NONLINEAR TIME SERIES MODELING OF SOME CANADIAN RIVER FLOW DATA

CENTRE FOR NEWFOUNDLAND STUDIES

TOTAL OF 10 PAGES ONLY MAY BE XEROXED

(Without Author's Permission)

DOUGLAS JAMES BATTEN

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-54860-0

Canada

NONLINEAR TIME SERIES MODELING OF SOME CANADIAN RIVER FLOW DATA

by

Douglas James Batten

A Practicum report submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Applied Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

January, 2000

St. John's, Newfoundland

Abstract

In hydrology the ability to model the average daily river flow for rivers plays an important role in the prediction of possible disasters such as flooding. The analysis of data and the accuracy of predictions rely on fitting suitable models to such data. In this practicum we investigate nonlinear time series modeling and in particular we study the theory of two approaches to model such time series. One approach assumes the underlying random structure of the time series is bilinear. The second approach uses wavelet smoothing techniques to decompose the time series into a wavelet smoothed component and a random component. The random component is then modeled by a suitable linear or bilinear process. By investigating the structure of the autocorrelation and third order cumulants, we find that the pure bilinear process is best for the data sets under study. Models were fitted to six time series data sets based on the average daily river flow variable for six rivers in Canada using both approaches. A simulation study was conducted to establish the suitability of the models by comparing its performance to the original time series. The bilinear approach was not favorable in modeling average daily river flow. However, the wavelet methodology illustrated an attractive technique to model such a time series.

Acknowledgments

I wish to thank my supervisor, Dr. Alwell Oyet (Department of Mathematics and Statistics, Memorial University of Newfoundland) for helpful comments and suggestions about my practicum. I wish to thank Dr. Leonard Lye (Faculty of Engineering and Applied Science, Memorial University of Newfoundland) for the data and background material for this research.

I acknowledge the Department of Mathematics and Statistics for the financial support in the form of Teaching Assistantships and Sessional Lecturer.

I also wish to thank my parents for their continuous encouragement and support throughout my university career.

Finally, I am grateful of the support from my twin brother Dennis who has been with me every step of the way.

Contents

Abstract i

Acknowledgments ii

List of Tables vi

List of Figures viii

1 Preliminaries 1

1.1 Introduction 1

1.2 Bilinear Models 4

1.3 Stationarity and Invertibility Conditions 5

1.3.1 Stationarity 6

1.3.2 Invertibility 8

2 The River Flow Data 10

2.1 Introduction 10

2.2 Data 11

2.3 Tests for Linearity 12

2.3.1 Approach Based On Examining Squares Of Time Series Data . . . 13
2.3.2 Approach Based On Tukey's One-Degree-Of-Freedom Test For Non-Additivity . . . 14
2.3.3 Tsay's Approach Based On Column Stacking . . . 15
2.4 Results . . . 17

3 Bilinear Time Series Model **21**
3.1 Introduction . . . 21
3.2 Estimation of the Parameters of the Bilinear Time Series Model 22
3.2.1 Newton-Raphson Iterative Technique . . . 23
3.2.2 Partial Derivatives . . . 24
3.2.3 Initial Estimates . . . 26
3.2.4 Model Selection . . . 27
3.3 Model Fitting to River Flow Data . . . 28
3.4 Results of Simulation Study . . . 32

4 Wavelet Filtering **41**
4.1 Introduction . . . 41
4.2 Wavelet Smoothing . . . 44
4.3 Estimation . . . 46
4.4 Model Identification Based on Third Order Moments and Cumulants . . . 48
4.4.1 Results . . . 50
4.5 Model Fitting to River Flow Data . . . 55
4.6 Results of Simulation Study . . . 58

5 Conclusion 69

Bibliography 71

A Data Sets 74

B Time Series Plots 87

List of Tables

2.1 Computed F-statistics and p-values for Peace River 18
2.2 Computed F-statistics and p-values for Castle River 18
2.3 Computed F-statistics and p-values for South River 19
2.4 Computed F-statistics and p-values for Salmonier River 19
2.5 Computed F-statistics and p-values for Gander River 20
2.6 Computed F-statistics and p-values for Moberly River 20

3.1 Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Peace River) 33
3.2 Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (South River) 34
3.3 Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Salmonier River) 36
3.4 Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Gander River) 37
3.5 Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Moberly River) 38

4.1 $C(1, k_2)$ Pattern for arbitrary q 49

4.2 Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Peace River) 59
4.3 Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Castle River) 60
4.4 Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (South River) 62
4.5 Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Salmonier River) 63
4.6 Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Gander River) 65
4.7 Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Moberly River) 66

A.1 River Flow Data For Peace River 75
A.2 River Flow Data For Castle River 77
A.3 River Flow Data For South River 79
A.4 River Flow Data For Salmonier River 81
A.5 River Flow Data For Gander River 83
A.6 River Flow Data For Moberly River 85

List of Figures

3.1 Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Peace River 34

3.2 Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for South River 35

3.3 Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Salmonier River . . . 36

3.4 Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Gander River 38

3.5 Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Moberly River 39

4.1 Standardized cumulant trace of Peace River 51

4.2 Standardized cumulant trace of Castle River 52

4.3 Standardized cumulant trace of South River 53

4.4 Standardized cumulant trace of Salmonier River 53

4.5 Standardized cumulant trace of Gander River 54

4.6 Standardized cumulant trace of Moberly River 55

4.7 Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Peace River 60
4.8 Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Castle River 61
4.9 Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for South River 63
4.10 Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Salmonier River 64
4.11 Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Gander River 66
4.12 Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Moberly River 67

B.1 Time series plot for Peace River 88
B.2 Time series plot for Castle River 89
B.3 Time series plot for South River 90
B.4 Time series plot for Salmonier River 91
B.5 Time series plot for Gander River 92
B.6 Time series plot for Moberly River 93

Chapter 1

Preliminaries

1.1 Introduction

A time series consists of observations that are collected in time. Some examples of a time series are daily temperatures in St. John's from 1980 to 1999; monthly kilometers flown by an airline company; and company profits in successive years. In most instances, the time series is considered linear and with the theory of linear models well defined along with its easy interpretation, it is not uncommon to fit linear models to most data. A typical example is the maximum daily or yearly flow of Canadian rivers. Hydrologists have used linear models in describing most of these flow data. Using tests for linearity, we show in this work that some of these flow data are generated by non-linear processes.

Although, in many situations, linear models do provide adequate approximations to the 'true' process generating the data, it may not necessarily be the best model. A series may be generated by an underlying random struc-

ture that is non-linear, and perhaps non-Gaussian. When considering such a series it would be insitefull to consider a non-linear approach. This has prompted the development of many non-linear time series models that have the capability to deal with such series. Some nonlinear models currently in use are:

a. Bilinear Time Series Models: Since, in this practicum, we are restricting our work to bilinear models the structure of this model will be given in Section 1.2.

b. Threshold Models: There are several groups of threshold models and one such group is known as the piecewise linear models. The general structure of this group of threshold models described by Tong (1990) is as follows: Let $\mathbf{X_t}$ be a k-dimensional time series and, for each t, let J_t be an indicator random variable, taking integer values $\{1, 2, \ldots, l\}$. The model is then given by

$$\mathbf{X_t} = \mathbf{B}^{(J_t)}\mathbf{X}_t + \mathbf{A}^{(J_t)}\mathbf{X}_{t-1} + \mathbf{H}^{(J_t)}\epsilon_t + \mathbf{c}^{(J_t)} \tag{1.1}$$

where, for $J_t = j$, $\mathbf{A}^{(j)}$ and $\mathbf{H}^{(j)}$ are coefficient matrices of size $k \times k$, $\mathbf{C}^{(j)}$ is a $k \times 1$ vector of constants, and ϵ_j with zero mean and a covariance matrix.

c. Fractional Autoregressive Models: To illustrate this group of models we will give the structure of the fractional autoregressive models of order one and it will be clear to see how this model may be extended to include higher-order lags. The structure described by several authors,

see Tong (1990), is

$$X_t = \frac{a_0 + \sum_{j=1}^{p} a_j X_{t-1}^j}{b_0 + \sum_{j=1}^{q} b_j X_{t-1}^j} + \epsilon_t \tag{1.2}$$

where ϵ_t is a sequence of independently and identically distributed random variables, $0 \leq p \leq q+1 < \infty$, $a_p \neq 0$ and $b_q \neq 0$.

The focus of this practicum is directed towards fitting models to river flow data consisting of daily averages. The ability to model daily flow data would be an enormous asset in hydrology. The Average daily flow variable is a variable that is used for the assessment of water supply reliability and for study of inflows into reservoirs. If a daily flow variable is successfully modeled then the storage size required of a reservoir can be determined.

This practicum is organized in the following way. In Chapter 1 we will introduce some concepts that will be essential in understanding the chapters that follow in this practicum. Section 1.2 discusses the bilinear model and describes the components which make up the model. Section 1.3 will outline the stationarity and invertibility conditions of the bilinear time series model.

The data used throughout this practicum will be discussed in Chapter 2 along with linearity tests of the data. In Chapter 3 we will discuss the estimation procedure for fitting bilinear models; fit bilinear models to the Canadian river flow data; and use these models to simulate our own data. Chapter 4 will consist of wavelet smoothing to attempt to make the time series simpler. We will then fit appropriate models to the new time series and simulate data based on the new models. Finally, the two methods used in this practicum will be compared in Chapter 5 along with a general discussion

of the results.

1.2 Bilinear Models

Non-linear models often involve complex calculations and are, at times, very difficult to analyze and interpret. The bilinear time series model studied by Granger and Andersen (1978) and Subba Rao (1981), is one non-linear time series model that is simpler than most.

Let e_t, $t \in Z$ be a sequence of independent and identically distributed random variables with $\mathrm{E}(e_t) = 0$ and $\mathrm{E}(e_t{}^2) = \sigma^2 < \infty$. Let $a_1, a_2, \ldots, a_p$, $c_1, c_2, \ldots, c_q$ and $b_{i,j}$, $1 \leq i \leq m$, $1 \leq j \leq k$ be real constants. If a time series X_t satisfies the difference equation

$$X_t + \sum_{j=1}^{p} a_j X_{t-j} = \sum_{j=1}^{q} c_j e_{t-j} + \sum_{i=1}^{m} \sum_{j=1}^{k} b_{ij} X_{t-i} e_{t-j} + e_t, \qquad (1.3)$$

then X_t is defined as a bilinear process that satisfies a bilinear time series model (1.3) denoted by Subba Rao (1981) as $BL(p, q, m, k)$. The model given by (1.3) along with simpler forms of the model have been discussed by Granger and Andersen (1978), Subba Rao (1981), Quinn (1982), Bhaskara Rao, Subba Rao and Walker (1983), and Sesay and Subba Rao (1991), amongst others.

The bilinear model given in (1.3) is a non-linear model, but the structural nature of this model is similar to that of linear models. The model given in (1.3) can be broken down into three components. The first part is the autoregressive (AR) part, the second is defined as the moving average (MA) part and the third part of the process is the pure or completely bilinear part.

In (1.3), if we let $b_{ij} = 0$ for all i and j, we obtain the autoregressive-moving average model $ARMA(p, q)$. Therefore, the structural theory of the bilinear model is analogous to that of the autoregressive model, the moving average model, and the mixed autoregressive-moving average model.

If $b_{ij} = 0$ for all $i < j$ in (1.3), the model is said to be super-diagonal. In (1.3), if $b_{ij} = 0$ for all $i \geq j$, the model is referred to as a sub-diagonal model, and it is known as a diagonal model if $b_{ij} = 0$ for all $i \neq j$.

1.3 Stationarity and Invertibility Conditions

If a time series model is to be useful in interpreting and forecasting it is essential, based on the Box-Jenkins methodology, that the time series is both stationary and invertible. In general, a time series X_t is said to be stationary if the statistical properties of the time series remain unchanged with time. A time series X_t is said to be strictly stationary if for any set of times t_i, $i = 1, 2, \ldots, m$ and any positive integer m the joint probability distribution of $\{X_{t_1}, X_{t_2}, \ldots X_{t_m}\}$ is identical to the joint probability distribution of $\{X_{t_1+h}, X_{t_2+h}, \ldots X_{t_m+h}\}$ for any integer h. If the joint moment up to order 2 exist and remain unchanged with time, then the time series X_t is said to be 2^{nd} order stationary or weakly stationary. The invertibility of a time series X_t simply implies that having knowledge of $\{X_h\}$, $h \leq t$ is equivalent to having knowledge of e_h, $h \leq t$.

Phan and Tran (1981) derived these conditions for the first order bilinear model. Subba Rao (1981) obtained the conditions for asymptotic stationarity and invertibility of a time series satisfying the model $BL(p, 0, p, 1)$. Granger

and Andersen (1978) and Quinn (1982) have derived these conditions for simpler bilinear models than that of Subba Rao (1981).

This report will mainly deal with a time series X_t that satisfies the bilinear model $BL(p, 0, m, k)$ and hence, to fix ideas, we shall discuss stationary and invertible conditions for the model $BL(p, 0, p, 1)$ given by

$$X_t + \sum_{j=1}^{p} a_j X_{t-j} = e_t + \sum_{i=1}^{p} b_{i1} X_{t-i} e_{t-1} \tag{1.4}$$

Interested readers can refer to Liu and Brockwell (1988) for a more general discussion.

1.3.1 Stationarity

The stationarity condition for a process satisfying

$$X_t + a_1 X_{t-1} = e_t + b_{11} X_{t-1} e_{t-1} \tag{1.5}$$

where $\mathrm{E}(e_t) = 0$ and $\mathrm{E}(e_t^2) = \sigma^2 < \infty$ has been established by Phan and Tran (1981). They have shown that there exists a strictly stationary process X_t if $a_1^2 + \sigma^2 b_{11}^2 < 1$. Subba Rao (1981) established conditions for strict stationarity of a process X_t satisfying (1.4). To discuss this condition we must first write the model in matrix notation. Let

$$\mathbf{X_t} = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}_{p \times 1}$$

and let us define the following matrices

$$A = \begin{pmatrix} -a_1 & -a_2 & \cdots & -a_{p-1} & -a_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{p\times p}, B = \begin{pmatrix} b_{11} & b_{21} & b_{31} & \cdots & b_{p1} \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{p\times p}$$

and

$$C = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{p\times 1}.$$

Equation (1.4) can now be rewritten in the form

$$\mathbf{X_t} = A\mathbf{X_{t-1}} + B\mathbf{X_{t-1}}e_{t-1} + Ce_t \tag{1.6}$$

where $\mathrm{E}(e_t) = 0$ and $\mathrm{E}(e_t^2) = \sigma^2 < \infty$. Subba Rao (1981) has shown that if

$$\rho(A \otimes A + \sigma^2 B \otimes B) < 1 \tag{1.7}$$

where $\otimes$ is the Kronecker product and $\rho(\cdot)$ is the spectral radius or the maximum eigenvalue function, then there exists a strictly stationary process X_t satisfying model (1.6). For example, if $a_1 = -0.01$, $a_2 = -0.2$, $b_{11} = 0.1$, $b_{12} = 0.08$ and $\sigma = .5$ then,

$$A = \begin{pmatrix} 0.01 & 0.2 \\ 1 & 0 \end{pmatrix}, B = \begin{pmatrix} 0.1 & 0.08 \\ 0 & 0 \end{pmatrix}$$

And

$$C = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Then, using the Kronecker product

$$A \otimes A = \begin{pmatrix} 0.0004 & 0.002 & 0.002 & 0.04 \\ 0.0002 & 0 & 0.2 & 0 \\ 0.0002 & 0.2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

and

$$B \otimes B = \begin{pmatrix} 0.01 & 0.008 & 0.008 & 0.0064 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and the maximum eigenvalue of $(A \otimes A + \sigma^2 B \otimes B)$ is equal to $0.2146096 < 1$.

1.3.2 Invertibility

To state the sufficient condition for invertibility of the bilinear model $BL(p, 0, p, 1)$, as discussed by Subba Rao (1981), we must consider the notation defined in Section 1.3.1 and define

$$H = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The following condition

$$H'B\mathrm{E}[x_t x_t']B'H < (H'C)^2 \tag{1.8}$$

is a sufficient condition for invertibility of model (1.6).

Chapter 2

The River Flow Data

2.1 Introduction

The data sets analyzed in this chapter and the remainder of this practicum are presented in Tables A.1-A.6 in **Appendix A**. The data in Tables A.1-A.6 consists of mean daily river flow measurements collected by Environment Canada for the years 1995 and 1996. Since the year 1996 was a leap year, each data set consists of 731 observations. There are gaging stations located at specific points in each river which measure the water level throughout each day. The water level is then converted to flow rates of m^3/s and the mean river flow for that day is then recorded.

Section 2.1 consists of a description of each data set used in this practicum. Section 2.3 describes three approaches for testing a time series for linearity. In Section 2.4, the results from the tests for linearity are given.

2.2 Data

The data in Table A.1 contain mean daily river flow measurements from Peace River at Hudson Hope in British Columbia. Table A.2 presents the mean daily river flow measurements for Castle River in Alberta. The river flow measurements contained in Tables A.3, A.4 and A.5 were taken from South River near Holyrood, Salmonier River near Lamaline and Gander River at Big Chute in Newfoundland respectively. The daily river flow measurements in Table A.6 were measured from Moberly River near Fort St. John in British Columbia.

The data sets are presented in the form of time series plots in **Appendix B**. Figure B.1 is the time series plot for Peace River. The river flow measurements range from a minimum of 329 m^3/s to a maximum of 5190 m^3/s. There does not appear to be any pattern in the time series over the two years. The daily river flow in the summer of 1996 more than doubled the maximum of any other day throughout the two years. The time series plot for Castle River is displayed in figure B.2. The minimum river flow measurement for Castle River is 1.62 m^3/s and the maximum is 812 m^3/s. Over the two years, there did appear to be a pattern in the time series. The river flow was low during the first 4 months, then the river flow begins to increase over the next 2 months and starts to decrease again after 6 months. The last four months of the year is similar to the first four.

The time series plots for the three rivers in Newfoundland are displayed in Figures B.3, B.4 and B.5. South River B.3 and Salmonier river B.4 portrayed similar patterns to each other, but the actual time series never depicted any patterns over the two years. South River had a minimum river flow of 0.069

m^3/s and a maximum of 13 m^3/s and Salmonier River had a minimum of 0.043 m^3/s and a maximum of 63.5 m^3/s. Gander River, displayed in Figure B.5, never showed any particular pattern. For the two years, Gander River had a minimum river flow of 23.5 m^3/s and a maximum flow of 669 m^3/s.

Displayed in Figure B.6 is the time series plot for Moberly River. This plot did display a pattern. The mean daily river flow dramatically increased in the summer months and remain very low during the remaining months of the year. Moberly river had a minimum river flow of 1.24 m^3/s and a maximum flow level of 87.1 m^3/s.

In this practicum we selected only two years of data for convenience to illustrate the techniques being implemented. Although we looked for patterns over the two year period, a time series consisting of a ten year period would be more valuable to a researcher when attempting to detect patterns in river flow data for a particular river, especially when trying to distinguish between dry and wet years.

2.3 Tests for Linearity

Before any formal modeling was performed on the data, we found it necessary to test the time series for linearity. The statistical approach and methods we have adopted will establish whether the time series is linear or non-linear. Tong (1990) discussed both informal graphical methods and formal tests in order to distinguish a time series as either linear or non-linear. Although the graphical methods are useful, it was deemed sufficient to consider only the formal tests. Three of the tests were selected for application to the data

given in **Appendix A**. The tests are described in Section 2.3.1, 2.3.2. and 2.3.3. The first two tests are known as Portmanteau tests and the third is a test with a specific alternative. The null hypothesis is always that the time series X_t is linear.

2.3.1 Approach Based On Examining Squares Of Time Series Data

The approach, based on squares of time series, was proposed by McLeod and Li (1983). This test for linearity was motivated by the fact that

$$\rho_\tau(X_t^2) = \{\rho_\tau(X_t)\}^2, \quad \text{for all } \tau$$

where

$$\rho_\tau = Z_t = corr(Z_{t+\tau}, Z_t)$$

provided that X_t is a stationary Gaussian time series.

It is a useful test for detecting non-linearity, and moreover the non-linearity may be in the direction of bilinearity. The test proposed by McLeod and Li (1983) is performed as follows. Let $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_N$ be the fitted residuals from an $ARMA$ model. The sample autocorrelation of the squared residuals r_k is then given by

$$r_k = \frac{\sum_{j=1}^{N-k}(\hat{\varepsilon}_j^2 - \hat{\sigma}^2)(\hat{\varepsilon}_{j+k}^2 - \hat{\sigma}^2)}{\sum_{j=1}^{N}(\hat{\varepsilon}_j^2 - \hat{\sigma}^2)} \tag{2.1}$$

where

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{N} \hat{\epsilon}_j^2}{N}.$$

Now, the test statistic is then given by

$$Q = \frac{N(N+2)\sum_{k=1}^{m} r_k^2}{(N-k)} \tag{2.2}$$

where $Q \sim \chi_m^2$ for some integer m.

2.3.2 Approach Based On Tukey's One-Degree-Of-Freedom Test For Non-Additivity

The second test for linearity we will consider was proposed by Keenan (1985). Let $(X_1, X_2, \ldots, X_N)$ denote a time series. The algorithm for Keenan's test for linearity based on Tukey's one-degree-of freedom test for non-additivity proceeds as follows:

1. Regress X_t on $\{1, X_{t-1}, X_{t-2}, \ldots, X_{t-M}\}$, where M is a fixed positive integer. From this model, calculate the fitted values $\{\hat{X}_t\}$, the fitted residuals, $\hat{e}_t$, $t = M+1, M+2, \ldots, N$, and the residual sum of squares, $\text{RSS} = \sum \hat{e}_t{}^2$.

2. Regress $\hat{X}_t^2$ on $\{1, X_{t-1}, X_{t-2}, \ldots, X_{t-M}\}$ and calculate the fitted residuals, $\{\hat{\xi}_t\}$, $t = M+1, M+2, \ldots, N$.

3. Regress $(\hat{e}_{M+1}, \hat{e}_{M+2}, \ldots, \hat{e}_N)$ on $(\hat{\xi}_{M+1}, \hat{\xi}_{M+2}, \ldots, \hat{\xi}_N)$ and calculate,

$$\eta = \eta_0 \left(\sum_{t=M+1}^{N} \hat{\xi}_t^2 \right)^{\frac{1}{2}}$$

where η_0 is the regression coefficient. The test statistic is now calculated as,

$$F = \frac{\eta^2 (N - 2M - 2)}{\text{RSS} - \eta^2} \tag{2.3}$$

where the null distribution of F is $F_{1,N-2M-2}$.

2.3.3 Tsay's Approach Based On Column Stacking

Tsay (1986) devised an approach to test for linearity that is an improvement to Keenan's test. The test is based on column stacking and the power of the test is increased over Keenan's test. This third test is implemented as follows:

1. Regress X_t on $\{1, X_{t-1}, X_{t-2}, \ldots, X_{t-M}\}$ and calculate the fitted residuals $\{\hat{e}_t\}$, $t = M+1, M+2, \ldots, N$. The regression model will be denoted by

$$X_t = \mathbf{Y_t b} + e_t \tag{2.4}$$

where $\mathbf{Y_t} = (1, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-M})$ and $\mathbf{b} = (b_0, b_1, \ldots, b_M)$, where M is a fixed positive integer.

2. Let $\mathbf{Z_t}$ be a row vector with dimension $m = \frac{1}{2}M(M+1)$. Using only the

measurements on or below the main diagonal and applying the usual column stacking operation, $\mathbf{Z_t}$ is obtained from the symmetric matrix $\mathbf{U_t'U_t}$, where $\mathbf{U_t} = (\hat{X}_{t-1}, \hat{X}_{t-2}, \ldots, \hat{X}_{t-M})$. For example, if

$$U_1'U_1 = \begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{pmatrix}_{4\times 4}$$

Then, applying the usual column stacking operation $Z_1 = \{u_{11}, u_{21}, u_{31}, u_{41}, u_{22}, u_{32}, u_{42}, u_{33}, u_{43}, u_{44}\}$. Using multivariate regression, regress $\mathbf{Z_t}$ on $\{1, X_{t-1}, X_{t-2}, \ldots, X_{t-M}\}$, where the regression model is given by

$$\mathbf{Z_t} = \mathbf{Y_t H} + \xi_t. \tag{2.5}$$

Finally, from the fitted model we obtain the fitted residual vector $\{\hat{\xi}_{M+1}, \hat{\xi}_{M+2}, \ldots, \hat{\xi}_N\}$.

3. Regress $(\hat{e}_{M+1}, \hat{e}_{M+2}, \ldots, \hat{e}_N)$ on $(\hat{\xi}_{M+1}, \hat{\xi}_{M+2}, \ldots, \hat{\xi}_N)$ and obtain the least squares residuals $\{\hat{\alpha}_t\}$, $t = M+1, M+2, \ldots, N$. Now, the test statistic is calculated as,

$$F = \frac{(\sum \hat{\xi}_t \hat{e}_t)(\sum \hat{\xi}_t' \hat{\xi}_t)^{-1}(\sum \hat{\xi}_t \hat{e}_t)'(N - M - m - 1)}{m \sum \hat{\alpha}_t^2} \tag{2.6}$$

where the summations range from $t = M+1$ to $t = N$. Asymptotically, the null distribution of F is $F_{\frac{1}{2}M(M+1), N-\frac{1}{2}M(M+3)-1}$.

2.4 Results

We applied the tests for linearity described in Sections 2.3.1, 2.3.2 and 2.3.3 to the six data sets given in **Appendix A**. In all three tests, M is a positive fixed integer with no specific rule on how to choose an appropriate value. Therefore, all three test were ran with $M = 4, 5, 7, 10$. The test statistics along with their respective p-values for the six rivers are shown in Tables 2.1-2.6 below.

With the exception of Gander River results displayed in Table 2.5, the p-values exhibited in the tables consistently agree with each other. For all rivers with the exception of Gander River, the results provide significant evidence that the river flow time series is non-linear and this explains why we are using non-linear models to describe the data. Gander River had an inconsistency in the results and this discrepancy was a consequence of Keenan's Test. For all four values of M, the p-values for Keenan's test were not significant suggesting that the time series X_t is linear. The test devised by McLeod and Li had significant p-values for all four values of M which were in agreement with Tsay's test. Since the bilinear model is made up of a linear component and a pure bilinear component it appears to be a reasonable model for studying such a series.

Table 2.1: Computed F-statistics and p-values for Peace River

	McLeod and Li		Keenan's Test		Tsay's Test	
M	Stat	P-value	Stat	P-value	Stat	P-value
4	2624.18	0	4.907	.02706	33.719	0
5	3191.92	0	5.093	.02432	25.156	0
7	4267.33	0	5.133	.02377	15.535	0
10	5732.45	0	4.662	.03177	10.161	0

Table 2.2: Computed F-statistics and p-values for Castle River

	McLeod and Li		Keenan's Test		Tsay's Test	
M	Stat	P-value	Stat	P-value	Stat	P-value
4	340.500	0	444.305	0	296.897	0
5	343.834	0	402.314	0	207.055	0
7	350.014	0	381.226	0	137.412	0
10	353.602	0	404.158	0	113.998	0

Table 2.3: Computed F-statistics and p-values for South River

	McLeod and Li		Keenan's Test		Tsay's Test	
M	Stat	P-value	Stat	P-value	Stat	P-value
4	314.848	0	92.193	0	56.682	0
5	316.360	0	54.586	0	40.279	0
7	316.446	0	49.729	0	22.435	0
10	316.452	0	63.686	0	15.886	0

Table 2.4: Computed F-statistics and p-values for Salmonier River

	McLeod and Li		Keenan's Test		Tsay's Test	
M	Stat	P-value	Stat	P-value	Stat	P-value
4	368.808	0	81.240	0	9.003	0
5	368.893	0	81.304	0	6.339	0
7	370.807	0	75.193	0	3.749	0
10	377.266	0	86.321	0	3.050	0

Table 2.5: Computed F-statistics and p-values for Gander River

	McLeod and Li		Keenan's Test		Tsay's Test	
M	Stat	P-value	Stat	P-value	Stat	P-value
4	2680.57	0	1.2899	0.2564	6.7751	0
5	3224.16	0	0.8769	0.3494	13.8336	0
7	4123.24	0	1.2292	0.2679	7.7224	0
10	5035.81	0	0.8604	0.3530	5.6089	0

Table 2.6: Computed F-statistics and p-values for Moberly River

	McLeod and Li		Keenan's Test		Tsay's Test	
M	Stat	P-value	Stat	P-value	Stat	P-value
4	2682.11	0	20.5615	0	15.5925	0
5	3229.63	0	25.7950	0	12.4396	0
7	4146.93	0	35.3119	0	9.7567	0
10	5119.35	0	29.9219	0	7.7252	0

Chapter 3

Bilinear Time Series Model

3.1 Introduction

The bilinear time series model was introduced in Chapter 1, where some of the theoretical concepts behind the model were discussed. The purpose of this chapter is to discuss techniques for estimating the parameters of a bilinear time series model and describe the methodology used to choose the best model. We will also apply the techniques discussed to modeling Canadian river flow data.

Section 3.2 will outline the underlying theory of the estimation procedure used to estimate the parameters of bilinear time series models. The Newton-Raphson technique is also described, along with the estimating equations required for this procedure. The concepts of Section 3.2 will be applied in the subsequent sections. In Section 3.3 we fit bilinear time series models to five of the six data sets from the river flow data discussed in Chapter 2. We attempted to fit bilinear models to all six data sets but were unsuccessful in

doing so with the data from Castle River. We did, however, overcome the difficulties with the data from Castle River using the methods described in Chapter 4. Finally, in Section 3.4, based on simulated data from the fitted models, we study the sampling properties of the mean and standard deviation and compare these results with the original data.

3.2 Estimation of the Parameters of the Bilinear Time Series Model

In this section we will obtain the parameter estimates of the bilinear time series model given by

$$X_t + \sum_{j=1}^{p} a_j X_{t-j} + \alpha = \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij} X_{t-i} e_{t-j} + e_t \tag{3.1}$$

where e_t are assumed to be independently and identically distributed as $N(0, \sigma_e^2)$. The model given above is the same as $BL(p, 0, m, k)$ with an extra parameter α added. Subba Rao (1981) suggests that this additional parameter, which affects only the mean, is very useful when fitting bilinear models to raw data. The parameters which will be estimated in model (3.1) are $\{a_i, 1 \leq i \leq p\}$, $\{b_{ij}, 1 \leq i \leq m, 1 \leq j \leq k\}$, α, and σ_e^2 for n observations for a total of $p + mk + 2$ parameters.

As with all models which involve lagged values of the X_t, we cannot evaluate the residuals for an initial stretch of data. We therefore consider the conditional likelihood based on $\{X_{\gamma+1}, X_{\gamma+2}, \ldots, X_n\}$ given $\{X_1, X_2 \ldots, X_\gamma\}$ where $\gamma = \{max(p, m, k) + 1\}$. The joint pdf of $\{e_\gamma, e_{\gamma+1}, \ldots, e_n\}$ is given by

$$f(x) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{\left(\frac{n-\gamma+1}{2}\right)} \exp\left[\frac{-1}{2\sigma_e^2}\sum_{t=\gamma}^{n} e_t^2\right] \tag{3.2}$$

which is also the likelihood function of $\{X_{\gamma+1}, X_{\gamma+2}, \ldots, X_n\}$. The conditional maximum likelihood estimates of $\theta = (\alpha, a_1, a_2, \ldots, a_p, b_{11}, b_{12}, \ldots, b_{mk})' = (\theta_1, \theta_2, \ldots, \theta_{p+mk+1})'$ are found by maximizing the likelihood function or by minimizing

$$Q(\theta) = \sum_{t=\gamma}^{n} e_t^2 \tag{3.3}$$

with respect to θ, the least squares fit. To minimize θ we must solve

$$\frac{\partial Q}{\partial \theta} \mid_{\theta_{min}} = 0 \tag{3.4}$$

which is done through the Newton-Raphson iterative technique.

3.2.1 Newton-Raphson Iterative Technique

The Newton-Raphson iterative procedure is based on the Taylor series expansion. It requires an initial guess for the first set of parameters and then the function is approximated in the neighborhood of that guess. The Newton-Raphson iterative equation to be used to minimize $Q(\theta)$ can be obtained by the following method. First, let $q = p + mk + 1$,

$$G(\theta) = \begin{pmatrix} \frac{\partial Q(\theta)}{\partial \theta_1} \\ \frac{\partial Q(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial Q(\theta)}{\partial \theta_q} \end{pmatrix}$$

be a vector of first-order partial derivatives and

$$H(\theta) = \left[\frac{\partial^2 Q(\theta)}{\partial\theta_i \partial\theta_j}\right]$$

be a matrix of second-order partial derivatives. To solve equation (3.2), we use the Taylor series expansion, and expand $G(\theta)$ near $\hat{\theta} = \theta$ and assuming third order and higher terms are negligible we obtain

$$G(\theta_0) + H(\theta_0)(\theta - \theta_0) = 0. \tag{3.5}$$

Next, we solve (3.5) for θ to obtain

$$\theta = \theta_0 - H^{-1}(\theta_0)G(\theta_0)$$

Therefore, in general, the $(i+1)^{th}$ iteration is given by

$$\theta^{(i+1)} = \theta^{(i)} - H(\theta^{(i)})G(\theta^{(i)}) \tag{3.6}$$

known as the Newton-Raphson iterative equation, and hence we obtain the parameter estimation technique for the model.

3.2.2 Partial Derivatives

The first- and second-order partial derivatives of $Q(\theta)$ are given by

$$\frac{\partial Q(\theta)}{\partial\theta_i} = 2\sum_{t=\gamma}^{n} e_t \frac{\partial e_t}{\partial\theta_i}, \quad i = 1, 2, \ldots, q, \tag{3.7}$$

$$\frac{\partial^2 Q(\theta)}{\partial\theta_i\partial\theta_j} = 2\sum_{t=\gamma}^{n}\frac{\partial e_t}{\partial\theta_i}\frac{\partial e_t}{\partial\theta_j} + 2\sum_{t=\gamma}^{n} e_t\frac{\partial^2 e_t}{\partial\theta_i\partial\theta_j},$$
$$i = 1, 2, \ldots, q; \quad j = 1, 2, \ldots, q.$$

Solving model (3.1) for e_t, it is clear that the partial derivatives of e_t must satisfy the following recursive equations:

$$\frac{\partial e_t}{\partial\alpha} = 1 - \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial e_{t-j}}{\partial\alpha} \tag{3.8}$$

$$\frac{\partial e_t}{\partial a_i} = X_{t-i} - \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial e_{t-j}}{\partial a_i}, \quad i = 1, 2, \ldots, p \tag{3.9}$$

$$\frac{\partial e_t}{\partial b_{rs}} = -X_{t-r}e_{t-s} - \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial e_{t-j}}{\partial b_{rs}}, \tag{3.10}$$
$$r = 1, 2, \ldots, m; \quad s = 1, 2, \ldots, k$$

$$\frac{\partial^2 e_t}{\partial\alpha^2} = -\sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial^2 e_{t-j}}{\partial\alpha^2} \tag{3.11}$$

$$\frac{\partial^2 e_t}{\partial a_i\partial a_{i'}} = -\sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial^2 e_{t-j}}{\partial a_i\partial a_{i'}}, \quad i, i' = 1, 2, \ldots, p \tag{3.12}$$

$$\frac{\partial^2 e_t}{\partial\alpha\partial b_{rs}} = -X_{t-r}\frac{\partial e_{t-s}}{\partial\alpha} - \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial^2 e_{t-j}}{\partial\alpha\partial b_{rs}}, \tag{3.13}$$
$$r = 1, 2, \ldots, m; \quad s = 1, 2, \ldots, k$$

$$\frac{\partial^2 e_t}{\partial a_i\partial b_{rs}} = -X_{t-r}\frac{\partial e_{t-s}}{\partial a_i} - \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial^2 e_{t-j}}{\partial a_i\partial b_{rs}}, \tag{3.14}$$
$$i = 1, 2, \ldots, p; \quad r = 1, 2, \ldots, m; \quad s = 1, 2, \ldots, k$$

$$\frac{\partial^2 e_t}{\partial b_{rs}\partial b_{r's'}} = -X_{t-r}\frac{\partial e_{t-s}}{\partial b_{r's'}} - X_{t-r'}\frac{\partial e_{t-s'}}{\partial b_{rs}} - \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij}X_{t-i}\frac{\partial^2 e_{t-j}}{\partial b_{rs}\partial b_{r's'}}, \quad (3.15)$$
$$r, r' = 1, 2, \ldots, m; \quad s, s' = 1, 2, \ldots, k$$

where we assume the initial conditions $e_t = 0$, $t = 1, 2, \ldots, \gamma - 1$, and also

$$\frac{\partial e_t}{\partial \theta_i} = 0 \text{ and } \frac{\partial^2 e_t}{\partial \theta_i \partial \theta_j} = 0, \quad i, j = 1, 2, \ldots, q; \quad t = 1, 2, \ldots, \gamma - 1$$

A direct result of the initial assumptions together with equation (3.9) leads to all second-order partial derivatives with respect to α and a_i, $i = 1, 2, \ldots, p$ equaling zero. Using the recursive equations (3.7), (3.8), (3.9) and (3.14), the first and second order derivatives of $Q(\theta)$ can be evaluated for a given set of initial values of α, $\{a_i\}$ and $\{b_{ij}\}$. The first and second order derivatives of $Q(\theta)$ can now be used in the Newton-Raphson iterative technique discussed in Section 3.2.1.

3.2.3 Initial Estimates

A fundamental component of the Newton-Raphson iterative technique is the initial estimates of the parameters. If a poor set of initial parameters are used, it is highly probable that one will not achieve convergence. There are different strategies one may consider when tackling this problem. The approach we found very effective is more of a step-up approach: fit a basic

model to the data first, then add additional parameters, one at a time.

To fit a bilinear time series model of order $BL(p, 0, m, k)$, we first fit a bilinear model of order $BL(2, 0, 1, 1)$, constant α included. To fit this model we fit an $AR(2)$ model and use the estimated parameters as initial estimates for the autoregressive part of the bilinear model, and for the moving-average part, we set $b_{11} = 0$. We next fit the bilinear model of order $BL(2, 0, 1, 2)$ using the estimates from the $BL(2, 0, 1, 1)$ as initial estimates and set $b_{12} = 0$. We continue this until all b_{ij}, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, k$ parameters are estimated. Once the pure bilinear part of the model is fitted, we fit the autoregressive part in a similar fashion.

3.2.4 Model Selection

Once we begin to fit bilinear time series models to the data, we must consider the order of the bilinear model that best represents the data. To choose the order of the model, we will consider three criteria: the Akaike Information Criterion (AIC), a bias-corrected version of the AIC known as the $AICC$ suggested by Hurvich and Tsai (1989) and the BIC which attempts to correct the over-fitting nature of the AIC. These criteria are defined by:

$$AIC = (N - M) \log \hat{\sigma}_e^2 + 2(p + mk + 1) \tag{3.16}$$

$$AICC = (N - M) \log \hat{\sigma}_e^2 + \frac{2(p + mk + 1)n}{N - (p + mk + 2)} \tag{3.17}$$

$$BIC = (N - (p + mk + 1)) \log \left[\frac{N \hat{\sigma}_e^2}{N - (p + mk + 1)} \right]$$

$$+ n(1 + \log\sqrt{2\pi}) + (p + mk + 1) \times \log\left[\frac{(\sum_{t=1}^{N} X_t^2 - N\hat{\sigma}_e^2)}{(p + mk + 1)}\right]. \quad (3.18)$$

where,

$$\hat{\sigma}_e^2 = \frac{1}{N - M}\sum_{t=M+1}^{N} \hat{e}_t^2.$$

It is essential that the above functions are calculated based on the same number of observations for each model. The number of observations used in the calculation of the above criterion is given by $(N - M)$, therefore M should be selected such that $(N - M)$ will remain constant for each model fitted.

Based on the AIC, $AICC$ and BIC criteria, we choose the model with the smallest AIC, $AICC$ and BIC values. Therefore, we continue to fit models until the information criteria increase, then we choose the previous model. The AIC, $AICC$ and BIC do not always coincide with each other. If two of three criteria agree, then the choice of the model will be made based on these two criteria. If it so happens that all three criteria disagree, then a simulation study can be conducted and the choice of the model will be made based on the sampling properties of the estimates.

3.3 Model Fitting to River Flow Data

Bilinear models were fitted to data from Peace River, South River, Salmonier River, Gander River and Moberly River. It was found that the methods described in the previous section could not be used to estimate the parameters

for the data from Castle River for models of order higher than $BL(2,0,1,1)$. For this reason, we never included Castle River in the model fitting of this section or in the simulation results of the next section. The reason for the difficulty is believed to be a consequence of the large variability in the data. With a minimum river flow of 1.62 m^3/s and a maximum river flow of 812 m^3/s, Castle River had a range of 810.38 m^3/s. Castle River, with a mean of 20.47 m^3/s, had a variance of 1699.128 m^3/s.

When modeling the other five rivers, it was found that working with the actual time series was very tedious. On many occasions, the inverse of Fisher's Information matrix was unmanageable numerically. This problem was easily solved through a simple transformation of the time series. The transformation, known as standardizing, is given by

$$X_t' = \frac{X_t - \mu_{X_t}}{\sigma_{X_t}}. \tag{3.19}$$

Standardizing the time series will, in fact, leave the estimates of the parameters of the models unchanged with the exception of α, which is simply an estimate of the mean. Therefore, the models discussed in this section will be based on the standardized series.

The bilinear models were fitted based on the three criteria AIC, $AICC$ and BIC discussed in the previous section. For Peace River it was found that the AIC, $AICC$ and BIC were minimized when $p = 5$, $m = 1$ and $k = 2$. The values of AIC, $AICC$ and BIC were -2608.7665, -2908.5670, and -1185.3044 respectively. The estimated values of the model coefficients where $\alpha = 0.0088$, $a_1 = -1.1881$, $a_2 = 0.4594$, $a_3 = -0.1598$, $a_4 = 0.0657$, $a_5 = -0.1553$, $b_{11} = 0.2185$ and $b_{12} = 0.0867$. Therefore, the fitted model for

Peace River is

$$\begin{aligned} X_t' &- 1.1881X_{t-1}' + 0.4594X_{t-2}' - 0.1598X_{t-3}' \\ &+ 0.06578X_{t-4}' - 0.1553X_{t-5}' + 0.0088 \\ &= 0.2185X_{t-1}'e_{t-1} + 0.0867X_{t-1}'e_{t-2} + e_t \end{aligned} \tag{3.20}$$

where the value of $\hat{\sigma}^2 = 0.0262$.

The AIC, $AICC$ and BIC for South River were minimized when $p = 3$, $m = 1$ and $k = 1$ and the values of the criteria were -368.0355, -367.9527 and 1047.6431 respectively. The estimated values of the coefficients of the model are $\alpha = -0.0399$, $a_1 = -0.9835$, $a_2 = 0.2791$, $a_3 = -0.1773$ and $b_{11} = -0.0641$. The bilinear model for South River is as follows

$$\begin{aligned} X_t' &- 0.9835X_{t-1}' + 0.2791X_{t-2}' - 0.1773X_{t-3}' \\ &- 0.0399 = -0.0641X_{t-1}'e_{t-1} + e_t \end{aligned} \tag{3.21}$$

where the value of $\hat{\sigma}^2 = 0.592$.

The BIC criteria for the Salmonier River was inconsistent with the AIC and the $AICC$. The selection of the order of the model for this river was more involved than the others. We found that the BIC was minimized when $p = 2$, $m = 1$ and $k = 1$ while the AIC and the $AICC$ were minimized when $p = 3$, $m = 1$ and $k = 1$. We then performed a simulation study for both models and studied sampling properties of specific statistics. We found that the best model of the two was the model of order $p = 3$, $m = 1$ and $k = 1$. The AIC, $AICC$ and BIC for this model were -473.7456, -473.6628

and 942.1053 respectively. The parameter estimates of the model are $\alpha = -0.0449$, $a_1 = -1.2251$, $a_2 = 0.4812$, $a_3 = -0.0776$ and $b_{11} = -0.0866$. The fitted model for Salmonier River is

$$\begin{aligned} X'_t &- 1.2251X'_{t-1} + 0.4812X'_{t-2} - 0.0776X'_{t-3} \\ &- 0.0449 = -0.0866X'_{t-1}e_{t-1} + e_t \end{aligned} \tag{3.22}$$

where the value of $\hat{\sigma}^2 = 0.5112$.

For Gander River, it was when $p = 5$, $m = 1$ and $k = 1$ that the AIC, $AICC$ and BIC were minimized and the values were -3566.6934, -3566.5385 and -2153.2538 respectively. The bilinear model had the following parameter estimates: $\alpha = -0.0009$, $a_1 = -2.3217$, $a_2 = 2.1609$, $a_3 = -1.1658$, $a_4 = 0.4121$, $a_5 = -0.0719$ and $b_{11} = -0.0617$. Therefore, the fitted model for Gander River is

$$\begin{aligned} X'_t &- 2.3217X'_{t-1} + 2.1609X'_{t-2} - 1.1658X'_{t-3} \\ &+ 0.4121X'_{t-4} - 0.0719X'_{t-5} + 0.0009 \\ &= -0.0617X'_{t-1}e_{t-1} + e_t \end{aligned} \tag{3.23}$$

where the value of $\hat{\sigma}^2 = 0.007$.

For Moberly River, it was found that the AIC, $AICC$ and BIC were minimized when $p = 3$, $m = 1$ and $k = 1$. The values for the criteria where $AIC = -4014.0864$, $AICC = -4014.0137$ and $BIC = -2619.2296$. The estimated values of the coefficients of the model are $\alpha = -0.0007$, $a_1 = -2.0568$, $a_2 = 1.3421$, $a_3 = -0.2777$ and $b_{11} = -0.125$. The bilinear model

for Moberly River is as follows

$$\begin{aligned} X_t' &- 2.0568X_{t-1}' + 1.3421X_{t-2}' - 0.2777X_{t-3}' \\ &- 0.0007 = -0.125X_{t-1}'e_{t-1} + e_t \end{aligned} \tag{3.24}$$

where the value of $\hat{\sigma}^2 = 0.0038$.

3.4 Results of Simulation Study

In the previous section we fitted bilinear time series models to five Canadian rivers based on the river flow variable. It is important to see if the behavior of the fitted models exhibit properties similar to that of the actual flow data. An investigation of the behavior of the fitted models can be achieved by conducting a simulation study. For each of the five rivers we simulated $n = 731$ observations from the fitted model

$$X_t' = -\sum_{j=1}^{p} a_j X_{t-j}' - \alpha + \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij} X_{t-i}' e_{t-j} + e_t \tag{3.25}$$

where $e_t \sim N(0, \hat{\sigma}_e^2)$. This was repeated 2000 times. The transformation of the time series was then reversed in order to be comparable to the original series as follows

$$\tilde{X}_t = X_t' \sigma_{X_t} + \mu_{X_t}. \tag{3.26}$$

Next, we examined the sampling properties of the mean, standard deviation and the maximum value of the simulated time series $\tilde{X}_t$.

The results of the simulation study for each river are displayed in Tables 3.1-3.5. There are four rows in each table were the first row contains sampling properties of the original time series X_t. The second, third and fourth rows contain sampling properties of the mean of $\bar{X}_t$, the standard deviation of $\bar{X}_t$ and the maximum of $\bar{X}_t$ for 2000 simulations. For each river, we investigated whether this process was able to simulate the original time series by constructing a time series plot of the actual time series X_t overlaid with one of the 2000 simulated bilinear time series.

1. Peace River:

Table 3.1: Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Peace River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	329	694	1180	1278	1490	5190	950.6962
Mean($\bar{X}_t$)	604.6	1121	1270	1294	1445	2354	251.6964
St.Dev($\bar{X}_t$)	334.8	511.6	583.3	607.8	678.8	1545	135.9731
Max($\bar{X}_t$)	1677	2737	3112	3222	3585	7541	720.485

The mean and standard deviation of X_t are 1279 and 950.6962 respectively. The mean of the means of $\bar{X}_t = 1294$ is close to the mean of the actual time series X_t. The mean of the standard deviation of $\bar{X}_t$ is equal to 607.8 which is different from the actual series suggesting that the sampling properties of the standard deviation of $\bar{X}_t$ are poor. The sampling properties for the maximum value were also poor with values $\max(X_t) = 5190$ and the mean of $\max(\bar{X}_t) = 3222$.

Figure 3.1: Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Peace River

Figure 3.1 displays time series plots of both the actual series X_t and the simulated series $\tilde{X}_t$. It is observed that the simulated series from the fitted bilinear model follows the general pattern of the original series, but there is immense variation between the two series mainly within the interval of approximately 100 to 380 days. The variation between the two series leads to the questioning of the usefulness of this model.

2. South River:

Table 3.2: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (South River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	0.069	0.2605	0.416	0.6855	0.752	13	0.964
Mean($\tilde{X}_t$)	-0.133	0.5591	0.6992	0.7009	0.8573	1.448	0.2307
St.Dev($\tilde{X}_t$)	1.149	1.433	1.514	1.519	1.597	1.993	0.1203
Max($\tilde{X}_t$)	3.542	4.425	4.727	4.766	5.072	7.302	0.4907

The sampling properties of the mean of the means of $\bar{X}_t = 0.7009$ appears to behave much the same as the mean of $X_t = 0.6855$. As did Peace River, there seem to be discrepancies with respect to the mean of the standard deviations and the mean of the maximum of $\bar{X}_t$. The standard deviation of the series X_t is 0.964 and the standard deviation of $\bar{X}_t$ is equal to 1.519. The original series had a maximum value of 13 while the mean of $\max(\bar{X}_t)$ was equal to 4.766.

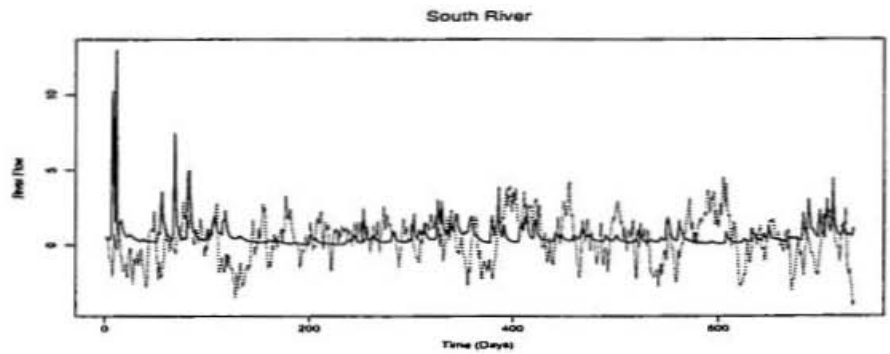

Figure 3.2: Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for South River

The time series plots of the actual series X_t and the simulated series $\bar{X}_t$ is found in figure 3.2. It is easily seen that the simulated time series plot does not, in any significant interval, model the actual time series.

3. Salmonier River:

Table 3.3: Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Salmonier River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	0.043	1.025	2.58	4.208	5	63.5	5.9732
Mean($\bar{X}_t$)	1.11	3.694	4.238	4.247	4.847	7.193	0.8899
St.Dev($\bar{X}_t$)	7.49	8.766	9.116	9.148	9.516	11.22	0.5599
Max($\bar{X}_t$)	21.73	26.41	27.92	28.16	29.72	39.53	2.497814

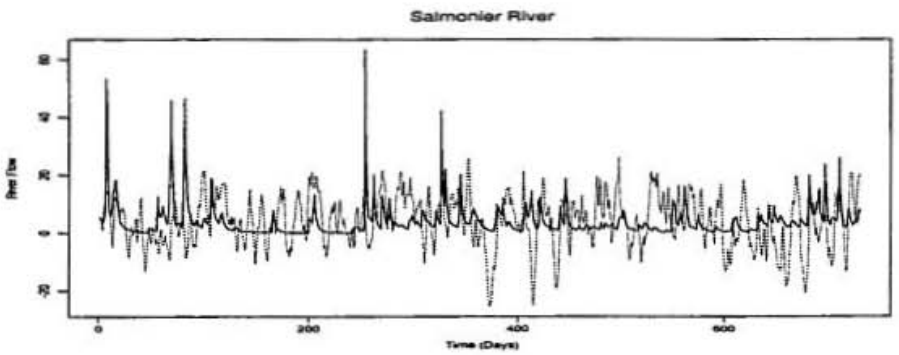

Figure 3.3: Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Salmonier River

Once again, while the sampling property of the mean of $\bar{X}_t$ appeared adequate, the mean of the standard deviation and the mean of the maximum value of $\bar{X}_t$ were not in accordance to the standard deviation and the maximum value of X_t. The values for the mean, the standard deviation and the maximum value of X_t were 4.208, 5.9732 and 63.5

respectively and the mean of the means of $\bar{X}_t = 4.247$, the mean of the standard deviation of $\bar{X}_t = 9.148$ and the mean of $\max(\bar{X}_t) = 28.16$. Figure 3.3 exhibit similar results to that of Figure 3.2. There are large variations between the plot of the actual time series and the plot of the simulated time series throughout the entire two year period.

4. Gander River:

Table 3.4: Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Gander River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	23.5	76.35	108	130.1	160	669	94.1295
Mean($\bar{X}_t$)	55.36	118.1	131	131	144.9	197.3	20.819
St.Dev($\bar{X}_t$)	67.95	92.06	98.92	99.59	106.7	138.9	10.8221
Max($\bar{X}_t$)	259.8	344.8	369.7	373	397.6	532.9	40.35806

The sampling properties of the mean and standard deviation for 2000 simulations for Gander River exhibited properties almost identical to the actual time series. However, Figure 3.4 display results suggesting the fitted bilinear model is not useful in simulating the data. Over the two year period, the simulated series continuously varies from the actual series. The mean of the time series X_t is equal to 130.1 while the mean of the means of $\bar{X}_t$ is 131. The standard deviation of the actual series is equal to 94.13 which is very similar to the mean of the standard deviations of $\bar{X}_t = 99.59$. The mean of the maximum value

of the simulated series $\tilde{X}_t$ was equal to 373 which was almost half of that of the original series X_t which had maximum value of 669.

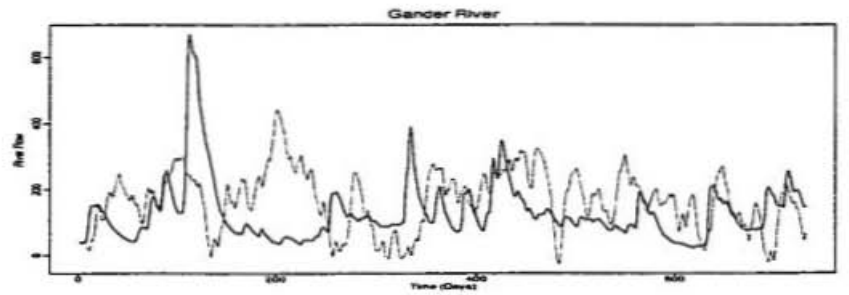

Figure 3.4: Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Gander River

5. Moberly River:

Table 3.5: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Moberly River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	1.24	1.84	3.24	13.17	15.25	87.1	18.741
Mean($\tilde{X}_t$)	-7.403	9.93	13.35	13.25	17.01	29.72	5.5015
St.Dev($\tilde{X}_t$)	12.12	17.28	19.11	19.39	21.25	32.19	3.047
Max($\tilde{X}_t$)	26.46	50.29	54.68	54.85	59.51	81.01	6.982

The fitted model for Moberly River simulated the best results for the mean and standard deviation, but the maximum value of the simulated series was off by a large margin. However, the simulated series does

not appear to follow the pattern of the original series as can be seen from Figure 3.5. The simulated series consistently deviates from the path of the original series. The mean of the time series X_t is 13.17 as compared to the mean of the means of $\tilde{X}_t$ which has a value equal to 13.25. The values for the standard deviation of X_t and the mean of the standard deviation of $\tilde{X}_t$ were also very close equaling 18.741 and 19.39 respectively. The maximum of X_t was 87.1 and the mean of $\max(\tilde{X}_t)$ was 54.85.

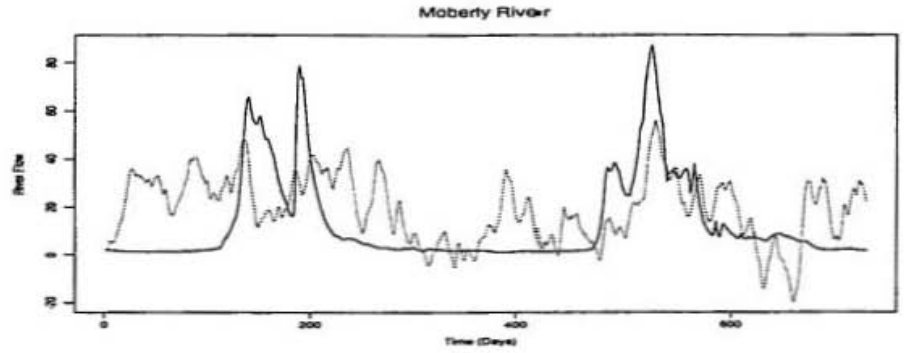

Figure 3.5: Plot of actual time series (solid line) with overlaid simulated bilinear time series plot (dotted line) for Moberly River

There were instances based on the bilinear approach where the sampling properties of the simulated time series followed closely to that of the actual time series. One statistic in particular, $\max(X_t)$, had very poor sampling properties for all five rivers. The simulation size is one possible reason for the poor sampling properties. If the simulation size was increased from $n = 2000$ to $n = 5000$ then it is possible that the estimation of the $\max(X_t)$ would be more accurate. A second possible reason is that the bilinear model is poor at

modeling extreme values. However, as a whole the simulation study for the five rivers easily explained that the fitted models were very poor. In Chapter 4, we combine the wavelet filtering approach and third order cumulants to improve the fit when modeling river flow data.

Chapter 4

Wavelet Filtering

4.1 Introduction

This chapter will consist primarily of decomposing a time series X_t, using wavelet smoothing techniques, into a wavelet smoothed version and a random component which we describe with pure bilinear models. Essentially, a time series X_t passes through a filter where the time series is decomposed into two components, a deterministic component and a random component. It is the random component that is of interest and we will fit a diagonal pure bilinear process, denoted by $DPBL(q)$, to this component of the non-stationary time series. The filtering process highlights any hidden non-stationarity and simplifies the structure of the random component. The diagonal pure bilinear process is given as

$$X_t = \sum_{i=1}^{q} \theta_i X_{t-i} e_{t-i} + e_t \tag{4.1}$$

where e_t is a sequence of independent and identically distributed Gaussian random variables with mean $\mu = 0$ and constant variance σ^2 and θ_i, $1 \leq i \leq q$ are constants. The model in (4.1) is a special case of the general bilinear autoregressive moving average process of Granger and Anderson (1978).

A wavelet system is the collection of dilated and translated versions of a scaling function $\phi(x)$ and a primary wavelet $\psi(x)$ defined by

$$\phi_{j,k}(x) = 2^{-\frac{j}{2}}\phi(2^{-j}x - k) \tag{4.2}$$

and

$$\psi_{j,k}(x) = 2^{-\frac{j}{2}}\psi(2^{-j}x - k), \quad j,k \in \mathcal{Z} \tag{4.3}$$

respectively. The functions $\phi(x)$ and $\psi(x)$ are choosen to satisfy the equations

$$\phi(x) = \sqrt{2}\sum_{p\in\mathcal{Z}} h_p\phi(2x - p) \tag{4.4}$$

and

$$\psi(x) = \sqrt{2}\sum_{r\in\mathcal{Z}} g_r\phi(2x - r), \quad g_r = (-1)^r h_{-r+1} \tag{4.5}$$

for a sequence $\{h_r\}$ of constants, called filter coefficients, with

$$\int \phi(x)dx = 1, \quad \int \psi(x)dx = 0, \quad \int \psi(x)dx = 1.$$

The Haar wavelet basis is the simplest example of a wavelet system on $\mathcal{L}_2(S)$. The scaling function is:

$$\phi(x) = I_{[0,1]}(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

The refining relations for the Haar wavelet basis are

$$\phi(x) = \phi(2x-1) + \phi(2x)$$

and

$$\psi(x) = \phi(2x) - \phi(2x-1).$$

In Section 4.2 we discuss the methodology behind wavelet filtering and the estimation of the filter coefficients will be described in section 4.3. Section 4.4 will consist of the method used to determine what models will be fitted to the data. The approach used in section 4.4 is based on pattern recognition of third order cumulants. In section 4.5 we will fit diagonal pure bilinear models denoted by $DPBL(q)$ to the Canadian river flow data discussed in Chapter 2. From the fitted models in section 4.5, we will simulate data in Section 4.6 and compare the sampling properties of the mean, standard deviation,

minimum value and maximum value to that of the original time series.

4.2 Wavelet Smoothing

Smoothing techniques can be very efficient, but the performance of the filter relies greatly on the choice of the filter coefficients; see Brockwell and Davis (1996). To begin this procedure, the user must decide upon a specific wavelet filter. The time series X_t is then allowed to pass through a linear wavelet filter which decomposes X_t into a non-random wavelet smoothed version $\hat{\eta}(t)$ and a random component $W(t)$. $W(t)$ is the remainder of the time series X_t after $\hat{\eta}(t)$ has been removed. It has some autocorrelation structure, but the underlying aspect of $W(t)$ is that it is still a time series and moreover, it has a simpler structure than X_t.

The series X_t is constructed from linear combinations of $\eta^{(j)}$ at various levels of j given by

$$X_t = \eta^{(\phi)} + \eta^{(0)} + \eta^{(1)} + \ldots + \eta^{(m)} + \ldots = \eta(t;m) + W(t) \qquad (4.7)$$

where η^{ϕ} is a multiple of a scaling function $\phi(t)$ and $\eta^{(j)}$ is a linear combination of 2^j dilated and translated versions of a primary wavelet function denoted by $\psi(t)$. The linear combination $\eta(t;m)$ can be written in terms of $\phi(t)$ and $\psi^{-j,k}(t) = 2^{j/2}\psi(2^j t - k)$ as

$$\eta(t;m) = d\phi(t) + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1} c_{jk}\ {}_N\psi^{-j,k}(t) \tag{4.8}$$

The work completed in this chapter will involve the Daubechies wavelet system generated by ${}_5\phi(x)$ and ${}_5\psi(x)$. Along with (4.8) the series X_t is given by

$$X_t = \sum_{j=1}^{N_*} q_j(t)\omega_j + W(t) \quad \text{where} \quad N_* = 2^{m+1}. \tag{4.9}$$

The components of the $N_* \times 1$ vector $\omega = (\omega_1, \ldots, \omega_{N_*})'$ are the filter coefficients $\{d, c_{jk}\}$ which will be determined from n realizations of the time series $\{X_t\}$. The vector $\mathbf{q} = (q_1(t), \ldots, q_{N_*}(t))'$ is comprised from the wavelet system chosen for the filtering process. We can see from (4.9) that the nonlinear time series $\{X_t\}$ is broken down into and described by two components. The first component is a non-random wavelet smoothed version and the second is a random process.

The rational behind using wavelet smoothing techniques is twofold. The first is to avoid the problem of trying to select a suitable nonlinear technique, from among many, to apply to our nonlinear process. The second reason is to avoid the often cumbersome and sometimes near impossible problem of estimating the parameters in complicated nonlinear models. In this section we gave a brief discussion on the methodology of wavelets and in the next section, as discussed by Oyet (1999), we will outline some theory behind the estimation of the filter coefficients. For more elaborate discussions

on the properties, uses and applications of wavelets see Daubechies (1992), Strang (1989), Alpert (1992), Antoniadis, Gregoire and McKeague (1994) and Härdle, Kerkyacharian, Picard and Tsybakov (1998).

4.3 Estimation

Without loss of generality, we will assume that the space of all possible values of t has been normalized to the $[0, 1]$ interval. Given n realizations of a nonlinear time series X_t then, from Oyet (1999), the smoothed version is evaluated as

$$\hat{\eta}(r; m) = \int_0^1 h(r; t) X_t \upsilon(t)\ d\xi(t) \tag{4.10}$$

where

$$h(r; t) = \mathbf{q}'(r)\mathbf{B}^{-1}\mathbf{q}(t); \quad \mathbf{B} = \mathbf{B}(\upsilon, \xi) = \int_0^1 \mathbf{q}(t)\mathbf{q}'(t)\upsilon(t)\ d\xi(t)$$

and $\xi(t)$ is the empirical distribution function of $\{t_i\}_{i=1}^n$. It can be established that $\hat{\eta}(r)$ is unbiased with variance

$$V(\hat{\eta}(r, m)) = \frac{R_w(0)}{n}\mathbf{q}'(r)\mathbf{B}^{-1}\mathbf{D}_1\mathbf{B}^{-1}\mathbf{q}(r) + 2\mathbf{q}'(r)\mathbf{B}^{-1}\mathbf{D}_2\mathbf{B}^{-1}\mathbf{q}(r)$$

where

$$\mathbf{D}_1 \ = \int_0^1 \mathbf{q}(t)\mathbf{q}'(t)\upsilon^2(t)\ d\xi(t)$$

$$\mathbf{D}_2 = \int_0^1 \int_{\{t>s\}} \mathbf{q}(t)\mathbf{q}'(s)v(t)v(s)R_w(t-s)\ d\xi(t)d\xi(s)$$

and $R_w(\cdot)$ is the autocovariance function of $W(t)$. At this point, it must be determined how to choose the most appropriate values for $v(t)$. The approach taken by Oyet (1999) is to choose $v(t)$ such that the Integrated Variance of $\hat{\eta}(r)$ denoted by $IV(\hat{\eta}(r))$ is minimized. The Integrated Variance of $\hat{\eta}(r)$ is given by

$$\begin{aligned} IV(\hat{\eta}(r;m)) &= \int_0^1 V(\hat{\eta}(r;m))\ dr = \frac{R_w(0)}{n} tr\{\mathbf{D}_1\mathbf{H}^{-1}\} + 2tr\{\mathbf{D}_2\mathbf{H}^{-1}\} \\ &= \frac{R_w(0)}{n}\int_0^1 \|\mathbf{H}^{-\frac{1}{2}}\mathbf{q}(t)\|^2 v(t)\ d\xi(t) \\ &\qquad + 2\int_0^1 \int_{\{t>s\}} \mathbf{q}'(s)\mathbf{H}^{-1}\mathbf{q}(t) \\ &\qquad\qquad \times v(t)v(s)R_w(t-s)\ d\xi(t)d\xi(s) \end{aligned} \tag{4.11}$$

where $\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{A} = \int_0^1 \mathbf{q}(r)\mathbf{q}'(r)\ dr$.

The IV given in (4.11) can be minimized by finding an appropriate weight function $v(t)$. The approach taken by Oyet (1999) is to search for an absolutely continuous measure which minimizes the IV loss function by allowing the measure ξ to be extended to the space of all distribution functions. The weight suggested by Oyet (1999) is

$$v_0(t;u) = \frac{u}{\|\mathbf{A}^{-\frac{1}{2}}\mathbf{q}(t)\|}; \quad u_Q = \int_0^1 \|\mathbf{A}^{-\frac{1}{2}}\mathbf{q}(t)\|\ dt \tag{4.12}$$

Then $v_0(t;u_Q)$ minimizes (4.11) under the constraints that $m(t)v(t) = 1$ and

$\int_0^1 m(t)dt = 1$.

4.4 Model Identification Based on Third Order Moments and Cumulants

Before models are fitted to the random component $W(t)$ of the series X_t using the $DPBL(q)$ process given by (4.1), the order of the process will be determined using third order moments and cumulants. If we assume third order stationarity, the third order cumulant depends only on k_1 and k_2 for all admissible integers t, k_1 and k_2 and is given by

$$C(k_1, k_2) = m(k_1, k_2) - \mu[R(k_1) + R(k_2) + R(k_1 - k_2)] - \mu^3$$

where $m(k_1, k_2) = \mathrm{E}(X_t X_{t+k_1} X_{t+k_2})$. It has been shown by Gabr (1988) that the cumulants $C(k_1, k_2)$ of a real valued process X_t have the following symmetric relationship:

$$C(k_1, k_2) = C(k_2, k_1) = C(-k_1, k_2 - k_1) = C(k_1 - k_2, -k_2).$$

From this relation, once the values in the upper half of the first quadrant of the Euclidean plane are known, then all the values of $C(k_1, k_2)$ are defined.

Table 4.1: $C(1, k_2)$ Pattern for arbitrary q

k_1 \ k_2	1	2	3	$\cdots$	q	$q+1$	$q+2$	$q+3$	$\cdots$
1	NZ	NZ	NZ	$\cdots$	NZ	NZ	0	0	$\cdots$

From Oyet (1999) we have that $C(k_1, k_2) = 0$ for $k_1 \leq q, k_2 - k_1 > q$ and $k_1 > q, k_2 - k_1 \geq q$ and $C(k_1, k_2)$ nonzero when $k_2 > k_1$. Based on these results, it has been shown that the third order cumulants define a pattern in the upper half of the first quadrant of the $k_1 k_2$ plane as shown in Table 4.1. This pattern is then easily extended to the entire Euclidean plane from the symmetric relationship satisfied by the cumulants. A useful pattern then, for detecting the order of a $DPBL(q)$ as can be seen in Table 4.1 is: $C(1, k_2) = 0$, for $k_2 = q + 2, q + 3, \ldots$ and nonzero elsewhere for an arbitrary value of q.

To determine the order of the model we will investigate the behavior of the standardized cumulants, given by

$$\rho(1, k_2) = \frac{C(1, k_2)}{C(0, 0)} \tag{4.13}$$

for a given finite sample time series X_t satisfying (4.1). We estimate the third order cumulants in (4.13) by

$$C(k_1, k_2) = \frac{1}{n - k_1 - k_2} \sum_{t=1}^{n-k_1-k_2} (X_t - \bar{X})(X_{t+k_1} - \bar{X})(X_{t+k_2} - \bar{X})$$

where,

$$\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t.$$

The order of the best model will be $k_2^* - 1$, where k_2^* is that value of k_2 at which $\rho(1, k_2)$ cuts-off. The cut-off point refers to the point at which $\rho(1, k_2) = 0$. Since the sample estimates of $\rho(1, k_2)$ will not be exactly zero, we will use the standardized cumulant trace, a plot of the absolute values of $\hat{\rho}(1, k_2)$ versus k_2, to determine the point where $\hat{\rho}(1, k_2)$ cuts off. Therefore, k_2^* will be the value of k_2 where the standardized cumulant trace begins to stabilize and hence, the order of the model will be $k_2^* - 1$.

4.4.1 Results

The model fitting and simulations will be conducted on the time series $W(t)$ given by (4.7). For this reason, the cumulant traces portrayed in this section were plotted based on $W(t)$. For all six Canadian river flow data given in Appendix A, the standardized cumulant traces were plotted and the order of the best model was selected based on the approach discussed earlier in this section.

Two of the six cumulant traces showed a distinct cut-off point while the other four did not show any clear point of stability. For the four rivers which did not have a distinct cut-off point we attempted to improve the cumulant trace through differencing. Depending on the river we used either first order, second order or third order differencing. Given a time series X_t, the first, second and third order differences are given by

$$Y_t = X_t - X_{t-1}, \quad t = 2, 3, \ldots, n \tag{4.14}$$

$$Y_t = X_t - 2X_{t-1} + X_{t-2}, \quad t = 3, 4, \ldots, n \tag{4.15}$$

$$Y_t = X_t - 3X_{t-1} + 3X_{t-2} - X_{t-3}, \quad t = 4, 5, \ldots, n \tag{4.16}$$

respectively. If at this stage the cut-off point was not 100% clear, we narrowed the possibilities down to two or three successive choices. Models were then fitted to all points and based on the AIC criteria discussed in section 3.2.4 the best model was selected. The standardized cumulant traces for each river are shown in Figures 4.1 - 4.6.

1. Model order for Peace River:

   The standardized cumulant trace in Figure 4.1 was based on $W(t)$. The trace appears to stabilize after $k_2^* = 6$, therefore the order of the model to be fitted is $DPBL(k_2^* - 1) = DPBL(5)$.

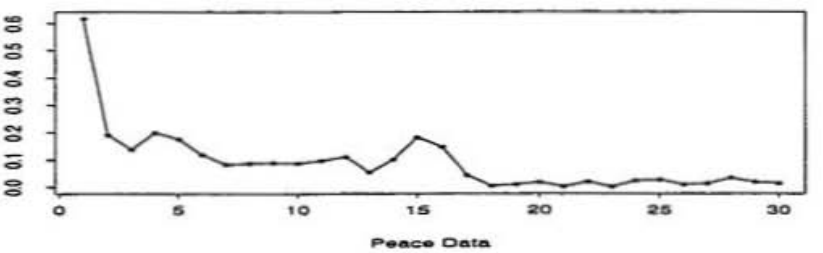

Figure 4.1: Standardized cumulant trace of Peace River

2. Model order for Castle River:

   The standardized cumulant trace in Figure 4.2 was also based solely on $W(t)$. The cut-off point in this cumulant trace is taken to be $k_2^* = 2$. The model to be fitted for this time series is $DPBL(1)$.

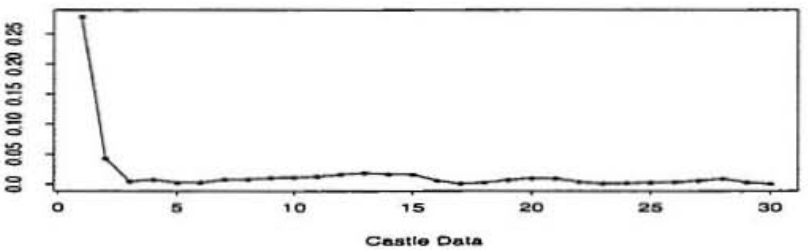

Figure 4.2: Standardized cumulant trace of Castle River

3. Model order for South River:

   When the standardized cumulant trace was constructed for $W(t)$ we could not find any apparent cut-off point to distinguish a suitable model. First order differencing was then implemented on $W(t)$ to construct the trace found in Figure 4.3. It appears that the cut-off point is at either $k_2^* = 4$ or $k_2^* = 5$. All models of order less than and equal to 4 will be fitted and the best model will be determined based on the AIC.

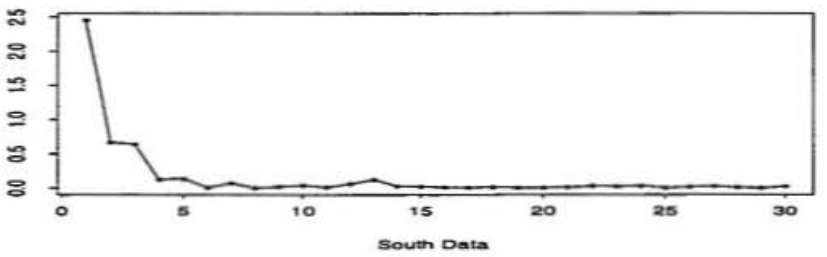

Figure 4.3: Standardized cumulant trace of South River

4. Model order for Salmonier River:

   First order differencing was also implemented on the time series $W(t)$ after a cut-off point could not be determined from the standardized cumulant trace. From Figure 4.4, which is the cumulant trace based on the differencing, shows that after $k_2^* = 4$ the trace begins to stabilize. Therefore, we will fit a $DPBL(3)$ to this series.

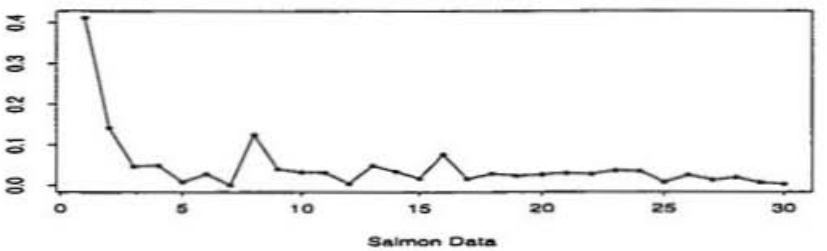

Figure 4.4: Standardized cumulant trace of Salmonier River

5. Model order for Gander River:

   In this case the trace for $W(t)$ highlighted certain regular fluctuations which suggested seasonal effects which were not apparent in the original series. For this reason, third order differencing was performed on the time series $W(t)$ and then a second standardized cumulant trace was plotted as shown in Figure 4.5. The cut-off point was not clear-cut so the values 2, 3 and 4 were selected for k_2^*. The models $DPBL(1)$, $DPBL(2)$ and $DPBL(3)$ will be fitted and the best model will then be established using the AIC criteria.

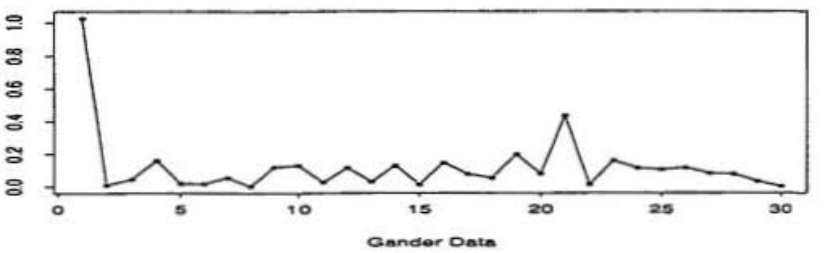

Figure 4.5: Standardized cumulant trace of Gander River

6. Model order for Moberly River:

   The standardized cumulant trace of $W(t)$ for this case also highlighted seasonal effects which were not detected in the original series. Hence, the standardized cumulant trace shown in Figure 4.6 was constructed from the second order differencing of $W(t)$. The value of the cut-off point selected is $k_2^* = 5$, hence a $DPBL(4)$ process will be used to fit a model to this time series.

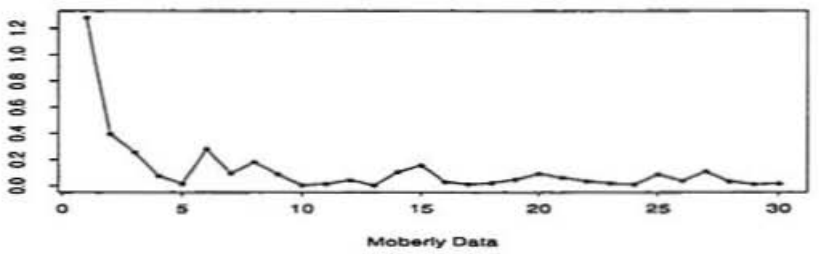

Figure 4.6: Standardized cumulant trace of Moberly River

4.5 Model Fitting to River Flow Data

Diagonal pure bilinear models, given by (4.1), were fitted to all six time series given in Appendix A after the wavelet smoothed version had been removed. The models were fitted to the time series $W(t)$ or $W'(t)$, where $W'(t)$ is $W(t)$ after implementing the appropriate differencing for each time series as discussed in the previous section along with the transformation given by (3.19). For the time series where it was suggested to fit more than one model, the best model was selected based on the AIC criteria.

For Peace River we fitted a $DPBL(5)$ and the estimated coefficients of the model are $\theta_1 = -0.1327$, $\theta_2 = 0.0825$, $\theta_3 = -0.0228$, $\theta_4 = 0.0399$ and $\theta_5 = 0.0285$. Therefore the fitted model for Peace River is

$$\begin{aligned} W'(t) &= -0.1327W'(t-1)e_{t-1} + 0.0825W'(t-2)e_{t-2} \\ &- 0.0228W'(t-3)e_{t-3} + 0.0399W'(t-4)e_{t-4} \end{aligned}$$

$$+ \quad 0.0285W'(t-5)e_{t-5} + e_t \tag{4.17}$$

where $\hat{\sigma}^2 = 0.8031$.

A $DPBL(1)$ was fitted to Castle River. The estimate value of the model parameter is $\theta_1 = 0.0079$ and the fitted model is given by

$$W'(t) = 0.0079W'(t-1)e_{t-1} + e_t \tag{4.18}$$

where $\hat{\sigma}^2 = 0.99$.

Diagonal pure bilinear models of order less than and equal to four were fitted to the time series for South River. The AIC was minimized for the $DPBL(1)$ model with a value of -133.7435. The parameter estimate for the model is $\theta_1 = -0.0636$ and the resulting model is

$$W'(t) = -0.0636W'(t-1)e_{t-1} + e_t \tag{4.19}$$

where $\hat{\sigma}^2 = 0.8282$.

For Salmonier River, we fitted a $DPBL(3)$ model to the time series. The estimates of the coefficients for the model are $\theta_1 = -0.0548$, $\theta_2 = -0.0390$ and $\theta_3 = -0.0133$. The model is given by

$$\begin{aligned} W'(t) &= -0.0548W'(t-1)e_{t-1} - 0.0390W'(t-2)e_{t-2} \\ &- 0.0133W'(t-3)e_{t-3} + e_t \end{aligned} \tag{4.20}$$

where $\hat{\sigma}^2 = 0.8015$.

The previous section proposed following three possible models for Gander River: $DPBL(1)$, $DPBL(2)$ and $DPBL(3)$. The $DPBL(2)$ model recorded the smallest value for the $AIC = -8.2037$. The estimates for the model coefficients are $\theta_1 = -0.0345$ and $\theta_2 = 0.0222$ leading to the following model

$$\begin{aligned} W'(t) &= -0.0345W'(t-1)e_{t-1} + 0.0222W'(t-2)e_{t-2} \\ &+ e_t \end{aligned} \tag{4.21}$$

where $\hat{\sigma}^2 = 0.9831$.

Two models fitted to Moberly river were $DPBL(3)$ and $DPBL(4)$. Based on the AIC criteria the better model was the $DPBL(4)$. The $AIC = -48.9466$ and the estimates of the parameters for the model are $\theta_1 = -0.0282$, $\theta_2 = -0.0016$, $\theta_3 = -0.0162$ and $\theta_4 = -0.0103$. The model is given by

$$\begin{aligned} W'(t) &= -0.0282W'(t-1)e_{t-1} - 0.0016W'(t-2)e_{t-2} \\ &- 0.0162W'(t-3)e_{t-3} - 0.0103W'(t-4)e_{t-4} \\ &+ e_t \end{aligned} \tag{4.22}$$

where $\hat{\sigma}^2 = 0.9288$.

4.6 Results of Simulation Study

At this stage we must consider whether the wavelet smoothing technique along with using the diagonal pure bilinear process to model $W(t)$ sustains specific properties of the original process, mainly the mean, standard deviation and maximum value. The properties were studied by simulating $n = 2000$ time series based on $W(t)$ or $W'(t)$. The transformation of the time series was then reversed, based on (3.26), along with any differencing that may have been invoked. We then combined the wavelet smoothed version with the simulated random linear component to obtain the following time series,

$$\tilde{X}_t = \tilde{\eta}(t;5) + W(t). \tag{4.23}$$

Next, we calculated the mean, standard deviation and maximum value for each of the $n = 2000$ simulated series and compared those sampling properties to that of the original series. Finally, we investigated whether this process was able to simulate the original time series by overlaying the time series plot of the original series with that of one of the 2000 simulated series.

Enumerated below are the results for each of the six rivers. For each river, properties of X_t are given together with some properties of the mean of $\tilde{X}_t$, of the standard deviation of $\tilde{X}_t$, and of the maximum of $\tilde{X}_t$.

1. Peace River

Table 4.2: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Peace River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	329	694	1180	1278	1490	5190	950.6962
Mean($\tilde{X}_t$)	1250	1272	1277	1277	1283	1305	8.1848
St.Dev($\tilde{X}_t$)	918.6	942.5	948.4	948.2	953.6	974.2	8.226
Max($\tilde{X}_t$)	4944	5312	5402	5410	5500	6126	141.6315

The results from the above tables show that both the mean of the means of $\tilde{X}_t$ and the mean of the standard deviations of $\tilde{X}_t$ are almost identical to that of the mean and standard deviation of the original series. The original series had a mean and standard deviation equal to 1278 and 950.6962 respectively as compared to the mean of the means and the mean of the standard deviations of $\{\tilde{X}_t\}$ which were equal to 1277 and 948.2 respectively. The mean of the max($\tilde{X}_t$) = 5410 is greater than the maximum value of the original which is equal to 5190. This is mainly due to the large range of values in the original series.

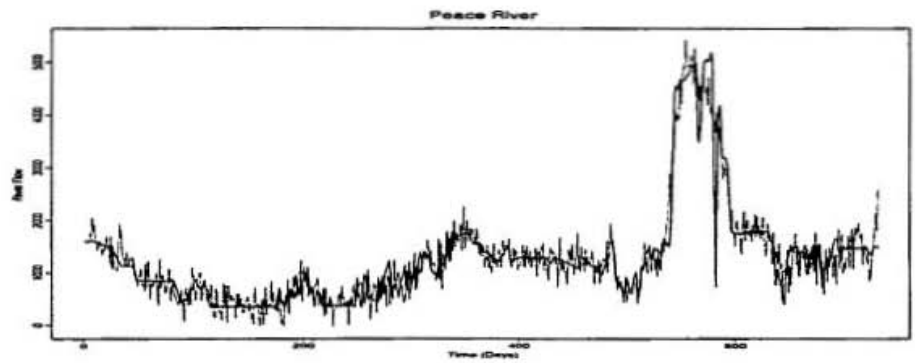

Figure 4.7: Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Peace River

Figure 4.7 consists of time series plots of both the actual series X_t and the simulated series $\tilde{X}_t$. From the plot it is observed that the simulated series follows the pattern of the original series. There are some slight variations from the original time series, but the variations are centered around the original series.

2. Castle River

Table 4.3: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Castle River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	1.62	4.135	8.1	20.47	21.2	812	41.2205
Mean($\tilde{X}_t$)	16.52	20.03	20.7	20.67	21.32	24.09	1.0052
St.Dev($\tilde{X}_t$)	38.39	40.79	41.41	41.41	42.04	44.16	0.8877
Max($\tilde{X}_t$)	199.1	228.8	239.2	240.4	250.7	306.1	16.7219

In Chapter 3, we were unsuccessful in fitting models to this time series when assuming that the series followed a general bilinear process. Based on the techniques used in this chapter, we see that the procedure has preserved the properties of the mean and standard deviation. The results of the simulation of $n = 2000$ time series are: the mean of the means of $\tilde{X}_t$ was equal to 20.67 and the mean of the standard deviations of $\tilde{X}_t$ was equal to 41.41. These values are almost identical to the original time series which had a mean of 20.47 with a standard deviation of 41.2205. The mean of the $\max(\tilde{X}_t) = 240.4$ was extremely smaller than the maximum value of $X_t = 812$.

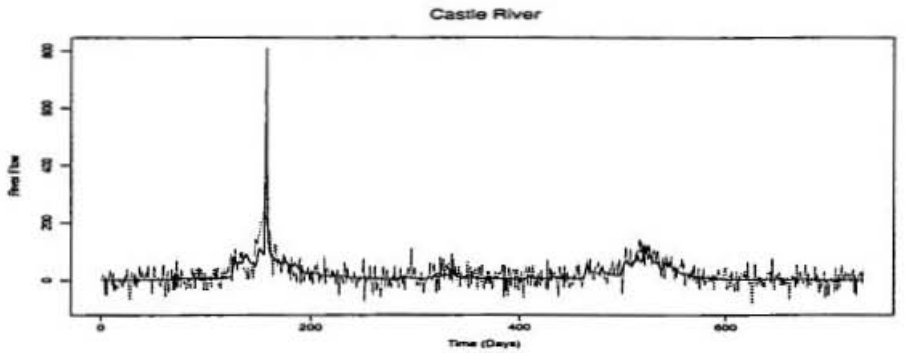

Figure 4.8: Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Castle River

Figure 4.8 displays time series plots of the original series X_t and the simulated series $\tilde{X}_t$. The fitted series followed the pattern of the original series in its entirety with only one exception. The simulated series was unable to peak at the same magnitude as the original series. The original series had flow rates of 396, 812 and 244 m^3/s on day 157, 158 and 159 respectively. The simulated series peaked at around 200

m^3/s. This also explains the large difference between the mean of the max($\tilde{X}_t$) and the max(X_t).

3. South River

Table 4.4: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (South River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	0.069	0.2605	0.416	0.6855	0.752	13	0.964
Mean($\tilde{X}_t$)	0.5394	0.6196	0.6395	0.6405	0.6616	0.7564	0.0303
St.Dev($\tilde{X}_t$)	1.167	1.247	1.265	1.265	1.283	1.34	0.0264
Max($\tilde{X}_t$)	9.976	12.39	12.95	12.96	13.54	15.44	0.8083

The mean of the means of $\tilde{X}_t$ for South River was equal to 0.6855 which was very close the the mean of X_t which had a value of 0.6405. The mean of the standard deviations of $\tilde{X}_t$ was equal to 0.964 which was slightly different from the standard deviation of the actual time series which was equal to 1.265. The values of the max($\tilde{X}_t$) and the max(X_t) were very close with values 12.96 and 13.0 respectively.

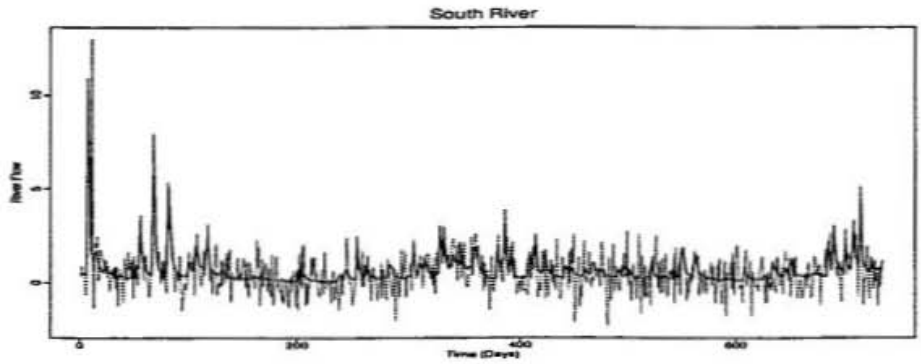

Figure 4.9: Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for South River

The time series plot of the original series X_t and the simulated series $\tilde{X}_t$ are displayed in Figure 4.9. The simulated series did not follow the pattern of the original series to the same extent as the previous two rivers.

4. Salmonier River

Table 4.5: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Salmonier River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	0.043	1.025	2.58	4.208	5	63.5	5.9732
Mean($\tilde{X}_t$)	3.259	3.677	3.784	3.788	3.897	4.323	0.1617
St.Dev($\tilde{X}_t$)	6.91	7.356	7.456	7.457	7.561	7.988	0.1544
Max($\tilde{X}_t$)	50.51	60.52	63.26	63.48	66.38	75.96	4.3091

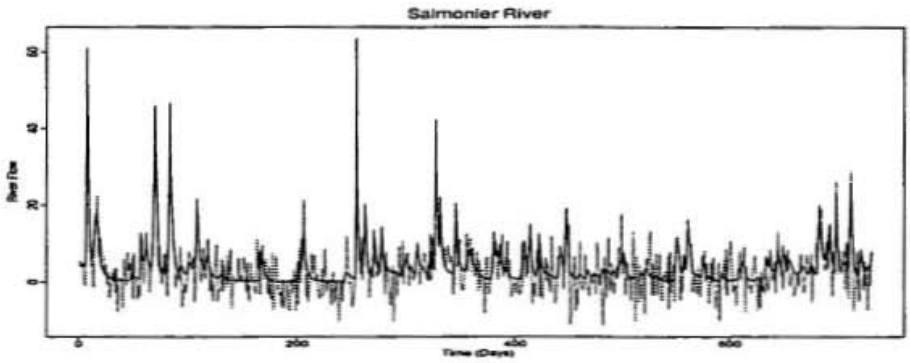

Figure 4.10: Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Salmonier River

The results of the mean and standard deviation for Salmonier River were not as precise as the results recorded for the other five rivers, however, the sampling properties did appear to be maintained. The mean of the actual time series and the mean of the means of the $\tilde{X}_t$ were 4.208 and 3.788 respectively. The standard deviation of X_t and the mean of the standard deviations of $\tilde{X}_t$ were 5.9732 and 7.457 respectively. The maximum value of $X_t = 63.5$ was almost identical to $\max(\tilde{X}_t)$ which recorded a value of 63.48. The plot of the simulated series $\tilde{X}_t$ in Figure 4.10 did not resemble the original series X_t. Although the sampling properties were preserved in the simulation, the fact that the simulated plot did not follow the pattern of the original series questions the appropriateness of the model.

5. Gander River

Table 4.6: Sampling Properties of X_t, Mean($\bar{X}_t$), St.Dev($\bar{X}_t$) and Max($\bar{X}_t$) (Gander River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	23.5	76.35	108	130.1	160	669	94.1295
Mean($\bar{X}_t$)	127.1	129.1	129.5	129.5	130	132.3	0.7032
St.Dev($\bar{X}_t$)	97.26	99.27	99.75	99.74	100.2	102.4	0.69997
Max($\bar{X}_t$)	600.5	652	664.6	664.9	677.9	753.8	18.8715

Based on the results of the mean, standard deviation and maximum value of the simulation study, we see that the procedures used in this chapter sustained these properties for the Gander River. The original series had a mean of 130.1 while the mean of the means of $\bar{X}_t$ was 129.5. The standard deviation of X_t was equal to 94.1295 while for $\bar{X}_t$ the mean of the standard deviations was 99.74. The maximum of X_t was equal to 669 while the mean of max($\bar{X}_t$) was 664.9.

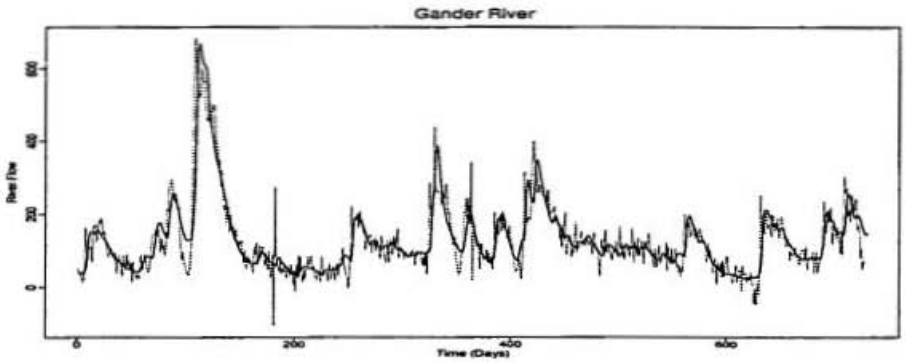

Figure 4.11: Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Gander River

With only some slight deviations from the original series, the simulated series in Figure 4.11 modeled the original series effectively.

6. Moberly River

Table 4.7: Sampling Properties of X_t, Mean($\tilde{X}_t$), St.Dev($\tilde{X}_t$) and Max($\tilde{X}_t$) (Moberly River)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev
X_t	1.24	1.84	3.24	13.17	15.25	87.1	18.7409
Mean($\tilde{X}_t$)	12.94	13.07	13.1	13.1	13.13	13.29	0.0465
St.Dev($\tilde{X}_t$)	18.89	19	19.04	19.04	19.07	19.17	0.0463
Max($\tilde{X}_t$)	86.69	88.71	89.36	89.4	90.06	92.72	0.9966

The results of Moberly River also showed convincing evidence that the method of wavelet smoothing worked extremely well when fitting models to river flow data. For Moberly River, the mean of X_t was 13.17

as compared to the mean of the means of $\tilde{X}_t$ which was equal to 13.1. The standard deviation of X_t was equal to 18.7409 which was very close the the mean of the standard deviation of $\tilde{X}_t$ which had a value of 19.04. The results for the maximum value were $\max(X_t) = 87.1$ and $\max(\tilde{X}_t) = 89.4$.

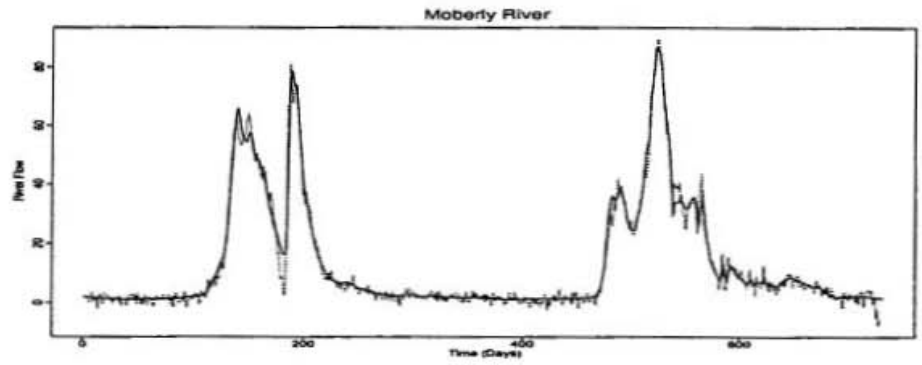

Figure 4.12: Plot of actual time series (solid line) with overlaid simulated diagonal pure bilinear time series plot (dotted line) for Moberly River

Figure 4.12 displays time series plots of the original series X_t and the simulated series $\tilde{X}_t$. The plot of the simulated series imitated the original series with the exception of some small variations that were mainly centered around the original series.

The simulated time series plots for Peace River, Gander River and Moberly River simulated their respective original time series extremely close. The simulated time series plot for Castle River was also very close to that of the original series, with the only exception being the point in the time series where a minimal number of points peaked the series far above any other interval in the series. The simulated time series plots for South River and

Salmonier River depicted similar patterns to that of their original series but were not as exact as the other four. The main reason for this is the number of ups and downs throughout the original series made it more difficult to simulate the series exactly.

The effectiveness of the procedures to model river flow data as presented in this chapter was based on two characteristics. The first component was concerned with the ability of the fitted model to preserve sampling properties of the original series, and the second pertained to the models capability to simulate the original time series. Overall, the sampling properties and the simulated plots for all six rivers gave convincing evidence that the procedures discussed throughout this chapter were very effective in modeling a two year period of river flow data.

Chapter 5

Conclusion

In many instances when analyzing time series data the time series is considered to follow a linear process and this is the best approach provided the assumption of linearity is correct. A major problem that arises when modeling time series data is assuming the time series follows a linear process when in fact it is non-linear. In many situations a time series may be generated by an underlying random structure that is non-linear and if this non-linearity is overlooked the fitted models will have no meaning. The time series considered in this practicum were tested for linearity in Chapter 2 and in all instances the time series were found to be non-linear.

In this practicum we fit time series models to the river flow variable for six rivers based on two approaches. The first approach discussed in Chapter 3 assumed the time series followed a bilinear process. The second approach discussed in Chapter 4 involved decomposing the time series into a wavelet smoothed version and a random component, using wavelet smoothing techniques, where the random component was assumed to follow a pure bilinear

process. Simulation studies for both approaches were conducted to identify whether the fitted models behaved in a similar manner to the original time series.

For both approaches certain sampling properties of the simulated series were compared to the original series. A time series plot which consisted of the simulated series and the original series was also constructed for each river. When assuming the time series followed a bilinear process there were some cases where the sampling properties of the simulated series were very close to that of the original series. However, in neither instance did the simulated plot provide encouraging evidence that the times series had an underlying structure that was bilinear. The techniques employed in Chapter 4, however, were very successful in modeling the mean river flow time series. The simulated results where almost identical to their respective original series.

The ability to appropriately model the river flow variable is a huge resource in the field of hydrology. The work completed in this practicum and especially the methods from Chapter 4, if applied to at least a 10 year river flow series, will aid hydrologists in making improved forecasts and predictions.

Bibliography

[1] Alpert, B. K. (1992), "Wavelets and Other Bases for Fast Numerical Linear Algbra," in: *Wavelets: A Tutorial in Theory and Applications,* ed. C. K. Chui, Boston: Academic Press, pp. 181-216.

[2] Antoniadis, A., Gregoire, G. and McKeague, I. W. (1994). "Wavelet Methods for Curve Estimation," *J. Amer. Stat. Ass.,* 89, 1340-52.

[3] Bhaskara Rao, M., Subba Rao, T., and Walker, A. M. (1983). "On the Existence of some Bilinear Time Series Models," *J. Time Ser. Anal.,* 4(2), 95-110.

[4] Brockwell, P. J. and Davis, R. A. (1996). *Introduction to Time Series and Forecasting,* New York: Springer-Verlag.

[5] Daubechies, I. (1992), *Ten Lectures On Wavelets,* Society for Industrial and Applied Mathematics.

[6] Gabr, M. M. (1988). "On the Third-Order Moment Structure and Bispectral Analysis of Some Bilinear Time Series," *J. Time Ser. Anal.,* 9(1), 11-20.

[7] Granger, C. W. J. and Andersen, A . P. (1978). *An Introduction to bilinear Time Series Models,* Göttingen: Vandenhoeck & Ruprecht.

[8] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation and Statistical Applications,* Springer-Verlag, New York.

[9] Hurvich, C. M. and Tsai, C. L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika,* 76, 297-307.

[10] Keenan, D. M. (1985). "A Tukey Non-additivity-type test for Time Series Nonlinearity," *Biometrika,* 72, 49-44.

[11] Liu, J. and Brockwell, P. J. (1988). "On the General Bilinear Time Series Model," *J. Appl. Prob.,* 25, 553-64.

[12] McLeod, A. I. and Li, W. K. (1983). "Diagnostic Checking ARMA Time Series Models using Squared-residual Autocorrelations," *J. Time Ser. Anal.,* 4, 269-73.

[13] Oyet, A.J. (1999). "Nonlinear Time Series Modelling: Pattern Recognition and Wavelet Filtering," *Technical Report,* Depertment of Mathematics and Statistics, Memorial University of Newfoundland, 1999-04.

[14] Phan, D. T. and Tran, L. T. (1981). "On the First Order Bilinear Time Series Model," *J. Appl. Prob.,* 18, 617-27.

[15] Quinn, B. G. (1982). "Stationarity and Invertibility of Simple Bilinear Models," *Stoch. Proc. Appl.,* 12, 225-30.

[16] Sesay, S. A. O., and Subba Rao, T. (1991). "Difference Equations for Higher Order Moments and Cumulants for the Bilinear Time Series Model BL(p,0,p,1)," *J. Time Ser. Anal.*, 12, 159-76.

[17] Strang, G. (1989). "Wavelets and Dilation Equations: A Brief Introduction," *SIAM Review*, 31, 614-27.

[18] Subba Rao, T. (1981). "On the Theory of Bilinear Time Series Models," *J. R. Stat. Soc.*, 43, **B**, 244-55.

[19] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, New York.

[20] Tsay, R. S. (1986). "Nonlinearity Test for Time Series," *Biometrika*, 73, 461-6.

Appendix A

Data Sets

Table A.1: River Flow Data For Peace River

1610 1590 1590 1590 1600 1600 1610 1620 1600 1590 1600 1580 1560 1580
1560 1520 1530 1530 1530 1520 1490 1490 1490 1490 1450 1430 1420 1360
1370 1310 1280 1220 1210 1140 1130 1130 1130 1140 1120 1120 1140 1190
1250 1280 1270 1230 1120 987 846 851 851 839 847 845 849 858 847 840 847
846 843 842 841 845 842 842 845 838 836 840 847 829 834 832 834 831 833
830 883 876 834 830 823 762 708 661 578 470 475 480 480 482 485 486 488
600 660 690 680 675 770 840 870 800 760 710 690 688 730 818 803 712 671
671 422 338 363 436 341 342 412 345 347 350 348 348 347 344 335 348 350
344 345 346 345 343 347 350 353 352 352 349 432 356 361 393 338 340 338
336 353 343 342 341 342 345 352 369 344 355 330 329 484 344 342 341 341
342 342 402 363 343 338 336 336 337 545 659 670 418 466 380 347 432 577
492 536 680 539 648 511 640 527 562 752 579 588 915 1010 995 1020 830
878 1090 915 689 614 656 550 348 401 659 568 578 613 481 392 347 344 385
381 356 354 354 355 354 354 378 378 360 357 357 387 356 355 388 426 353
350 448 403 387 367 532 353 355 354 575 902 517 589 351 346 517 935 966
869 643 495 496 526 580 468 490 457 474 476 443 546 716 529 934 605 902
1030 1060 1140 1230 1100 744 666 665 717 794 681 660 534 657 824 899 921
1040 1140 739 663 688 1210 956 735 846 931 1140 728 719 969 1260 1400
1300 1020 1350 1410 1350 1250 1350 1320 1250 1100 1050 1000 970 990 950
930 900 1070 1150 1180 1170 1160 1150 1160 1300 1700 1450 1230 1230 1650
1680 1760 1640 1700 1700 1700 1710 1710 1740 1740 1740 1740 1740 1740
1740 1740 1740 1630 1590 1550 1530 1530 1530 1530 1530 1530 1530 1510
1390 1390 1390 1380 1410 1380 1340 1340 1300 1200 1200 1230 1210 1220
1220 1220 1240 1230 1340 1330 1340 1410 1460 1470 1470 1420 1190 1180
1210 1210 1270 1270 1260 1250 1280 1280 1280 1280 1300 1310 1270 1270
1280 1290 1270 1270 1280 1260 1270 1270 1260 1250 1260 1260 1260 1250
1310 1310 1270 1260 1280 1280 1280 1200 1170 1190 1240 1290 1280 1250
1250 1260 1260 1250 1250 1240 1220 1200 1200 1190 1190 1180 1170 1150
1140 1130 1120 1100 1100 1110 1120 1130 1140 1150 1160 1170 1180 1180
1180 1170 1160 1150 1140 1130 1110 1100 1100 1100 1100 1120 1100 950 960
1000 1070 1120 1200 1600 1570 1570 1560 1550 1610 1320 1000 850 780 770
790 820 725 640 595 600 650 900 825 700 600 605 612 785 939 1010 921 566
525 737 1120 1050 1050 1140 1350 1370 1440 1430 1520 1310 1250 1420 1410
1460 1420 1440 1410 1390 1380 1310 1510 1660 1610 1530 1570 1500 1500
1820 2510 3420 4220 4550 4460 4510 4590 4660 4650 4660 4660 4660 4670
4730 4740 4760 4780 4880 4950 4940 4930 4920 4980 4710 3540 3480 3680

4160 4650 4920 5010 5020 5060 5010 5010 5100 5190 5060 3690 2320 729
2000 3800 4150 4180 3800 3250 3190 3180 3180 3110 3010 2800 2390 1820
1770 1750 1750 1760 1750 1750 1740 1730 1730 1750 1780 1740 1760 1780
1770 1730 1770 1800 1710 1780 1770 1750 1800 1760 1790 1740 1730 1810
1730 1770 1780 1780 1770 1770 1770 1650 1610 1580 1600 1330 760 897 1070
918 796 720 593 416 388 924 1200 1300 1300 1420 1540 1410 1470 1490 1500
1500 1500 949 925 1350 1420 1420 1250 1380 967 965 1140 1380 1260 1410
1530 1020 1240 1330 1530 1500 1540 1110 726 698 875 1120 1200 1300 1100
1140 1230 1340 1440 1540 1620 1650 1650 1600 1480 1400 1340 1380 1440
1500 1450 1450 1470 1490 1500 1490 1460 1450 1470 1460 1450 1460 1480
1460 1450 1460 1490 1500 1510 1500 1440 1400 1380 1340 1340 1390 1450
1500 1500 1490 1470

Table A.2: River Flow Data For Castle River

2.36 2.30 2.26 2.22 2.19 2.15 2.10 2.12 2.16 2.28 2.26 2.20 2.17 2.13 2.10
2.08 2.05 2.03 2.02 2.03 2.05 2.05 2.05 2.14 2.18 2.23 2.30 2.40 2.50 2.60 2.75
2.95 3.12 3.10 3.06 2.98 2.88 2.84 2.78 2.74 2.69 2.64 2.58 2.53 2.50 2.47 2.45
2.47 2.52 2.85 3.30 3.70 3.92 4.00 4.03 4.00 3.90 3.79 3.68 3.51 3.50 3.50 3.52
3.62 3.90 4.46 5.09 6.10 6.75 8.34 9.48 8.66 8.18 8.50 8.71 8.25 8.01 7.79 7.64
7.98 7.61 7.27 7.06 6.71 6.29 6.18 5.91 5.85 5.83 5.78 6.04 6.30 6.26 6.30 7.23
7.29 7.28 7.56 7.70 7.45 7.47 7.39 7.53 7.80 7.64 7.22 7.27 7.18 7.21 7.68 8.00
8.60 9.00 9.60 10.90 10.80 11.00 10.90 10.60 10.60 10.10 10.70 11.00 11.80
14.60 43.00 70.50 78.10 60.00 54.70 58.30 64.00 60.30 57.60 65.70 75.80 78.00
91.90 89.20 76.50 74.30 64.40 59.10 56.60 60.50 55.50 49.60 52.40 69.00 90.80
111.00 111.00 102.00 95.90 89.30 90.90 396.00 812.00 244.00 146.00 105.00
88.20 83.00 81.10 80.60 73.50 70.00 70.60 68.30 62.40 62.90 56.30 64.60 65.40
64.10 61.80 59.40 61.20 54.30 50.80 48.10 48.50 47.60 52.00 61.80 54.20 48.70
45.70 41.50 38.70 38.70 34.90 31.40 28.40 26.00 24.70 24.40 22.20 20.90 19.90
19.40 32.20 29.00 26.60 28.50 25.80 23.20 22.70 20.70 19.00 17.90 17.20 16.70
18.60 17.50 16.40 15.40 14.30 14.80 15.20 14.40 13.70 16.90 17.50 16.50 15.80
15.10 14.80 14.30 14.10 13.60 13.10 12.30 11.90 11.40 10.90 10.70 10.50 10.00
9.77 9.60 9.82 9.58 9.19 8.73 8.34 8.24 8.20 8.33 8.36 7.87 7.64 7.46 7.37 7.20
6.88 6.88 6.71 6.68 6.70 6.82 6.46 6.26 6.07 5.88 5.74 5.54 5.36 5.23 5.05 4.95
5.33 4.99 5.05 4.90 4.81 5.07 4.86 4.75 4.67 4.60 4.57 4.76 8.90 10.60 11.10
11.00 10.50 10.60 11.80 11.90 11.40 10.80 10.40 10.20 9.68 9.22 8.82 8.66 8.25
7.90 7.60 6.61 6.39 5.78 5.65 5.94 5.68 5.62 5.52 5.32 6.00 10.90 9.35 9.44 9.85
9.45 11.30 15.60 21.10 22.70 24.50 24.10 21.30 19.40 17.60 16.20 19.60 22.90
21.60 17.30 15.70 20.60 42.30 38.80 31.30 25.60 22.70 19.10 16.10 14.90 10.90
11.30 11.40 11.60 11.80 12.00 11.80 11.70 10.90 9.95 10.30 9.70 8.72 8.10 8.11
8.30 8.70 9.00 9.30 9.40 9.50 9.60 8.50 7.50 7.00 6.99 6.18 5.41 4.74 4.15 4.19
4.88 5.68 6.60 7.68 8.95 7.85 7.40 6.97 6.57 6.19 5.83 5.63 5.45 5.26 5.08 4.90
4.74 4.58 4.42 4.26 4.12 3.98 3.86 3.74 3.63 3.53 3.41 3.31 3.41 3.49 3.97 4.69
5.53 6.54 7.71 9.82 12.50 15.90 14.80 13.70 9.37 9.04 9.51 8.92 8.42 7.96 8.38
7.74 7.60 7.47 7.36 7.23 7.10 7.02 6.94 6.84 6.76 6.68 6.70 8.98 12.00 11.20
10.80 10.40 10.00 9.80 9.57 9.35 9.10 8.89 8.69 8.53 8.39 8.24 8.10 7.96 7.82
7.92 7.69 7.53 7.38 7.31 7.27 7.21 7.17 7.13 7.09 7.05 8.05 10.40 14.20 22.90
38.60 56.40 50.30 41.20 33.70 29.40 27.70 29.40 30.60 28.80 27.40 25.70 24.10
22.70 22.40 43.60 39.30 33.50 29.70 26.60 24.90 23.90 23.70 23.30 22.20 22.00
21.40 21.90 22.10 21.00 19.70 18.80 18.80 20.20 25.10 37.60 48.20 57.10 64.90
69.20 65.90 59.70 53.90 52.00 64.90 63.40 61.30 64.80 66.90 82.70 88.90 79.60

70.90 64.00 64.70 76.10 118.00 125.00 101.00 94.50 117.00 123.00 103.00 86.30
76.20 67.10 65.30 63.70 63.10 65.50 61.60 50.30 40.50 35.50 46.40 45.10 56.90
63.00 57.30 55.00 55.40 52.40 48.40 45.50 45.10 45.00 45.90 42.20 35.90 31.60
28.80 27.30 26.80 25.60 23.50 21.70 20.40 19.50 19.10 18.80 18.00 16.30 15.10
14.70 14.20 13.70 13.50 13.60 13.70 13.10 11.90 11.80 12.20 10.10 9.61 9.08
8.76 8.42 8.74 8.33 7.82 7.50 7.25 7.04 6.86 6.59 6.52 6.41 6.24 6.06 5.96 5.59
5.42 5.31 5.24 5.04 4.89 4.58 4.51 4.45 4.47 4.34 4.29 4.09 3.76 3.78 3.73 3.74
3.90 4.13 3.94 3.77 3.71 3.70 3.64 3.53 3.49 3.47 3.45 3.38 3.35 3.66 5.15 4.46
4.17 4.19 4.27 4.19 4.15 4.14 4.10 4.00 3.89 3.96 4.07 4.20 4.26 4.29 4.27 4.33
4.32 4.25 4.22 4.16 4.09 4.04 4.03 4.01 3.98 4.04 4.00 3.90 3.91 3.95 3.82 3.68
3.64 3.67 3.66 3.59 3.48 3.45 3.46 3.46 3.45 3.44 3.44 3.28 3.13 3.01 2.95 2.82
2.56 2.46 2.35 2.23 2.14 2.02 1.95 1.87 1.80 1.74 1.71 1.69 1.68 1.62 1.66 1.65
1.66 1.67 1.68 1.72 1.83 1.97 2.07 2.05 2.03 1.98 1.94 1.92 1.91 1.92 1.94 1.96
2.01 2.10 2.12 2.11 2.10 2.09 2.08 2.07 2.06 2.04 2.04 2.03 2.04 2.04 2.04 2.00
1.97 1.96 1.94 1.96 1.97 1.98 1.99

Table A.3: River Flow Data For South River

0.560 0.510 0.475 0.450 0.430 0.418 0.740 9.490 10.300 3.510 1.520 13.000
0.900 0.650 1.510 1.620 1.690 1.250 0.948 0.809 0.688 0.596 0.538 0.654 0.730
0.693 0.627 0.560 0.500 0.460 0.420 0.385 0.350 0.320 0.295 0.272 0.340 0.440
0.390 0.340 0.320 0.290 0.270 0.260 0.245 0.235 0.280 0.320 0.280 0.260 0.240
0.260 0.940 0.850 0.740 2.790 3.580 1.640 1.550 1.200 0.922 0.775 0.650 0.570
0.500 0.440 0.390 0.900 7.500 6.160 2.600 1.280 1.250 0.681 0.690 0.540 0.500
0.470 0.440 0.410 0.600 4.580 5.060 3.430 2.330 1.370 1.100 0.922 0.754 0.615
0.542 0.500 0.460 0.420 0.400 0.380 0.370 0.355 0.340 0.650 1.140 0.884 0.705
0.637 0.726 1.230 1.980 1.950 1.420 1.130 0.921 0.848 0.848 0.780 0.685 1.640
1.670 2.350 1.800 1.150 0.850 0.687 0.605 0.533 0.500 0.479 0.456 0.440 0.440
0.475 0.581 0.642 0.601 0.532 0.494 0.451 0.397 0.346 0.312 0.293 0.279 0.267
0.257 0.284 0.279 0.257 0.244 0.215 0.195 0.191 0.191 0.199 0.204 0.201 0.213
0.227 0.216 0.192 0.182 0.325 0.356 0.310 0.268 0.247 0.245 0.278 0.395 0.392
0.313 0.262 0.233 0.228 0.228 0.193 0.166 0.158 0.159 0.149 0.137 0.131 0.122
0.115 0.107 0.131 0.176 0.174 0.162 0.141 0.125 0.101 0.089 0.087 0.090 0.088
0.087 0.081 0.078 0.082 0.081 0.084 0.544 0.714 0.488 0.360 0.378 0.474 0.414
0.301 0.240 0.195 0.168 0.153 0.143 0.139 0.144 0.132 0.116 0.103 0.096 0.094
0.090 0.083 0.076 0.069 0.074 0.076 0.074 0.072 0.072 0.071 0.069 0.073 0.076
0.097 0.102 0.113 0.212 0.405 0.371 0.316 0.271 0.228 0.372 0.735 0.576 0.469
0.399 0.301 0.245 0.207 0.178 0.164 0.169 2.440 1.520 0.793 0.535 0.450 0.420
0.366 0.337 0.643 0.687 0.511 0.409 0.342 0.303 0.278 0.262 0.249 0.236 0.231
0.222 0.209 0.201 0.220 0.265 0.286 0.278 0.266 0.254 0.649 0.977 0.842 0.770
0.550 0.360 0.290 0.295 0.300 0.270 0.262 0.255 0.242 0.226 0.230 0.234 0.231
0.378 0.405 0.358 0.310 1.310 2.110 1.070 0.644 0.447 0.390 1.050 1.030 0.743
0.677 0.972 1.280 1.070 0.731 0.584 0.503 0.450 0.413 0.394 0.416 0.426 0.393
0.491 0.522 3.020 2.490 1.710 1.400 2.950 1.770 1.200 0.910 0.717 0.764 0.984
0.853 1.530 1.350 1.040 1.130 1.040 1.900 2.140 1.950 1.240 1.080 0.920 0.825
0.750 0.690 0.875 0.850 0.825 0.910 1.100 1.660 1.880 1.950 1.290 0.973 0.769
0.637 0.590 0.540 0.460 0.380 0.320 0.280 0.250 0.232 0.217 0.198 0.186 0.175
0.170 0.260 2.000 1.920 0.950 0.600 1.190 1.160 1.540 3.870 1.800 0.911 0.617
0.541 1.400 1.180 0.721 0.529 0.420 0.340 0.298 0.270 0.245 0.235 0.222 0.210
0.202 0.228 0.245 0.320 1.090 1.260 1.620 0.932 1.050 1.880 1.420 2.470 1.810
1.000 0.660 0.509 0.464 0.490 0.612 1.450 1.710 1.270 0.903 0.658 0.539 0.470
0.410 0.380 0.350 0.320 0.340 0.362 0.330 0.300 0.278 0.265 0.255 0.249 0.311
0.836 0.867 0.711 0.591 0.512 0.582 0.758 0.672 0.508 0.414 0.383 0.381 0.349
0.310 0.298 0.307 0.338 0.402 0.707 0.662 0.604 0.505 0.443 0.453 0.464 0.623

0.988 1.080 0.916 0.750 0.632 0.580 0.661 0.713 0.688 0.586 0.480 0.422 0.385
0.354 0.329 0.307 0.288 0.272 0.290 0.296 0.294 0.285 0.265 0.422 0.710 0.560
0.439 0.373 0.317 0.294 0.313 0.365 0.377 0.428 0.713 0.606 0.456 0.365 0.312
0.291 0.293 0.291 0.281 0.259 0.228 0.210 0.199 0.189 0.187 0.210 0.307 0.318
0.291 0.252 0.775 1.070 0.691 0.505 0.423 0.356 0.302 0.285 0.262 0.241 0.234
0.241 0.343 0.362 0.329 0.284 0.261 0.247 0.234 0.225 0.220 0.300 0.420 0.525
0.394 0.310 0.288 0.419 1.670 1.840 1.630 1.220 0.876 0.616 0.471 0.391 0.368
0.338 0.294 0.647 1.660 1.090 1.070 0.835 0.590 0.469 0.432 0.396 0.381 0.333
0.302 0.280 0.262 0.250 0.236 0.226 0.212 0.199 0.190 0.188 0.186 0.191 0.176
0.156 0.144 0.134 0.124 0.166 0.170 0.159 0.178 0.224 0.212 0.192 0.184 0.154
0.139 0.125 0.130 0.127 0.118 0.132 0.127 0.121 0.135 0.229 0.905 0.693 0.463
0.376 0.310 0.248 0.224 0.201 0.189 0.174 0.172 0.173 0.182 0.192 0.190 0.188
0.227 0.413 0.408 0.327 0.277 0.258 0.256 0.254 0.260 1.150 1.030 0.656 0.734
0.695 0.502 0.396 0.356 0.316 0.283 0.265 0.774 0.759 0.503 0.429 0.532 1.310
1.080 0.706 0.491 0.405 0.407 0.416 0.405 0.379 0.368 0.354 0.333 0.315 0.296
0.285 0.282 0.303 0.365 0.383 0.389 0.434 0.533 0.491 0.414 0.423 0.506 0.480
0.411 0.371 0.340 0.315 0.301 0.558 2.150 1.630 0.957 0.670 1.170 3.050 2.380
1.600 1.130 0.974 1.010 0.808 0.649 0.569 0.504 1.830 2.340 1.410 0.942 0.718
0.682 1.030 1.790 2.780 2.490 1.530 1.080 0.880 1.100 4.440 3.000 1.570 1.110
0.897 0.776 0.684 0.623 0.615 0.825 1.320 0.906 0.747 0.635 0.591 0.774 0.729
0.684 0.583 0.714 1.140

Table A.4: River Flow Data For Salmonier River

5.600 5.000 4.650 4.300 3.850 5.000 8.000 53.500 45.000 16.200 8.000 5.600
7.000 9.500 13.900 17.200 18.600 14.800 10.300 6.650 4.000 3.400 2.750 2.250
1.950 1.700 1.500 1.300 1.150 1.050 0.940 0.860 0.790 0.740 0.675 0.630 0.580
0.540 0.505 0.478 0.445 0.422 0.400 0.380 0.360 0.345 0.325 0.900 1.750 1.500
1.320 1.170 1.100 1.020 0.910 0.830 12.800 8.000 5.600 5.600 7.500 9.550
7.250 5.800 4.400 3.200 4.200 9.000 20.000 46.000 30.000 18.000 6.000 3.800
2.400 2.850 3.550 3.400 3.200 2.800 2.600 18.000 46.600 28.200 17.900 9.880
7.700 6.760 5.120 4.110 3.060 2.710 3.900 4.460 4.090 3.000 2.570 2.030 2.030
3.660 5.230 4.900 3.970 3.400 3.980 5.150 6.960 19.000 17.300 10.300 7.010
5.720 4.640 3.950 3.270 4.820 6.540 7.180 4.790 3.490 2.690 2.090 1.640 1.440
1.310 1.160 0.976 0.947 1.370 2.020 2.340 1.940 1.610 1.320 0.958 0.797 0.693
0.592 0.553 0.545 0.554 0.568 0.644 0.621 0.556 0.519 0.502 0.496 0.428 0.300
0.342 0.382 0.386 0.391 0.411 0.411 0.379 0.315 0.287 0.403 0.452 0.366 0.238
0.282 0.511 5.160 8.010 5.580 3.350 2.140 1.580 1.200 0.850 0.609 0.531 0.471
0.399 0.290 0.221 0.178 0.206 0.211 0.163 0.143 0.109 0.108 0.109 0.104 0.094
0.078 0.083 0.083 0.079 0.076 0.068 0.072 0.081 0.070 0.043 0.124 5.340 4.500
4.160 3.250 9.310 13.400 7.550 4.930 3.490 2.340 1.690 1.360 1.050 0.824 0.636
0.495 0.428 0.401 0.354 0.264 0.238 0.227 0.213 0.200 0.203 0.210 0.228 0.212
0.207 0.218 0.195 0.173 0.150 0.124 0.111 0.112 0.115 0.117 0.124 0.128 0.143
0.142 0.438 2.530 1.960 2.250 2.140 1.690 1.310 1.110 0.884 0.765 2.640 63.500
24.200 9.310 5.350 5.340 4.980 3.780 7.080 20.300 11.800 6.610 4.470 2.990
2.390 1.800 1.490 13.000 10.100 7.640 4.670 3.400 2.660 7.960 12.000 7.810
5.600 3.410 2.540 3.850 3.710 2.970 2.450 2.710 2.770 2.390 2.510 2.690 2.370
1.980 1.660 1.520 1.360 3.450 4.910 4.320 6.330 5.250 3.920 3.030 3.220 4.250
3.510 2.770 2.300 1.930 8.130 7.680 6.210 4.690 4.210 4.310 3.880 3.120 2.570
2.550 2.240 2.020 3.630 7.820 6.330 4.880 3.610 3.040 42.400 24.400 13.500
9.090 22.100 13.200 8.800 5.680 4.350 3.710 3.300 3.000 2.700 2.500 2.300
2.200 2.080 4.800 20.400 17.000 9.480 5.750 4.870 4.100 3.450 3.100 2.800
3.300 4.200 5.400 6.900 8.500 7.250 6.200 5.400 4.700 4.100 3.550 3.150 2.800
2.250 1.900 1.650 1.400 1.200 1.100 0.910 0.815 0.740 0.690 0.650 2.000 4.000
10.000 7.000 9.000 8.000 6.750 6.000 9.800 7.000 4.600 2.800 2.150 4.000 3.600
3.000 2.400 1.750 1.400 1.200 1.060 0.950 0.850 0.775 0.700 0.650 0.625 0.600
4.000 9.970 10.700 9.560 5.670 5.430 7.080 10.300 14.800 9.710 4.500 3.030
2.040 1.530 1.550 3.560 12.500 10.500 8.320 4.950 3.680 2.950 2.700 2.400
2.200 1.950 1.750 1.600 1.450 1.350 1.250 1.150 1.070 1.000 1.150 3.100 9.290
7.540 5.060 3.490 2.670 19.000 14.000 10.000 4.200 2.600 2.000 1.600 1.450

1.280 1.120 1.000 0.950 1.500 2.400 1.700 1.400 1.200 1.080 0.860 1.600 3.200
2.600 2.100 1.750 1.320 1.600 1.800 2.200 1.900 1.700 1.480 1.300 1.200 1.050
3.600 3.000 1.800 1.450 3.400 3.000 2.520 2.120 2.000 2.060 2.680 3.090 2.340
1.760 1.460 1.150 2.530 4.140 5.640 5.140 5.010 7.900 6.530 4.510 3.070 2.120
1.600 1.510 1.320 1.220 1.190 1.010 0.873 0.740 0.674 0.603 0.779 2.700 2.180
1.750 1.460 2.710 3.560 3.230 2.700 2.280 1.680 1.280 1.150 0.910 0.923 1.990
1.850 1.650 1.210 0.970 0.799 0.650 0.482 0.375 0.312 0.255 0.424 1.060 0.836
0.666 0.575 0.536 0.546 6.190 11.400 10.700 9.440 6.340 3.890 2.500 3.690
7.120 5.070 3.230 7.640 16.300 10.800 9.360 6.640 4.100 2.580 2.120 1.960
2.100 1.920 1.590 1.300 1.310 6.330 4.880 3.480 2.390 1.680 1.260 1.530 1.550
1.210 0.949 0.774 0.633 0.554 0.465 0.489 0.492 0.391 0.843 2.300 1.390 1.050
0.969 0.917 0.760 0.582 0.485 0.412 0.363 0.719 0.813 0.600 0.538 0.817 3.810
2.720 4.040 5.000 5.800 4.500 2.920 2.050 1.430 0.974 0.767 0.607 0.565 0.535
0.500 0.450 0.650 1.380 1.250 1.150 1.100 1.030 0.960 0.930 0.940 6.750 5.400
4.000 6.500 5.000 4.000 3.100 3.200 2.900 2.280 2.450 8.400 6.000 4.400 3.400
3.800 9.500 6.500 5.200 4.200 3.250 5.200 6.000 5.000 4.200 3.600 2.920 2.550
2.180 1.930 2.250 2.800 3.400 4.200 4.000 3.600 3.200 2.550 2.200 2.000 3.000
4.950 4.200 3.650 3.180 2.800 2.450 2.180 7.000 20.100 14.000 10.000 7.500
5.950 7.250 6.300 10.500 8.000 14.000 15.200 7.430 5.540 4.750 3.560 23.600
18.000 7.830 5.180 3.550 2.570 3.740 5.280 5.240 4.670 4.890 3.780 4.670
13.000 26.000 15.600 8.750 5.530 3.350 2.290 1.770 2.270 3.650 8.620 8.070
5.620 3.830 2.730 2.630 4.000 3.600 3.450 4.400 8.000 7.000

Table A.5: River Flow Data For Gander River

40.8 40.2 39.2 39.3 39.1 38.8 40.0 48.2 66.3 103.0 134.0 149.0 152.0 152.0
152.0 148.0 151.0 155.0 155.0 153.0 150.0 145.0 138.0 135.0 132.0 128.0 122.0
116.0 109.0 104.0 98.3 92.5 88.5 83.7 80.6 78.4 74.8 71.1 68.2 65.9 63.3 60.5
58.4 55.5 52.9 51.4 51.2 48.3 47.0 46.0 44.2 42.6 42.5 41.6 41.3 43.9 44.7 47.5
56.5 66.3 75.4 80.7 83.5 84.9 84.3 83.2 82.2 80.0 81.1 91.4 115.0 141.0 161.0
172.0 177.0 175.0 169.0 159.0 152.0 146.0 139.0 134.0 143.0 167.0 204.0 233.0
251.0 258.0 256.0 246.0 232.0 217.0 203.0 189.0 179.0 165.0 152.0 143.0 136.0
130.0 128.0 128.0 129.0 130.0 136.0 156.0 203.0 294.0 431.0 575.0 653.0 669.0
658.0 634.0 615.0 612.0 607.0 603.0 593.0 549.0 499.0 464.0 445.0 426.0 403.0
381.0 364.0 349.0 337.0 325.0 312.0 300.0 289.0 275.0 263.0 248.0 233.0 216.0
201.0 185.0 173.0 161.0 150.0 140.0 131.0 126.0 120.0 111.0 103.0 98.0 93.1
90.1 86.9 81.3 78.3 75.9 72.4 70.9 68.2 65.8 65.3 66.8 67.9 65.9 65.2 66.4 74.0
85.2 92.7 96.8 95.0 92.5 88.9 84.9 80.2 76.3 72.6 70.8 68.8 65.2 62.0 59.7 58.4
64.4 78.7 82.5 77.1 71.3 66.1 63.0 60.6 56.6 53.1 50.0 48.7 45.9 41.9 39.3 39.2
39.2 37.9 36.7 36.6 38.2 41.5 47.0 52.8 54.6 55.6 56.7 56.7 54.5 53.6 52.3 52.4
51.8 50.2 47.7 45.9 44.8 42.3 39.9 39.4 38.2 35.3 33.6 39.4 45.9 47.2 47.3 48.2
48.1 46.8 47.1 47.3 46.6 46.8 48.0 52.7 57.6 59.5 59.3 57.7 63.4 74.2 83.1 87.0
87.3 85.2 80.9 77.6 73.3 71.5 98.4 158.0 187.0 191.0 187.0 190.0 194.0 189.0
180.0 176.0 169.0 160.0 152.0 141.0 132.0 126.0 116.0 120.0 122.0 128.0 126.0
120.0 116.0 113.0 112.0 111.0 110.0 105.0 104.0 106.0 111.0 114.0 113.0 113.0
114.0 117.0 119.0 121.0 121.0 116.0 111.0 105.0 103.0 101.0 103.0 101.0 97.3
94.5 93.8 88.4 87.3 86.9 86.9 86.5 86.7 86.4 87.0 88.8 89.0 89.8 89.7 93.2 95.1
95.3 93.7 92.5 91.8 90.2 90.8 94.4 96.8 97.1 97.8 114.0 189.0 237.0 255.0 285.0
363.0 389.0 375.0 342.0 308.0 275.0 253.0 228.0 210.0 193.0 179.0 171.0 163.0
153.0 145.0 140.0 133.0 126.0 118.0 113.0 105.0 100.0 97.5 96.9 98.7 102.0
113.0 134.0 163.0 186.0 198.0 210.0 205.0 192.0 175.0 158.0 143.0 132.0 123.0
114.0 107.0 100.0 92.5 88.5 82.7 77.9 76.4 72.4 70.8 69.3 74.6 73.4 77.0 84.9
115.0 163.0 185.0 189.0 189.0 189.0 196.0 201.0 191.0 178.0 163.0 149.0 137.0
125.0 113.0 104.0 96.0 87.5 83.2 76.6 73.1 73.8 83.5 98.2 119.0 129.0 129.0
130.0 170.0 258.0 294.0 291.0 272.0 251.0 233.0 231.0 261.0 312.0 346.0 348.0
340.0 322.0 305.0 290.0 275.0 255.0 240.0 228.0 216.0 204.0 192.0 185.0 175.0
180.0 189.0 195.0 190.0 181.0 172.0 168.0 166.0 162.0 154.0 140.0 132.0 139.0
142.0 140.0 132.0 128.0 122.0 115.0 110.0 107.0 109.0 115.0 124.0 122.0 121.0
122.0 125.0 126.0 127.0 133.0 139.0 141.0 139.0 135.0 129.0 124.0 118.0 112.0
108.0 105.0 102.0 97.2 91.4 89.5 86.8 85.2 92.7 105.0 113.0 116.0 116.0 114.0
112.0 111.0 109.0 105.0 102.0 96.0 91.4 90.7 87.8 85.6 93.0 104.0 113.0 115.0

115.0 113.0 109.0 108.0 108.0 109.0 109.0 110.0 110.0 109.0 105.0 101.0 102.0
106.0 108.0 109.0 109.0 110.0 109.0 106.0 101.0 100.0 96.3 92.4 89.4 86.2 82.9
78.0 75.0 75.2 76.2 76.6 81.1 84.7 89.3 89.8 87.7 83.4 80.1 78.5 78.3 76.5 74.4
71.2 67.6 68.9 76.3 85.9 92.3 93.0 92.9 92.1 88.8 85.1 82.1 83.0 114.0 167.0
192.0 196.0 191.0 182.0 172.0 165.0 160.0 154.0 146.0 137.0 126.0 122.0 127.0
131.0 130.0 124.0 117.0 108.0 101.0 93.7 86.3 81.0 74.9 69.8 64.6 59.3 57.5
55.9 53.2 49.4 47.0 46.4 43.8 40.8 39.8 39.1 37.6 37.0 36.0 36.8 36.9 35.5 35.2
34.9 34.5 32.8 31.6 30.2 29.1 28.3 27.4 26.0 25.0 23.9 23.5 23.5 25.8 26.3 27.8
28.1 28.0 27.3 26.7 26.2 26.4 30.0 33.8 37.9 39.9 64.1 149.0 199.0 208.0 212.0
212.0 207.0 199.0 192.0 186.0 170.0 166.0 172.0 173.0 167.0 160.0 155.0 160.0
163.0 161.0 156.0 150.0 142.0 131.0 122.0 121.0 119.0 115.0 110.0 105.0 98.7
93.8 90.9 88.1 84.4 82.3 78.7 77.7 77.4 76.3 75.7 77.8 76.3 78.3 79.6 80.6 78.9
76.9 76.1 76.2 78.5 80.4 80.9 78.9 77.7 78.9 87.2 101.0 120.0 161.0 198.0 207.0
204.0 198.0 190.0 185.0 182.0 178.0 168.0 158.0 151.0 149.0 149.0 151.0 153.0
150.0 144.0 144.0 173.0 228.0 251.0 256.0 252.0 239.0 223.0 206.0 192.0 190.0
196.0 196.0 197.0 190.0 187.0 177.0 166.0 156.0 145.0 143.0 148.0

Table A.6: River Flow Data For Moberly River

2.05 2.03 2.00 1.97 1.93 1.85 1.82 1.79 1.76 1.75 1.70 1.65 1.64 1.62 1.59 1.58
1.57 1.56 1.56 1.55 1.52 1.48 1.45 1.42 1.40 1.39 1.38 1.37 1.36 1.35 1.34 1.33
1.33 1.32 1.32 1.31 1.29 1.28 1.27 1.26 1.25 1.24 1.25 1.24 1.25 1.25 1.26 1.27
1.28 1.29 1.29 1.30 1.30 1.30 1.31 1.31 1.32 1.33 1.34 1.36 1.35 1.34 1.33 1.32
1.33 1.34 1.35 1.36 1.38 1.42 1.45 1.47 1.49 1.50 1.51 1.50 1.51 1.51 1.52 1.52
1.53 1.54 1.55 1.58 1.62 1.64 1.67 1.72 1.75 1.79 1.82 1.85 1.86 1.88 1.92 1.97
2.01 2.10 2.15 2.20 2.21 2.24 2.29 2.36 2.42 2.48 2.53 2.63 2.72 2.85 3.00 3.20
3.60 4.50 5.50 6.40 7.19 7.37 8.18 8.95 9.72 10.40 11.70 12.70 13.80 14.70
15.40 16.70 18.90 22.00 25.90 31.20 36.60 41.90 47.80 53.60 58.50 62.60 65.40
66.00 64.40 61.50 58.80 57.00 56.10 55.20 54.60 54.50 55.80 57.30 57.90 56.80
54.80 52.60 50.50 49.20 48.40 48.50 48.10 46.90 45.60 44.20 43.10 41.70 40.50
38.20 36.10 33.80 31.90 30.30 28.90 27.20 25.60 24.80 23.80 22.60 21.40 19.80
18.60 17.40 16.50 16.30 16.40 17.70 26.60 53.30 66.70 76.50 78.90 77.30 73.50
74.00 73.40 69.10 64.00 58.80 53.60 49.10 45.00 40.50 37.80 35.20 33.00 30.70
28.60 26.70 25.10 23.40 21.80 19.90 18.70 17.80 16.40 15.10 14.30 13.40 12.50
11.50 10.80 10.20 9.68 9.09 8.33 8.02 8.00 7.92 7.63 7.62 7.24 6.83 6.59 6.36
6.45 6.61 6.76 6.93 6.93 6.94 6.84 6.82 6.57 6.57 6.62 6.56 6.16 6.17 5.67 5.38
5.23 5.04 4.89 4.74 4.64 4.70 4.84 4.62 4.20 4.08 3.99 3.89 3.77 3.64 3.55 3.41
3.33 3.23 3.12 2.98 2.91 2.88 3.00 3.01 2.95 2.94 2.95 2.84 2.74 2.65 2.62 2.57
2.55 2.57 2.65 2.43 2.29 2.30 2.32 2.43 2.56 2.58 2.62 2.85 2.81 2.77 2.78 2.69
2.70 2.80 2.79 3.11 3.24 2.95 2.68 2.50 2.37 2.32 2.30 2.25 2.00 1.87 1.86 1.92
2.15 2.40 2.50 2.55 2.54 2.52 2.47 2.41 2.38 2.35 2.32 2.29 2.26 2.23 2.20 2.17
2.14 2.10 2.03 1.99 1.98 1.99 2.00 1.98 1.96 1.90 1.88 1.85 1.84 1.83 1.84 1.85
1.86 1.88 1.90 1.92 1.94 1.95 1.97 1.98 1.99 2.00 2.03 2.04 2.02 2.00 1.99 1.97
1.95 1.93 1.90 1.88 1.87 1.87 1.84 1.78 1.75 1.74 1.74 1.73 1.72 1.70 1.65 1.60
1.54 1.45 1.39 1.35 1.30 1.26 1.25 1.26 1.27 1.29 1.31 1.32 1.33 1.32 1.30 1.29
1.27 1.28 1.29 1.30 1.32 1.34 1.35 1.39 1.40 1.43 1.45 1.49 1.48 1.47 1.47 1.50
1.54 1.55 1.54 1.51 1.50 1.47 1.45 1.42 1.41 1.41 1.40 1.40 1.40 1.40 1.41 1.42
1.43 1.44 1.45 1.47 1.48 1.48 1.49 1.52 1.55 1.56 1.59 1.60 1.61 1.61 1.61 1.61
1.60 1.59 1.58 1.56 1.54 1.52 1.51 1.50 1.50 1.50 1.52 1.55 1.60 1.62 1.69 1.72
1.78 1.80 1.82 1.85 1.89 1.92 1.99 2.05 2.10 2.20 2.40 2.60 3.40 4.20 4.90 5.80
7.00 9.00 11.00 14.00 18.00 22.00 27.00 32.00 35.00 36.40 36.10 34.70 35.40
35.60 35.70 37.70 38.50 37.80 36.80 35.50 34.80 33.00 31.00 28.70 27.30 26.20
25.20 25.10 24.60 24.90 24.60 25.90 28.60 29.90 31.30 33.30 36.20 38.80 41.90
45.60 50.50 53.40 54.30 57.00 64.10 71.00 75.20 78.00 80.70 84.10 86.40 87.10
85.00 82.10 78.60 74.50 70.40 66.40 62.50 59.10 57.40 54.00 49.00 42.00 36.00

30.20 33.00 34.30 34.20 34.00 34.00 34.90 35.00 34.10 33.60 32.40 32.00 31.90
32.00 32.70 33.50 34.10 34.90 35.50 35.90 35.50 34.00 30.60 25.50 27.00 31.00
34.00 38.00 34.00 31.00 27.00 23.80 21.00 18.80 17.00 15.00 13.40 12.50 11.50
10.80 10.00 9.60 9.00 8.50 7.70 8.50 10.00 14.00 11.50 9.00 7.60 7.20 9.50
11.40 13.00 12.20 11.80 11.40 10.60 10.40 10.00 9.50 8.80 8.40 8.20 8.00 7.60
7.20 6.91 6.61 6.55 7.85 8.63 7.77 7.11 6.71 6.56 6.54 7.10 7.33 7.25 7.11 6.96
6.79 6.70 9.14 6.95 6.78 6.71 6.59 6.61 6.57 6.45 6.23 6.02 5.90 5.74 5.64 6.28
6.44 6.50 7.21 7.64 7.67 8.00 8.06 8.41 8.58 8.66 8.74 8.89 8.67 8.38 8.24 7.77
7.66 7.58 7.94 7.29 7.01 6.94 7.06 6.77 6.57 6.37 6.25 6.01 5.70 5.60 5.58 5.60
5.60 5.60 5.59 5.50 5.42 5.20 4.90 4.59 4.20 4.00 3.65 3.40 3.15 2.95 2.80 2.60
2.45 2.32 2.28 2.23 2.20 2.18 2.17 2.16 2.13 2.11 2.10 2.09 2.08 2.08 2.09 2.10
2.13 2.15 2.18 2.19 2.20 2.22 2.25 2.26 2.29 2.31 2.32 2.33 2.32 2.32 2.31 2.30
2.28 2.25 2.20 2.18 2.13 2.08 2.00 1.96 1.90 1.88 1.84 1.80 1.79 1.77 1.75

Appendix B

Time Series Plots

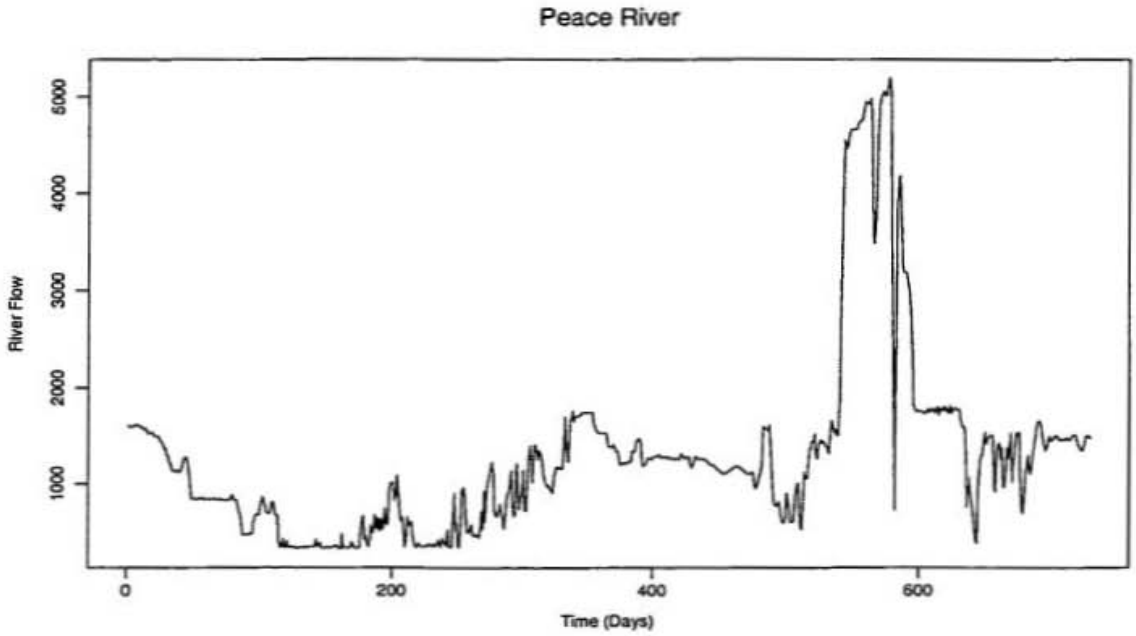

Figure B.1: Time series plot for Peace River

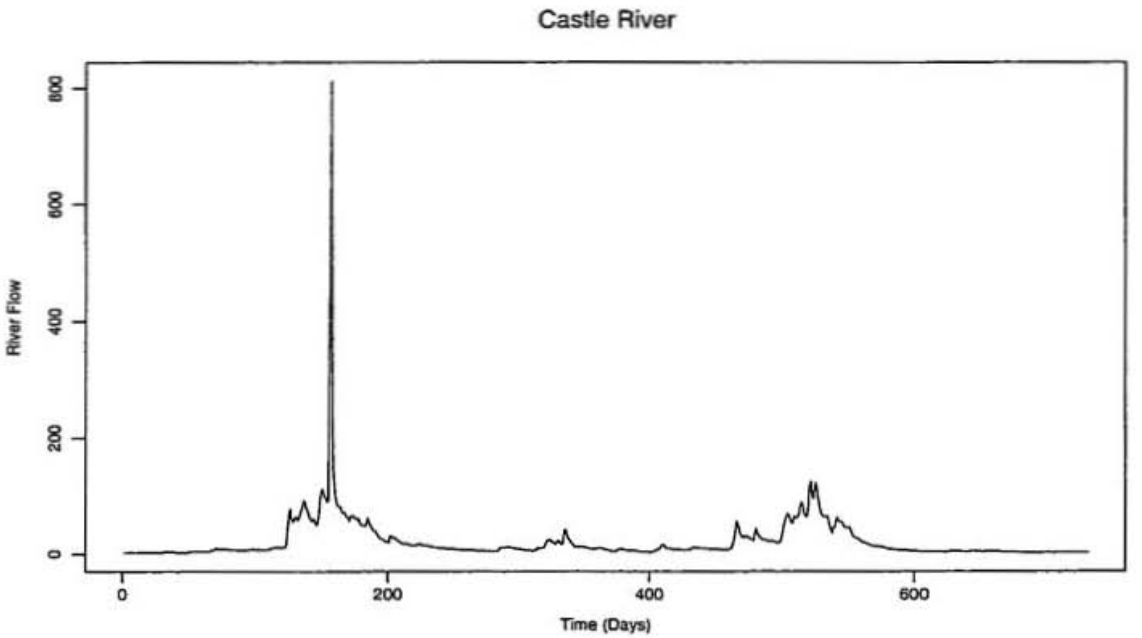

Figure B.2: Time series plot for Castle River

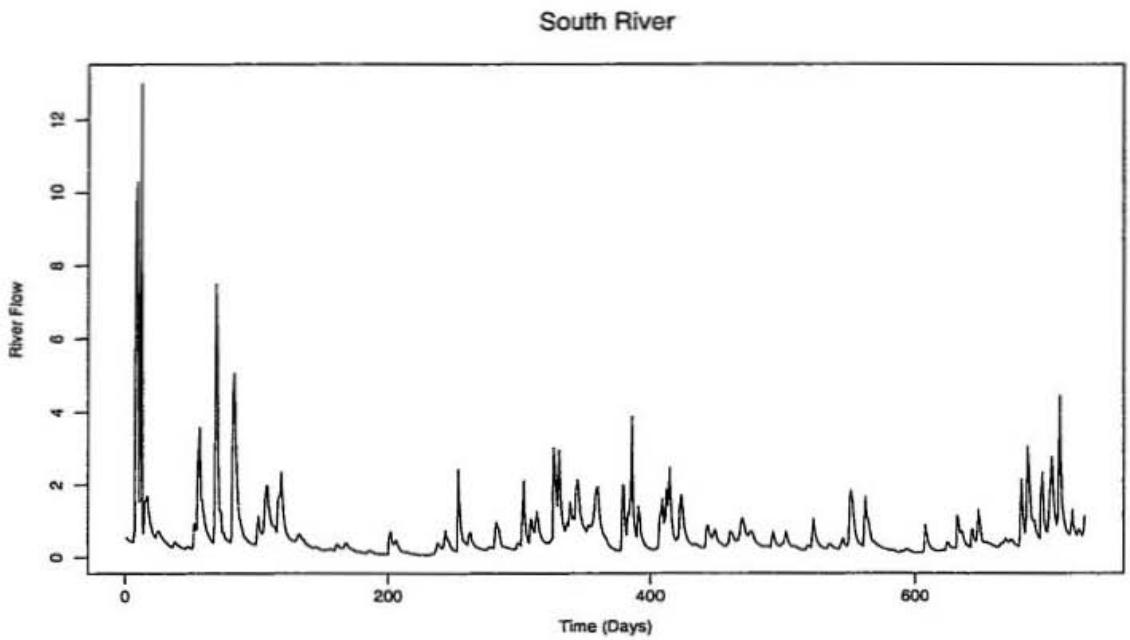

Figure B.3: Time series plot for South River

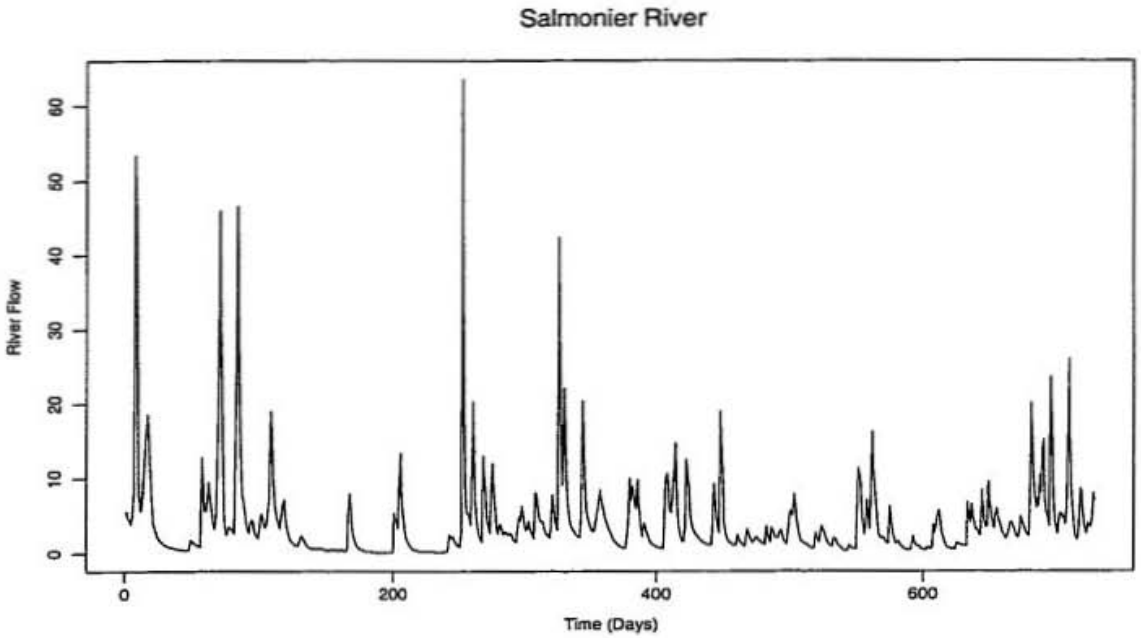

Figure B.4: Time series plot for Salmonier River

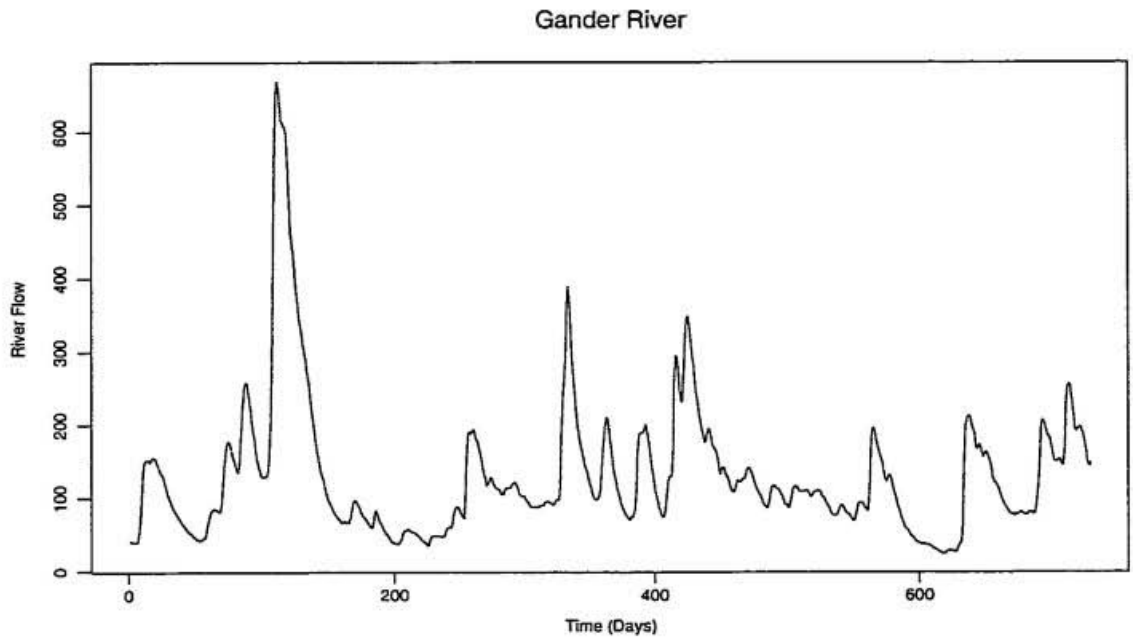

Figure B.5: Time series plot for Gander River

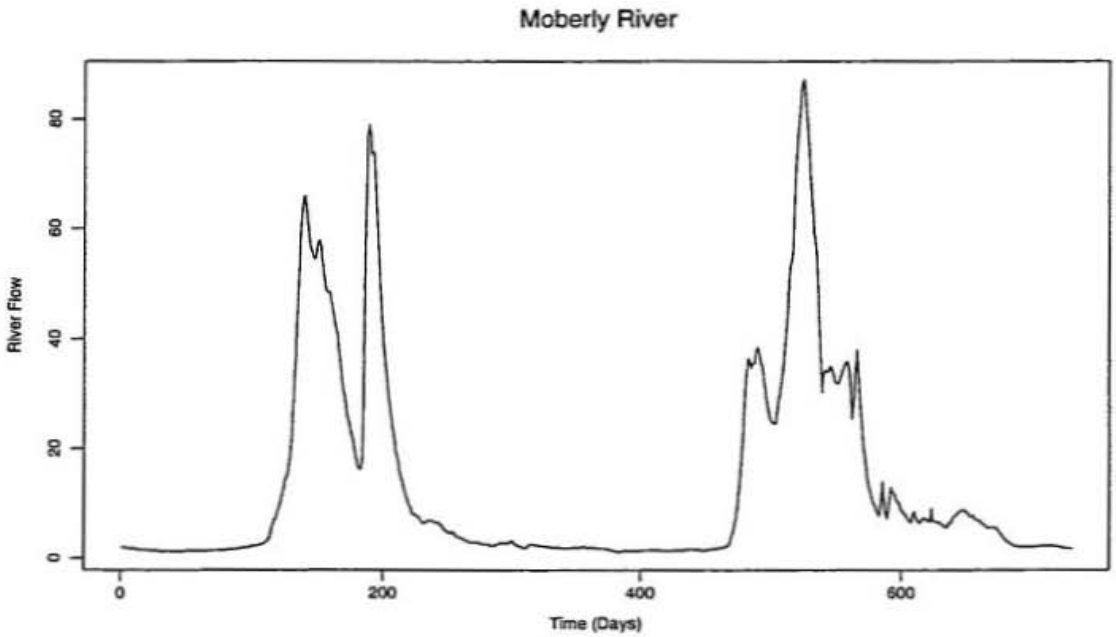

Figure B.6: Time series plot for Moberly River