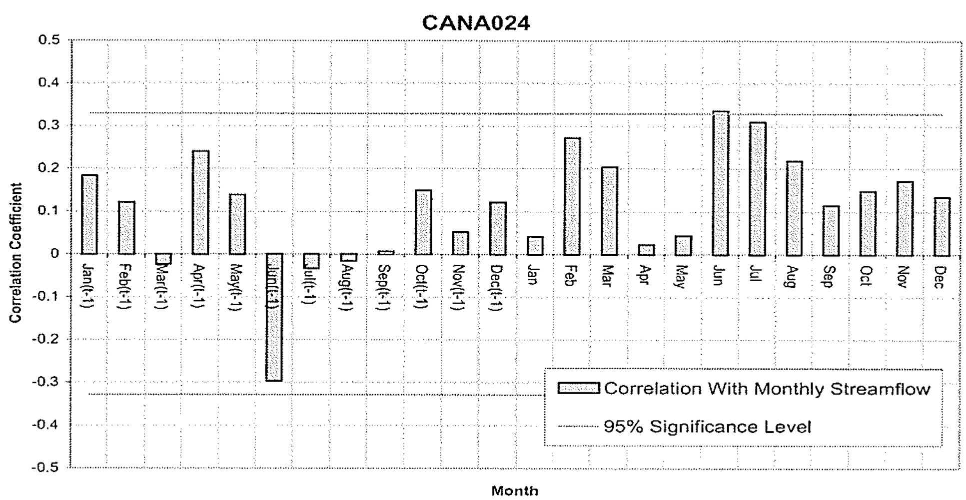


Figure E.3: Correlation between CANA024 and Monthly Flows of 05AA022

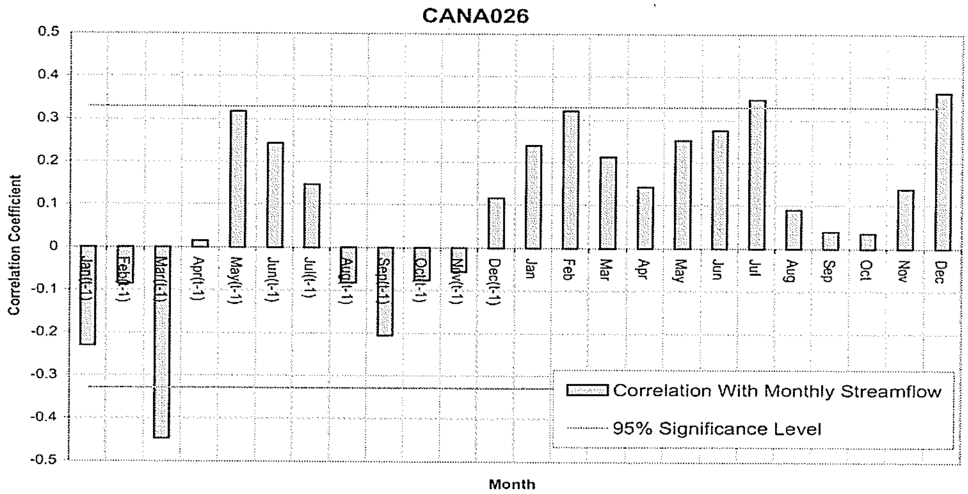


Figure E.4: Correlation between CANA026 and Monthly Flows of 05AA022

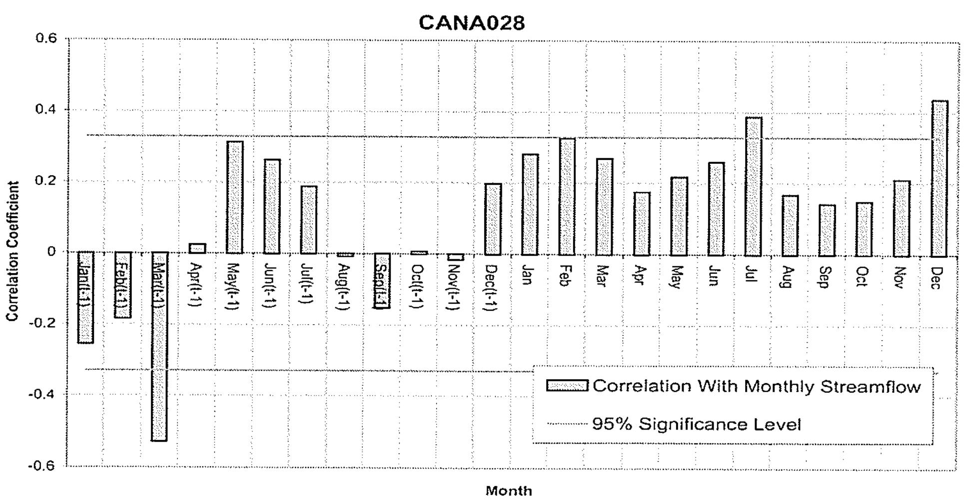


Figure E.5: Correlation between CANA028 and Monthly Flows of 05AA022

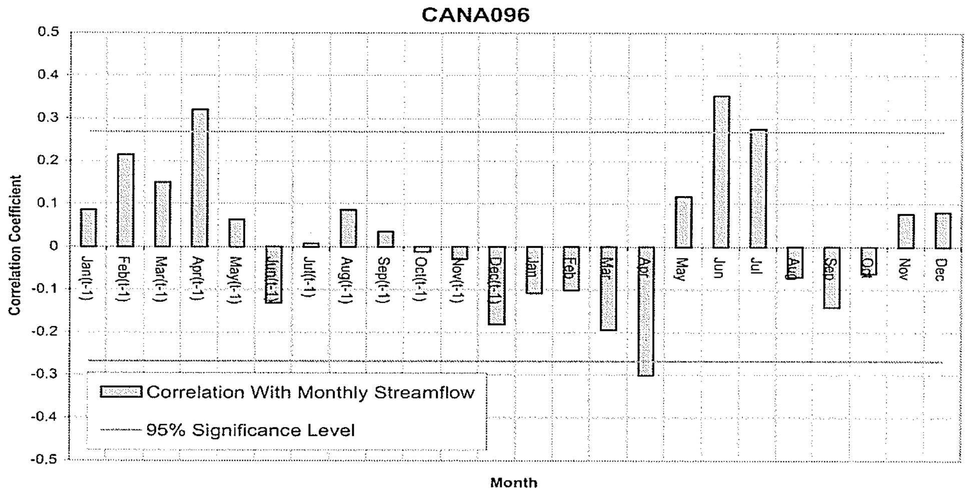


Figure E.6: Correlation between CANA096 and Monthly Flows of 05AA022

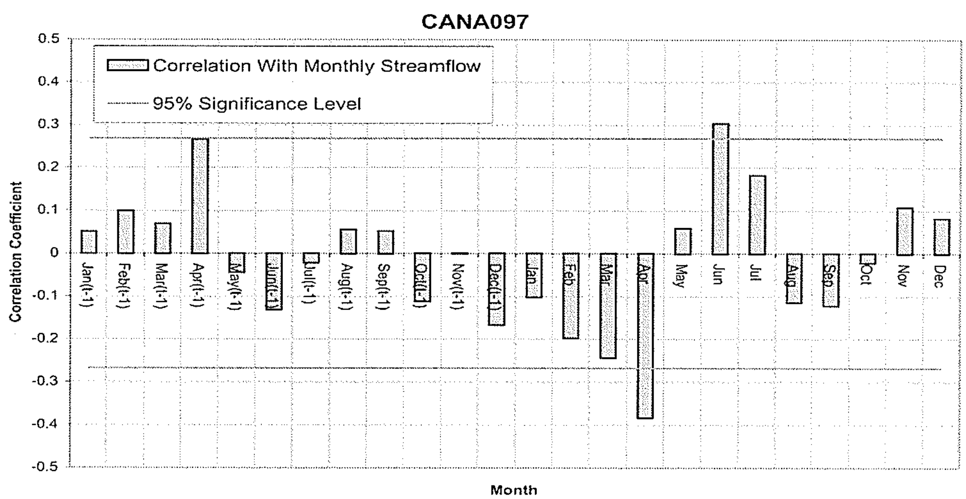


Figure E.7: Correlation between CANA097 and Monthly Flows of 05AA022

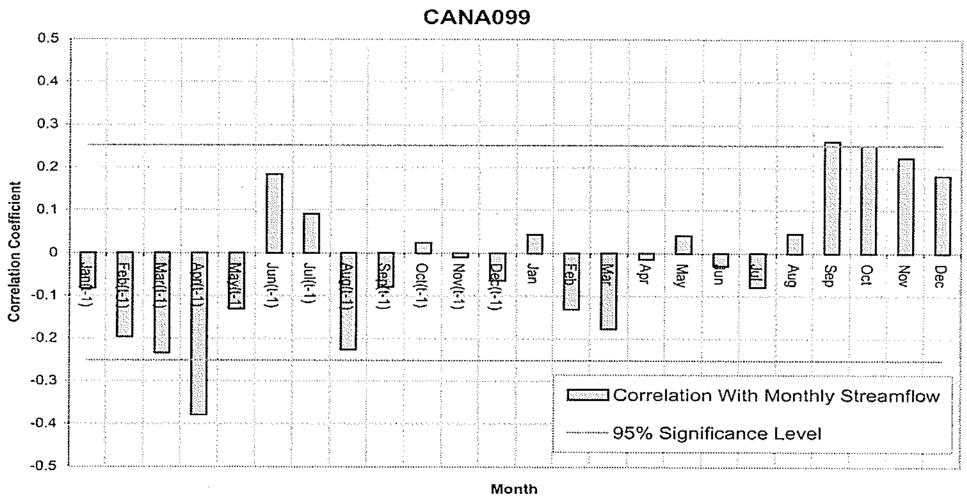


Figure E.8: Correlation between CANA099 and Monthly Flows of 05AA022

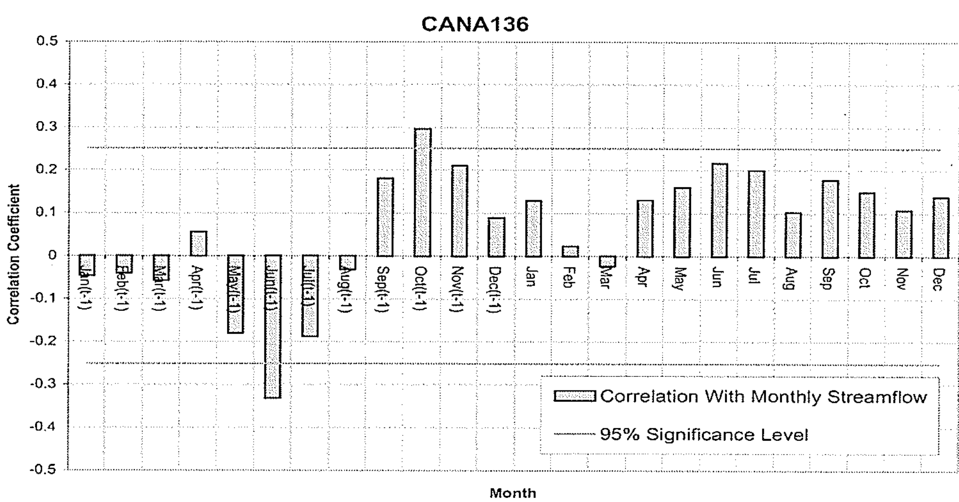


Figure E.9: Correlation between CANA136 and Monthly Flows of 05AA022

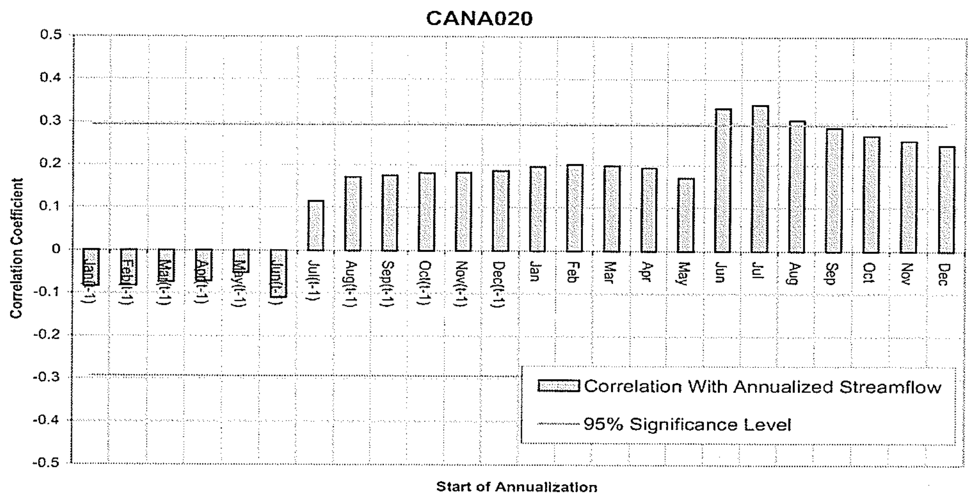


Figure E.10: Correlation between CANA020 and Annualized Flows of 05AA022

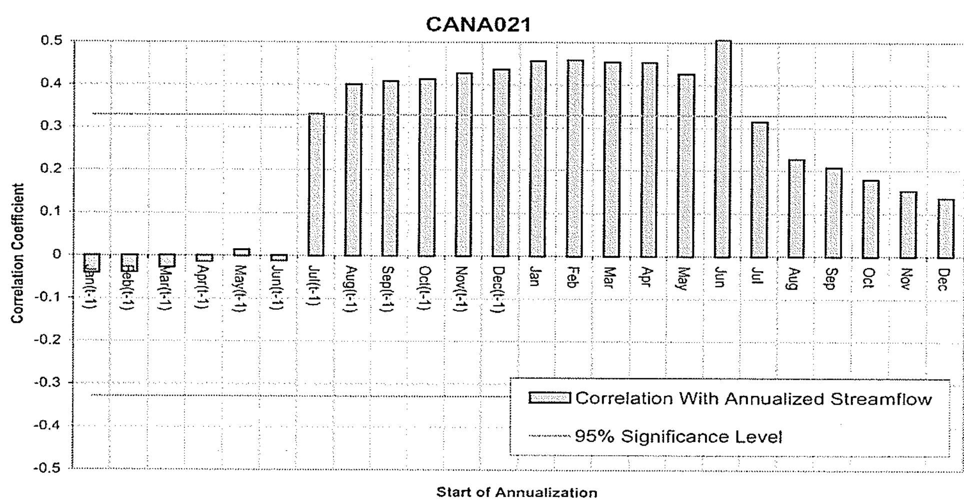


Figure E.11: Correlation between CANA021 and Annualized Flows of 05AA022

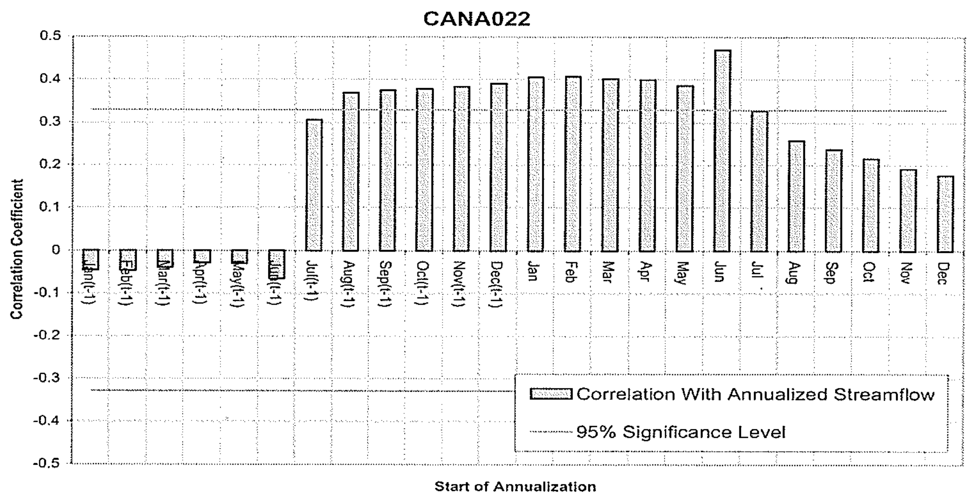


Figure E.12: Correlation between CANA022 and Annualized Flows of 05AA022

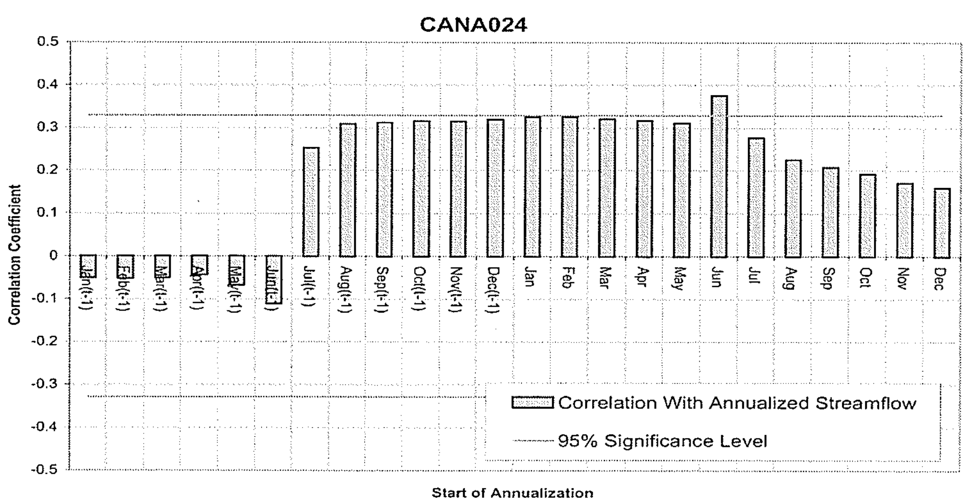


Figure E.13: Correlation between CANA024 and Annualized Flows of 05AA022

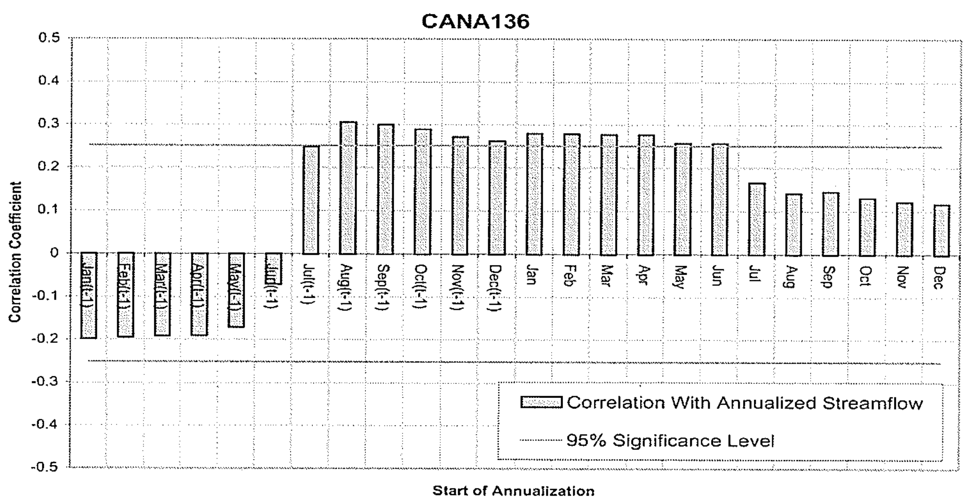


Figure E.14: Correlation between CANA136 and Flows of Annualized 05AA022

Appendix F

Gauged and Reconstructed Streamflow Record for 07BE001

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1805		15339
1806		11772
1807		12905
1808		14099
1809		13915
1810		13582
1811		10557
1812		13303
1813		9966
1814		13741
1815		11565
1816		13560
1817		14781
1818		13473
1819		14984
1820		13343
1821		11773
1822		13629
1823		11610
1824		14709
1825		14004
1826		13594
1827		16403
1828		16226
1829		10638
1830		12239
1831		10972
1832		14175
1833		13149
1834		14897
1835		10368
1836		12964
1837		12173
1838		16750
1839		13429
1840		11122
1841		11055
1842		12861
1843		12042
1844		15604
1845		11820
1846		12271
1847		13634
1848		11293
1849		14120
1850		10554
1851		11935
1852		14231
1853		13208
1854		15020

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1855		11906
1856		15453
1857		10376
1858		15737
1859		12126
1860		14137
1861		11093
1862		13681
1863		11399
1864		15640
1865		14309
1866		11893
1867		12175
1868		12830
1869		11359
1870		15142
1871		14608
1872		14940
1873		15081
1874		15186
1875		14421
1876		12752
1877		18755
1878		14084
1879		20402
1880		11803
1881		12866
1882		12381
1883		13689
1884		17029
1885		15515
1886		11865
1887		11213
1888		12455
1889		15096
1890		12107
1891		13477
1892		12068
1893		13871
1894		11345
1895		13394
1896		14144
1897		15350
1898		14429
1899		15330
1900		15382
1901		13327
1902		15501
1903		14285
1904		10189

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1905		13580
1906		14403
1907		15374
1908		12541
1909		13417
1910		12277
1911		15527
1912		13123
1913		13284
1914	13378	13182
1915	14734	14484
1916	11928	12813
1917	13347	10225
1918	11272	11797
1919	9416	13983
1920	16671	9726
1921	13602	13106
1922	9504	13114
1923	12524	12905
1924	13094	12424
1925	15415	13833
1926	12438	13407
1927	15661	12612
1928	14624	11759
1929	10911	11239
1930		13557
1931		12850
1932		13568
1933		12822
1934		11563
1935		12119
1936		12936
1937		13980
1938		10705
1939		12036
1940		14083
1941		13905
1942		13885
1943		12172
1944		15783
1945		14161
1946		14762
1947		15568
1948		12297
1949		16877
1950		13433
1951		15868
1952	12715	17231
1953	14596	14294
1954	23337	13478

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1955	13284	12139
1956	12811	11749
1957	12306	15762
1958	12122	13407
1959	10713	10000
1960	11738	14015
1961	10171	12353
1962	14242	13278
1963	13020	16635
1964	14041	13216
1965	21382	13216
1966	14776	
1967	11956	
1968	10267	
1969	12172	
1970	10628	
1971	18344	
1972	17136	
1973	13498	
1974	18038	
1975	11711	
1976	13263	
1977	17862	
1978	16784	
1979	14687	
1980	15683	
1981	11191	
1982	16399	
1983	11719	
1984	12645	
1985	12049	
1986	15976	
1987	10221	
1988	9472	
1989	15626	
1990	15122	
1991	15203	
1992	9700	
1993	9837	
1994	13491	

Appendix G

Gauged and Reconstructed Streamflow Record for 05AA023

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1571		576
1572		423
1573		681
1574		732
1575		459
1576		513
1577		437
1578		525
1579		459
1580		559
1581		578
1582		528
1583		680
1584		547
1585		592
1586		466
1587		796
1588		556
1589		670
1590		704
1591		574
1592		333
1593		483
1594		516
1595		499
1596		593
1597		590
1598		571
1599		416
1600		536
1601		475
1602		510
1603		451
1604		332
1605		488
1606		446
1607		608
1608		632
1609		515
1610		490
1611		563
1612		527
1613		525
1614		370
1615		500
1616		492
1617		395
1618		384
1619		429
1620		503

Year	Gauged Data (Mm^3)	Reconstructed Data (Mm^3)
1621		436
1622		305
1623		550
1624		497
1625		369
1626		371
1627		542
1628		397
1629		472
1630		688
1631		720
1632		537
1633		414
1634		268
1635		627
1636		421
1637		496
1638		465
1639		632
1640		637
1641		468
1642		576
1643		629
1644		460
1645		334
1646		572
1647		398
1648		641
1649		451
1650		407
1651		500
1652		533
1653		604
1654		517
1655		449
1656		245
1657		559
1658		566
1659		485
1660		295
1661		662
1662		431
1663		459
1664		664
1665		638
1666		600
1667		704
1668		778
1669		368
1670		604

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1671		752
1672		754
1673		288
1674		796
1675		597
1676		392
1677		538
1678		449
1679		677
1680		610
1681		475
1682		472
1683		292
1684		656
1685		533
1686		546
1687		502
1688		635
1689		561
1690		563
1691		334
1692		589
1693		621
1694		402
1695		497
1696		526
1697		691
1698		443
1699		524
1700		473
1701		532
1702		405
1703		596
1704		508
1705		130
1706		567
1707		527
1708		621
1709		520
1710		549
1711		547
1712		684
1713		610
1714		433
1715		485
1716		216
1717		333
1718		572
1719		316
1720		318

Year	Gauged Data (Mm^3)	Reconstructed Data (Mm^3)
1721		751
1722		689
1723		698
1724		409
1725		606
1726		653
1727		613
1728		435
1729		333
1730		622
1731		564
1732		704
1733		497
1734		594
1735		503
1736		546
1737		445
1738		584
1739		590
1740		370
1741		607
1742		543
1743		397
1744		672
1745		543
1746		420
1747		446
1748		544
1749		732
1750		366
1751		607
1752		632
1753		625
1754		508
1755		460
1756		407
1757		477
1758		240
1759		561
1760		629
1761		561
1762		545
1763		213
1764		583
1765		250
1766		702
1767		431
1768		568
1769		655
1770		541

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1771		299
1772		412
1773		682
1774		557
1775		410
1776		523
1777		648
1778		613
1779		511
1780		721
1781		557
1782		403
1783		471
1784		478
1785		682
1786		737
1787		453
1788		610
1789		349
1790		256
1791		656
1792		234
1793		309
1794		535
1795		340
1796		631
1797		444
1798		572
1799		243
1800		802
1801		529
1802		480
1803		615
1804		559
1805		594
1806		609
1807		342
1808		318
1809		694
1810		579
1811		427
1812		526
1813		402
1814		169
1815		759
1816		319
1817		474
1818		662
1819		522
1820		548

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1821		426
1822		529
1823		512
1824		605
1825		707
1826		620
1827		617
1828		724
1829		629
1830		182
1831		551
1832		595
1833		525
1834		526
1835		704
1836		492
1837		494
1838		798
1839		653
1840		477
1841		333
1842		445
1843		327
1844		637
1845		543
1846		315
1847		563
1848		451
1849		511
1850		538
1851		377
1852		597
1853		441
1854		676
1855		660
1856		448
1857		540
1858		378
1859		693
1860		432
1861		471
1862		174
1863		572
1864		393
1865		691
1866		400
1867		469
1868		374
1869		331
1870		486

Year	Gauged Data (Mm^3)	Reconstructed Data (Mm^3)
1871		690
1872		584
1873		594
1874		601
1875		471
1876		619
1877		634
1878		659
1879		650
1880		605
1881		306
1882		495
1883		423
1884		703
1885		679
1886		511
1887		465
1888		277
1889		677
1890		539
1891		350
1892		538
1893		484
1894		470
1895		331
1896		592
1897		799
1898		790
1899		665
1900		755
1901		712
1902		610
1903		748
1904		522
1905		319
1906		615
1907		719
1908		530
1909		260
1910		743
1911	460	506
1912	495	633
1913	401	532
1914	669	580
1915	765	683
1916	544	547
1917	342	339
1918	315	256
1919	506	544
1920	350	379

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1921	361	316
1922	663	568
1923	430	464
1924	436	535
1925	393	496
1926	774	669
1927	658	528
1928	444	561
1929	425	434
1930		408
1931		655
1932		477
1933		353
1934		531
1935		373
1936		327
1937		582
1938		356
1939		430
1940		550
1941		552
1942		652
1943		411
1944		616
1945		563
1946		706
1947		596
1948		368
1949		479
1950	970	545
1951	527	514
1952	833	645
1953	700	535
1954	587	571
1955	616	351
1956	452	495
1957	517	528
1958	637	654
1959	465	465
1960	514	424
1961	394	516
1962	584	617
1963	599	682
1964	590	598
1965	541	557
1966	786	
1967	505	
1968	650	
1969	431	
1970	563	

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1971	923	
1972	368	
1973	789	
1974	677	
1975	608	
1976	252	
1977	619	
1978	385	
1979	422	
1980	688	
1981	428	
1982	300	
1983	268	
1984	342	
1985	548	
1986	345	
1987	267	
1988	404	
1989	745	
1990	752	

Appendix H

Gauged and Reconstructed Streamflow Record for 05AA022

Year	Gauged Data (Mm^3)	Reconstructed Data (Mm^3)
1639		408
1640		491
1641		516
1642		442
1643		571
1644		538
1645		448
1646		387
1647		479
1648		358
1649		572
1650		434
1651		385
1652		515
1653		493
1654		587
1655		498
1656		428
1657		254
1658		545
1659		590
1660		526
1661		377
1662		574
1663		389
1664		483
1665		615
1666		530
1667		438
1668		571
1669		623
1670		361
1671		519
1672		707
1673		669
1674		339
1675		624
1676		566
1677		472
1678		482
1679		418
1680		573
1681		481
1682		400
1683		310
1684		298
1685		561
1686		441
1687		417
1688		452

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1689		601
1690		571
1691		586
1692		373
1693		525
1694		490
1695		403
1696		499
1697		527
1698		653
1699		451
1700		496
1701		395
1702		420
1703		345
1704		534
1705		463
1706		216
1707		499
1708		436
1709		557
1710		514
1711		559
1712		552
1713		656
1714		552
1715		376
1716		422
1717		314
1718		453
1719		548
1720		307
1721		387
1722		676
1723		516
1724		498
1725		386
1726		530
1727		506
1728		539
1729		455
1730		441
1731		497
1732		407
1733		597
1734		524
1735		603
1736		548
1737		484
1738		449

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1739		587
1740		557
1741		392
1742		542
1743		459
1744		366
1745		589
1746		392
1747		426
1748		483
1749		530
1750		560
1751		292
1752		581
1753		578
1754		559
1755		432
1756		442
1757		372
1758		443
1759		245
1760		552
1761		516
1762		536
1763		533
1764		382
1765		586
1766		279
1767		663
1768		450
1769		538
1770		559
1771		482
1772		312
1773		437
1774		568
1775		514
1776		411
1777		483
1778		552
1779		474
1780		441
1781		652
1782		515
1783		467
1784		523
1785		493
1786		612
1787		702
1788		534

Year	Gauged Data (Mm^3)	Reconstructed Data (Mm^3)
1789		647
1790		371
1791		336
1792		554
1793		227
1794		376
1795		493
1796		370
1797		523
1798		385
1799		471
1800		229
1801		733
1802		466
1803		471
1804		571
1805		535
1806		563
1807		591
1808		422
1809		417
1810		624
1811		510
1812		467
1813		472
1814		311
1815		211
1816		646
1817		332
1818		477
1819		614
1820		533
1821		523
1822		472
1823		508
1824		416
1825		452
1826		578
1827		513
1828		486
1829		617
1830		667
1831		326
1832		488
1833		409
1834		466
1835		524
1836		607
1837		309
1838		327

Year	Gauged Data (Mm^3)	Reconstructed Data (Mm^3)
1839		632
1840		576
1841		420
1842		342
1843		437
1844		325
1845		604
1846		541
1847		369
1848		542
1849		435
1850		497
1851		488
1852		409
1853		430
1854		358
1855		660
1856		549
1857		471
1858		499
1859		395
1860		551
1861		366
1862		438
1863		359
1864		558
1865		377
1866		615
1867		377
1868		500
1869		380
1870		369
1871		497
1872		588
1873		488
1874		555
1875		622
1876		489
1877		598
1878		508
1879		588
1880		602
1881		560
1882		435
1883		574
1884		398
1885		535
1886		560
1887		508
1888		494

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1889		246
1890		557
1891		419
1892		391
1893		498
1894		410
1895		447
1896		383
1897		569
1898		627
1899		539
1900		530
1901		640
1902		605
1903		527
1904		667
1905		447
1906		323
1907		496
1908		617
1909		472
1910		396
1911	641	666
1912	382	441
1913	449	509
1914	451	485
1915	451	459
1916	535	518
1917	445	454
1918	352	479
1919	247	298
1920	462	476
1921	380	280
1922	431	347
1923	597	504
1924	448	409
1925	301	468
1926	365	472
1927	824	525
1928	456	468
1929	397	438
1930		362
1931		355
1932		552
1933		479
1934		421
1935		566
1936		404
1937		363
1938		527

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1939		317
1940		438
1941		420
1942		449
1943		497
1944		433
1945	527	554
1946	542	429
1947	637	532
1948	602	543
1949	383	466
1950	736	587
1951	747	576
1952	371	471
1953	679	567
1954	585	615
1955	592	683
1956	514	413
1957	454	458
1958	440	497
1959	646	594
1960	467	489
1961	478	421
1962	344	412
1963	448	505
1964	611	625
1965	604	593
1966	496	
1967	595	
1968	567	
1969	453	
1970	514	
1971	554	
1972	555	
1973	419	
1974	600	
1975	795	
1976	345	
1977	265	
1978	451	
1979	399	
1980	488	
1981	380	
1982	461	
1983	329	
1984	370	
1985	510	
1986	436	
1987	300	
1988	303	

Year	Gauged Data	Reconstructed Data
	(Mm^3)	(Mm^3)
1989	541	
1990	594	