Analysis of System Drought for Manitoba Hydro Using Stochastic Methods

by

Bertuğ Akıntuğ

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil Engineering

University of Manitoba

Winnipeg, Manitoba

Canada

Copyright ©2006 by Bertuğ Akıntuğ

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

Analysis of System Drought for Manitoba Hydro Using Stochastic Methods

BY

Bertuğ Akintuğ

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

OF

DOCTOR OF PHILOSOPHY

Bertuğ Akintuğ © 2006

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Abstract

Stochastic time series models are commonly used in the analysis of large-scale water resources systems. In the stochastic approach, synthetic flow scenarios are generated and used for the analysis of complex events such as multi-year droughts. Conclusions drawn from such analyses are only plausible to the extent that the underlying time series model realistically represents the natural variability of flows. Traditionally, hydrologists have favoured autoregressive moving average (ARMA) models to describe annual flows. In this research project, a class of model called Markov-Switching (MS) model (also referred to as a Hidden Markov model) is presented as an alternative to conventional ARMA models. The basic assumption underlying this model is that a limited number of flow regimes exists and that each flow year can be classified as belonging to one of these regimes. The persistence of and switching between regimes is described by a Markov chain. Within each regime, it is assumed that annual flows follow a normal distribution with mean and variance that depend on the regime. The simplicity of this model makes it possible to derive a number of model characteristics analytically such as moments, autocorrelation, and crosscorrelation. Model estimation is possible with the maximum likelihood method implemented using the Expec-

tation Maximization (EM) algorithm. The uncertainty in the model parameters can be assessed through Bayesian inference using Markov Chain Monte Carlo (MCMC) methods.

A Markov-Switching disaggregation (MSD) model is also proposed in this research project to disaggregate higher-level flows generated using the MS model into lower-level flows. The MSD model preserves the additivity property because for a given year both the higher-level and lower-level variables are generated from normal distributions.

The 2-state MS and MSD models are applied to Manitoba Hydro's system along with more conventional first order autoregressive and disaggregation models and parameter and missing data uncertainty are identified in the analysis of system drought.

Acknowledgements

First of all, I would like to thank my supervisor Assoc. Prof. Dr. Peter Rasmussen for his support. During my study, I have walked into his office with problems so many times and left his office with motivation and encouragement. His guidance and comments influenced the quality of this thesis.

It was my pleasure to have Prof. Dr. Jose D. Salas as my external examiner. I am also very grateful to the other members of my examining committee, Prof. Dr. A. Thavaneswaran, and Assoc. Prof. Dr. K. Snelgrove for their useful comments.

I gratefully acknowledge the financial support provided by Manitoba Hydro and Natural Sciences and Engineering Research Council of Canada (NSERC) for this research project. Special thanks to Bill Girling, Bruce Hinton, and Harold Surminski from Manitoba Hydro for helpful advises.

I also gratefully acknowledge the bursaries and scholarships provided by the University of Manitoba and the Government of Manitoba during my study.

A very special thank to my family. I am forever grateful to my wife Nesibe and my sons Hasan and Tecelli. I appreciate for their love, understanding, and patience. I would like to dedicate this thesis to my wife Nesibe because it would not be possible

to get this degree without her support. As she always does, she stood beside me and gave me encouragement to carry on during this long PhD journey. Most importantly, she sacrificed from her PhD study during this period. Now it is my turn to give all my support to her for her PhD study in the near future.

Contents

Page

Contents v

List of Tables xii

List of Figures xiv

1 Introduction 1

1.1 Introduction 1

1.2 Objectives of the Study 9

1.3 Structure of the Report 11

2 Manitoba Hydro's System 12

2.1 Introduction 12

2.2 Major Rivers and Hydrological Components in Manitoba Hydro's System 14

2.2.1 Winnipeg River 15

2.2.2 Saskatchewan River 15

2.2.3 Churchill River 16

2.2.4 Partial Inflow Available as Outflow (PIAO) 17

2.2.5 Local Flows 18

2.3 Regulation of Reservoirs in the System 19

2.3.1 Regulation of Lake Winnipeg 19

2.3.2 Regulation of Southern Indian Lake 25

2.4 Hydro-power Generating Stations 27

3 Drought Definition for Manitoba Hydro **29**

3.1 Introduction 29

3.2 General Definition of Drought 30

3.3 Definition of Hydrological Drought 31

3.4 Definition of Energy Drought 34

3.5 Definition of Energy Drought for Manitoba Hydro's System 36

4 Traditional Stochastic Models for Data Generation **41**

4.1 Introduction 41

4.2 Marginal Distributions and Data Transformation 44

4.2.1 2-Parameter Lognormal Transformation 45

4.2.2 3-Parameter Lognormal Transformation 46

4.2.3 Box-Cox Transformation 46

4.2.4 Selection of the Transformation Method 47

4.3 Generation of Aggregated Flows 47

4.3.1 Single-Site Autoregressive Model 48

4.3.2 Parameter Uncertainty through Bayesian Inference 50

4.3.3 Incorporating Parameter Uncertainty into the Single-Site AR(1) Model 53

4.3.4 Multi-Site Autoregressive Model 57

4.3.5 Incorporating Parameter Uncertainty into the Multi-Site AR(1) Model 60

4.4 Disaggregation of Aggregated Flows 64

4.4.1 Valencia-Schaake Disaggregation Model 67

4.4.2 Mejia and Rousselle Disaggregation Model 70

4.4.3 Lane's Condensed Disaggregation Model 71

4.4.4 Streamflow Adjustment Procedure 74

5 Single-Site Markov-Switching Model **77**

5.1 Introduction 77

5.2 Formulation of the Single-site MS model 82

5.3 Theoretical Properties of the MS model 83

5.3.1 Stationary Probabilities of the Markov Chain 84

5.3.2 Marginal Distribution of Observed Variables 85

5.3.3 Moments of the MS Model 86

5.3.4 Autocorrelation Structure of the MS Model 88

5.4 Parameter Estimation of the MS Model 90

5.5 Estimation of State Probability Sequences 98

5.5.1 Forward-Backward Algorithm 99

5.5.2 Kim's Algorithm 102

5.6 Identification of the MS Model 105

5.7 Incorporating Parameter Uncertainty into the Single-Site MS Model . . 106

5.7.1 Markov Chain Monte Carlo Methods 106

5.7.2 Application of the Gibbs Sampler to the Single-Site MS Model . 110

6 Multi-Site Markov-Switching Model 121

6.1 Introduction 121

6.2 Formulation of the Multi-Site MS model 122

6.3 Parameter Estimation of the Multi-Site MS Model 123

6.4 Crosscorrelation Structure of the MS Model 128

6.5 Incorporating Parameter and Data Uncertainty into Multi-Site MS Model131

6.5.1 Application of the Gibbs Sampler to the Multi-Site MS Model . 131

7 Disaggregation Using Markov-Switching Model 139

7.1 Introduction 139

7.2 Model Formulation 140

7.3 Parameter Estimation 141

7.4 Preliminary Analysis 143

7.5 Discussion 149

8 Application to Manitoba Hydro's System 150

8.1 Introduction 150

8.2 Historical Data Sets 151

8.3 Selection of Modeling Sites and Water Year 152

8.3.1 Basin Sites 153

8.3.2 Sub-basin Sites 154

8.3.3 Key Sites 156

8.4 Stochastic Modeling Frameworks 156

8.5 Application of the Traditional Models 160

8.5.1 Marginal Distributions and Data Transformation 160

8.5.2 ARMA-Type Model Selection 162

8.6 Application of the 2-state Single-site MS Model to Manitoba Hydro Data 163

8.6.1 Incorporating Parameter Uncertainty 169

8.7 Application of the 2-State Multi-site MS Model to Manitoba Hydro Data 175

8.7.1 Incorporating Parameter and Data Uncertainty 180

8.8 Application of the 2-State MS Disaggregation Model 183

8.9 Verification of Annual Models 185

8.10 Energy Drought Frequency Analysis for Manitoba Hydro's System . . . 187

8.10.1 Frequency Analysis Results and Discussion 188

9 Conclusions **198**

References **204**

A Evidence of Long-term Cycles in Manitoba Hydro's Annual Streamflow **217**

A.1 Introduction 217

B Probability Density Functions and Normal Probability Plots for the Niagara River Monthly Flows **220**

B.1 Introduction 220

C Comparison of Simulated and Observed Cross-Correlations for the Niagara River Monthly Flows **227**

C.1 Introduction 227

D Probability Plots of the Simulated Niagara River Monthly Flows **234**

D.1 Introduction 234

E The Single-site 2-state MS Model: Fitted Marginal Distributions **241**

E.1 Introduction 241

F The Single-site 2-state MS Model: State Probabilities and Regime Changes in Observed Streamflow **245**

F.1 Introduction 245

G The Single-site 2-state MS Model: Autocorrelation Function Comparison **249**

G.1 Introduction 249

H Multi-site MS Model Probability Plots **254**

H.1 Introduction 254

I Autocorrelation Function Comparison of MS Models **258**

I.1 Introduction 258

J The Multi-site 2-state MS Model: Posterior Distributions 262

J.1 Introduction 262

K The Spatial 2-state MS Disaggregation Model: Fitted Marginal Distributions 266

K.1 Introduction 266

L The Spatial 2-state MS Disaggregation Model: Regime Changes in Observed Streamflow 270

L.1 Introduction 270

M Verification of Stochastic Models 274

M.1 Modeling Framework 1.1 274

M.2 Modeling Framework 1.3 276

M.3 Modeling Framework 2.1 278

M.4 Modeling Framework 2.3 280

N Conditional Drought Frequency Curves 282

N.1 Introduction 282

List of Tables

Table		Page
2.1	Major hydrologic components of Manitoba Hydro's system.	15
2.2	Local Flows on the Burntwood and Nelson River.	18
2.3	Maximum preferred release during winter months.	22
2.4	Minimum outflows of Lake Winnipeg.	24
2.5	Licensed release limits for Notigi and Missi control stations.	26
2.6	Manitoba Hydro's generating stations.	28
3.1	Flow-to-power factors and inflow sites for generating stations.	36
3.2	Monthly energy demands for Manitoba Hydro's system.	38
3.3	Historical drought statistics of Manitoba Hydro's system.	40
8.1	Mean, standard deviation, and coefficient of skewness of observed flows	154
8.2	Coefficients of NR5 and NR6 Local Flows	156
8.3	Selected models in Framework-1.	160
8.4	Selected models in Framework-2.	161
8.5	Selected marginal distributions for basin sites and key sites.	162

8.6 AIC values for annual ARMA-type model identification. 163

8.7 BIC values for MS Model identification. 164

8.8 Estimated ML-parameters and associated model properties of the single-site 2-state MS model. 166

8.9 Estimated ML-parameters of the multi-site MS model and moments. . 177

8.10 Modeled and observed cross-correlation of Manitoba Hydro annual runoff using the multi-site MS model. 179

8.11 ML-parameters and moments of the MS spatial disaggregation model . 184

8.12 Modeled and observed cross-correlation of Manitoba Hydro annual runoff using the MSD model 185

8.13 Statistics of calculated return periods of given drought severities using model MF-1.1. 191

8.14 Return Period (T) of critical drought (3309 GWh) using all modeling frameworks and data sets. 194

List of Figures

Figure Page

2.1 Major drainage basins contributing to Manitoba Hydro's system . . . 13

2.2 Manitoba Hydro flow system 14

2.3 Schematic diagram of Manitoba Hydro generating stations. 27

3.1 Drought parameters using Theory of Runs. 32

3.2 Duration and severity of a drought event for Manitoba Hydro. 39

4.1 Stochastic modeling approach. 42

4.2 The Bayes theorem. 52

5.1 Cycles observed in PDO-Index and Columbia River 78

5.2 Cycles observed in Annual Aggregated Basin Flows 79

7.1 Niagara River mixture pdf and probability plot for January flows. . . . 144

7.2 Niagara River simulated monthly mean values using VSD and MSD models. 145

7.3 Niagara River simulated monthly std. deviations using VSD and MSD models. 145

7.4 Niagara River simulated monthly skewness using VSD and MSD models. 146

7.5 Niagara River simulated and observed cross-correlation pairs for January flows 147

7.6 Niagara River simulated January flows probability plot. 148

8.1 Observed and missing Manitoba Hydro streaflow data. 151

8.2 Stochastic Modeling Framework-1 158

8.3 Stochastic Modeling Framework-2 159

8.4 The MS model pdf and probability plot for Annual Aggregated Basin Flows 167

8.5 State probabilities and regime shifts for annual Aggregated Basin Flows. 167

8.6 Autocorrelation function for annual Aggregated Basin Flows 169

8.7 2-State MS model parameters simulated using Gibbs sampling. 173

8.8 Posterior pdf of 2-State MS model parameters for Aggregated Basin Flows 173

8.9 Correlograms for the simulated parameters of the 2-state MS model for Aggregated Basin Flows. 174

8.10 Saskatchewan River. Normal probability plot of observed annual flows and fitted multi-site MS model. 177

8.11 Saskatchewan River. Observed and multi-site MS model autocorrelation. 178

8.12 Drought frequency curves obtained from all modeling frameworks (1912-98). 189

8.13 Drought frequency curves obtained from all modeling frameworks (1930-98). 190

8.14 Conditional drought frequency curves obtained from MF-2.5 (1912-98). 193

8.15 Aggregated Basin Flows (1912-98). Streamflow drought frequency curves obtained using the Modeling Framework-1. 196

A.1 Churchill River annual flows 5-yr running average. 218

A.2 Saskatchewan River annual flows 5-yr running average. 218

A.3 Lake Winnipeg annual PIAO 5-yr running average. 218

A.4 Winnipeg River annual flows 5-yr running average. 219

A.5 Aggregated annual Local Flows 5-yr running average. 219

A.6 Aggregated annual Basin Flows 5-yr running average. 219

B.1 Niagara River mixture pdf and probability plot for January flows. . . . 221

B.2 Niagara River mixture pdf and probability plot for February flows. . . . 221

B.3 Niagara River mixture pdf and probability plot for March flows. 222

B.4 Niagara River mixture pdf and probability plot for April flows. 222

B.5 Niagara River mixture pdf and probability plot for May flows. 223

B.6 Niagara River mixture pdf and probability plot for June flows. 223

B.7 Niagara River mixture pdf and probability plot for July flows. 224

B.8 Niagara River mixture pdf and probability plot for August flows. 224

B.9 Niagara River mixture pdf and probability plot for September flows. . . 225

B.10 Niagara River mixture pdf and probability plot for October flows. . . . 225

B.11 Niagara River mixture pdf and probability plot for November flows. . . 226

B.12 Niagara River mixture pdf and probability plot for December flows. . . 226

C.1 Niagara River simulated and observed cross-correlations for January flows. 228

C.2 Niagara River simulated and observed cross-correlations for February flows. 228

C.3 Niagara River simulated and observed cross-correlations for March flows. 229

C.4 Niagara River simulated and observed cross-correlations for April flows. 229

C.5 Niagara River simulated and observed cross-correlations for May flows. 230

C.6 Niagara River simulated and observed cross-correlations for June flows. 230

C.7 Niagara River simulated and observed cross-correlations for July flows. 231

C.8 Niagara River simulated and observed cross-correlations for August flows. 231

C.9 Niagara River simulated and observed cross-correlations for September flows. 232

C.10 Niagara River simulated and observed cross-correlations for October flows. 232

C.11 Niagara River simulated and observed cross-correlations for November flows. 233

C.12 Niagara River simulated and observed cross-correlations for December flows. 233

D.1 Niagara River probability plot for simulated January flows. 235

D.2 Niagara River probability plot for simulated February flows. 235

D.3 Niagara River probability plot for simulated March flows. 236

D.4 Niagara River probability plot for simulated April flows. 236

D.5 Niagara River probability plot for simulated May flows. 237

D.6 Niagara River probability plot for simulated June flows. 237

D.7 Niagara River probability plot for simulated July flows. 238

D.8 Niagara River probability plot for simulated August flows. 238

D.9 Niagara River probability plot for simulated September flows. 239

D.10 Niagara River probability plot for simulated October flows. 239

D.11 Niagara River probability plot for simulated November flows. 240

D.12 Niagara River probability plot for simulated December flows. 240

E.1 Churchill River. The 2-state MS model pdf and probability plot. 242

E.2 Saskatchewan River. The 2-state MS model pdf and probability plot. . 242

E.3 Lake Winnipeg PIAO. The 2-state MS model pdf and probability plot. 243

E.4 Winnipeg River. The 2-state MS model pdf and probability plot. . . . 243

E.5 Aggregated Local Flows. The 2-state MS model pdf and probability plot.244

E.6 Aggregated Basin Flows. The 2-state MS model pdf and probability plot.244

F.1 Churchill River. State probabilities and regime shifts 246

F.2 Saskatchewan River. State probabilities and regime shifts 246

F.3 Lake Winnipeg PIAO. State probabilities and regime shifts 247

F.4 Winnipeg River. State probabilities and regime shifts 247

F.5 Aggregated Local Flows. State probabilities and regime shifts 248

G.1 Churchill River. Observed and modeled autocorrelation functions. . . . 250

G.2 Saskatchewan River. Observed and modeled autocorrelation functions. 251

G.3 Lake Winnipeg PIAO. Observed and modeled autocorrelation functions. 251

G.4 Winnipeg River. Observed and modeled autocorrelation functions. . . . 252

G.5 Aggregated Local Flows. Observed and modeled autocorrelation functions. 252

G.6 Aggregated Basin Flows. Observed and modeled autocorrelation functions. 253

H.1 Churchill River. Normal probability plot and fitted multi-site MS model 255

H.2 Saskatchewan River. Normal probability plot and fitted multi-site MS model. 255

H.3 Lake Winnipeg PIAO. Normal probability plot and fitted multi-site MS model. 256

H.4 Winnipeg River. Normal probability plot and fitted multi-site MS model.256

H.5 Aggregated Local Flows. Normal probability plot and fitted multi-site MS model. 257

I.1 Churchill River. Observed and modeled autocorrelation functions. . . . 259

I.2 Saskatchewan River. Observed and modeled autocorrelation functions. 259

I.3 Lake Winnipeg PIAO. Observed and modeled autocorrelation functions. 260

I.4 Winnipeg River. Observed and modeled autocorrelation functions. . . . 260

I.5 Aggregated Local Flows. Observed and modeled autocorrelation functions. 261

J.1 Posterior distribution of transition probabilities for Multi-site MS model. 263

J.2 Churchill River annual flows. Posterior distribution of state means and standard deviations for Multi-site MS model. 263

J.3 Saskatchewan River annual flows. Posterior distribution of state means and standard deviations for Multi-site MS model. 264

J.4 Lake Winnipeg PIAO. Posterior distribution of state means and standard deviations for Multi-site MS model. 264

J.5 Winnipeg River annual flows. Posterior distribution of state means and standard deviations for Multi-site MS model. 265

J.6 Annual Aggregated Local Flows. Posterior distribution of state means and standard deviations for Multi-site MS model. 265

K.1 Churchill River. The 2-state MS disaggregation model pdf and probability plot. 267

K.2 Saskatchewan River. The 2-state MS disaggregation model pdf and probability plot. 267

K.3 Lake Winnipeg PIAO. The 2-state MS disaggregation model pdf and probability plot. 268

K.4 Winnipeg River. The 2-state MS disaggregation model pdf and probability plot. 268

K.5 Aggregated Local Flows. The 2-state MS disaggregation model pdf and probability plot. 269

L.1 The MSD model state probabilities 271

L.2 Churchill River. Regime shifts 271

L.3 Saskatchewan River. Regime shifts 272

L.4 Lake Winnipeg PIAO. Regime shifts 272

L.5 Winnipeg River. Regime shifts 273

L.6 Local River. Regime shifts 273

N.1 Conditional Drought Frequency Curves obtained from MF-1.1 (1912-98). 283

N.2 Conditional Drought Frequency Curves obtained from MF-1.2 (1912-98). 284

N.3 Conditional Drought Frequency Curves obtained from MF-1.3 (1912-98). 285

N.4 Conditional Drought Frequency Curves obtained from MF-1.4 (1912-98). 286

N.5 Conditional Drought Frequency Curves obtained from MF-2.1 (1912-98). 287

N.6 Conditional Drought Frequency Curves obtained from MF-2.2 (1912-98). 288

N.7 Conditional Drought Frequency Curves obtained from MF-2.3 (1912-98). 289

N.8 Conditional Drought Frequency Curves obtained from MF-2.4 (1912-98). 290

N.9 Conditional Drought Frequency Curves obtained from MF-2.5 (1912-98). 291

N.10 Conditional Drought Frequency Curves obtained from MF-1.1 (1930-98). 292

N.11 Conditional Drought Frequency Curves obtained from MF-1.2 (1930-98). 293

N.12 Conditional Drought Frequency Curves obtained from MF-1.3 (1930-98). 294

N.13 Conditional Drought Frequency Curves obtained from MF-1.4 (1930-98). 295

N.14 Conditional Drought Frequency Curves obtained from MF-2.1 (1930-98). 296

N.15 Conditional Drought Frequency Curves obtained from MF-2.2 (1930-98). 297

N.16 Conditional Drought Frequency Curves obtained from MF-2.3 (1930-98). 298

N.17 Conditional Drought Frequency Curves obtained from MF-2.4 (1930-98). 299

N.18 Conditional Drought Frequency Curves obtained from MF-2.5 (1930-98). 300

Chapter 1

Introduction

1.1 Introduction

Energy supply is one of the most important problems facing the world today. The availability of energy, preferably cheap energy, is a premise for industrial growth. Electricity is commonly generated in hydro, thermal, and nuclear plants. There are environmental concerns regarding nuclear power. The environmental impacts of the Chernobyl nuclear disaster in 1986, the world's worst nuclear power accident, will persist many years into the future. Fossil fuels (natural gas, coal, and oil) are the primary sources of thermal power. However, fossil fuel sources are non-renewable and will be exhausted in a foreseeable future. Hydropower is renewable, clean, and cheap, but is subject to uncertainty in resources because the energy generation of a hydropower station is directly dependent on the availability of water. Extended low flows or droughts reduce the system output and eventually may make it impossible to

meet the required energy demand. Hence, understanding the frequency and severity of droughts is important in order to quantify the reliability of energy supply.

Manitoba Hydro is responsible for the production of electricity to the benefit of the population of Manitoba. Over 96% of 5143 MW total installed capacity of Manitoba Hydro is derived from 14 hydropower stations in Manitoba. Approximately 60% of its production is consumed by the province and the rest is exported, primarily to the US. Therefore, greater emphasis is being placed on Manitoba Hydro's capability to meet long-term firm power export contracts under extreme drought conditions. One characteristic that distinguishes hydropower from thermal and nuclear power is the uncertainty related to the natural variability of the energy supply. River runoff varies in time and space and cannot be predicted in advance. Since river runoff is the source of hydropower energy, variability in energy supply is inevitable. Because of the significant storage capacity in Manitoba Hydro's system, short-term (e.g. daily, weekly) fluctuations in inflow are not a concern for Manitoba Hydro. However, longer periods (e.g. months, years) of extended low flows may lead to situations where thermal resources and imports would be required to meet contracts. Therefore, it is of high importance to be able to quantify the reliability of energy supply. The overall supply reliability is a function of various technical factors and the availability of water. The need to determine the reliability of inflow is the primary motivation for this study.

For a large hydropower system, several factors increase the complexity in defining drought. Firstly, droughts are inherently regional in nature but only to a certain extent. The Nelson and Churchill River drainage basins cover an area of 1.34×10^6 km^2

with distinct hydrological and climatologic characteristics. Hence, in a typical drought study it will be necessary to divide the system into homogeneous sub-basins.

Secondly, the large storage capacity also increases the complexity in defining drought. Lake Winnipeg has a surface area of 24,420 km^2 which translates into a storage capacity equivalent to six months of average outflow from the lake. Therefore, the regulation of this lake plays a vital role in enhancing the system capability during drought conditions.

Thirdly, for a hydropower system, the drought must be defined in terms of energy potential. This requires that flows at each generating station be converted to energy. To assess the system capability, the overall energy potential of the system must be compared with the overall energy demand.

Fourthly, relatively long data series are required to accurate model droughts. In many cases, such long series are not available and one must often rely on data filling and extension.

Hydropower utilities have traditionally employed the *critical drought* as basis for system development. The critical drought is defined as the most severe extended low flow period in the record of historical inflows. In the past, hydropower systems were often developed to be able to meet a fixed demand under inflow conditions identical to the critical drought. However, the critical drought approach does not provide a complete assessment of system reliability. There are indeed obvious limitations associated with the critical drought approach. Firstly, no probability of occurrence can be assigned to the critical drought, and for that reason, it is not possible to make state-

ments about supply reliability. Secondly, the critical drought is a function of record length. A long record of inflow is likely to contain a more severe drought episode than a short record. A sensible planning criterion should be independent of record length. The critical drought period currently employed by Manitoba Hydro occurred in 1939-41; another severe drought occurred in the late eighties.

Alternatively, a *stochastic approach* may be used to overcome some of the limitations of the critical drought approach. The stochastic approach requires the estimation of a stochastic time series model. Stochastic models attempt to reproduce the key statistics of historical data such as means, variances, autocorrelations, and crosscorrelations. In this approach, historical data are used to estimate the parameters of the model. Once the model is estimated, large set of equally likely synthetic traces can be generated. These traces show many possible hydrologic conditions that do not explicitly appear in the historical record. For example, if a definition of the critical drought can be developed, the simulated probability distribution of the critical drought can be determined using these traces. This information could potentially be used to assess the return period of the historical critical drought and other drought scenarios in general and can be used in the assessment of system reliability.

Stochastic models of hydrologic time series have played a major role in the water resources literature over the past four decades. Stochastic models have been reasonably successful as a practical tool for analysis, forecasting, and control. Since they have been used for decades, much experience has been accumulated in their application. Autoregressive (AR), autoregressive moving average (ARMA), periodic autoregressive

moving average (PARMA), and disaggregation models are the most commonly used stochastic models in the hydrology literature and software packages of these models are available (e.g. S-PLUS, SPIGOT, and SAMS2000 [*Grygier and Stedinger*, 1990; *Salas et al.*, 2000]).

In the last decade, Manitoba Hydro has funded a number of research projects aimed to define the probability of system drought using a stochastic approach. These projects used observed monthly flows at selected key points to develop a stochastic streamflow generation model. *DeWit* [1995] developed the first stochastic model for Manitoba Hydro's system and employed SPIGOT, a synthetic streamflow generation software package, to generate synthetic streamflow sequences. He worked directly with flows and estimated that the return period of the historical critical drought (1939-41) to be 381 years. *Rangarajan* [1998] improved the modeling framework in SPIGOT and modified the regulation rules for Lake Winnipeg. He defined energy drought by converting water flow at the generating stations into equivalent energy flow. The study initiated by *Rangarajan* [1998] was completed by *Akıntuğ* [2002] who estimated that the return period of the historical critical drought to be 397 years.

Previous studies did not include any treatment of model, parameter, and data uncertainty and left a number of questions open. Although they were a step in the right direction, the assessment of the probability of system drought requires further investigation by integrating uncertainty. There is a general need to quantify uncertainties in the estimate of drought frequency. Estimated return periods depend strongly on various modeling choices and parameter estimates. In this research project, the

uncertainties associated with the structure of stochastic time series models and their parameter estimates will be investigated. The uncertainty associated with filled-in data will also be considered. The integration and quantification of model, parameter, and data uncertainty into an overall assessment of system reliability should provide a better foundation for making future planning decisions.

Low frequency variability of annual hydrological data have been observed in North American climate regimes. In addition to observed data, paleoclimatic studies have proved the existence of long-term wet and dry periods in North America. The preliminary streamflow data analysis of Manitoba Hydro's system also revealed apparent long-term wet and dry periods. The existence of cycles such as those observed in the Nelson and Churchill River Basin cannot be explained by pure randomness.

In the literature, numerous studies have identified global climatic mechanisms that influence the North American climate at annual and decadal time scales. Climate states are potentially linked to low-frequency oceanic circulations such as the North Atlantic Oscillation (NAO), the Southern Oscillation (SOI), and the Pacific Decadal Oscillation (PDO). There is a complex interaction between these circulations. However, it is believed that they influence each other and the cumulative effect of these circulations produces a *quasi-cyclic* forcing mechanism in the hydrological cycle.

Model uncertainty due to inadequate representation of cycles is neglected in most studies. Model uncertainty refers to the sensitivity of predictions to the specification of a correct model structure. Since Manitoba Hydro is primarily interested in multi-year droughts, this study will mainly focus on the uncertainties associated with the

annual stochastic time series model. In the previous studies by Manitoba Hydro, the lag-one autoregressive [AR(1)] model was selected for annual streamflow generation in SPIGOT. However, one may question whether the AR-assumption is the best choice if indeed low-frequency climate variability is present. A significant part of the research reported in this thesis was motivated by doubt about the ability of the AR(1) model to adequately reproduce the persistence patterns evident in Manitoba Hydro's streamflow data. Streamflow data from Manitoba Hydro's system show periods where dry and wet spells have persisted for a significant number of years. These cyclic observations motivated the search for alternative stochastic time series models, in particular models that can mimic the regime-like behavior of hydrologic time series.

It is not clear if conventional stochastic models such as AR and ARMA can adequately reproduce the cycles present in Manitoba Hydro's data. The reproduction of cycles of wet and dry years in synthetic streamflow generation is obviously extremely important when the model is used to determine drought frequency. An annual stochastic model type called *Markov-Switching model* is selected in this research because it presumably should be able to better represent wet and dry cycles in the historical data. In the literature, this type of model is also referred to as hidden Markov models or Markov mixture models. The idea of a Markov-Switching model is not new in the hydrology literature. It was applied thirty years ago by *Jackson* [1975]. A major challenge at the time was the estimation of the model parameters. The Markov-Switching model framework has been successfully employed in different fields, including speech recognition, econometrics, ion channels, image analysis, and DNA composition. There

has also been a number of recent applications in hydrology, including *Wilks* [1998], *Hughes et al.* [1999], *Lu and Berliner* [1999], *Thyer and Kuczera* [2000, 2003a, b], and *Kehagias* [2004].

The parameters of a stochastic model are estimated from historical records. Parameter uncertainty refers to the uncertainty due to the use of a limited amount of data for model estimation. Estimated parameters have sampling errors whose magnitude depends on the length of the historical data records. In most practical applications, this sampling uncertainty is neglected although it may have significant impact on the conclusions drawn from the study [*Stedinger and Taylor*, 1982b]. Parameter uncertainty will here be included in the Markov-Switching model framework through Bayesian inference. In the Bayesian approach, the unknown parameters of the stochastic model are treated as random variables instead of fixed quantities. This approach quantifies the parameter uncertainty by deriving the distribution of the model parameters.

In a multi-site analysis, the length of observed data at each site must be equal. To make best use of available data, it is common practice to extend the shorter series using record extension techniques. Missing streamflow data of Manitoba Hydro have been filled-in and the records have been extended by Manitoba Hydro prior to this study [*Girling*, 1988, 1990]. However, since filled-in data are not observed values, the uncertainty resulting from using estimated values rather than observations must be quantified and taken into account. In this research project, an attempt is made to quantify uncertainty associated with reconstructed data.

1.2 Objectives of the Study

The objectives of this research project are specified in the light of the discussion above. The overall objective of the project is to develop a comprehensive probabilistic framework incorporating model, parameter, and data uncertainty in order to derive a better estimate of drought frequency and its associated precision. The developed framework will be applied specifically to Manitoba Hydro, but the proposed techniques are general and can be transferred to any other hydropower system. More specific objectives are:

- *To incorporate parameter uncertainty into a multi-site AR(1) model:* In this project, the performance of a multi-site Markov-Switching model with parameter uncertainty will be compared with a multi-site AR(1) model with parameter uncertainty for Manitoba Hydro's system. Although parameter uncertainty was integrated into the single-site AR(1) model by *Stedinger and Taylor* [1982b], a preliminary analysis revealed that the extension to multi-site models is somewhat more involved.

- *To develop a single- and multi-site Markov-Switching model for annual streamflow time series:* Since both single-site and multi-site modeling frameworks are required in this project, a single-site and multi-site Markov-Switching model, which presumably should be abe to better mimic the regime-like behavior of observed annual streamflow data, will be developed.

- *To explore the theoretical properties of Markov-Switching models:* The theoretical

properties such as autocorrelation structure, moments, and marginal distribution of Markov-Switching models will be explored.

- *To identify and implement an efficient parameter estimation method for the Markov-Switching model:* The lack of efficient estimation methods for some years prevented the Markov-Switching model to gain popularity. However, significant advances in parameter estimation techniques for Markov-Switching model have been made in recent years. An efficient parameter estimation method will be explored.

- *To incorporate parameter uncertainty into single- and multi-site Markov-Switching models:* Parameter uncertainty will be integrated into Markov-Switching models through numerical Bayesian methods.

- *To incorporate reconstructed data uncertainty into multi-site Markov-Switching model:* The uncertainty associated with reconstructed data will be taken into consideration in the multi-site Markov-Switching model through the use of numerical Bayesian methods.

- *To develop a disaggregation model using a Markov-Switching modeling approach:* A variation of the classical disaggregation model exploiting the characteristics of the Markov-Switching framework will be developed.

- *To incorporate model, parameter, and data uncertainty in the estimation of the probability of system drought:* The importance of model, parameter, and data

uncertainty will be analyzed by comparing the results of the drought frequency analysis with and without considering uncertainties.

- *To perform a drought frequency analysis:* A comprehensive energy drought frequency analysis will be performed for the Manitoba Hydro's system.

1.3 Structure of the Report

The thesis is organized as follows: Manitoba Hydro's system and hydrological data are described in Chapter 2. Chapter 3 explains the definition of hydrological and energy drought for Manitoba Hydro. A review of the AR(1) and disaggregation models are given in Chapter 4. The integration of parameter uncertainty into the multi-site AR(1) model is also given in this chapter. A detailed explanation of the Markov-Switching modeling framework which forms the basis of the thesis is explained in Chapter 5 and its multi-site extension is given in Chapter 6. In Chapter 7, the disaggregation model using the Markov-Switching modeling approach is presented. The application of the developed models to Manitoba Hydro's system is provided in Chapter 8. Finally, the conclusions are given in Chapter 9.

Chapter 2

Manitoba Hydro's System

2.1 Introduction

The system of Manitoba Hydro comprises the Nelson River and the Upper Churchill River drainage basins. The Nelson River Basin is the 20^{th} largest on a global scale. The Nelson River drains Lake Winnipeg into Hudson Bay. Several large river systems, notably the Saskatchewan, the Assiniboine, the Red, and the Winnipeg River systems contribute water to Lake Winnipeg and are therefore part of the greater Nelson River drainage basin. The Upper Churchill River drainage basin is actually separate from the Nelson River drainage basin but has become part of Manitoba Hydro's system as a result of the diversion of water from the Churchill River into the Nelson River at Southern Indian Lake. Figure 2.1 illustrates the major drainage basins contributing to Manitoba Hydro's system. Despite the relatively flat relief of Northern Manitoba, the Nelson River provides 80% of the hydropower production of Manitoba Hydro

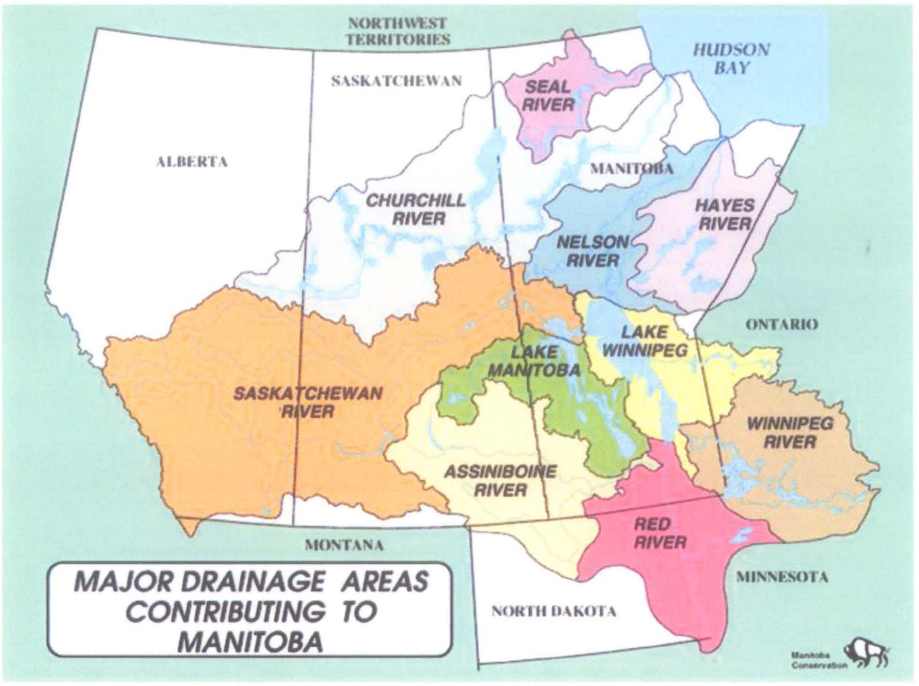


Figure 2.1: Major drainage basins contributing to Manitoba Hydro's system (from Manitoba Water Stewardship web site).

because of its high flow volume. Lake Winnipeg and Southern Indian Lake are the two major reservoirs in the system regulated according to the needs of Manitoba Hydro. The generating stations are located on three rivers namely the Winnipeg River, the Saskatchewan River, and the Nelson River. A brief explanation of rivers, reservoirs, and hydropower generating stations in Manitoba Hydro's system are given in this chapter.

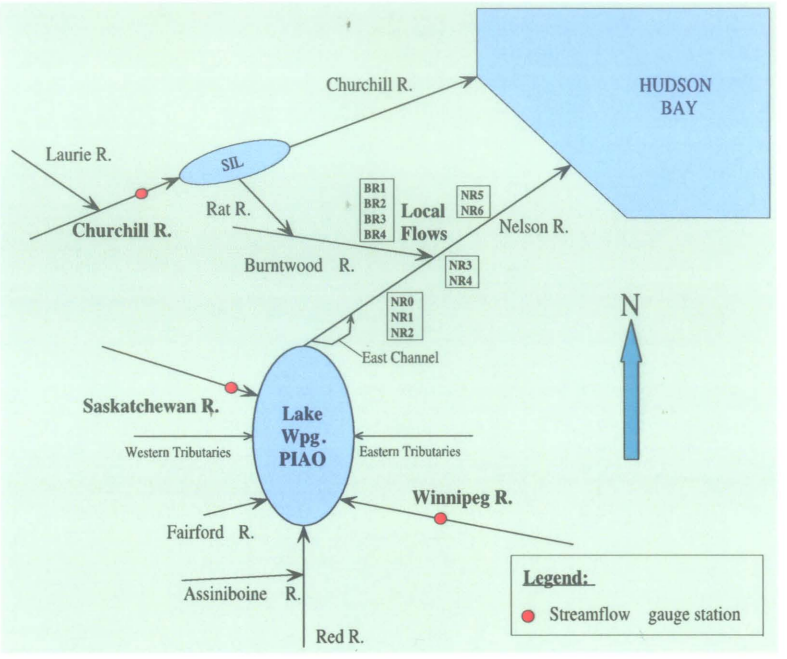


Figure 2.2: Schematic diagram of Manitoba Hydro flow system. Streamflow gauge stations and other hydrological components of the system.

2.2 Major Rivers and Hydrological Components in Manitoba Hydro's System

The Nelson River drainage basin includes five major contributing rivers namely the Saskatchewan River, the Winnipeg River, the Red River, the Assiniboine River, and the Churchill River. The Upper Churchill River Basin is considered part of the system because of the Churchill River diversion at Southern Indian Lake (SIL). A schematic diagram of the river system and the location of streamflow stations are given in Figure 2.2. Manitoba Hydro's system is usually divided into five distinct hydrologic components, described in the following. The average annual flows and the relative contribution to the system of each hydrologic component are given in Table 2.1.

Table 2.1: Major hydrologic components of Manitoba Hydro's system and their contributions to the system.

No	Site	Location	Av. Annual Flow 1912-98 (cfs)	Contribution 1912-98 (%)
1	Winnipeg River	Slave Falls	31,650	27.6
2	Saskatchewan River	Grand Rapids	20,100	17.5
3	Churchill River*	Southern Indian Lake	31,390	27.4
4	Local Flows	Burntwood and Nelson River	13,140	11.4
5	Lake Winnipeg PIAO	Lake Winnipeg	18,500	16.1

* Diverted flow

2.2.1 Winnipeg River

The Winnipeg River system is about 765 km long and runs in a northerly and westerly direction from its ultimate source near Lake Superior to the southern shore of Lake Winnipeg. The drainage basin area of Winnipeg River is $150,000$ km^2. Two thirds of this drainage basin is in northwestern Ontario, one-fifth spreads south into northern Minnesota, while the rest of it is in eastern Manitoba, where the river flows into Lake Winnipeg at Traverse Bay.

Streamflow records are available from 1912 at Slave Falls with no missing records. The average annual flow is 31,650 cfs. There are six hydropower generation stations on this river with about 560 MW total capacity (Great Falls, Seven Sisters, Pine Falls, McArthur, Slave Falls, and Pointe DuBoise).

2.2.2 Saskatchewan River

The Saskatchewan River originates from the Canadian Rockies and flows east into Manitoba where it passes through Cedar Lake before emptying into Lake Winnipeg.

It is the fourth longest river system in Canada, travelling almost 2000 km and its drainage area is $334,100$ km^2. Streamflow records are available from 1912 at Grand Rapids station which is located close to Lake Winnipeg. The average annual flow is 20,100 cfs. A single generation station with a capacity of 480 MW is located at Grand Rapids. The Grand Rapids Generation Station was the first hydroelectric generation station built in Northern Manitoba.

2.2.3 Churchill River

The drainage basin area of the Churchill River is $283,350$ km^2. It is located north of the Nelson and Saskatchewan River basins. Manitoba Hydro diverts a significant portion of the Churchill River flow at Southern Indian Lake through the Rat River and Burntwood River into the Nelson River. This increases the flow in the lower reach of the Nelson River where Manitoba Hydro's three largest power generation stations, Kettle, Long Spruce, and Limestone, are located. The diverted flow is regulated by the Notigi Control Structure on the Rat River and by the Missi Control Structure at the northern end of Southern Indian Lake. Flow records for Churchill River at its entrance to Southern Indian Lake (SIL) are available from 1930. The average annual flow of the Churchill River is 35,550 cfs of which an average of 31,390 cfs is diverted into the Nelson River.

2.2.4 Partial Inflow Available as Outflow (PIAO)

In addition to the Saskatchewan and the Winnipeg River, several other gauged and ungauged streams flow into Lake Winnipeg from east and west. The total inflow available as outflow (TIAO) is defined by a simple water balance equation that involves the summation of the Lake Winnipeg outflow (O_{LWpg}) over a suitable time period and its storage change (ΔS_{LWpg}) during the same period:

$$\begin{aligned} TIAO &= \text{(Total Inflow)} - \text{(Total Loss)} \\ &= O_{LWpg} \pm \Delta S_{LWpg} \end{aligned} \tag{2.1}$$

The *partial inflow available as outflow (PIAO)* is obtained by subtracting the inflows of the Winnipeg River, I_{Wpg}, and the Saskatchewan River, I_{Sask}, from the TIAO (*Girling* [1990]).

$$PIAO = TIAO - (I_{Wpg} + I_{Sask}) \tag{2.2}$$

The Lake Winnipeg PIAO is used as one of the hydrological components in the stochastic modeling of Manitoba Hydro's system. Unlike the other components, the PIAO does not represent the flow conditions at a specific point on a particular river. The PIAO takes into account losses due to evaporation and seepage from the lake. The effect of evaporation is significant during dry weather because Lake Winnipeg has a surface area of $24,390$ km^2. The PIAO is an important component of Manitoba Hydro's system because the effect of evaporation must be considered in a drought

Table 2.2: Local Flows on the Burntwood and Nelson River.

No	Index	River	Location	Av. Annual Flow 1912-98 (cfs)
1	BR1	Burntwood	South Bay - Notigi	1,068
2	BR2	Burntwood	Notigi - Wuskwatim	1,772
3	BR3	Burntwood	Wuskwatim - Manasan	170
4	BR4A	Burntwood	Manasan - Thompson	232
5	BR4B	Burntwood	Thompson - First Rapids	117
6	NR0	Nelson	Lake Winnipeg - Jenpeg	971
7	NR1	Nelson	Jenpeg - Bladder Rapids	1,494
8	NR2	Nelson	Bladder Rapids - Kelsey	1,514
9	NR3	Nelson	Kelsey - Birthday	4,515
10	NR4	Nelson	Birthday - Kettle	425
11	NR5	Nelson	Kettle - Long Spruce	459
12	NR6	Nelson	Long Spruce - Limestone	181

study. The PIAO consists of the Red River, the Assiniboine River, and the Fairford River, gauged and ungauged tributaries from the western and eastern side of Lake Winnipeg, and the water loss due to evaporation and seepage.

2.2.5 Local Flows

In addition to the four major components described in the previous section, the local flows of the Burntwood (BR) and the Nelson River (NR) are also considered in Manitoba Hydro's system. The sites are named according to the river they are located on. The aggregation of the 12 local flows given in Table 2.2 are modeled as a single artificial basin site. This aggregated basin site is then disaggregated into sub-basin sites in the stochastic modeling framework described in Chapter 8. Table 2.2 also indicates the location of local flows on the Nelson and Burntwood River and their average annual flows.

2.3 Regulation of Reservoirs in the System

In the mid-1960's, provincial power planners made a long-term decision on the future supply of electricity for the province of Manitoba. The decision was to build a series of hydroelectric generating stations on the Nelson River. Three generating stations on the lower Nelson River were constructed. The regulation of Lake Winnipeg was required to allow for greater flows into the Nelson River when needed. In order to increase the power production potential of the lower Nelson River by as much as 40%, most of the Churchill River flow was diverted into the lower Nelson River at Southern Indian Lake through the Rat River and the Burntwood River.

Developing an accurate model for the management of Lake Winnipeg and Southern Indian Lake is beyond the scope of this study. Therefore Lake Winnipeg and Southern Indian Lake outflows are estimated by using a simplified representation of the system. *DeWit* [1995] developed a heuristic-based model for the regulation of Lake Winnipeg and Southern Indian Lake that was further refined by *Rangarajan* [1998]. In this study, DeWit's and Rangarajan's heuristic-based regulation models, explained in the following sections, are employed.

2.3.1 Regulation of Lake Winnipeg

In February 1966, the province of Manitoba and the government of Canada came to an agreement allowing Manitoba Hydro to regulate Lake Winnipeg as a natural reservoir for hydroelectric development on the Nelson River. Regulation is necessary because the natural flow pattern from Lake Winnipeg into the Nelson River is opposite to

the energy needs of the province. Demand is higher than supply during winter, while maximum inflow occurs during spring and early summer where the demand is relatively low. Therefore, it is necessary to decrease the outflow from Lake Winnipeg in spring and early summer in order to have more available outflow in the fall and winter. The Lake Winnipeg Regulation project guarantees winter outflows so that the winter demand for electricity can be met.

The license allows Manitoba Hydro to set the outflows as required for power production purposes along the Nelson River as long as the level of the lake is kept between 711 ft and 715 ft above sea level. During wet and dry periods, the lake level may rise above or drop below these levels, however, according to the agreement, the lake will be operated to return to the licensed operating range as quickly as inflow conditions permit.

The storage capacity between 711 ft and 715 ft of Lake Winnipeg corresponds to approximately six months of average outflow requirements. Hence, the regulation of Lake Winnipeg has a significant influence on drought and flood conditions. There are several other constraints on the operation of the lake. Beside maximum and minimum lake levels, Manitoba Hydro must also respect maximum and minimum release flows. In addition to these considerations, the management of the reservoir might be affected by other factors such as long range forecasts of precipitation conditions and power demand. For example, if the spring is predicted wetter than normal, Manitoba Hydro may release more than the usual portion to provide room for flood control.

Surface elevation and release flows are the two major constraints for the operation

of Lake Winnipeg.

- *Surface Elevation*: Although, the lake elevation mostly fluctuates between 713 ft and 714 ft, the allowable operating range is between 711 ft and 715 ft. These are the licensed minimum and maximum lake elevations.

- *Minimum and Maximum Release (R)*: The preferred minimum flow is $R_{min} = 25,000$ cfs while the preferred maximum flow is $R_{min} = 150,000$ cfs. Sometimes releases may exceed the maximum or go below the minimum.

The storage capacity of Lake Winnipeg is 400,000 cfs-month between maximum and minimum lake elevations, which means 1 ft of elevation is equal to 100,000 cfs-month. In other words, there is enough water to meet the minimum release for 16 months at full storage.

As given in (2.1), total inflow available as outflow (TIAO) is calculated as the outflow plus or minus storage change in Lake Winnipeg. TIAO is equal to PIAO plus Winnipeg and Saskatchewan River inflow. The TIAO takes into consideration the effects of losses due to evaporation and seepage in Lake Winnipeg. The outflow from Lake Winnipeg depends on TIAO. The heuristic release rules applied in this study consider two conditions:

1. $\boldsymbol{TIAO \geq R_{max}}$**:**

   The excess flow ($TIAO - R_{max}$) will be stored fully or partly in the lake. If the lake reaches the maximum level of 715 ft, the reminder will be released from Lake Winnipeg. If the storage is full at the end of the previous month, the

Table 2.3: Maximum preferred release during winter months.

		R_{max} (kcfs)				
Storage	Elevation	Nov	Dec	Jan	Feb	Mar
0	709	45.0	50.0	50.0	50.0	45.0
200,000	711	45.0	50.0	50.0	50.0	45.0
300,000	712	45.0	50.0	50.0	50.0	45.0
400,000	713	45.0	50.0	50.0	50.0	45.0
500,000	714	66.2	79.0	74.4	74.4	66.2
600,000	715	96.8	108.2	104.0	104.0	96.8
700,000	716	114.7	123.4	118.1	118.1	114.7
1,000,000	719	186.3	184.2	147.5	174.5	186.3

lake outflow will be equal to TIAO. There is an assumption in this model that the absolute maximum storage level is 715 ft since this elevation is the licensed maximum.

2. ***TIAO < R_{max}:***

- **November through March**

  During winter months the lake outflow is set to the maximum preferred release even during drought months. These preferred maximum releases are given in Table 2.3 (from *Rangarajan* [1998]). Depending on the previous month's storage, the release outflow is obtained from this table using linear interpolation. The storage at the end of month m can be calculated using the water balance equation:

$$Storage_m = Storage_{m-1} + TIAO_m - Outflow_m \tag{2.3}$$

  If this storage exceeds 714.75 ft, the excess water is also released.

- **April through July**

  During spring, the lake level should be sufficiently low to allow room for the incoming spring runoff. The percentage of TIAO (P_1) that is keep in storage is influenced by the difference between the maximum storage and the previous month's storage:

$$P_1 = \frac{1.25(575,000 - Storage_{m-1})}{375,000} \tag{2.4}$$

  where $575,000\,\text{ft}^3$ and $375,000\,\text{ft}^3$ correspond to the maximum storage and the average storage, respectively. After calculating the percentage of TIAO to keep in storage, the storage at the end of the month is calculated as

$$Storage_m = Storage_{m-1} + (TIAO_m \times P_1) \tag{2.5}$$

  If this new storage exceeds the maximum level (714.75 ft), the excess water is released as well.

  Minimum-dry and minimum-wet outflows are given in Table 2.4. If the previous month's water level is greater than 713.5 ft, the minimum outflow is set equal to minimum wet, otherwise it is equal to minimum dry. If the calculated outflow of that particular month is less than its minimum value, the outflow is set equal to the minimum.

- **August through October**

  During the end of summer and fall, Manitoba Hydro starts to use the Lake

Table 2.4: Minimum outflows of Lake Winnipeg.

Period (Month)	Minimum Dry Outflow (cfs-month)	Minimum Wet Outflow (cfs-month)	Period (Month)	Minimum Dry Outflow (cfs-month)	Minimum Wet Outflow (cfs-month)
Jan	50,000	50,000	Jul	25,000	40,000
Feb	50,000	50,000	Aug	25,000	40,000
Mar	45,000	45,000	Sep	25,000	40,000
Apr	30,000	40,000	Oct	30,000	40,000
May	25,000	40,000	Nov	40,000	45,000
Jun	25,000	40,000	Dec	50,000	50,000

Winnipeg storage to meet its demand. In this case, in addition to TIAO a percentage of TIAO is also released from the lake. This percentage of TIAO (P_2) is influenced by the difference between the previous month's storage and the minimum storage:

$$P_2 = \frac{0.50(Storage_{m-1} - 200,000)}{375,000} \tag{2.6}$$

where $200,000\,\text{ft}^3$ and $375,000\,\text{ft}^3$ correspond to the minimum storage and the average storage, respectively. After calculating the percentage of TIAO to release from storage, the storage at the end of the month is calculated as

$$Storage_m = Storage_{m-1} - (TIAO_m \times P_2) \tag{2.7}$$

Manitoba Hydro uses a deterministic long-term planning model (SPLASH) to assess the adequacy of their generation system. Monthly historical flow records are used to simulate the rule curves of Lake Winnipeg. *Rangarajan* [1998] adjusted the constants 1.25 and 0.50 in (2.4) and (2.6) in order to mimic the SPLASH model outflows

during the most severe drought event (1938-42).

2.3.2 Regulation of Southern Indian Lake

The regulation of Southern Indian Lake is not as complicated as Lake Winnipeg. In this project, it will be based on the heuristic rules suggested by *DeWit* [1995]. Manitoba Hydro operates two control structures on this lake. These control structures allow the diversion of a significant portion of the Churchill River flows into the lower reach of the Nelson River where the three largest generation stations, Kettle, Long Spruce, and Limestone, are located. This diversion enhances energy generation and the energy supply reliability of the system. The operation of this diversion depends on the various agreements between Manitoba Hydro and local communities and other parties affected by the regulation of Southern Indian Lake. There are several constraints in this agreement such as maximum and minimum releases and lake surface elevations.

DeWit [1995] considered three major constraints on the regulation of Southern Indian Lake. These three major constraints, given below, are considered in order to determine the release flow through the Notigi control station to the Nelson River:

1. The specified minimum release at Missi must be maintained.

2. The Notigi release should be kept at licensed maximum or as high as possible.

3. The Lake surface elevation of Southern Indian Lake should be between 843.5 ft and 847.5 ft above see level.

Table 2.5: Licensed release limits for Notigi and Missi control stations.

Period (Month)	Max. Notigi Release (cfs)	Min. Missi Release (cfs)	Period (Month)	Max. Notigi Release (cfs)	Min. Missi Release (cfs)
Jan	34,000	4,000	Jul	35,000	500
Feb	34,000	3,000	Aug	35,000	500
Mar	34,000	2,000	Sep	35,000	500
Apr	34,000	2,000	Oct	35,000	2,000
May	35,000	1,000	Nov	34,000	6,000
Jun	35,000	500	Dec	34,000	5,000

The storage volume between the maximum and minimum lake elevation is 54,000 cfs-month (Bruce Hinton, personal communication). The licensed limits of Notigi and Missi are given in Table 2.5

Based on the three constraints above, Notigi releases can be calculated for three basic flow conditions:

1. **Churchill River inflow is greater than the summation of Notigi licensed maximum and Missi licensed minimum:**

   Notigi releases will be equal to the maximum licensed limit and excess water will be either fully or partly stored in Southern Indian Lake with the reminder being released from Missi.

2. **Churchill River inflow is less than the summation of Notogi licensed maximum and Missi licensed minimum, and greater than Missi licensed minimum:**

   Missi release will be the licensed minimum and the storage in Southern Indian Lake will be used to bring Notigi releases up to the licensed maximum or as high as possible.

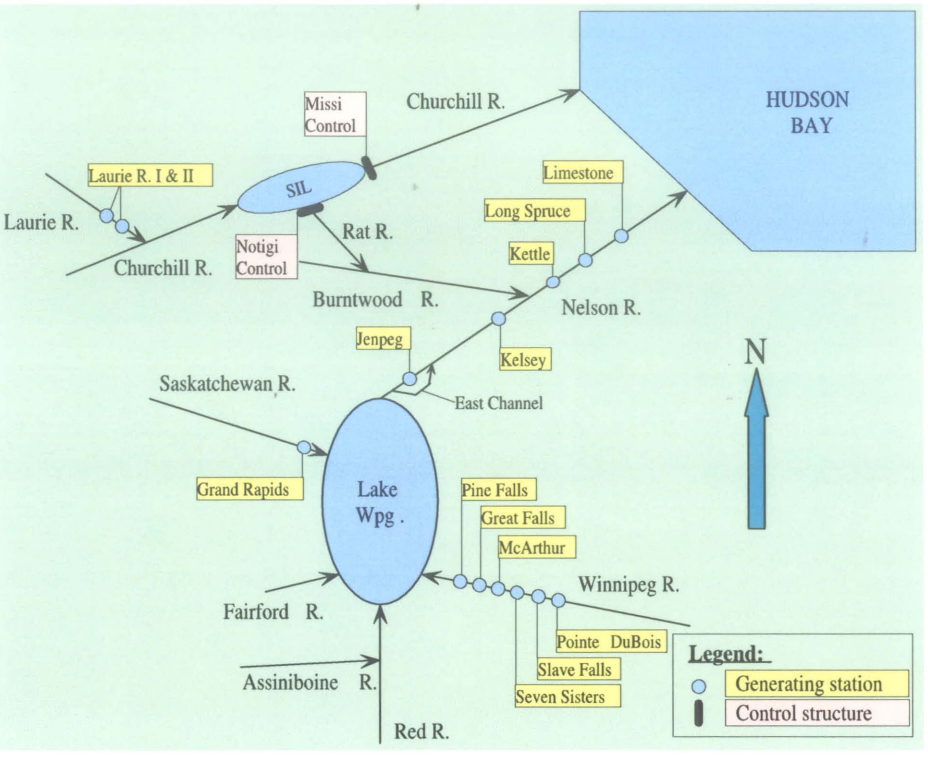


Figure 2.3: Schematic diagram of Manitoba Hydro generating stations.

3. **Churchill River inflow is less than Missi license minimum:**

   The storage in Southern Indian Lake will be used to bring Missi release up to the licensed minimum. Any remaining storage will be used to increase Notigi releases.

2.4 Hydro-power Generating Stations

There are fourteen hydro-power and two thermal generating stations in Manitoba Hydro's system. The installed capacity of these generating stations are given in Table 2.6 and a diagram showing the approximate location of each generating station is given

Table 2.6: Manitoba Hydro's generating stations.

Generating Station	Capacity (MW)	Generating Station	Capacity (MW)
Limestone	1294	Slave Falls	68
Long Spruce	1020	McArthur	54
Kettle	1224	Seven Sisters	155
Kelsey	215	Pointe DeBoise	72
Jenpeg	97	Laurie River I	5
Grand Rapids	480	Laurie River II	6
Pine Falls	88	Brandon	97
Great Falls	129	Selkirk	139

in Figure 2.3. The Missi and the Notigi control structures which regulate Southern Indian Lake, as well as the East Channel which diverts part of the Lake Winnipeg outflows around the Jenpeg Generating Station into Cross Lake are also shown in this figure.

As seen from Figure 2.3, the majority of the generating stations are located on the Nelson River and the Winnipeg River. The total installed electric capacity of Manitoba Hydro's system is 5143 MW. The thermal stations in Brandon and Selkirk account for 4% of this amount. The remaining 96% of the capacity are associated with the 14 hydropower generating stations with about 77% of the total capacity coming from the stations on the Nelson River, 9% from six stations on the Winnipeg River, and 10% from one station on the Saskatchewan River. In addition, two small hydropower stations are located on the Laurie River but because of their location and size, these are not considered in the drought frequency analysis.

Chapter 3

Drought Definition for Manitoba Hydro

3.1 Introduction

The various drought definitions employed in practice reflect differences in regions, needs, and disciplinary approach. Since there are widely diverse views on the interpretation of droughts among the scientists of different disciplines, it is not possible to give a universal definition.

In this chapter, different drought definitions encountered in the literature are briefly presented. The emphasis is given to the definition of hydrological drought and energy drought. After identifying a suitable definition of energy drought for Manitoba Hydro's system, historical energy drought events are extracted from observed streamflow data and discussed.

3.2 General Definition of Drought

In general, a drought may be defined in terms of meteorological, agricultural, hydrological, energy, and socioeconomic factors.

Meteorological drought is usually a period of negative departure of precipitation from a given threshold value. Different thresholds such as 50%, 75%, 90%, and 100% of the long-term average precipitation may be considered. Due to climatic differences, what is considered a drought in one location may not be a drought in another location. Meteorological observations are the first indicators of drought.

Agricultural drought occurs when the amount of soil moisture no longer meets the need of a particular crop at a particular time. Agricultural drought is usually the first economic sector to be affected by drought.

Hydrological drought refers to a situation when surface and subsurface water supplies such as streamflow, reservoir elevations, and groundwater tables are below given threshold values. As in the case of meteorological droughts, threshold values may be defined as different percentages of the long-term average. There is a time lag between meteorological drought and hydrological drought. When low precipitation occurs over an extended period of time, this shortage will be reflected in declining surface and subsurface water levels.

Hydropower energy drought occurs when the generated energy does not meet the energy demand in a particular time period. Since streamflow water is the source of hydropower energy, energy drought is closely related to hydrological drought.

Socioeconomic drought occurs when physical water shortage begins to affect people.

In this situation, the demand for an economic good exceeds supply as a result of a weather-related shortfall in water supply. According to this definition, energy drought can be considered a socioeconomic drought.

The first step in a scientific analysis is to define and specify the components of the problem under investigation. Therefore, a clear definition of drought is required before a drought frequency analysis can be undertaken. Since the performance of a hydropower system is related to the availability of streamflow, hydrological drought and energy drought will be considered in this research project.

3.3 Definition of Hydrological Drought

The theory of runs represents perhaps the best framework for the definition of hydrologic droughts. *Yevjevich* [1967] recommended the theory of runs for defining hydrologic droughts and for studying their statistical properties. The theory of runs has since been used by many researchers [*Sen*, 1977, 1980a, b; *Chander et al.*, 1981; *Zelenhasic and Salvai*, 1987; *Mathier et al.*, 1992; *Sharma*, 2000; *Shiau and Shen*, 2001; *Zelenhasic*, 2002].

The runs of the time series of a stochastic variable can be defined in several ways. A time series x_t can be converted into a binary series by selecting a threshold value x_0 and recording observations as above and below the threshold. A run is defined as a consecutive sequence of above-threshold or below-threshold observations. The parameter x_0 can be variable in time.

Figure 3.1 illustrates some common drought variables using concepts from the

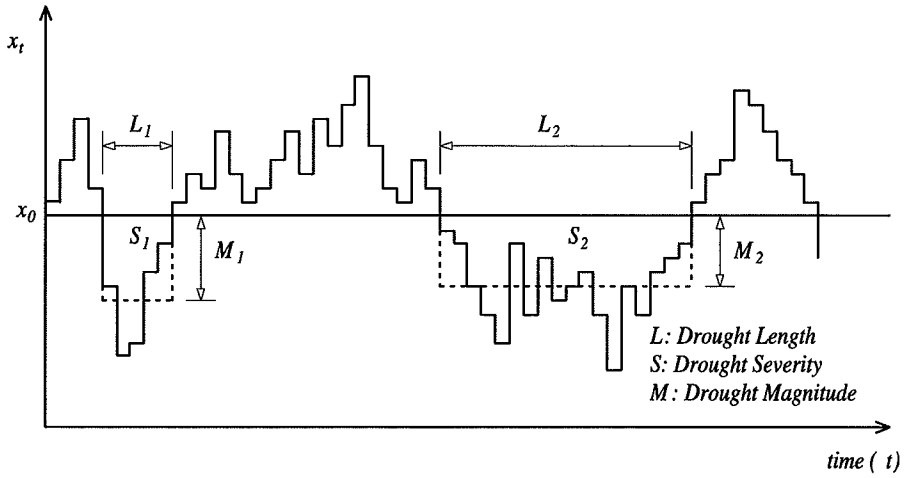


Figure 3.1: Drought parameters using Theory of Runs.

theory of runs. For a given time series of x_t and a selected truncation level x_0, there are three drought parameters that are often considered:

1. *Drought Duration* (L_d): Negative run length, i.e. the time elapsed between a downcross and the next upcross.

2. *Drought Severity* (S_d): Negative run sum, i.e. the sum of all negative deviations between a downcross and an upcross.

3. *Drought Magnitude* (M_d): The average negative deviation from the truncation level x_0, ($M_d = S_d/L_d$).

These three parameters have been extensively used in the definition of hydrologic droughts. Since drought magnitude is completely determined by the other two, drought duration and drought severity can be considered the two primary parameters of a hydrologic drought analysis. It has been noted in the literature that there are strong

positive correlations between drought duration and drought severity [*Sen*, 1977; *Chang and Stenson*, 1990; *Sharma et al.*, 1997].

The main advantage of using runs in the definition of hydrologic drought is the possibility of determining their properties analytically or by data generation methods. For example, it is possible to determine analytically the probability density function (pdf) of the drought severities of given drought durations. When the analytical approach becomes too complex, an appropriate data generation method can be employed to obtain properties of runs such as drought severities and drought durations frequencies.

Based on the concept of runs, *Dracup et al.* [1980] outlined four decisions to be made in arriving at a viable drought definition: the nature of the water deficit (e.g. precipitation, soil moisture, streamflow), the basic time unit of the data (e.g. month, season, year), the truncation level which distinguishes low flows from high flows (e.g. mean, median, mode), and the regionalization and/or standardization approach.

In this project, the analysis of hydropower system drought requires the modeling of flows at a monthly time step. The main advantage of using a monthly time resolution is sensitivity because a drought event may initiate in one year and continue to the subsequent year. If only annual flows were analyzed, such drought events would not appear in the drought analysis. The seasonal variability of the energy demand is another reason for using a monthly time step.

The truncation level divides the time series into deficits (droughts) and surpluses. A water shortage exists when the flow is below the truncation level. Therefore, the selection of the truncation level is one of the most important factors in the definition

and analysis of drought. There is no universal definition of truncation level. It may be a constant [*Yevjevich*, 1967] or a function of time [*Mathier et al.*, 1991, 1992]. It typically depends on the type of drought under study and the objectives that water users wish to attain. Since this study will involve drought analysis of monthly time series, a seasonally variable truncation level will be considered.

3.4 Definition of Energy Drought

In drought analysis for hydropower systems, drought should be defined in terms of energy. For a fixed head, the produced energy at a hydropower station is proportional to the amount of turbined water. Hence, monthly or annual streamflow at a hydropower station can be converted into energy by multiplying by a factor, assuming a constant head. In general, an energy drought occurs when energy supply is less than energy demand. In large hydropower systems comprising multiple stations and covering different drainage basins, the deficit at one station does not necessarily result in system drought. An energy drought occurs only when the total energy demand exceeds the total energy supply of the system. The energy time series at monthly or annual time steps may be used in the assessment of drought frequency.

Energy drought, as proposed by *Rangarajan* [1998] for Manitoba Hydro's system, is defined in the following. The energy flow (MWh) at generating station i during month m in year t may be defined as

$$E_{tm}^{(i)} = q_{tm}^{(i)} \times f_i \times h_m \tag{3.1}$$

where $q_{tm}^{(i)}$ is the monthly average flow and f_i is the flow-to-power conversion factor at generation station i, $i = 1, 2, ..., n$, and h_m is the total time (in hours) in month m. Monthly energy flow at each generation station can be calculated and summed to give the monthly total energy flow E_{tm} of the system:

$$E_{tm} = \sum_{i=1}^{n} E_{tm}^{(i)} \tag{3.2}$$

Monthly energy deficits of the system are defined using the monthly net energy demand D_m as a truncation level. When the total energy flow E_{tm} is less than the net energy demand D_m, an energy deficit ED_{tm} occurs. Mathematically

$$ED_{tm} = \begin{cases} D_m - E_{tm} & \text{if} \quad E_{tm} < D_m \\ 0 & \text{if} \quad E_{tm} \geq D_m \end{cases} \tag{3.3}$$

for $t = 1, 2, \ldots, T$, and $m = 1, 2, \ldots, 12$.

Based on the monthly time series of ED_{tm}, energy drought duration and severity can be calculated using the theory of runs as described in Section 3.3. In this case, the duration L_d of an energy drought is the number of consecutive months where total energy flow is lower than net energy demand and the energy drought severity is the summation of energy deficits during the L_d months.

Table 3.1: Flow-to-power factors and inflow sites for generating stations.

Generation Station	Flow-to-Power Factor (MW/kcfs)	Inflow Site
Point DuBois	3.16	Winnipeg River at Slave Falls
Slave Falls	2.01	Winnipeg River at Slave Falls
Seven Sisters	4.24	Winnipeg River at Slave Falls
McArthur	1.75	Winnipeg River at Slave Falls
Great Falls	3.93	Winnipeg River at Slave Falls
Pine Falls	2.87	Winnipeg River at Slave Falls
Grand Rapids	9.26	Saskatchewan River at Grand Rapids
Jenpeg	2.16	Lake Winnipeg Outflow
Kelsey	4.05	Lake Winnipeg Outflow + NR0 + NR1 + NR2
Kettle	7.11	Kelsey flows + Notigi Release + NR3 + NR4 + BR1 + BR2 + BR3 + BR4A + BR4B
Long Spruce	6.09	Kettle flows + NR5
Limestone	7.89	Long Spruce flows + NR6

3.5 Definition of Energy Drought for Manitoba Hydro's System

In this project, energy drought has been tailored to the specifics of Manitoba Hydro's system. Calculation of the energy flow at each station requires information about forebay and tailrace elevations. Since Manitoba Hydro's generating stations have relatively little variation in forebay-levels, a pragmatic approach is used which assumes that a given amount of flow at a particular site will produce a certain amount of power. Manitoba Hydro has developed flow-to-power conversion factors at each of the generation stations based on historical records. Although some of the water might be spilled, it is assumed that all flow passes through the turbines. This assumption is realistic for a drought study since all available water would be used during drought periods. Flow-to-power factors for each hydropower station are given in Table 3.1.

Although the six Winnipeg River generating stations are listed individually in Table 3.1, they are treated as one station having an inflow equivalent to the Winnipeg River flow at Slave Falls and a flow-to-power factor 17.96 MW/kcfs which is the summation of six factors.

DeWit [1995] mentioned that another special case occurs when calculating the energy flow at the Jenpeg generating station. The flow at Jenpeg is a combination of Lake Winnipeg outflow and local flow NR0. However, a portion of these flows bypasses the generating station through the East Channel. To incorporate this condition into the model, it is assumed that none of the NR0 flow passes through Jenpeg's generating station while all of the Lake Winnipeg outflow does.

The monthly flows of the Saskatchewan River at Grand Rapids, the Winnipeg River at Slave Falls, the PIAO of Lake Winnipeg, and the Churchill River at Southern Indian Lake are known. The outflow from Lake Winnipeg and Southern Indian Lake are calculated according to the heuristic rules given in Section 2.3. With the knowledge of local inflows to the Burntwood and Nelson Rivers, the monthly flow at each generating station can be obtained and converted to energy flow using (3.1). The total energy flow is obtained using (3.2).

The monthly net energy demand of the system for the load year 2005 was obtained from the SPLASH model used at Manitoba Hydro (Bruce Hinton, personal communication, 2003). In the calculation of the monthly net energy demand, domestic (Province of Manitoba) and export energy demands as well as energy imports and thermal energy generation of each month are taken into consideration. The net energy

Table 3.2: Monthly energy demands for Manitoba Hydro's system.

Month	Manitoba Load (GWh)	Load Adjustment (GWh)	Firm Export (GWh)	Firm Import (GWh)	Thermal Generation (GWh)	Net Hydro Demand (GWh)
Jan	2357.99	29.95	420.63	369.02	380.58	2058.97
Feb	2022.00	27.35	383.62	336.34	346.82	1749.81
Mar	2063.00	28.76	420.63	369.02	380.58	1762.79
Apr	1678.00	22.94	515.62	486.53	313.52	1416.51
May	1626.00	20.75	492.42	286.99	361.53	1490.65
Jun	1579.98	19.83	480.58	277.78	344.04	1458.57
Jul	1643.00	19.59	498.42	286.99	352.02	1521.99
Aug	1673.00	20.04	499.42	286.99	353.80	1551.67
Sep	1537.00	20.57	481.58	277.78	252.96	1508.40
Oct	1723.99	22.20	495.42	286.99	332.72	1621.90
Nov	1950.98	26.23	410.11	357.16	366.21	1663.94
Dec	2306.99	29.50	421.63	369.02	380.58	2008.52

Personal communication with Bruce Hinton, 2003 Manitoba Hydro

demands given in Table 3.2 are employed as threshold values.

The theory of runs was used to define energy drought events and associated characteristics for Manitoba Hydro's system. However, the definition of drought *duration* is adapted to the specifics of Manitoba Hydro. Based on the analysis of storage levels computed by the SPLASH model, drought duration is defined as follows. The duration of a drought event is not the number of consecutive deficits, but rather the number of periods when the surface elevation of Lake Winnipeg is lower than 714 ft which is the average elevation. This means that after identifying an energy deficit, the lake level should be checked backward and forward from that particular point to determine the beginning and the end of the drought event. A drought event is assumed to start when the lake elevation drops below 714 ft and terminate when it reaches 714 ft again. After obtaining the drought length T_d, the energy drought *severity* for Manitoba Hydro is obtained as the summation of energy deficits during this period. Figure 3.2 illustrates

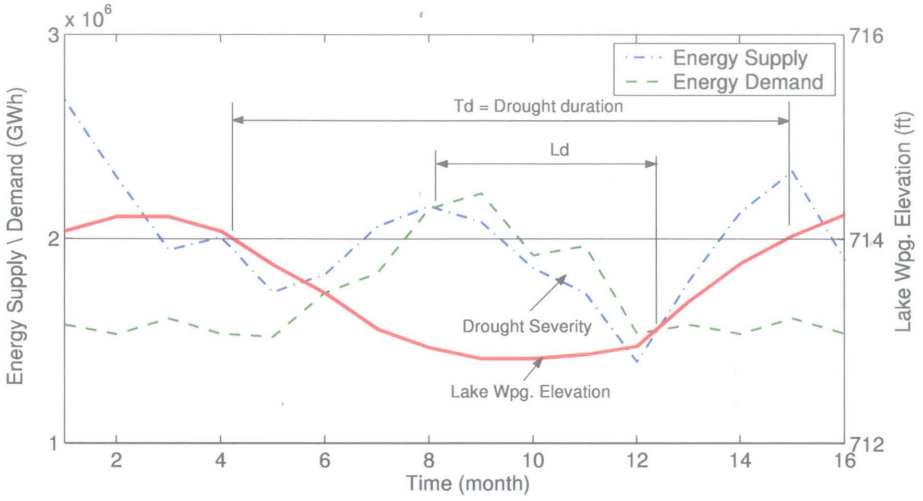


Figure 3.2: Duration and severity of a drought event for Manitoba Hydro.

these concepts. In the analysis, it was observed that in many cases a couple of spring months separates two consecutive drought events because of snow melt. Therefore, when the time between two consecutive drought events are less than or equal to four months, these drought events are combined into one drought event.

As mentioned earlier in this chapter, the *magnitude* of a particular drought event is equal to the ratio of severity and duration for that event. The magnitude of a drought event is generally a poor indicator for a hydropower system because two drought events could have the same magnitude and yet have very different power production capabilities. Duration and severity are probably the two most effective parameters to described drought events.

Based on the energy drought definition for Manitoba Hydro given above, energy drought events were extracted from the historical record. A total of five drought events listed in Table 3.3, were identified from the 1912-98 record data. The most severe

Table 3.3: Historical drought statistics of Manitoba Hydro's system.

Event No	Start Date	End Date	Duration (months)	Severity (GWh)	Magnitude (GWh/month)
1	Sep. 1929	May. 1933	45	57.8	1.3
2	**Nov. 1938**	**Apr. 1942**	**42**	**3309.0**	**78.8**
3	Nov. 1960	May. 1962	19	208.7	11.0
4	Sep. 1976	Mar. 1978	19	104.0	5.5
5	Oct. 1987	Apr. 1992	55	2780.7	50.6

drought event occurred between November 1938 and April 1942, with a duration of 42 months and a severity of 3,309 GWh.

Chapter 4

Traditional Stochastic Models for Data Generation

4.1 Introduction

The objective of drought frequency analysis is to assign a probability of occurrence to a specified drought event or alternatively to estimate the drought event having a prescribed probability of occurrence. Once the drought is defined, the frequency of droughts can be estimated in several ways, using empirical, analytical, or stochastic techniques.

Since it is not possible to predict future hydrologic variables with certainty, stochastic and probabilistic theories are proper tools to employ in the modeling of hydrologic time series. The limitations of using a short historical flow trace in the drought analysis can be overcome by using a stochastic time series model. The stochastic approach

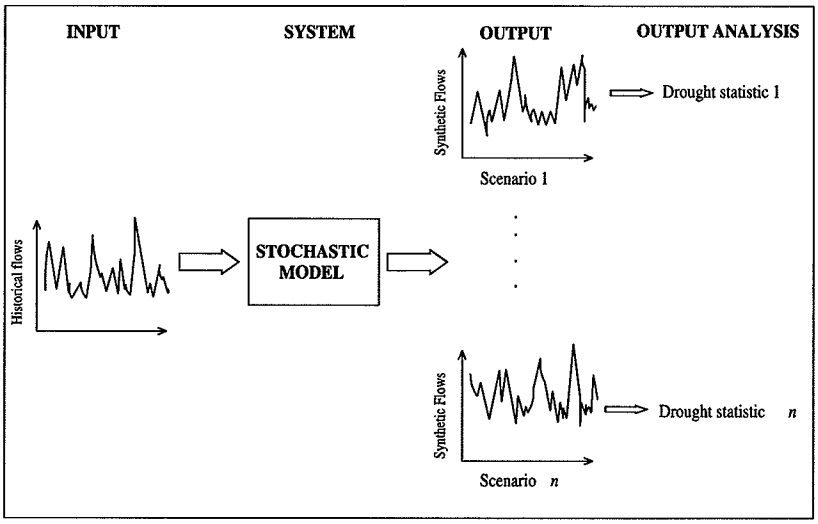


Figure 4.1: Stochastic modeling approach.

is employed in this research project.

Stochastic streamflow models have played a major role in the analysis and planning of water resources systems throughout the world. Stochastic time series models should reproduce important statistics of historical data such as means, variances, autocorrelation, and cross-correlation. The aim of synthetic hydrologic simulations is to produce a large set of equally likely traces that are statistically similar to the historical data. Stochastic models do not generate any new information independent of the data, but they allow one to examine complex characteristics of flows such as multi-year droughts that cannot be obtained by conventional statistical analysis.

Figure 4.1 illustrates the typical approach to stochastic modeling. Historical streamflows are used to choose and verify a stochastic model. The stochastic time series model is then used to generate a number of flow scenarios of given length. From each scenario, one can extract the drought statistic of interest and perform a frequency analysis. The

outcome of the frequency analysis is useful for determining the probability of system drought.

In selecting a stochastic model, it is important to identify which characteristics of the streamflow time series being modeled are important and which are not. The model should preserve the overall mean and variance of the data. Furthermore, there is often considerable persistence or long memory in historical streamflow time series that is particularly important to preserve in drought studies. Obviously, the historical records should guide one in the model development.

The stochastic approach appears to be the most viable technique to analyze complex hydrological characteristics. In this project, the streamflow series at several gauging stations in the Churchill and Nelson River Basin are considered for probabilistic drought analysis of Manitoba Hydro's system. Therefore, the analysis and modeling of both univariate and multivariate time series are required.

In this project, the synthetic data generation will be done in two steps: (1) generation of annual flows, and (2) disaggregation of generated annual flows into monthly flows. This is called a disaggregation approach. The main advantage of this approach is that the generated flows preserve the historical statistical properties at both annual and monthly levels.

The majority of stochastic models involves an assumption that the underlying process is normal. For such processes, a necessary condition is that the data be normally distributed. However, streamflow data are usually not normally distributed. In this case, observed data must be transformed into normal before model calibration. Some

common transformations are described in the next section. Since first order autoregressive [AR(1)] models were selected for the modeling of annual flows in previous studies at Manitoba Hydro, single-site and multi-site AR(1) models are described next. The integration of parameter uncertainty into AR(1) models is also presented in this chapter for both single- and multi-site cases. Finally, the traditional temporal and spatial disaggregation models employed in this research project are presented.

4.2 Marginal Distributions and Data Transformation

Most of the models for generating stochastic processes deal directly with normally distributed random variables. Unfortunately, streamflow distributions tend to be positively skewed. *Salas et al.* [1988] proposed three main approaches for dealing with skewed hydrologic time series.

1. To transform the skewed series into normal before modeling the series;

2. To model the original skewed series and handle the skewness through the probability distribution of the uncorrelated residuals;

3. To find a relationship between the first two moments of the original skewed series and those of the normal series so that the moments of the original skewed series are preserved. *Loucks et al.* [1981] gave an example of this approach using an autoregressive Markov model for annual flows.

Since the best techniques are developed for normal processes, it is generally simpler to transform the skewed variables into normal rather than finding similar procedures for the non-normal variables. However, when transforming the original series into normal, biases in the statistical properties of the generated flows may occur. For example, the mean of the transformed series may be reproduced but not the mean of the original series. *Salas et al.* [1988] suggested to use the first approach if biases are small. Otherwise, other approaches should be envisioned.

A disadvantage of the first approach is revealed when annual flows are disaggregated into monthly flows. When the transformed flows are modeled, the generated monthly flows generally fail to sum up to the previously generated annual flows.

In this project, the first approach has been employed. The 2-parameter lognormal, 3-parameter lognormal, and Box-Cox transformation methods are widely used to transform streamflows to an approximate normal distribution [*Grygier and Stedinger*, 1990, 2001; *Salas*, 1993; *Stedinger*, 1980; *Loucks et al.*, 1981; *Thyer et al.*, 2002; *Salas et al.*, 2000]. It should be noted that in a drought analysis, a good fit in the lower tail of the marginal distribution is more important than the fit in the upper tail.

4.2.1 2-Parameter Lognormal Transformation

In hydrology, flows are often adequately modeled by a 2-parameter lognormal distribution. A logarithmic transformation of the observed flow q_t at time step t, i.e.

$$x_t = \ln(q_t) \tag{4.1}$$

then produces a normally distributed observation x_t.

4.2.2 3-Parameter Lognormal Transformation

This transformation is the same as the 2-parameter lognormal transformation, except that it has a non-zero lower bound τ. q_t is transformed as

$$x_t = \ln(q_t - \tau) \tag{4.2}$$

where q_t must be greater than its lower bound τ. *Stedinger* [1980] and *Stedinger and Taylor* [1982a] suggest procedures for estimating the lower bound τ.

4.2.3 Box-Cox Transformation

When observed flows do not follow a normal or a lognormal distribution, a Box-Cox transformation may be applied. The Box-Cox transformation of q_t is

$$x_t = \begin{cases} (q_t^{\lambda} - 1)/\lambda & \text{if} \quad \lambda \neq 0 \\ \ln(q_t) & \text{if} \quad \lambda = 0 \end{cases} \tag{4.3}$$

where λ is a parameter chosen to ensure that the transformed hydrological data x_t are approximately normally distributed. To define a measure of normality of x_t, one may look at the correlation of point coordinates in a quantile-quantile probability plot, see next section for details. The higher the correlation coefficient, the better a normal distribution fits the data. The optimal value of λ is determined numerically as the

value that maximizes the probability plot correlation coefficient.

4.2.4 Selection of the Transformation Method

In order to decide which transformation method is most appropriate, the Filliben probability plot correlation coefficient test statistic is employed [*Filliben*, 1975; *Vogel*, 1986]. The Filliben statistic r measures the correlation between the transformed ordered flows x_i and their corresponding normal quantiles M_i which can be obtained as

$$\hat{r} = \frac{\sum_{i=1}^{T}(x_i - \bar{x})(M_i - \bar{M})}{\left[\sum_{i=1}^{T}(x_i - \bar{x})^2 \sum_{i=1}^{T}(M_i - \bar{M})^2\right]^{1/2}} \tag{4.4}$$

Among the given transformation methods, the candidate with the highest r-value should be selected.

4.3 Generation of Aggregated Flows

Autoregressive (AR) and autoregressive moving average (ARMA) models are commonly used for generation of aggregated flows such as annual flows. These models reproduce important statistical properties of the annual streamflow series under consideration. In the previous studies of Manitoba Hydro's system, it has been found that based on the Akaike Information Criterion (AIC), the AR(1) model is the best choice between competing ARMA models. The AR(1) model was also used in this study as a basis for comparison with the proposed MS model in the assessment of the probability of system drought. Since both singe-site and multi-site models are

required in the stochastic model frameworks, (see Chapter 8), both univariate and multivariate AR(1) models together with methods for assessing parameter uncertainty are explained in the following sections.

4.3.1 Single-Site Autoregressive Model

The autoregressive (AR) model is one of the most popular models for annual time series in hydrology. This is in part due to its simplicity. Let x_t be a normally distributed variable with mean μ_x and variance σ_x^2. The autoregressive model of order p, denoted AR(p), may be written as

$$x_t = \mu_x + \sum_{j=1}^{p} \phi_j (x_{t-j} - \mu_x) + \varepsilon_t \tag{4.5}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is an uncorrelated (in time) normal random variable with mean zero and variance σ_ε^2. The coefficients $\phi_1, \ldots, \phi_p$ are autoregression coefficients. The parameters of the model are μ_x, σ_x, $\phi_1, \ldots, \phi_p$, and σ_ε^2.

The mean and the variances of an AR(p) process are

$$E[x] = \mu_x \tag{4.6}$$

$$V[x] = \sigma_x^2 = \frac{\sigma_\varepsilon^2}{(1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p)} \tag{4.7}$$

respectively. The autocorrelation function of AR(p) can be calculated as

$$\rho_0 = 1 \tag{4.8a}$$

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}, \quad k \geq 1 \tag{4.8b}$$

where ρ_k is the lag-k autocorrelation coefficient of the variable x_t.

The AR(1) model

$$x_t = \mu_x + \phi_1 (x_{t-1} - \mu_x) + \varepsilon_t \tag{4.9}$$

is a particularly popular special case of (4.5). The parameters of an AR(1) model can be estimated from observed data using the method of moments:

$$\mu_x = \overline{x} = \frac{1}{T} \sum_{t=1}^{T} x_t \tag{4.10}$$

$$\sigma_x = s_x = \frac{1}{T-1} \sum_{t=1}^{T} (x_t - \overline{x})^2 \tag{4.11}$$

$$\phi = r_1 = \frac{\sum_{t=1}^{T-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \tag{4.12}$$

$$\sigma_\varepsilon^2 = \sigma_x^2 (1 - \rho_1^2) \tag{4.13}$$

The autocorrelation function of the AR(1) model can be obtained from (4.8b) as

$$\rho_k = \phi_1^k, \quad k \geq 0 \tag{4.14}$$

The AR(1) model preserves the historical mean, standard deviation, and lag-1 serial correlation coefficient of the transformed variables.

In the previous drought studies for Manitoba Hydro, the impact of uncertainty in the parameters of the AR(1) model was not considered. However, *Stedinger and Taylor* [1982b] found that the impact of parameter uncertainty on derived storage capacity-reliability relationship for reservoir systems was as important as choosing an appropriate stochastic time series model.

In this project, parameter uncertainty will be investigated primarily at the annual level, while the uncertainty of the parameters of the temporal and spatial disaggregation components of the model will be ignored. The rationale for this is that the droughts of interest here are multi-year droughts, so that the effect of the annual model is much more significant than the disaggregation models. The full parameter uncertainty is integrated into the AR(1) model through Bayesian Inference as explained in Section 4.3.3. A general explanation of Bayesian Inference is presented in the next section. A more detailed account can be found in for example *Gelman et al.* [1995] and *Box and Tiao* [1992].

4.3.2 Parameter Uncertainty through Bayesian Inference

In the frequentist approach, the probability of an event is defined by the frequency of that event. According to frequentists, the parameters of a stochastic model are unknown constants that can be estimated using available observed data. The assumption is that the population parameters are equal to the sample estimates. The method of moments and the method of maximum likelihood are the most popular techniques for parameter estimation. However, when the amount of observed data is limited, which

is often the case in hydrology, parameter uncertainty can be considerable and needs to be taken into account.

The Bayesian approach provides a useful mechanism for quantifying parameter uncertainty. In the Bayesian approach, there are two sources of information for learning about unknown parameters: (1) prior information, based for example on expert knowledge, and (2) observed data. The unknown parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$ is treated as a random variable described by a probability density function which expresses the uncertainty about $\boldsymbol{\theta}$. Prior to observing the data, this density function is denoted $p(\boldsymbol{\theta})$ and is called the *prior distribution*; it contains all available knowledge about the parameters other than that associated with the observations. After observing the data $\boldsymbol{X} = (x_1, \ldots, x_T)'$, the density is denoted $p(\boldsymbol{\theta}|\boldsymbol{X})$ and is referred to as the *posterior distribution.* The process of collecting data, and thus acquiring knowledge about $\boldsymbol{\theta}$, is reflected in a reduction in uncertainty. Therefore, the posterior distribution is more concentrated than the prior. Using Bayes' theorem, the posterior distribution of $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{X})} \tag{4.15}$$

where $p(\boldsymbol{X}|\boldsymbol{\theta})$ is the likelihood function $l(\boldsymbol{\theta}|\boldsymbol{X})$, and $p(\boldsymbol{X}) = \int p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the normalizing constant of the posterior distribution necessary to ensure that the posterior distribution integrates to one. $p(\boldsymbol{X})$ is called the marginal distribution of the data or prior predictive distribution. The quantity $p(\boldsymbol{X})$ does not depend on $\boldsymbol{\theta}$, and with fixed $\boldsymbol{X}$, it can be considered a constant so that Bayes' theorem can be written in the more

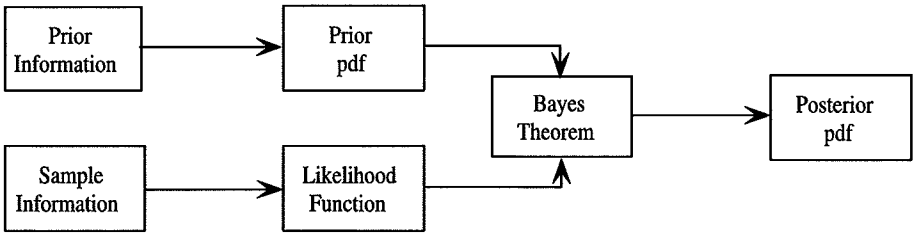


Figure 4.2: The Bayes theorem.

compact form

$$p(\boldsymbol{\theta}|\boldsymbol{X}) \propto l(\boldsymbol{\theta}|\boldsymbol{X})p(\boldsymbol{\theta}) \tag{4.16}$$

which shows that the posterior distribution is *proportional* to the likelihood function multiplied by the prior distribution. In the posterior distribution of the model parameters, the likelihood function provides the contribution from the observed data and the prior distribution provides the contribution from prior knowledge about model parameters.

In the Bayesian approach, the unconditional probability density of a future variable x_f, given observed variables $\boldsymbol{X}$, can be obtained as [*Zellner*, 1971]

$$p(x|\boldsymbol{X}) = \int_{\boldsymbol{\theta}} p(x|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{X})d\boldsymbol{\theta} \tag{4.17}$$

This resulting $p(x|\boldsymbol{X})$ is also known as the posterior predictive or Bayesian probability density function and can be used for simulation of random realizations of x.

Once the posterior distribution of the model parameters has been derived, sets of parameters can be drawn from that distribution [*Vicens et al.*, 1975a, b; *Stedinger and Taylor*, 1982b; *Stedinger et al.*, 1985; *Box and Tiao*, 1992]. In the case of the AR(1)

model, the posterior density of the model parameters can be derived analytically. However, an analytical expression for the posterior distribution may not be possible for complex models. In such cases, Markov chain Monte Carlo (MCMC) simulation methods can be employed to draw samples from the posterior distribution. In this research, an MCMC method is used in the quantification of the parameter uncertainty of the MS model discussed in Chapters 5 and 6.

In streamflow simulation, parameter uncertainty may be incorporated by generating from the Bayesian distribution of streamflow. This is done by first sampling the model parameters from the posterior distributions and then using the parameter set to simulate one streamflow sequence of desired length. This process is repeated for each sequence.

4.3.3 Incorporating Parameter Uncertainty into the Single-Site AR(1) Model

The analysis of parameter uncertainty in univariate and multivariate regression models was developed four decades ago, see for example *Zellner* [1971]. In the hydrological literature, *Vicens et al.* [1975b] defined the single-site AR(1) model as a univariate normal linear regression model and incorporated parameter uncertainty. *Stedinger and Taylor* [1982b] examined the effect of the parameter uncertainty of the annual AR(1) model on estimates of monthly reservoir system reliability. They found that streamflow models may underestimate the frequency of severe droughts if parameter uncertainty is not considered. Despite these conditions, the quantification of parameter

uncertainty is often found lacking in the application of stochastic hydrological models.

For the purpose of parameter estimation, (4.9) may be written as a univariate regression model

$$x_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t \tag{4.18}$$

where β_0, β_1 and σ_ε^2 are the unknown model parameters. The above model can also be expressed in matrix forms as

$$\boldsymbol{X} = \boldsymbol{\omega}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4.19a}$$

$$\begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_T \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{T-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix} \tag{4.19b}$$

The joint likelihood function of $\boldsymbol{\beta}$ and σ_ε^2 given the observed data $\boldsymbol{X}$ can be written as [*Zellner*, 1971]:

$$l(\boldsymbol{\beta}, \sigma_\varepsilon | \boldsymbol{X}) \propto \frac{1}{\sigma_\varepsilon^T} \exp\left\{ -\frac{1}{2\sigma_\varepsilon^2} \left[\nu s_\varepsilon^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \boldsymbol{\omega}' \boldsymbol{\omega} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right] \right\} \tag{4.20}$$

where $\nu = T - 2$ is the degrees of freedom and $\hat{\boldsymbol{\beta}}$ is the standard maximum likelihood estimator of the unknown model parameter vector $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\omega}'\boldsymbol{\omega})^{-1} \boldsymbol{\omega}' \boldsymbol{X} \tag{4.21}$$

The term s_ε^2 is the observed variance of the residuals calculated as

$$s_\varepsilon^2 = \frac{1}{\nu}(\boldsymbol{X} - \boldsymbol{\omega}\hat{\boldsymbol{\beta}})'(\boldsymbol{X} - \boldsymbol{\omega}\hat{\boldsymbol{\beta}}) \tag{4.22}$$

To obtain the posterior density function of the model parameters, $p(\boldsymbol{\beta}, \sigma_\varepsilon|\boldsymbol{X})$, the likelihood function in (4.20) must be multiplied by the prior density function, $p(\boldsymbol{\beta}, \sigma_\varepsilon)$.

If little is known a priori about the model parameters, a diffuse joint prior density function of $\boldsymbol{\beta}$ and σ_ε may be used. A common choice is Jeffrey's noninformative prior:

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma_\varepsilon) &= p(\boldsymbol{\beta})p(\sigma_\varepsilon) \\ &\propto \frac{1}{\sigma_\varepsilon} \end{aligned} \tag{4.23}$$

where $-\infty < \boldsymbol{\beta} < \infty$ and $0 < \sigma_\varepsilon < \infty$. With this prior, the joint posterior density function becomes

$$p(\boldsymbol{\beta}, \sigma_\varepsilon|\boldsymbol{X}) \propto \frac{1}{\sigma_\varepsilon^{T+1}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[\nu s_\varepsilon^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{\omega}'\boldsymbol{\omega}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\} \tag{4.24}$$

Using the conditional probability theorem, the above equation can also be written as

$$p(\boldsymbol{\beta}, \sigma_\varepsilon|\boldsymbol{X}) = p(\boldsymbol{\beta}|\sigma_\varepsilon, \boldsymbol{X})p(\sigma_\varepsilon|\boldsymbol{X}) \tag{4.25}$$

The marginal posterior density of σ_ε can be obtained by integrating (4.24) with

respect to the elements of $\boldsymbol{\beta}$

$$p(\sigma_\varepsilon|\boldsymbol{X}) \propto \frac{1}{\sigma_\varepsilon^{\nu+1}} \exp\left(-\frac{\nu s_\varepsilon^2}{2\sigma_\varepsilon}\right) \tag{4.26}$$

which is the kernel of an inverse-Gamma distribution with parameters $(\nu/2, \nu s_\varepsilon^2/2)$ or a scaled inverse-χ^2-distribution with parameters (ν, s_ε^2).

Once the marginal posterior density of σ_ε, $p(\sigma_\varepsilon|\boldsymbol{X})$, and the joint posterior density of $\boldsymbol{\beta}$ and σ_ε, $p(\boldsymbol{\beta}, \sigma_\varepsilon|\boldsymbol{X})$, are known, the conditional density function of $\boldsymbol{\beta}$ given σ_ε and $\boldsymbol{X}$ can be obtained from (4.25) as

$$p(\boldsymbol{\beta}|\sigma_\varepsilon, \boldsymbol{X}) \propto \frac{1}{\sigma_\varepsilon^{T-\nu}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'(\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right]\right\} \tag{4.27}$$

which is a multivariate (bivariate) normal distribution with mean $\hat{\boldsymbol{\beta}}$ and covariance matrix $\sigma_\varepsilon^2(\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}$.

In summary, sampling from the joint posterior distribution can be done as follows

1. Calculate $\hat{\boldsymbol{\beta}}$ from (4.21) and s_ε^2 from (4.22).

2. Draw σ_ε^2 from the scaled inverse-χ^2-distribution

$$\sigma_\varepsilon^2 \sim Inv - \chi^2(\nu, s_\varepsilon^2) \tag{4.28}$$

In practice, one can simulate σ_ε^2 as

$$\sigma_\varepsilon^2 = \frac{\nu s_\varepsilon^2}{\tilde{\chi}^2} \tag{4.29}$$

where $\tilde{\chi}^2$ is a random variable drawn from a standard χ^2-distribution with ν degrees of freedom.

3. Given the simulated σ_ε^2, draw $\boldsymbol{\beta}$ from the multivariate normal distribution

$$\boldsymbol{\beta} \sim N(\hat{\boldsymbol{\beta}}, \sigma_\varepsilon^2(\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}) \tag{4.30}$$

For each set of simulated model parameters, one streamflow sequence of desired length can be generated. Parameter simulation should be repeated for the generation of each new streamflow sequence.

4.3.4 Multi-Site Autoregressive Model

In the analysis of multiple time series, vector and matrix notation is needed but the basic principles are similar to univariate analysis. The main properties of multi-site AR models are the mean, the cross-covariance, and the autocovariance structure.

Let $q_t^{(n)}$ be the observed annual flow at site n $(n = 1, \ldots, N)$ in year t $(t = 1, \ldots, T)$. If the observed annual flows do not follow a normal distribution, they must be transformed as described in Section 4.2. The transformation function may be different for each site. Let $x_t^{(n)}$ be the normally distributed transformed annual flow. Before estimating the model parameters, it is common practice to standardize the transformed flows. For example, the elements of the standardized annual sequence

of the first site can be obtained as

$$z_t^{(1)} = \frac{x_t^{(1)} - \bar{x}^{(1)}}{s_x^{(1)}} \tag{4.31}$$

where $\bar{x}^{(1)}$ is the sample mean and $s_x^{(1)}$ is the sample standard deviation of the transformed flows at site 1.

The multi-site AR(1) model suggested by *Matalas* [1967] is defined as

$$\boldsymbol{Z}_t = \boldsymbol{A}\boldsymbol{Z}_{t-1} + \boldsymbol{B}\boldsymbol{\epsilon}_t \tag{4.32}$$

in which $\boldsymbol{A}$ and $\boldsymbol{B}$ are $(N \times N)$ parameter matrices and $\boldsymbol{Z}_t$ and $\boldsymbol{Z}_{t-1}$ are $(N \times 1)$ column vectors of observations at time t and $t-1$, respectively. The $(N \times 1)$ vector $\boldsymbol{\epsilon}_t$ contains uncorrelated normally distributed random variables with zero mean and unit variance. Hence, the covariance of $\boldsymbol{\epsilon}_t$ is the identity matrix, $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \boldsymbol{I}$. It is also assumed that the vector $\boldsymbol{\epsilon}_t$ is uncorrelated with $\boldsymbol{Z}_{t-1}$, i.e. $E[\boldsymbol{Z}_{t-1}\boldsymbol{\epsilon}_t'] = \boldsymbol{0}$.

The parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are generally obtained using the method of moments. By multiplying (4.32) by $\boldsymbol{Z}_t'$ and $\boldsymbol{Z}_{t-1}'$ and taking expectation, one can obtain the following expressions:

$$\hat{\boldsymbol{A}} = \boldsymbol{M}_1\boldsymbol{M}_0^{-1} \tag{4.33}$$

$$\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}' = \boldsymbol{M}_0 - \boldsymbol{M}_1\boldsymbol{M}_0^{-1}\boldsymbol{M}_1' \tag{4.34}$$

where $\boldsymbol{M}_0$ is the lag-zero cross-covariance matrix

$$\hat{\boldsymbol{M}}_0 = E[\boldsymbol{Z}_t \boldsymbol{Z}_t'] \tag{4.35}$$

and $\boldsymbol{M}_1$ is the lag-one cross-covariance matrix

$$\hat{\boldsymbol{M}}_1 = E[\boldsymbol{Z}_t \boldsymbol{Z}_{t-1}'] \tag{4.36}$$

The lag-k cross-covariance matrix for the multi-site AR(1) model is

$$\boldsymbol{M}_k = \boldsymbol{A}^k \boldsymbol{M}_0, \quad k \geq 0 \tag{4.37}$$

Determination of $\boldsymbol{B}$ from knowledge of $\boldsymbol{B}\boldsymbol{B}'$ is a classical problem in multivariate statistics. Methods to solve $\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}'$ for $\hat{\boldsymbol{B}}$ are given in *Salas et al.* [1988] and *Bras and Rodríguez-Iturbe* [1993]. Software such as MATLAB have the capability to solve the above estimation problem.

The multi-site AR(1) model has the property to preserve the historical means, standard deviations, and lag-zero and lag-one cross-correlations. However, the AR(1) model may not preserve the long-term persistence observed in the historical data. Because autoregressive models are "short memory" models which means that their autocorrelation functions decay very fast as the time lag increases, *Salas et al.* [1988] noted that AR(1) models produce smaller droughts for a given demand level. Therefore, if the observed data suggest long-term persistence, alternative models should be

used.

4.3.5 Incorporating Parameter Uncertainty into the Multi-Site AR(1) Model

Since this research project has a multi-site dimension, parameter uncertainty must be considered at the multi-site level. The quantification of parameter uncertainty in multi-site AR(1) models was presented by *Valdes and Rodriguez-Iturbe* [1977] in the hydrologic literature. Incorporating parameter uncertainty into multi-site AR(1) models is not straightforward. Unlike in the single-site case, the structure of the standard multivariate regression model cannot be used for the multi-site AR(1) model. However, a special case of the generalized traditional multivariate regression model can be employed to represent the multi-site AR(1) model structure [*Tiao and Zellner*, 1964].

Let $x_1^{(n)}, \ldots, x_T^{(n)}$ denote the normally distributed annual flow series at site n. The multi-site AR(1) model can be written

$$\begin{aligned} x_t^{(1)} &= \beta_{10} + \beta_{11} x_{t-1}^{(1)} + \beta_{12} x_{t-1}^{(2)} + \ldots \beta_{1N} x_{t-1}^{(N)} + \varepsilon_t^{(1)} \\ x_t^{(2)} &= \beta_{20} + \beta_{21} x_{t-1}^{(1)} + \beta_{22} x_{t-1}^{(2)} + \ldots \beta_{2N} x_{t-1}^{(N)} + \varepsilon_t^{(2)} \\ &\vdots \\ x_t^{(N)} &= \beta_{N0} + \beta_{N1} x_{t-1}^{(1)} + \beta_{N2} x_{t-1}^{(2)} + \ldots \beta_{NN} x_{t-1}^{(N)} + \varepsilon_t^{(N)} \end{aligned} \tag{4.38}$$

where β_{ij} are regression coefficients and ε_t are independent identically distributed zero

mean normal random variables with covariance matrix $\boldsymbol{\Sigma}$, $\varepsilon_t \sim N_N(0, \boldsymbol{\Sigma})$.

The above model can be expressed in a compact normal regression model

$$\boldsymbol{X} = \boldsymbol{\Omega}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4.39a}$$

$$\begin{bmatrix} \boldsymbol{X}^{(1)} \\ \vdots \\ \boldsymbol{X}^{(N)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\omega} & & \\ & \ddots & \\ & & \boldsymbol{\omega} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}^{(1)} \\ \vdots \\ \boldsymbol{\varepsilon}^{(N)} \end{bmatrix} \tag{4.39b}$$

where

$$\boldsymbol{X}^{(n)} = \begin{bmatrix} x_2^{(n)} \\ x_3^{(n)} \\ \vdots \\ x_T^{(n)} \end{bmatrix}, \quad \boldsymbol{\omega} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_1^{(N)} \\ 1 & x_2^{(1)} & \dots & x_2^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T-1}^{(1)} & \dots & x_{T-1}^{(N)} \end{bmatrix}, \quad \boldsymbol{\beta}_n = \begin{bmatrix} \beta_{n0} \\ \beta_{n1} \\ \vdots \\ \beta_{nN} \end{bmatrix}, \quad \boldsymbol{\varepsilon}^{(n)} = \begin{bmatrix} \varepsilon_2^{(n)} \\ \varepsilon_3^{(n)} \\ \vdots \\ \varepsilon_T^{(n)} \end{bmatrix} \tag{4.40}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are the model parameter matrices. Using (4.39a), the joint likelihood function of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ given the observation vector $\boldsymbol{X}$ can be written as

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \boldsymbol{X}) \propto |\boldsymbol{\Sigma}|^{\frac{1}{2}T} \exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\omega}'\boldsymbol{\omega}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right] \tag{4.41}$$

In (4.41), tr denotes the trace operator, $\otimes$ denotes Kronecker or direct matrix multiplication, and $\hat{\boldsymbol{\beta}}' = (\hat{\boldsymbol{\beta}}_1', \dots, \hat{\boldsymbol{\beta}}_N')$ is the standard maximum likelihood estimator of the

unknown model parameter vector $\boldsymbol{\beta}$, i.e.

$$\hat{\boldsymbol{\beta}}_n = (\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}\boldsymbol{\omega}'\boldsymbol{X}^{(n)}. \tag{4.42}$$

The matrix $\boldsymbol{S} = \{s_{ij}\}$ is proportional to the sample covariance matrix. The elements of $\boldsymbol{S}$ are defined as

$$s_{ij} = (\boldsymbol{X}^{(i)} - \boldsymbol{\omega}\hat{\boldsymbol{\beta}}_i)'(\boldsymbol{X}^{(j)} - \boldsymbol{\omega}\hat{\boldsymbol{\beta}}_j), \quad i,j = 1,\ldots,N \tag{4.43}$$

It is assumed that little is known a priori about the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ and that they are independently distributed. Diffuse priors are used when no information about the model parameters are known. For the multi-site case, *Zellner* [1971] gives Jeffrey's diffuse prior as

$$\begin{aligned} p(\boldsymbol{\beta},\boldsymbol{\Sigma}) &= p(\boldsymbol{\beta})p(\boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(N+1)} \end{aligned} \tag{4.44}$$

Multiplying the likelihood function in (4.41) by the prior distribution in (4.44), the posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ can be obtained as

$$p(\boldsymbol{\beta},\boldsymbol{\Sigma}|\boldsymbol{X}) \propto |\boldsymbol{\Sigma}|^{\frac{1}{2}(T+N+1)} \exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\omega}'\boldsymbol{\omega}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right] \tag{4.45}$$

The above equation can also be written

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\boldsymbol{X}) = p(\boldsymbol{\beta}|\boldsymbol{\Sigma}, \boldsymbol{X})p(\boldsymbol{\Sigma}|\boldsymbol{X}) \tag{4.46}$$

with

$$p(\boldsymbol{\beta}|\boldsymbol{\Sigma}, \boldsymbol{X}) \propto |\boldsymbol{\Sigma}|^{\frac{1}{2}k} \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\omega}'\boldsymbol{\omega}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right] \tag{4.47}$$

and

$$p(\boldsymbol{\Sigma}|\boldsymbol{X}) \propto |\boldsymbol{\Sigma}|^{\frac{1}{2}\nu} \exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right] \tag{4.48}$$

where k denotes the size of the $\boldsymbol{\beta}$ vector and $\nu = T - k + N + 1$ denotes the degrees of freedom. It can be seen that the conditional posterior distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\Sigma}$ in (4.47) is a multivariate normal distribution with mean $\hat{\boldsymbol{\beta}}$ and covariance matrix $\boldsymbol{\Sigma} \otimes (\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}$ and the posterior distribution of $\boldsymbol{\Sigma}$ in (4.48) is an inverted Wishart distribution.

Like in the single-site AR(1) model, the simulation of the model parameters is a two step process. After estimating the parameter matrices in (4.42) and (4.43), the covariance matrix $\boldsymbol{\Sigma}$ is first simulated from the inverted Wishart distribution:

$$\boldsymbol{\Sigma} \sim \text{Inv-Wishart}(\nu, \boldsymbol{S}) \tag{4.49}$$

and then, for given $\boldsymbol{\Sigma}$, $\boldsymbol{\beta}$ is simulated from a multivariate normal distribution:

$$\boldsymbol{\beta} \sim \mathrm{N}_d(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma} \otimes (\boldsymbol{\omega}'\boldsymbol{\omega})^{-1}/\nu) \tag{4.50}$$

When parameter uncertainty is considered as defined above, generated steamflow sequences reflect both the natural hydrologic variability of streamflows and uncertainty about the model parameters.

4.4 Disaggregation of Aggregated Flows

Among the various techniques for time series analysis developed in hydrology, the disaggregation approach seems to be the most suitable for this research project [*Valencia and Schaake*, 1973]. Disaggregation models can be used at *temporal* and *spatial* levels. In temporal disaggregation, generated annual flows are disaggregated into monthly flows. In spatial disaggregation, generated aggregated flows from a region are disaggregated into flows at individual sites within the same region. The generation of aggregated flows and the disaggregation procedure are independent.

Typical sequential linear models such as periodic autoregressive (PAR) and periodic autoregressive moving average (PARMA) or other types of models preserve the statistical characteristics at the monthly level but not at the annual level. The major advantage of disaggregation models is that they preserve historical statistics of both annual and monthly streamflows. Disaggregation models also have an additivity property when observed data are normally distributed. The seasonal values adds up to the annual values. In most cases, however, observed streamflows are not normally distributed. When the transformed flows are modeled, disaggregated monthly flows generally fail to sum to the generated annual flow. In this case, an adjustment procedure must be used to ensure that disaggregated flows add up to aggregated flows.

The first well-accepted disaggregation model was developed by *Valencia and Schaake* [1973]. A number of variations of this model have subsequently been proposed. The first disaggregation models were multivariate (simultaneous disaggregation of annual flows at several sites). *Valencia and Schaake* [1973], *Mejia and Rousselle* [1976], *Tao and Delleur* [1976], and *Hoshi and Burges* [1979] are typical examples of such models. These models attempt to reproduce all covariance properties between monthly flows as well as those between monthly and annual flows among all sites and time steps. This results in a huge number of parameters which is not desirable when the sample size is small. Since short record length is a common case in hydrology, attempts have been made to reduce the number of parameters by preserving a reduced number of statistics of the data.

One way of reducing the number of parameters is to use *condensed disaggregation models* such as those developed by *Lane* [1979], *Stedinger et al.* [1985], *Grygier and Stedinger* [1988], and *Grygier and Stedinger* [1990]. Condensed models reduce the number of parameters to be estimated by modeling only a reduced number correlations among the monthly flows. Lane's condensed model which was employed in this research project, is described later in Section 4.4.3.

As an alternative to condensed models, *Salas et al.* [1988], *Loucks et al.* [1981], and *Bras and Rodríguez-Iturbe* [1993] suggested that the disaggregation be done in steps (stages or cascades). For example, annual flows at one or more sites can be disaggregated to monthly flows in two or more steps. *Santos and Salas* [1992] analyzed the potential of the step disaggregation idea. In the Santos-Salas model, an annual flow

value may be disaggregated into 12 monthly flows by first disaggregating the annual flow into the first month flow and the sum of the remaining 11 months. Then the latter sum is disaggregated into the second month flow and the sum of the remaining 10 months, and so on until all monthly flows have been obtained. *Salas* [1993] notes that the Santos-Salas step model is very similar to the condensed model of *Stedinger et al.* [1985].

Another stepwise disaggregation model was proposed by *Koutsoyiannis* [1992]. This approach, called the *dynamic disaggregation model (DDM)*, is very similar to the Santos and Salas stepwise disaggregation procedure. However, there are some differences between them. At each steps, the DDM uses a nonlinear generation module that disaggregates a given amount into two parts. However, this module adds notable mathematical complexity. Therefore, *Koutsoyiannis and Manetas* [1996] presented a simpler method that keeps some ideas of the DDM approach. Their disaggregation method is based on three simple ideas. First, it starts by using a sequential PAR(1) model and keeps its formalism and parameter set. Second, it uses accurate adjusting procedures to allocate the error in the additive property. Third, it uses repetition in order to improve the approximation of statistics that are not explicitly preserved by the adjusting procedures.

Tarboton et al. [1998] developed a *nonparametric disaggregation model* (NPD). Their NPD model can capture the dependence structure present in the observed data without imposing arbitrary linearity or distributional assumptions. The necessary joint probability density functions are estimated directly from the historical data. These

methods circumvent the drawbacks of the parametric methods. The methods are data-driven and relatively automatic, so non-linear dependence will be incorporated to the extent suggested by the data. The empirical marginal distributions are reproduced and normalizing transformations are avoided. The real advantage of the NPD method is the ability to model complex relationships between aggregated and disaggregated flows. However, a disadvantage of the NPD approach is that it is computationally intensive.

In this research project, Lane's condensed disaggregation model [*Lane*, 1979] is employed to disaggregate generated annual flows at five sites into monthly flows at those sites. Furthermore, Mejia and Rousselle's model [*Mejia and Rousselle*, 1976] is employed to spatially disaggregate aggregated annual flows for the basin into annual flows for the five basin sites. The same model is also used to disaggregate generated aggregated monthly Local Flows at the Burtwood and the Nelson River into monthly flows at the individual local flow sites. In the following sections, the employed disaggregation models as well as Valencia-Schaake model are presented in more details.

4.4.1 Valencia-Schaake Disaggregation Model

The basic disaggregation model developed by *Valencia and Schaake* [1973] can be used both for temporal or spatial disaggregation. For example, in order to spatially disaggregate the aggregated zero mean annual basin flow (higher-level variable) x_t in year t into N zero mean annual basin site flows (lower-level variables) $\boldsymbol{y}_t = (y_t^{(1)}, \ldots, y_t^{(N)})'$

for the same year, the Valencia-Schaake disaggregation model can be written as

$$\boldsymbol{y}_t = \boldsymbol{A}x_t + \boldsymbol{B}\boldsymbol{\epsilon}_t \tag{4.51}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are parameter matrices with dimensions $(N \times 1)$ and $(N \times N)$, respectively, and $\boldsymbol{\epsilon}_t$ is an $(N \times 1)$ column vector of independent standard normal variables with zero mean and unit variance. As in the multi-site annual model, the covariance matrix of $\boldsymbol{\epsilon}_t$ is the identity matrix, $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \boldsymbol{I}$, and vector $\boldsymbol{\epsilon}_t$ is uncorrelated with x_t, $E[x_t\boldsymbol{\epsilon}_t'] = \boldsymbol{0}$.

The parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ may be estimated using the method of moments. Prior to parameter estimation, the marginal distribution of all observed aggregated annual flows and annual flows of each site must be transformed to normal. Then the transformed observed data sets must be centered by subtracting their sample mean. Using the method of moments, the estimates of parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be obtained as

$$\hat{\boldsymbol{A}} = \boldsymbol{S}_{YX}\boldsymbol{S}_{XX}^{-1} \tag{4.52}$$

$$\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}' = \boldsymbol{S}_{YY} - \boldsymbol{A}\boldsymbol{S}_{XY} \tag{4.53}$$

where

$$\boldsymbol{X} = [x_1 \dots x_T], \qquad \boldsymbol{Y} = \begin{bmatrix} y_1^{(1)} & \dots & y_T^{(1)} \\ \vdots & \ddots & \vdots \\ y_1^{(N)} & \dots & y_T^{(N)} \end{bmatrix} \tag{4.54}$$

and $\boldsymbol{S_{UV}}$ represents the sample covariance of the vectors $\boldsymbol{U}$ and $\boldsymbol{V}$. Because $\boldsymbol{U}$ and $\boldsymbol{V}$ have zero mean, the sample covariance matrices can be calculated as

$$\boldsymbol{S_{XX}} = (1/T)\boldsymbol{XX'} \tag{4.55a}$$

$$\boldsymbol{S_{YX}} = (1/T)\boldsymbol{YX'} \tag{4.55b}$$

$$\boldsymbol{S_{YY}} = (1/T)\boldsymbol{YY'} \tag{4.55c}$$

$$\boldsymbol{S_{XY}} = \boldsymbol{S'_{YX}} \tag{4.55d}$$

The *Valencia and Schaake* [1973] spatial disaggregation model in (4.51) preserves the correlation between all lower-level variables and between all lower-level and higher-level variables. However, when this model is used to disaggregate annual flows into monthly flows, the correlation coefficient between the last month of one year and the first month of the following year is not explicitly preserved. *Mejia and Rousselle* [1976] tried to overcome this problem by including a new parameter matrix and a vector of monthly values from the previous year at the right hand site of the Valencia and Schaake model in (4.51).

4.4.2 Mejia and Rousselle Disaggregation Model

When spatial disaggregation is considered as in (4.51), the Mejia and Rousselle model takes the form

$$\boldsymbol{y}_t = \boldsymbol{A}x_t + \boldsymbol{B}\boldsymbol{\epsilon}_t + \boldsymbol{C}\boldsymbol{y}_{t-1} \tag{4.56}$$

where $\boldsymbol{y}_{t-1}$ is an $(N \times 1)$ column matrix of disaggregated annual lower-level variables from the previous time step (year) for each site and $\boldsymbol{C}$ is a new $(N \times N)$ parameter matrix. All other terms remain the same as for the basic model in (4.51). The parameters $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ may be estimated using the method of moments as

$$\hat{\boldsymbol{A}} = \left[\boldsymbol{S}_{Y_tX_t} - \boldsymbol{S}_{Y_tY_{t-1}}\boldsymbol{S}^{-1}_{Y_{t-1}Y_{t-1}}\boldsymbol{S}'_{X_tY_{t-1}}\right] \times \left[\boldsymbol{S}_{X_tX_t} - \boldsymbol{S}_{X_tY_{t-1}}\boldsymbol{S}^{-1}_{Y_{t-1}Y_{t-1}}\boldsymbol{S}'_{X_tY_{t-1}}\right]^{-1} \tag{4.57a}$$

$$\hat{\boldsymbol{C}} = \left[\boldsymbol{S}_{Y_tY_{t-1}} - \hat{\boldsymbol{A}}\boldsymbol{S}_{X_tY_{t-1}}\right]\boldsymbol{S}^{-1}_{Y_{t-1}Y_{t-1}} \tag{4.57b}$$

$$\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}' = \boldsymbol{S}_{Y_tY_t} - \hat{\boldsymbol{A}}\boldsymbol{S}_{X_tY_t} - \hat{\boldsymbol{C}}\boldsymbol{S}'_{Y_tY_{t-1}} \tag{4.57c}$$

Lane [1981] showed that if one uses the above equations directly in the estimation of parameters, some important moments are not preserved. He suggested the following equations be used instead of the sample covariance matrices $\boldsymbol{S}_{X_tY_{t-1}}$ and $\boldsymbol{S}_{Y_tY_{t-1}}$:

$$\boldsymbol{S}^*_{X_tY_{t-1}} = \boldsymbol{S}_{X_tX_{t-1}}\boldsymbol{S}^{-1}_{X_tX_t}\boldsymbol{S}_{X_tY_t} \tag{4.58a}$$

$$\boldsymbol{S}^*_{Y_tY_{t-1}} = \boldsymbol{S}_{Y_tY_{t-1}} + \boldsymbol{S}_{Y_tX_t}\boldsymbol{S}^{-1}_{Y_tY_t}\left[\boldsymbol{S}^*_{X_tY_{t-1}} - \boldsymbol{S}_{X_tY_{t-1}}\right] \tag{4.58b}$$

In addition to the properties of the Valencia-Schaake model, this model also pre-

serves the lag-1 cross-correlations between the aggregated basin flows and disaggregated basin site flows, and between basin site flows themselves. Although this model is slightly more involved than the basic model, it is still quite straightforward. However, the problem of the excessive number of parameters in the Valencia and Schaake model is made worse especially when it is used as a temporal disaggregation model to disaggregate generated multi-site annual flows into monthly flows.

4.4.3 Lane's Condensed Disaggregation Model

The above disaggregation models have a large number of parameters when they are used for multi-site disaggregation of annual flows into monthly flows. A large number of parameters estimated from a small sample gives large estimation errors. Since short length of records is a common case in hydrology, attempts have been made to reduce the number of the parameters in disaggregation models by preserving only important characteristics of the model.

Lane [1979] developed a condensed version of the Valencia-Schaake disaggregation model by relaxing the constraint that the model reproduce the observed correlations between every pair of monthly flows. Lane's model reproduces only the concurrent and lag-1 month-to-month correlations, and the correlations between the monthly and annual flows. This condensed model is known as the LAST model [*Lane*, 1979]. Since in the condensed disaggregation models, the annual flows are disaggregated one-month-at-a-time, one of the important attributes of the original disaggregation model, the additivity property, is lost. To preserve the additivity property, one must

adjust the disaggregated monthly flows. Adjustment procedures are discussed in the next section. However, any adjustment may distort the marginal distribution of the generated monthly data.

In Lane's condensed multi-site temporal disaggregation model, zero mean annual flows at N sites, $\boldsymbol{x}_t = (x_t^{(1)}, \ldots, x_t^{(N)})'$, can be disaggregated into zero mean flows for month m at the N sites, $\boldsymbol{y}_{tm} = (y_{tm}^{(1)}, \ldots, y_{tm}^{(N)})'$. For the first month of the year, Lane's disaggregation model can be written as

$$\boldsymbol{y}_{t1} = \boldsymbol{A}_1 \boldsymbol{x}_t + \boldsymbol{B}_1 \boldsymbol{\epsilon}_{t1} \tag{4.59a}$$

and for $m = 2, \ldots, 12$ as

$$\boldsymbol{y}_{tm} = \boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{B}_m \boldsymbol{\epsilon}_{tm} + \boldsymbol{C}_m \boldsymbol{y}_{t,m-1} \tag{4.59b}$$

where $\boldsymbol{A}_m$, $\boldsymbol{B}_m$, and $\boldsymbol{C}_m$ are $(N \times N)$ parameter matrices for month m, $\boldsymbol{y}_{t,m-1}$ is an $(N \times 1)$ vector of previous month flows at all N sites, and $\boldsymbol{\epsilon}_{t,m}$ is an $(N \times 1)$ vector of independent (uncorrelated) zero-mean, unit variance normal random variables. When disaggregation of annual flows into monthly flows is considered, there are 12 sets of parameters $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ and 11 sets of parameters $\boldsymbol{C}_m$.

Prior to parameter estimation, the marginal distribution of all observed monthly and annual flows at each site must be transformed to normal if they are not already normally distributed. Different transformation models can be used for different months at the same site. Furthermore, the transformed flows should be centered.

For the purpose of parameter estimation, the above models are presented as

$$\boldsymbol{Y}_1 = \boldsymbol{A}_1\boldsymbol{X} + \boldsymbol{B}_1\boldsymbol{\epsilon}_1 \tag{4.60a}$$

and for $m = 2, \ldots, 12$ as

$$\boldsymbol{Y}_m = \boldsymbol{A}_m\boldsymbol{X} + \boldsymbol{B}_m\boldsymbol{\epsilon}_m + \boldsymbol{C}_m\boldsymbol{Y}_{m-1} \tag{4.60b}$$

where $\boldsymbol{Y}_m$ represents the transformed month m flows at all sites, $\boldsymbol{X}$ represents the transformed annual flows at all sites, and $\boldsymbol{Y}_{m-1}$ represents the transformed previous month flows at each site. The dimension of all matrices is $(N \times T)$ and can be formed as

$$\boldsymbol{X} = \begin{bmatrix} x_1^{(1)} & \cdots & x_T^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(N)} & \cdots & x_T^{(N)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{Y}_m = \begin{bmatrix} y_{1m}^{(1)} & \cdots & y_{Tm}^{(1)} \\ \vdots & \ddots & \vdots \\ y_{1m}^{(N)} & \cdots & y_{Tm}^{(N)} \end{bmatrix} \tag{4.61}$$

The parameter matrices $\boldsymbol{A}_m$, $\boldsymbol{B}_m$, and $\boldsymbol{C}_m$ may be estimated using the method of moments. For the first month, the parameters of the model given in (4.60a) are obtained as

$$\hat{\boldsymbol{A}}_1 = \boldsymbol{S}_{\boldsymbol{Y}_1\boldsymbol{X}}\boldsymbol{S}_{\boldsymbol{X}\boldsymbol{X}}^{-1} \tag{4.62}$$

$$\hat{\boldsymbol{B}}_1\hat{\boldsymbol{B}}_1' = \boldsymbol{S}_{\boldsymbol{Y}_1\boldsymbol{Y}_1} - \hat{\boldsymbol{A}}_1\boldsymbol{S}_{\boldsymbol{X}\boldsymbol{Y}_1} \tag{4.63}$$

and for month m, $m = 2, \ldots, 12$, the parameters of the model given in (4.60b) are obtained as

$$\hat{A}_m = \left[S_{Y_m X} - S_{Y_m Y_{m-1}} S^{-1}_{Y_{m-1} Y_{m-1}} S_{Y_{m-1} X}\right] \times \left[S_{XX} - S_{XY_{m-1}} S^{-1}_{Y_{m-1} Y_{m-1}} S_{Y_{m-1} X}\right]^{-1} \tag{4.64}$$

$$\hat{C}_m = \left[S_{Y_m Y_{m-1}} - \hat{A}_m S_{XY_{m-1}}\right] S^{-1}_{Y_{m-1} Y_{m-1}} \tag{4.65}$$

$$\hat{B}_m \hat{B}'_m = S_{Y_m Y_m} - \hat{A}_m S_{XY_m} - \hat{C}_m S_{Y_{m-1} Y_m} \tag{4.66}$$

The main advantage of Lane's model is the reduction of the number of parameters. It requires fewer parameters than the full Valencia-Schaake model because it does not explicitly preserve high-lag month-to-month correlations. A shortcoming of the model is that the monthly data do not exactly add up to give the annual time series because the months are not generated jointly. Lane concluded that the benefit of the parameter reduction far outweighs this shortcoming. In fact, it should be noted that the additivity problem is common to all disaggregation models if the data are transformed to normal. The solution to this problem is to adjust the seasonal flows so that they add up exactly to the annual values.

4.4.4 Streamflow Adjustment Procedure

When the transformed flows rather than the original observed flows are modeled, generated monthly flows generally fail to sum to the previously generated annual flow. Therefore, generated monthly flows need to be adjusted in order for their sum to equal

the generated annual flows. Several flow adjustment procedures are proposed in *Lane* [1979] and *Stedinger and Vogel* [1984] and a comparison of these procedures has been done by *Grygier and Stedinger* [1988]. In this project, two adjustment procedures are employed.

Proportional Adjustment

Grygier and Stedinger [1988] found that the *proportional procedure* distorted low flows less than other procedures. In the proportional procedure, in order to preserve the additivity property in year t, the corrections are allocated proportionally to the originally generated monthly flows q_{tm} by multiplying by a factor d_t.

$$q^*_{tm} = d_t q_{tm}, \qquad m = 1, \ldots, 12 \tag{4.67}$$

The factor d_t is calculated as

$$d_t = \frac{q_t}{\sum_{m=1}^{12} q_{tm}} \tag{4.68}$$

where q_t is the generated annual flow. After this adjustment the sum of adjusted monthly flows $(q^*_{t1} + \cdots + q^*_{t,12})$ will be exactly equal to q_t.

This procedure is also used in the software packages LAST [*Lane*, 1979], SPIGOT [*Grygier and Stedinger*, 1990], and SAMS2000 [*Salas et al.*, 2000].

Standard Deviation Adjustment

[*Lane*, 1979] proposed to adjust flows in proportional to their standard deviation

$$q^*_{tm} = q_{tm} + d_t s_m, \qquad m = 1, \ldots, 12 \tag{4.69}$$

where the factor d_t is calculated as

$$d_t = \frac{q_t - \sum_{m=1}^{12} q_{tm}}{\sum_{m=1}^{12} s_m} \tag{4.70}$$

where s_m is the standard deviation of the month m.

Chapter 5

Single-Site Markov-Switching Model

5.1 Introduction

In recent years, there has been a growing awareness of the existence of low-frequency climate signals that affect the hydrology of many parts of the world. Low-frequency climate variability can often be associated with oceanic circulations because the oceans provide the inertia needed to sustain variability at decadal time scales. Although our understanding of the physics that govern these oceanic circulations and their relationship to surface climate is incomplete, observations confirm their existence.

Several ocean driven mechanisms have been linked to the climate of North America. The El Niño-Southern Oscillation has received much attention, because of its strong impact on the global climate. El Niño is an intermittent phenomenon, occurring

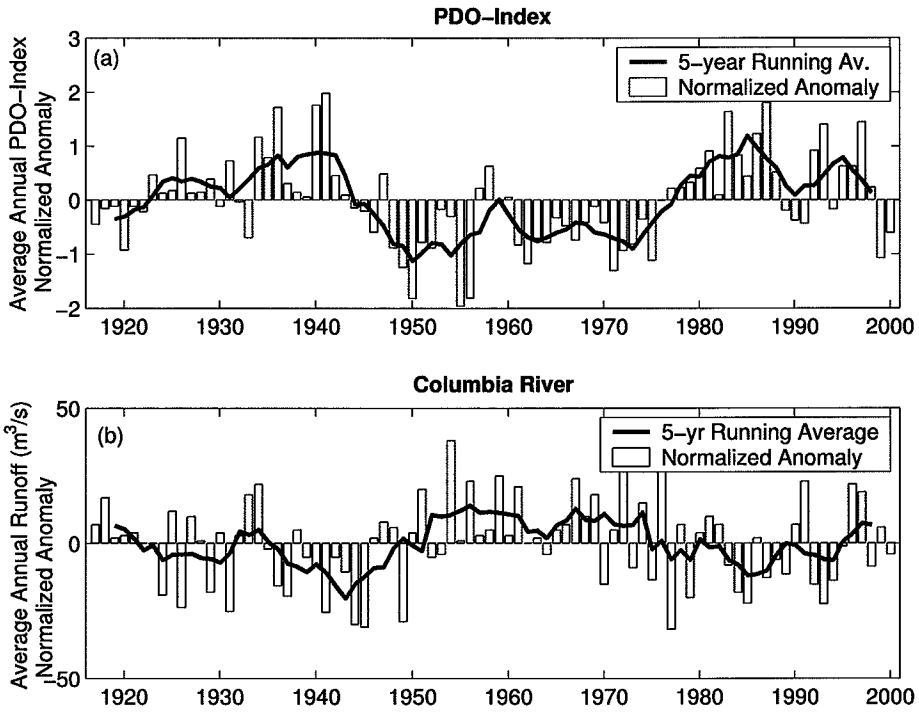


Figure 5.1: Cycles observed in PDO-Index and Columbia River

with irregular intervals of typically 4-7 years. Other phenomena tend to have a more regime-like occurrence, often sustained over periods of several years. Examples include the North Atlantic Oscillation (NAO) which has been found to have an impact on the North-Eastern part of North America and the Pacific Decadal Oscillation (PDO) which has been linked to the climate of the North Western part of North America [*Neal et al.*, 2002; *Spence*, 2002; *Hsieh et al.*, 2003]. As seen in Figure 5.1, the annual PDO-Index has obvious cycles, remaining for extended periods above or below the historical average. This is particularly obvious when considering the 5-year running average. Although not quite as obvious, a similar tendency is observed in the unregulated annual streamflow of the Columbia River (at Nicholson, British Columbia, Canada).

The existence of cycles such as those observed in Figure 5.1 cannot be explained by

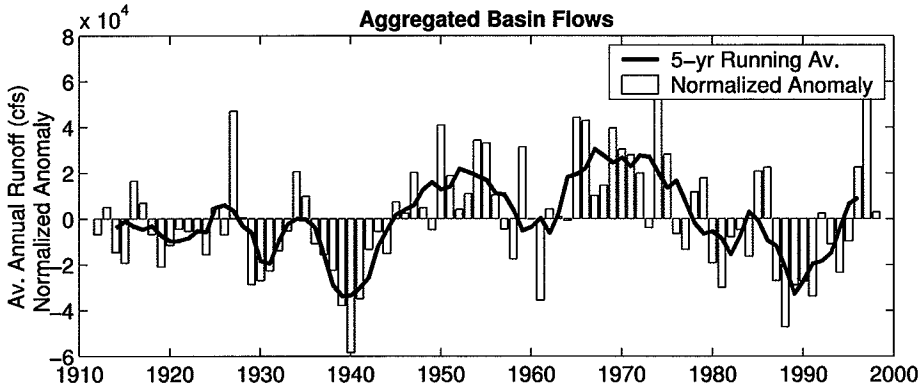


Figure 5.2: Annual Aggregated Basin Flows 1912-98. Normalized anomalies and 5-yr running average.

pure randomness. Moreover, one might suspect that traditional time series models such as ARMA-type models fail to provide a good description of the characteristic cycles. Stochastic time series models are often employed in drought and management studies. Failure to adequately reproduce cycles obviously may lead to erroneous conclusions. Therefore, as a preliminary analysis, the observed annual streamflow data for Manitoba Hydro were analyzed to identify if there is evidence of long-term wet and dry cycles. When the annual observed streamflow data of Manitoba Hydro are smoothed by a 5-yr running average, distinct wet and dry periods are revealed at all sites. Figure 5.2 illustrates these wet and dry periods for Aggregated Basin Flows, that is, the sum of the five sites of Manitoba Hydro's system. The 5-yr running average of each of the five sites of Manitoba Hydro system is given in Appendix A. The occurrence and persistence of wet and dry periods are clearly visible at all sites. They may be considered to be a result of the complex climate dynamics that influence the North American climate.

The above observation motivates a search for alternative time series models, in particular models that can mimic the regime-like behavior of Manitoba Hydro's streamflow data. In this research project, models that can be used in 'simulation mode', which effectively excludes the possibility of using climate indices such as PDO and NAO as conditioning variables, are of particular interest.

In stochastic time series models, the generated synthetic series should be statistically similar to the observed series, but should provide a wider range of scenarios. As mentioned previously, statistically similar implies that basic statistics such as the mean, the variance, and the autocorrelation function of the variable of interest are well preserved by the model. However, when cycles are prevalent, other criteria may be needed. In fact, the autocorrelation function alone may not be sufficient to adequately describe the persistence structure of the data, although in the literature *autocorrelation* and *persistence* are often used synonymously.

In this chapter, a class of models called Markov-Switching (MS) models for modeling single-site annual hydrologic time series is presented. In the literature, they are also referred to as hidden Markov models or Markov mixture models. The motivation for adopting this type of model in hydrology is the assumption that the climate regime of each year can be described by a state variable that can take only a limited number of values, typically 2-3. The state variable evolves in time according to a discrete Markov chain, described by a matrix of transition probabilities. The sequence of state variables for an observed series is not known. The assumption is that an observed hydrologic variable arises from a statistical population whose parameters depend on

the particular state in which the climate is. Hence, to simulate a synthetic series of for example annual runoff from an MS model with all parameters specified, one would first simulate the state series using the transition probabilities of the Markov chain and then generate the annual runoff for each year from a distribution whose parameter depends on the state for that year.

MS models have been applied in different fields, including speech recognition, econometrics, ion channels, image analysis, and DNA composition. In hydrology, the idea of a Markov-Switching model was applied thirty years ago by *Jackson* [1975]. A major challenge at the time was the estimation of model parameters. In recent years, significant progress in the area of MS processes has been made, including the development of efficient algorithms for estimating parameters. There have also been a number of recent applications in hydrology, including *Wilks* [1998], *Hughes et al.* [1999], *Lu and Berliner* [1999], *Thyer and Kuczera* [2000], *Thyer and Kuczera* [2003a, b], and *Kehagias* [2004].

The MS model also has close links to the Shifting Level (SL) model proposed by *Salas and Boes* [1980]. In fact, as pointed out by *Fortin et al.* [2004], the SL model can be considered a special case of the MS model.

In this chapter, several objectives of the research project are addressed. After formulation of the single-site MS model, the statistical properties of the model, including the marginal distribution, moments, and autocorrelation function, are developed and discussed. Two methods of implementing maximum likelihood estimators, one based on direct numerical maximization of the likelihood function and the other based on

the Expectation-Maximization (EM) algorithm are presented. Although the EM algorithm is somewhat elaborate, the method can be easily implemented on a computer. Finally parameter uncertainty is integrated into the MS model.

5.2 Formulation of the Single-site MS model

There are several variants of MS models. In this research project, a relatively simple version of the model which provides an adequate description of Manitoba Hydro's data is employed. More specifically, it is assumed that the process in question is stationary and switching between M climate regimes (or states). It may be useful to think of climate regimes as for example *dry*, *normal*, and *wet* spells. The particular regime is described by a state variable s_t which can take discrete values from 1 to M. The state variable is unobserved and is therefore referred to as a hidden variable. The observed variable q_t is assumed to have been drawn from a probability distribution whose parameters are conditional upon the particular state prevailing at time t. It is also assumed that given the state sequence, q_t is independent of previous observations. Although in principle there is no restriction on the type of distributions used to describe q_t, it is here assumed that q_t is normally distributed with mean and variance that are specific to the state. To generate q_t, one would first simulate the state series using the transition probabilities; then given s_t, q_t would be generated as

$$q_t = \mu_{s_t} + \sigma_{s_t}\epsilon_t \,, \quad t = 1, \ldots, T \tag{5.1}$$

where $\epsilon_t \sim N(0,1)$ is a standard normal variable, and μ_{s_t} and σ_{s_t} are the state dependent mean and standard deviation of a normal distribution.

The unobserved discrete state variable s_t may be dependent upon $s_{t-1}, s_{t-2}, \ldots, s_{t-r}$ in which case the process of s_t is an r-order Markov chain. In the above model, it is assumed that the state variable s_t follows an M-state first-order Markov chain with transition probability matrix

$$\boldsymbol{P} = [p_{ij}] = \Pr\{s_t = j | s_{t-1} = i\} , \quad i,j = 1, \ldots, M \tag{5.2}$$

where $\sum_{j=1}^{M} p_{ij} = 1$. Persistence in q_t is inherited exclusively from the characteristics of this Markov chain. Therefore, drought properties of synthetic runoff series generated from the MS model are closely linked to the transition probabilities of the Markov chain.

In summary, the MS model is described by the $(M \times M)$-matrix $\boldsymbol{P}$ of transition probabilities and the $2M$ parameters of the normal distribution of each state, $(\mu_1, \sigma_1, \mu_2, \sigma_2, \ldots, \mu_M, \sigma_M)$.

5.3 Theoretical Properties of the MS model

In this section, the statistical and stochastic properties of the MS model described above are examined. The focus will be on analytical expressions for the mean, the variance, and the coefficient of skewness of q_t, its statistical distribution over time, as well as its autocorrelation function. These are properties that are commonly used to

judge the adequacy of hydrological time series models.

5.3.1 Stationary Probabilities of the Markov Chain

Many of the following expressions involve the stationary probabilities, π_i, of the Markov chain, that is, the probability that the Markov chain will be in state i at some time into the future where the effect of the initial state has vanished. For an infinitely long sequence, π_i may alternatively be interpreted as the fraction of time the chain is in state i. Given the transition probability matrix $\boldsymbol{P}$, the stationary probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_M)'$ can be obtained by solving $\boldsymbol{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$ subject to $\sum_{i=1}^{M} \pi_i = 1$ which can also be written as $\mathbf{1}_M'\boldsymbol{\pi} = 1$, where $\mathbf{1}_M$ denotes an $(M \times 1)$ vector of 1's. Using these equations, the vector $\boldsymbol{\pi}$ must satisfy $\boldsymbol{A}\boldsymbol{\pi} = \boldsymbol{e}_{M+1}$, where $\boldsymbol{e}_{M+1}$ denotes the last column of the identity matrix, $\boldsymbol{I}_{M+1}$ and $\boldsymbol{A}$ denotes the $(M+1) \times M$ matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_M - \boldsymbol{P}' \\ \mathbf{1}_M' \end{bmatrix} \tag{5.3}$$

Then,

$$\boldsymbol{\pi} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{e}_{M+1} \tag{5.4}$$

For example, for the two state MS model ($M = 2$), the stationary probabilities are

$$\pi_1 = \frac{p_{21}}{p_{12} + p_{21}} \quad \text{and} \quad \pi_2 = \frac{p_{12}}{p_{12} + p_{21}} \tag{5.5}$$

5.3.2 Marginal Distribution of Observed Variables

Assuming that all state distributions are normal, then if the process is in state 1, the observed variable q_t will be drawn from a $N(\mu_1, \sigma_1^2)$ distribution, if the process is in state 2, it will be drawn from a $N(\mu_2, \sigma_2^2)$ distribution, and so on. The density of q_t conditional on the state variable $s_t = i$ can then be written

$$f(q_t|s_t = i, \boldsymbol{\psi}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(q_t - \mu_i)^2}{2\sigma_i^2}\right\}, \qquad i = 1, \ldots, M \tag{5.6}$$

where $\boldsymbol{\psi} = (\mu_1, \ldots, \mu_M, \sigma_1^2, \ldots, \sigma_M^2)'$ is a vector of population parameters. The joint density of the observed variable q_t and the unobserved variable s_t is the product of the conditional and marginal densities

$$p(q_t, s_t = i|\boldsymbol{\eta}) = f(q_t|s_t = i, \boldsymbol{\psi}) \Pr\{s_t = i; \boldsymbol{p}\} \tag{5.7}$$

where $\boldsymbol{p} = (p_{11}, p_{12}, \ldots, p_{MM})'$ denotes the $(M^2 \times 1)$ vector of Markov transition probabilities, the parameter vector $\boldsymbol{\eta} = (\boldsymbol{\psi}', \boldsymbol{p}')'$, and $\Pr\{s_t = i; \boldsymbol{p}\} = \pi_i$, where $\boldsymbol{p}$ is the vector of the Markov chain parameters. The marginal or unconditional density of q_t is obtained by summing over all possible states

$$f(q_t|\boldsymbol{\eta}) = \sum_{i=1}^{M} \pi_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(q_t - \mu_i)^2}{2\sigma_i^2}\right\} \tag{5.8}$$

which is a mixture of M normal densities.

There is an extensive statistical literature on mixture densities. Mixture density

functions have been found to be quite flexible compared to conventional distributions which of course is no surprise since they generally have more parameters. For example, it has been found that a mixture of 2-3 normal distributions can approximate a log-normal distribution quite well [*McLachlan and Peel*, 2000]. Thus there is hope that the MS model will provide a good reproduction of the marginal distribution of runoff, q_t, although the marginal distribution is a complex function of transition probabilities and M sets of normal density parameters. This point is further investigated in Chapter 8.

It is worth noting that the flexibility of the mixture density is likely to eliminate the need for transformation of the data. This is in contrast to more conventional time series models that rely on the normal assumption and which in many cases require prior transformation of the data.

5.3.3 Moments of the MS Model

The moments of q_t, in particular the mean value, the variance, and the coefficient of skewness, are of significant interest in hydrological time series models. Assuming that the state process starts with a random draw from the stationary distribution, order R central moments of the MS model can be obtained from the principle of conditional

expectation as follows:

$$\begin{aligned} E\left[(q_t-\mu)^R\right] &= E\left[E\left[(q_t-\mu)^R|s_t\right]\right] \\ &= E\left[E\left[\{(\mu_{s_t}-\mu)+\sigma_{s_t}\epsilon_t\}^R|s_t\right]\right] \\ &= \sum_{i=1}^{M}\pi_i\sum_{r=0}^{R}\binom{R}{r}(\mu_i-\mu)^{R-r}\sigma_i^r E[\epsilon_t^r] \end{aligned} \tag{5.9}$$

where the last expression has been obtained using the binomial formula. Under the assumption that $\epsilon_t \sim N(0,1)$, it can be shown that

$$E[\epsilon_t^r] = \begin{cases} 1 & \text{when } r=0 \\ \prod_{\ell=1}^{r/2}(2\ell-1) & \text{when } r \text{ is even} \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

The mean value, the variance, and the coefficient of skewness of q_t can be readily obtained using the general formula [*Timmermann*, 2000]:

$$\mu = E[q_t] = \sum_{i=1}^{M}\pi_i\,\mu_i \tag{5.11}$$

$$\sigma^2 = E[(q_t-\mu)^2] = \sum_{i=1}^{M}\pi_i\left[(\mu_i-\mu)^2+\sigma_i^2\right] \tag{5.12}$$

$$c_s = \frac{E[(q_t-\mu)^3]}{\sigma^3} = \sigma^{-3}\sum_{i=1}^{M}\pi_i\left[(\mu_i-\mu)^3+3\sigma_i^2(\mu_i-\mu)\right] \tag{5.13}$$

5.3.4 Autocorrelation Structure of the MS Model

Another model property of interest is the autocorrelation function of q_t. For a linear, normal model, the persistence structure of the process is entirely described by the autocorrelation function. For a nonlinear, non-normal model, the autocorrelation function will not describe all aspects of the persistence structure, but it remains a key property that should be reasonably preserved by the model.

Let $\gamma_r = E\left[(q_t - \mu)(q_{t-r} - \mu)\right]$ denote the lag-r autocovariance of q_t. To derive an analytical expression for this function, one can first substitute (5.1) into the definition of the autocovariance function:

$$\begin{aligned}
\gamma_r &= E\left[(q_t - \mu)(q_{t-r} - \mu)\right] \\
&= E\left[(\mu_{s_t} + \sigma_{s_t}\epsilon_t - \mu)(\mu_{s_{t-r}} + \sigma_{s_{t-r}}\epsilon_{t-r} - \mu)\right] \\
&= E\left[\mu_{s_t}\mu_{s_{t-r}}\right] - \mu^2\,, \quad r = 1, 2, \ldots
\end{aligned} \tag{5.14}$$

The last expression is obtained by noting that all product terms involving ϵ_t and ϵ_{t-r} vanish after taking expectation. Also, since $E[\mu_{s_t}] = E[\mu_{s_{t-r}}] = \mu$, we have $E[\mu_{s_t}\mu] = E[\mu_{s_{t-r}}\mu] = \mu^2$. It should be noted that (5.14) cannot be used to calculate the variance ($r = 0$), since for that case the term $E[\epsilon_t^2]$ must be considered. The variance must be obtained using (5.12). From the above expression, it can be seen that the autocovariance of q_t is entirely determined by the Markov chain and its associated

mean levels. The first term in the last line of (5.14) can be calculated as

$$\begin{aligned} E[\mu_{s_t}\mu_{s_{t-r}}] &= \sum_{i=1}^{M}\sum_{j=1}^{M}\mu_j\mu_i \Pr\{s_t = j, s_{t-r} = i\} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\mu_j\mu_i \Pr\{s_t = j | s_{t-r} = i\} \Pr\{s_{t-r} = i\} \end{aligned} \tag{5.15}$$

Here $\Pr\{s_{t-r}=i\}=\pi_i$ as defined earlier, and from the theory of Markov chains, it is well known that $\Pr\{s_t=j|s_{t-r}=i\}$ can be obtained as the (i,j)'th entry in the matrix $\boldsymbol{P}^r$, where $\boldsymbol{P}$ is the one-step transition matrix. If we define $\boldsymbol{\Pi}$ to be the diagonal matrix containing $\pi_1,\ldots,\pi_M$ on the diagonal and $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_M)'$, then the double summation in (5.15) can be written concisely as $E[\mu_{s_t}\mu_{s_{t-r}}] = \boldsymbol{\mu}'\boldsymbol{\Pi}\boldsymbol{P}^r\boldsymbol{\mu}$. Therefore

$$\gamma_r = \boldsymbol{\mu}'\boldsymbol{\Pi}\boldsymbol{P}^r\boldsymbol{\mu} - \mu^2\,, \quad r = 1,2,\ldots \tag{5.16}$$

In most situations, it is more useful to consider the autocorrelation function which may be obtained as

$$\rho_r = \frac{\gamma_r}{\sigma^2}, \quad r = 1,2,\ldots \tag{5.17}$$

where σ^2 is given in (5.12). Equations (5.16) and (5.17) suggest an exponentially decaying autocorrelation. *Salas and Boes* [1980] observed that the autocorrelation structure of their Shifting Level model was equivalent to that of an ARMA(1,1) model. The same applies to the MS model. More specifically, *Poskitt and Chung* [1996] showed that the autocorrelation structure of the M-state Markov chain process is equivalent to that of an ARMA$(M-1, M-1)$ process. For example, a 2-state MS model has

the same autocorrelation function as an ARMA(1,1) model. In other words, it is in principle possible to find an ARMA model that has exactly the same autocorrelation function as a given MS model. Of course, the models may differ in other aspects such as marginal distributions. However, even with identical autocorrelation functions and marginal distributions, the structure of MS models and ARMA models are fundamentally different and therefore may lead to different results in terms of more complex simulation statistics [*Akıntuğ and Rasmussen*, 2005b].

5.4 Parameter Estimation of the MS Model

The lack of efficient estimation methods for some years prevented the MS model from gaining popularity. In recent years, there has been significant interest in MS models and this has among other things led to the development of efficient algorithms for parameter estimation.

Because of the relatively high number of parameters in MS models, the method-of-moments is not likely to produce good estimates [*Andel*, 1993]. The method of maximum likelihood holds more promise. In this section, two possible implementations of the maximum likelihood method for MS models are briefly outlined.

The following notation will be needed. Let $\boldsymbol{Q}_T = (q_1, q_2, \ldots, q_T)'$ be a vector of observations and $\boldsymbol{S}_T = (s_1, s_2, \ldots, s_T)'$ be the associated (unobserved) state variables. The MS model in (5.1) is described by the parameters of the normal densities associated with each state, $\boldsymbol{\psi}$, and the transition probabilities vector, $\boldsymbol{p}$. Since the rows in $\boldsymbol{P}$ must sum to 1, not all elements of $\boldsymbol{P}$ need to be considered. For example, a 2-state Markov

chain is entirely described by two probabilities. In order to start the simulation, the algorithm described below also requires an assumption about the distribution of the first state variable, s_1. It will be assumed that the initial states are drawn from a probability distribution, whose parameter vector is $\boldsymbol{\varrho} = (\varrho_1, \varrho_2, \ldots, \varrho_M)'$, where $\varrho_i = \Pr\{s_1 = i|\boldsymbol{Q}_T\}$. Hence, the complete parameter set of the MS model is $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{p}', \boldsymbol{\varrho}')'$. These are the parameters that must be estimated from an observation sequence $\boldsymbol{Q}_T$.

Given the model parameters $\boldsymbol{\theta}$, the joint density of $(\boldsymbol{Q}_T, \boldsymbol{S}_T)$ can be expressed as

$$p(\boldsymbol{Q}_T, \boldsymbol{S}_T|\boldsymbol{\theta}) = f(\boldsymbol{Q}_T|\boldsymbol{S}_T, \boldsymbol{\theta}) \Pr\{\boldsymbol{S}_T|\boldsymbol{\theta}\} \tag{5.18}$$

where the two terms on the right-hand side are

$$f(\boldsymbol{Q}_T|\boldsymbol{S}_T, \boldsymbol{\theta}) = \prod_{t=1}^{T} f(q_t|\mu_{s_t}, \sigma_{s_t}) \,, \tag{5.19}$$

$$\Pr\{\boldsymbol{S}_T|\boldsymbol{\theta}\} = \varrho_{s_1} \prod_{t=1}^{T-1} \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} \tag{5.20}$$

Note that it specifically has been assumed that $f(q_t|\mu_{s_t}, \sigma_{s_t})$ is a normal distribution. If both $\boldsymbol{Q}_T$ and $\boldsymbol{S}_T$ had been observed, one could readily compute the complete-data likelihood function as

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{Q}_T, \boldsymbol{S}_T) &= p(\boldsymbol{Q}_T, \boldsymbol{S}_T|\boldsymbol{\theta}) \\ &= \varrho_{s_1} \prod_{t=1}^{T-1} \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} \prod_{t=1}^{T} f(q_t|\mu_{s_t}, \sigma_{s_t}) \end{aligned} \tag{5.21}$$

Since the likelihood function cannot be conditioned on the unobserved state vari-

ables, it is necessary to eliminate $\boldsymbol{S}_T$ from the joint density by marginalization. This requires summation over all possible state sequences:

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{Q}_T) &= p(\boldsymbol{Q}_T|\boldsymbol{\theta}) \\ &= \sum_{\text{All } \boldsymbol{S}_T} p(\boldsymbol{Q}_T, \boldsymbol{S}_T|\boldsymbol{\theta}) \\ &= \sum_{\text{All } \boldsymbol{S}_T} \varrho_{s_1} \prod_{t=1}^{T-1} \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} \prod_{t=1}^{T} f(q_t|\mu_{s_t}, \sigma_{s_t}) \end{aligned} \tag{5.22}$$

In the above form, the likelihood function is intractable because of the need to sum over all M^T possible state sequences. However, noting that for each t, the state variable s_t appears only in a few factors, (5.22) can be simplified to

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{Q}_T) = &\sum_{s_1=1}^{M} \varrho_{s_1} f(q_1|\mu_{s_1}, \sigma_{s_1}) \sum_{s_2=1}^{M} p_{s_1 s_2} f(q_2|\mu_{s_2}, \sigma_{s_2}) \\ &\cdots \sum_{s_T=1}^{M} p_{s_{T-1} s_T} f(q_T|\mu_{s_T}, \sigma_{s_T}) \end{aligned} \tag{5.23}$$

In matrix form, (5.23) can be written concisely as [*MacDonald and Zucchini*, 1997]

$$l(\boldsymbol{\theta}|\boldsymbol{Q}_T) = \boldsymbol{A}_1' \left(\prod_{t=2}^{T} \boldsymbol{A}_t \right) \mathbf{1}_M \tag{5.24}$$

where

$$\boldsymbol{A}_1 = \begin{bmatrix} \varrho_1 f(q_1|\mu_1, \sigma_1) \\ \vdots \\ \varrho_M f(q_1|\mu_M, \sigma_M) \end{bmatrix},$$

$$\boldsymbol{A}_t = \begin{bmatrix} p_{11} f(q_t|\mu_1, \sigma_1) & \cdots & p_{1M} f(q_t|\mu_M, \sigma_M) \\ \vdots & \ddots & \vdots \\ p_{M1} f(q_t|\mu_1, \sigma_1) & \cdots & p_{MM} f(q_t|\mu_M, \sigma_M) \end{bmatrix},$$

and $\mathbf{1}_M$ is a M-dimensional column vector of 1s. Calculation of the likelihood function via (5.24) is straightforward. A numerical procedure must be employed to find the $\boldsymbol{\theta}$ that maximizes (5.24).

An alternative way to implement the maximum likelihood method is the Expectation-Maximization (EM) algorithm. This method was originally devised by *Dempster et al.* [1977] to handle the case of missing data, but has been adapted to MS models [*Bilmes*, 1998; *McLachlan and Peel*, 2000]. The EM algorithm may have some advantages over direct numerical optimization. It is relatively easy to implement and it is found to be more robust than direct maximization of the likelihood function. In addition, the EM method produces some useful statistics as a byproduct.

The EM algorithm is an iterative approach that cycles through two steps called the expectation (E) step and the maximization (M) step until a local or global maximum of the likelihood function is found. In the EM implementation, the log-likelihood function is conditioned on both the observations $\boldsymbol{Q}_T$ and the unobserved state variables $\boldsymbol{S}_T$. The log-likelihood function is obtained by taking the logarithm of (5.21)

$$\log l(\boldsymbol{\theta}|\boldsymbol{Q}_T, \boldsymbol{S}_T) = \log \varrho_{s_1} + \sum_{t=1}^{T-1} \log \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} + \sum_{t=1}^{T} f(q_t|\mu_{s_t}, \sigma_{s_t}) \qquad (5.25)$$

$\boldsymbol{S}_T$ is unobserved, but given the observations $\boldsymbol{Q}_T$ and some estimate of $\boldsymbol{\theta}$, one

can calculate the *expectation* (over the possible M^T state sequences) of the complete likelihood function which is the first step (E-step) of the EM-algorithm. The key component in the implementation of the EM-algorithm is to define a function for the expected value of the log-likelihood function.

Given the observed flows $\boldsymbol{Q}_T$ and the MS model parameters, $\boldsymbol{\theta}$, the following probability sequences are of interest:

$$\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}, \quad i = 1, \ldots, M;\ t = 1, \ldots, T \tag{5.26}$$

$$\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}, \quad i, j = 1, \ldots, M;\ t = 1, \ldots, T-1 \tag{5.27}$$

Specifically, $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ is the probability of being in state i at time t and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ is the joint probability of being in state i at time t and state j at time $t+1$, given the observed data. These two probabilities are key elements in Markov-Switching models [*Rabiner and Juang*, 1986] and in the EM estimation procedure. Their estimation is discussed further in Section 5.5.

Once the above probabilities have been determined as discussed in Section 5.5, it

is straightforward to express the expected log-likelihood function as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= E[\log l(\boldsymbol{\theta}|\boldsymbol{Q}_T, \boldsymbol{S}_T)|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}] \\ &= \sum_{i=1}^{M} \Pr\{s_1 = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \log \varrho_i \\ &\quad + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{t=1}^{T-1} \Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \log p_{ij} \\ &\quad + \sum_{i=1}^{M}\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \log f(q_t|\mu_i, \sigma_i^2) \end{aligned} \tag{5.28}$$

where $\boldsymbol{\theta}^{(k)}$ is the current estimate of $\boldsymbol{\theta}$. The probabilities $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ are estimated using $\boldsymbol{\theta}^{(k)}$ in place of $\boldsymbol{\theta}$. The expectation is over the possible sequences of the unobserved $\boldsymbol{S}_T$, given the observation $\boldsymbol{Q}_T$ and the model parameters $\boldsymbol{\theta}^{(k)}$. Therefore, (5.28) is a function of $\boldsymbol{\theta}$ only. This completes the E-step of the EM-algorithm.

In the M-step of the EM-algorithm, the expectation of the log-likelihood function, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$, is maximized with respect to $\boldsymbol{\theta}$. This leads to a new estimate of the model parameters that improves the current estimate and becomes the next value of $\boldsymbol{\theta}$ in the iterative EM-algorithm. More specifically,

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \tag{5.29}$$

The E- and M-steps are repeated until the difference between $\boldsymbol{\theta}^{(k+1)}$ and $\boldsymbol{\theta}^{(k)}$ is smaller than some specified convergence criterion. The final value of $\boldsymbol{\theta}^{(k)}$ is a local maximum of the log-likelihood function. Generally, it will also be the global maximum. In our

applications of the model to hydrological and climatic data, cases have not yet been encountered where there were reasons to believe that a global maximum had not been found [*Akıntuğ and Rasmussen*, 2005b]. The estimation procedure appears to be quite stable. As a byproduct, the estimation also gives the probability distribution of $\boldsymbol{S}_T$.

A closer inspection of (5.28) shows that its maximization is greatly simplified by the fact that its three terms involve different parameters and therefore can be maximized individually. The first term involves only the parameter ϱ_i and can be maximized using the method of Lagrange multipliers to ensure that $\sum_{i=1}^{M} \varrho_i = 1$. This gives the intuitive result:

$$\varrho_i^{(k+1)} = \Pr\{s_1 = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}\,, \quad i = 1, \ldots M. \tag{5.30}$$

Similarly, transition probabilities can be obtained from the second term in (5.28). Using again the Lagrange method to ensure that $\sum_{j=1}^{M} p_{ij} = 1$, one finds

$$p_{ij}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}{\sum_{t=1}^{T-1} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}\,, \quad i, j = 1, \ldots M. \tag{5.31}$$

Finally, using the third term in (5.28), the parameters of the state distributions (normal distributions) can be obtained as

$$\mu_i^{(k+1)} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} q_t}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}\,, \quad i = 1, \ldots M. \tag{5.32}$$

$$(\sigma_i^2)^{(k+1)} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \left(q_t - \mu_i^{(k+1)}\right)^2}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}, \quad i = 1, \ldots M. \tag{5.33}$$

Note that in the above expressions $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ are estimated using the old parameter vector $\boldsymbol{\theta}^{(k)}$. It should be clear that the key to implementing the EM-algorithm is the calculation of $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$, described in detail in the following section.

In order to start the iterative estimation algorithm, initial parameter estimates are needed. A number of techniques can be used to select the initial estimates of the model parameters [*Rabiner*, 1989]. It was found that the estimation procedure is quite robust to the specification of starting values. A simple approach consists of using the empirical distribution of q_t to determine M equiprobable ranges. The initial parameters μ_i and σ_i are estimated from the members of each range. Similarly, transition probabilities may be obtained by counting the empirical frequency of transitions. The probability distribution of the first state may be set to $\varrho_i = 1/M$.

The EM algorithm for estimating the parameters of the MS model may be summarized as follows:

1. Select initial parameter values, $\boldsymbol{\theta}^{(1)}$, and set $k = 1$.

2. Using the current parameter values, calculate $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ according to the procedure given in the following section.

3. Obtain an improved estimate of the parameter $\boldsymbol{\theta}^{(k+1)}$ using Equations (5.30)-(5.33). Update k.

4. Compare the new parameter estimates with the previous ones and determine if the specified convergence criterion is satisfied. If not, return to point 2.

Hamilton [1990] suggested to stop when $|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}|$ is less than 10^{-8}, however, it was experienced in this project that this is not a good choice for all parameters. Different convergence criterions should be applied to distribution parameters and to transition probabilities because μ_i and σ_i are scale dependent whereas transition probabilities are limited to [0,1].

Both ML algorithms described above were employed to Manitoba Hydro's streamflow data and in the large majority of cases identical solutions were obtained. At rare occasions, one method may find a local optimum. A change of initial values typically would resolve this conflict. It was found that the EM algorithm is slightly more robust than direct optimization, albeit not necessarily faster.

5.5 Estimation of State Probability Sequences

The EM algorithm requires estimation of the probabilities $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$. Two possible methods to estimate these probabilities are outlined in the following sections.

5.5.1 Forward-Backward Algorithm

Define the forward variable $\alpha_i(t)$ to be the probability of observing the partial sequence $q_1, q_2, \ldots, q_t$, and ending up in state i at time t:

$$\alpha_i(t) = f(q_1, q_2, \ldots, q_t, s_t = i|\boldsymbol{\theta}) \tag{5.34}$$

and the backward variable $\beta_i(t)$ to be the probability of observing the rest of the sequence $q_{t+1}, q_{t+2}, \ldots, q_T$, given that the state at time t is i:

$$\beta_i(t) = f(q_{t+1}, q_{t+2}, \ldots, q_T|s_t = i, \boldsymbol{\theta}) \tag{5.35}$$

It is possible to express (5.26) and (5.27) in terms of $\alpha_i(t)$ and $\beta_i(t)$ as

$$\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\} = \frac{\alpha_i(t)\beta_i(t)}{\sum_{i=1}^{M} \alpha_i(t)\beta_i(t)} \tag{5.36}$$

$$\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\} = \frac{\alpha_i(t)p_{ij}f(q_{t+1}|\mu_i, \sigma_i)\beta_j(t+1)}{\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i(t)p_{ij}f(q_{t+1}|\mu_i, \sigma_i)\beta_j(t+1)} \tag{5.37}$$

To estimate $\alpha_i(t)$ and $\beta_i(t)$, the following algorithm can be used.

The Forward Algorithm

The forward variable, $\alpha_i(t)$, $i = 1, 2, \ldots, M$, can be calculated recursively as follows:

1. Initialization: For $t = 1$

$$\alpha_i(1) = \varrho_i f(q_1|\mu_i, \sigma_i), \quad i = 1, 2, \ldots, M \tag{5.38}$$

2. Induction: For $t = 2, 3, \ldots, T$

$$\alpha_j(t) = \left(\sum_{i=1}^{M} \alpha_i(t-1) p_{ij} \right) f(q_t|\mu_j, \sigma_j), \quad j = 1, 2, \ldots, M \tag{5.39}$$

The Backward Algorithm

The backward variable, $\beta_i(t)$, $i = 1, 2, \ldots, M$, can be calculated as follows:

1. Initialization: For $t = T$

$$\beta_j(T) = 1, \quad j = 1, 2, \ldots, M \tag{5.40}$$

2. Induction: For $T - 1, T - 2, \ldots, 1$

$$\beta_i(t) = \sum_{j=1}^{M} p_{ij} f(q_{t+1}|\mu_j, \sigma_j) \beta_j(t+1), \quad i = 1, 2, \ldots, M \tag{5.41}$$

More details about this algorithm are provided in *Rabiner and Juang* [1986], *Bilmes* [1998], and *McLachlan and Peel* [2000].

It was found that the calculation of the sequences $\alpha_i(t)$ and $\beta_i(t)$ as described above is numerically unstable. This is not surprising since they involve products of up to T terms, each one smaller than 1. Even for relatively small samples, the precision

capacity of the computer is quickly reached. Several re-scaling techniques have been proposed to overcome this problem [*Leroux and Puterman*, 1992; *Rabiner*, 1989]. One possibility, described by *Rabiner* [1989], is to re-scale $\alpha_i(t)$ and $\beta_i(t)$ by normalizing them over all states. The procedure starts by calculating $\alpha_i(1)$ as above and then scale it as

$$\hat{\alpha}_i(1) = \frac{\alpha_i(1)}{\sum_{j=1}^{M} \alpha_j(1)}, \quad i = 1, 2, \ldots, M \tag{5.42}$$

where $\hat{\alpha}_i(1)$ is the scaled value. For $t = 2, 3, \ldots, T$, $\alpha_i(t)$ is calculated recursively based on the re-scaled values

$$\begin{aligned} \alpha_i(t) &= \left(\sum_{j=1}^{M} \hat{\alpha}_j(t-1) p_{ji} \right) f(q_t | \mu_i, \sigma_i), \quad i = 1, 2, \ldots, M \\ \hat{\alpha}_i(t) &= \frac{\alpha_i(t)}{\sum_{j=1}^{M} \alpha_j(t)}, \quad i = 1, 2, \ldots, M. \end{aligned} \tag{5.43}$$

The same procedure is used to re-scale $\beta_i(t)$. For $t=T$, one can set $\beta_i(T) = 1$, $i = 1, 2, \ldots, M$, and calculate the scaled coefficient set $\hat{\beta}_i(T)$ as

$$\hat{\beta}_i(T) = \frac{\beta_i(T)}{\sum_{j=1}^{M} \beta_j(T)}, \quad i = 1, 2, \ldots, M. \tag{5.44}$$

Then for $t = T-1, T-2, \ldots, 1$, $\hat{\beta}_i(t)$ is calculated as

$$\begin{aligned} \beta_i(t) &= \sum_{j=1}^{M} p_{ij} f(q_{t+1} | \mu_j, \sigma_j) \hat{\beta}_j(t+1), \quad i = 1, 2, \ldots, M \\ \hat{\beta}_i(t) &= \frac{\beta_i(t)}{\sum_{j=1}^{M} \beta_j(t)}, \quad i = 1, 2, \ldots, M \end{aligned} \tag{5.45}$$

Because the calculation of $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ involves ratios of $\alpha_i(t)$ and $\beta_i(t)$, the re-scaling factors cancel out. Therefore, $\hat{\alpha}_i(t)$ and $\hat{\beta}_i(t)$ can be used directly in (5.36) and (5.37) in lieu of $\alpha_i(t)$ and $\beta_i(t)$.

It was found that the scaling procedure is necessary in most cases and that it eliminates any numerical instability.

5.5.2 Kim's Algorithm

Kim [1994] presented an alternative to the Forward-Backward algorithm explained above. This algorithm is valid when s_t depends on past observations only through the value of s_{t-1} and future observations only through the value of s_{t+1}, and the conditional density of q_t in (5.6) depends on $s_t, s_{t-1}, \ldots$ only through the current state s_t. Let $\boldsymbol{Q}_{t-1} = (q_1, q_2, \ldots, q_{t-1})$ represent observations up till time $t-1$ and $\boldsymbol{Q}_T = (q_1, q_2, \ldots, q_T)$ represent all observations. Given parameter estimates of the MS model, Kim's algorithm consists of two sub-algorithms. In the first sub-algorithm, called the *filtering algorithm*, the $(T \times M)$ matrix of filtered probabilities $\Pr\{s_t = j|\boldsymbol{Q}_t, \boldsymbol{\theta}\}$, for $j = 1, \ldots, M$, are calculated and in the second sub-algorithm, called the *smoothing algorithm*, the $(T \times M)$ matrix of smoothed probabilities $\Pr\{s_t = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ for $j = 1, \ldots, M$ are calculated.

Filtering Algorithm

Step 1: Given $\Pr\{s_{t-1} = i|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\}$, $i = 1, \ldots, M$, the probabilities $\Pr\{s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\}$, $j = 1, \ldots, M$, are calculated as

$$\begin{aligned}\Pr\{s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\} &= \sum_{i=1}^{M} \Pr\{s_t = j, s_{t-1} = i|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\} \\ &= \sum_{i=1}^{M} \Pr\{s_t = j|s_{t-1} = i\} \times \Pr\{s_{t-1} = i|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\} \end{aligned} \tag{5.46}$$

where $\Pr\{s_t = j|s_{t-1} = i\} = p_{ij}$ are transition probabilities.

Step 2: After observing q_t at time t, the probability term can be updated as

$$\begin{aligned}\Pr\{s_t = j|\boldsymbol{Q}_t, \boldsymbol{\theta}\} &= \Pr\{s_t = j|q_t, \boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\} \\ &= \frac{p(q_t, s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta})}{f(q_t|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta})} \\ &= \frac{f(q_t|s_t = j, \boldsymbol{Q}_{t-1}, \boldsymbol{\theta}) \times \Pr\{s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\}}{\sum_{j=1}^{M} f(q_t|s_t = j, \boldsymbol{Q}_{t-1}, \boldsymbol{\theta}) \times \Pr\{s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\}} \end{aligned} \tag{5.47}$$

where $\boldsymbol{Q}_t = \{\boldsymbol{Q}_{t-1}, q_t\}$

The above two steps may be iterated for $t = 1, 2, \ldots, T$ to obtain $\Pr\{s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\}$ and $\Pr\{s_t = j|\boldsymbol{Q}_t\}$. However, in order to initialize step 1, $\Pr\{s_0 = i|\boldsymbol{Q}_0\}$ is needed. The stationary probabilities π_i can be used instead of $\Pr\{s_0 = i|\boldsymbol{Q}_0\}$ at $t = 1$.

Smoothing Algorithm

After obtaining $\Pr\{s_t = j|\boldsymbol{Q}_{t-1}, \boldsymbol{\theta}\}$ and $\Pr\{s_t = j|\boldsymbol{Q}_t, \boldsymbol{\theta}\}$ in the filtering algorithm, inferences can be made on s_t using all the information in the sample and $\Pr\{s_t =$

$i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ for $t = 1, 2, \ldots, T$.

Step 1: Consider the following derivation of the joint probability that $s_t = j$ and $s_{t+1} = k$ based on the full data set:

$$\begin{aligned}
\Pr\{s_t = j, s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\} &= \Pr\{s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\} \times \Pr\{s_t = j|s_{t+1} = k, \boldsymbol{Q}_T, \boldsymbol{\theta}\} \\
&= \Pr\{s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\} \times \Pr\{s_t = j|s_{t+1} = k, \boldsymbol{Q}_t, \boldsymbol{\theta}\} \\
&= \frac{\Pr\{s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\} \times \Pr\{s_t = j, s_{t+1} = k, \boldsymbol{Q}_t, \boldsymbol{\theta}\}}{\Pr\{s_{t+1} = k, \boldsymbol{Q}_t, \boldsymbol{\theta}\}} \\
&= \frac{\Pr\{s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\} \times \Pr\{s_t = j|\boldsymbol{Q}_t, \boldsymbol{\theta}\} \times \Pr\{s_t = j|s_{t+1} = k\}}{\Pr\{s_{t+1} = k, \boldsymbol{Q}_t, \boldsymbol{\theta}\}}
\end{aligned} \tag{5.48}$$

Step 2: Once the joint probability $\Pr\{s_t = j, s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ has been calculated, state probabilities $\Pr\{s_t = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ can be obtained as

$$\Pr\{s_t = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\} = \sum_{k=1}^{M} \Pr\{s_t = j, s_{t+1} = k|\boldsymbol{Q}_T, \boldsymbol{\theta}\} \tag{5.49}$$

Given the last iteration of the filtering algorithm, $\Pr\{S_T = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$, the above two steps can be iterated backwards for $t = T-1, T-2, \ldots, 1$ to get the smoothed probabilities, $\Pr\{s_t = j|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$. Unlike the Forward-Backward algorithm, there is no numerical instability in Kim's method and its implementation is relatively easy.

5.6 Identification of the MS Model

In Section 5.4, it was assumed that the order of the MS model to be estimated was known. The optimal number of states is usually not known a priori and must be selected based on objective criteria. This is similar to identifying the optimal order (p, q) of an ARMA(p, q) model. In fact, because of the similarity between the autocorrelation functions of MS and ARMA models, one approach could be to estimate the number of states by examining the autocorrelation of the best-fit ARMA(p, q) model and use the information about p and q to choose the appropriate number of states in the MS model [*Zhang and Stine*, 2001].

In statistical analyses, model selection is often based on some measure related to the likelihood function. Examples include the likelihood ratio test [*Hansen*, 1992], the Akaike Information Criterion (AIC) [*Akaike*, 1974], and the Bayesian Information Criterion (BIC) [*Schwarz*, 1978]. *Fraley and Raftery* [1998] suggested that the BIC may be a good choice for the MS class of models. The BIC for the M-state MS model is defined as

$$BIC_M = -2 \log L(\hat{\boldsymbol{\theta}}_{ML} | \boldsymbol{Q}_T) + d_M \log(T) \tag{5.50}$$

where $L(\hat{\boldsymbol{\theta}}_{ML} | \boldsymbol{Q}_T)$ is the data likelihood function given in (5.24), evaluated at the maximum likelihood estimate of $\boldsymbol{\theta}$, d_M is the number of independent parameters in the M-state MS model, and T is the number of observed data. For a model with M states, $d_M = M^2 + M$. Preference is given to models with low BIC values.

5.7 Incorporating Parameter Uncertainty into the Single-Site MS Model

As in the case of the AR(1) model, parameter uncertainty for the single-site MS model will be approached from a Bayesian perspective. The posterior distribution of the AR(1) model parameters could be derived analytically and because it takes the form of a standard probability distribution, the parameters of the model can be drawn directly from that posterior distribution. In the Bayesian analysis of MS models, both the parameters of the model, $\boldsymbol{\theta}$, and the states variables s_t, $t = 1, 2, \ldots, T$, must be treated as random variables. However, in this case, a standard analytical expression of the joint posterior distribution of the parameters is not possible.

When the posterior distribution cannot be expressed in analytical form, Markov chain Monte Carlo (MCMC) methods may be used to draw samples from the posterior distribution [*Gelman et al.*, 1995; *Gemerman*, 1997; *Gilks et al.*, 1998]. A brief explanation of MCMC methods is given in the following section.

5.7.1 Markov Chain Monte Carlo Methods

It is often possible to represent the behavior of a physical system by describing all the different states the system may occupy and by indicating how the system moves from one state to another in time. A Markov chain is a sequence of random values whose value at a given time depends only upon the value at the previous time step. The controlling factor in a Markov chain is the transition probability, which is the

conditional probability for the system to go to a particular state, given the current state of the system.

Numerical methods which involve random sampling are called Monte Carlo methods. Monte Carlo methods have been used for centuries, but only in the past several decades has the technique gained the status of a full-fledged numerical method capable of addressing the most complex applications.

MCMC method together with the Bayesian framework is a powerful method in the quantification of parameter uncertainty. This stochastic simulation technique is able to reproduce the statistical joint distribution of the whole parameter set of any model. There has been an explosion in the use of MCMC in statistics and hydrology over recent years, primarily because of their application in Bayesian inference. The basic idea of MCMC methods is to simulate a Markov chain sequence where at each iteration a sample of the model parameters are generated so that any desired feature of the posterior distribution may be accurately summarized. Given certain conditions, the distribution of these samples converges to a stationary distribution which is the posterior distribution. MCMC methods make possible the use of flexible Bayesian models that would otherwise be computationally infeasible. The last decade has witnessed a burst of activity in applying Bayesian methods to parameter uncertainty in hydrology. Most of these applications have used MCMC methods to simulate posterior distributions [*Kuczera and Parent*, 1998; *Bates*, 2001; *Campbell and Bates*, 2001; *Thyer et al.*, 2002; *Bell et al.*, 2002].

There are many different MCMC techniques but the most fundamental are the

Metropolis-Hastings algorithm and the *Gibbs sampler*. The Gibbs sampling method seems the most efficient technique to apply to the MS model.

Gibbs Sampler

Where the form of the full conditional distributions of parameters are known, these may be used to obtain candidate parameters. The Gibbs sampling is appropriate when sampling from the marginal distributions is not convenient or possible. Details of the Gibbs sampler can be found in *Smith and Richman* [1993], *Chen et al.* [2000], and *Gemerman* [1997], and an excellent tutorial on this technique is provided by *Casella and George* [1992]. This technique has also been widely used in hydrology [*Sanso and Guenni*, 1999; *Thyer and Kuczera*, 2000].

In Gibbs sampling, each component of the model parameter vector is drawn from the distribution of that component conditioned on the data and the remaining parameters. This distribution is referred to as the full conditional distribution.

The Gibbs sampler is based on the following algorithm:

Step 1: Initialize the iteration counter to $k = 1$ and the parameter vector to $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})^T$

Step 2: Draw a new value $\boldsymbol{\theta}^{(k)}$ from the full conditional distributions

$$\begin{aligned}
\theta_1^{(k)} &\sim p(\theta_1|\theta_2^{(k-1)}, \theta_3^{(k-1)}, \ldots, \theta_d^{(k-1)}) \\
\theta_2^{(k)} &\sim p(\theta_2|\theta_1^{(k)}, \theta_3^{(k-1)}, \ldots, \theta_d^{(k-1)}) \\
&\vdots \quad \vdots \qquad \vdots \\
\theta_d^{(k)} &\sim p(\theta_d|\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_{d-1}^{(k)})
\end{aligned}$$

Step 3: Set $k = k + 1$, and return to step 2.

Since each simulated parameter vector depends only on the previous simulated parameter vector, and not on any other previous values or the iteration counter k, the Gibbs sampler algorithm defines a homogeneous Markov chain.

Although the Gibbs sampler appears to converge significantly faster in practice than the Metropolis algorithm (*Gelman et al.* [1995]), it requires the ability to randomly sample one parameter or blocks of parameters at a time from the univariate or multivariate conditional probability distribution, respectively. Therefore, the derivation of such conditional distributions is required in order to use the Gibbs sampler as a MCMC simulation technique.

5.7.2 Application of the Gibbs Sampler to the Single-Site MS Model

As explained above, the idea in the Gibbs sampler is to simulate, in turn, from the distribution of each parameter conditioned on the data and the remaining parameters. Thus, at iteration k each component of the parameter vector is sampled from the conditional distribution:

$$\theta_j^{(k)} \leftarrow p(\theta_j \mid \theta_1^{(k)}, \ldots, \theta_{j-1}^{(k)}, \theta_{j+1}^{(k-1)}, \ldots, \theta_d^{(k-1)}, \boldsymbol{Q}_T) \tag{5.51}$$

where d is the number of the components of the parameter vector. θ_j can refer to either a scalar or a subvector of the parameter vector $\boldsymbol{\theta}$. *Smith and Richman* [1993] noted that if the parameters are highly correlated, then the convergence of the Gibbs sampler could be very slow. Therefore, if possible highly correlated parameters should be blocked together as a subvector of $\boldsymbol{\theta}$ and sampled from a multivariate conditional distribution.

For the single-site MS model, Gibbs sampling can be implemented with the following simulation steps:

$$\boldsymbol{S}_T^{(k)} \leftarrow p(\boldsymbol{S}_T \mid \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\sigma}^{(k-1)}, \boldsymbol{P}^{(k-1)}, \boldsymbol{Q}_T) \tag{5.52}$$

$$\boldsymbol{P}^{(k)} \leftarrow p(\boldsymbol{P} \mid \boldsymbol{S}_T^{(k)}) \tag{5.53}$$

$$\mu_i^{(k)}, \sigma_i^{(k)} \leftarrow p(\mu_i, \sigma_i \mid \boldsymbol{S}_T^{(k)}, \mu_i^{(k-1)}, \sigma_i^{(k-1)}, \boldsymbol{Q}_T) \tag{5.54}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M)'$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_M)'$, and i refers to the state. The hidden state time series $\boldsymbol{S}_T$ is sampled first because once it is known, the sampling of the remaining quantities is relatively simple. The conditional distribution of the transition probabilities $\boldsymbol{P}$ is purely dependent upon knowledge of $\boldsymbol{S}_T$. Hence, in (5.53) the data $\boldsymbol{Q}_T$ is omitted in the conditioning. The description and the derivation of the above conditional densities are presented in the following sections

Simulating the State Probability Sequence

The state variable s_t, $t = 1, 2, \ldots, T$, can be simulated one by one from each of the following T conditional distributions

$$p(s_t|\boldsymbol{S}_{-t}, \boldsymbol{Q}_T, \boldsymbol{\theta}), \quad t = 1, 2, \ldots, T \tag{5.55}$$

where $\boldsymbol{S}_{-t} = (s_1, \ldots, s_{t-1}, s_{t+1}, \ldots, s_T)^T$. However, it is also possible to simulate the entire state sequence from the joint conditional distribution of $\boldsymbol{S}_T$, $p(\boldsymbol{S}_T|\boldsymbol{Q}_T, \boldsymbol{\theta})$ [*Chib*, 1996]. This density can be written by applying the conditional probability theorem repeatedly

$$\begin{aligned} p(\boldsymbol{S}_T|\boldsymbol{Q}_T, \boldsymbol{\theta}) =& p(s_1, \{s_2, \ldots, s_T\}|\boldsymbol{Q}_T, \boldsymbol{\theta}) \\ =& p(s_1|\{s_2, \ldots, s_T\}, \boldsymbol{Q}_T, \boldsymbol{\theta}) \times p(\{s_2, \ldots, s_T\}|\boldsymbol{Q}_T, \boldsymbol{\theta}) \\ =& p(s_1, \{s_2, \ldots, s_T\}|\boldsymbol{Q}_T, \boldsymbol{\theta}) \times p(s_2, |\{s_3, \ldots, s_T\}, \boldsymbol{Q}_T, \boldsymbol{\theta}) \\ & \times p(\{s_3, \ldots, s_T\}|\boldsymbol{Q}_T, \boldsymbol{\theta}) \end{aligned} \tag{5.56}$$

The summary of this recursion can be written as

$$p(\boldsymbol{S}_T|\boldsymbol{Q}_T,\boldsymbol{\theta}) = p(s_1,|\{s_2,\ldots,s_T\},\boldsymbol{Q}_T,\boldsymbol{\theta})\ldots p(s_t|\{s_{t+1},\ldots,s_T\},\boldsymbol{Q}_T,\boldsymbol{\theta})\ldots p(s_T|\boldsymbol{Q}_T,\boldsymbol{\theta}) \tag{5.57}$$

in which terms of the form $p(s_t|\{s_{t+1},\ldots,s_T\},\boldsymbol{Q}_T,\boldsymbol{\theta})$ occur repeatedly. Applying Bayes theorem, the conditional probability theorem, and the Markovian property of the states, *Chib* [1996] defined a simplified expression for the typical term

$$p(s_t = j|\{s_{t+1},\ldots,s_T\},\boldsymbol{Q}_T,\boldsymbol{\theta}) \propto p(s_{t+1} = k|s_t = j,\boldsymbol{\theta})p(s_t = j|\boldsymbol{Q}_t,\boldsymbol{\theta}) \tag{5.58}$$

where the first term on the right hand side is the transition probability of going from state j at time t to state k at time $t+1$, and the other term is the probability of $s_t = j$ given $\boldsymbol{Q}_t$ which can be obtained using the filtering algorithm given in Section 5.5.2. The normalizing constant of this function is the sum of the numbers obtained in (5.58) as s_t runs through $1,\ldots,M$, $\sum_{j=1}^{M} p(s_{t+1} = k|s_t = j,\boldsymbol{\theta})p(s_t = j|\boldsymbol{Q}_t,\boldsymbol{\theta})$.

Using the equations defined above, the simulation of the state sequence is as follows. Run the filtering algorithm to calculate $p(s_t = j|\boldsymbol{Q}_t,\boldsymbol{\theta})$. Once $p(s_T = j|\boldsymbol{Q}_T,\boldsymbol{\theta})$ is known from the last iteration of the filtering algorithm, it is straightforward to generate s_T. After simulation of s_T, the remaining states are simulated backwards for $T-1, T-2,\ldots,1$ using (5.58).

Simulating Transition Probabilities

Given the state sequence, $\boldsymbol{S}_T$, the derivation of the conditional posterior distribution of the unique elements of the transition probability matrix, $\boldsymbol{P}$, is straightforward because $\boldsymbol{P}$ is independent of the data, $\boldsymbol{Q}_T$, and the parameters of the model, $\boldsymbol{\theta}$.

Let the ith row of $\boldsymbol{P}$ be denoted $\boldsymbol{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iM})^T$. It is assumed that the form of the posterior conditional distribution of the rows of the transition probability matrix are equivalent. In determining the conditional posterior distribution $p(\boldsymbol{p}_i|\boldsymbol{S}_T)$, the likelihood function, $l(\boldsymbol{S}_T|\boldsymbol{p}_i)$, is required. The likelihood function for $\boldsymbol{p}_i$ has the kernel of a Multinomial distribution.

$$l(\boldsymbol{S}_T|\boldsymbol{p}_i) \propto \prod_{j=1}^{M} (p_{ij})^{n_{ij}} \tag{5.59}$$

where n_{ij} is the number of times that state i is followed by state j in the state sequence $\boldsymbol{S}_T$.

When a conjugate prior is used, the posterior distribution is from the same family as the prior. The Dirichlet distribution is a conjugate prior for the Multinominal likelihood function. An independent Dirichlet distribution for the prior distribution of $\boldsymbol{p}_i$ has the form

$$p(\boldsymbol{p}_i) \propto p_{i1}^{(\alpha_{i1}-1)} p_{i2}^{(\alpha_{i2}-1)} \ldots p_{iM}^{(\alpha_{iM}-1)} \tag{5.60}$$

where α_{ij} are the parameters of the prior Dirichlet distribution. When this prior is multiplied by the likelihood function in (5.59), the updated posterior distribution is

also Dirichlet:

$$p(\boldsymbol{p}_i|\boldsymbol{S}_T) \propto p_{i1}^{(\alpha_{i1}+n_{i1}-1)} p_{i2}^{(\alpha_{i2}+n_{i2}-1)} \dots p_{iM}^{(\alpha_{iM}+n_{iM}-1)} \tag{5.61}$$

In the Gibbs sampler, the simulation of $\boldsymbol{p}_i$ from a Dirichlet distribution must be repeated for each row of $\boldsymbol{P}$

$$\boldsymbol{p}_i|\boldsymbol{S}_T \sim Dir(\alpha_{i1} + n_{i1}, \dots, \alpha_{iM} + n_{iM}) \tag{5.62}$$

The prior parameters may be fixed as $\alpha_{i1} = \cdots = \alpha_{iM} = 1$, in which the prior distributions will be non-informative or diffuse. This is a common choice in Bayesian inference.

Simulating State Mean and Variance

In Bayesian inference, one must ensure that the posterior distribution is a proper probability distribution which means that the integral of its density must equal 1. The use of an improper noninformative prior does not preclude that a proper posterior can be obtained [*Gelman et al.*, 1995]. For example, assume that the prior distribution of the state variance is $p(\sigma^2) \propto 1/\sigma^2$ which is a noninformative distribution. This prior distribution is improper because over the range of $(0, \infty)$, it has an infinite integral. However, if there is at least one data point this improper prior gives a proper posterior distribution. Since there is a chance to simulate no data in a particular state, it was decided not to use an improper prior distribution of σ^2 in the Bayesian analysis of the

MS model.

Let the observed streamflow data coming from state i be denoted q_τ where $\tau = 1, 2, \ldots, n_i$. According to the assumptions, q_τ follows a normal distribution with mean μ_i and variance σ_i^2. The likelihood function for (μ_i, σ_i^2) is

$$l(\mu_i, \sigma_i^2 | \boldsymbol{Q}_i) \propto (\sigma_i^{-n_i}) \exp\left\{-\frac{1}{2\sigma_i^2}\left[(n_i - 1)s_i^2 + n_i(\bar{q}_i - \mu_i)^2\right]\right\} \tag{5.63}$$

where the sufficient statistics $\bar{q}_i = {n_i}^{-1} \sum_{\tau=1}^{n_i} q_\tau$ and $s_i^2 = (n_i - 1)^{-1} \sum_{\tau=1}^{n_i} (q_\tau - \bar{q}_i)^2$ are the sample average and the sample variance of observations in state i, respectively.

The joint prior distribution of (μ_i, σ_i^2) may be written as $p(\mu_i, \sigma_i^2) = p(\mu_i|\sigma_i^2)p(\sigma_i^2)$. The use of the normal distribution for $p(\mu_i|\sigma_i)$ and the scaled Inverse Chi-square distribution for $p(\sigma_i^2)$ result in the following conjugate prior distribution

$$p(\mu_i, \sigma_i^2) \propto \sigma_i^{-1}(\sigma_i^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[\nu_0\sigma_0^2 + \kappa_0(\mu_0 - \mu_i)^2\right]\right\} \tag{5.64}$$

where ν_0, κ_0, μ_0, and σ_0^2 are the parameters of the prior normal-inverse-χ^2-distribution.

In the Gibbs Sampler iterations, the hidden state of the observed variable q_t in year t is known from the simulated state sequence $\boldsymbol{S}_T$. When no data in a particular state is simulated, the state mean and variance are simulated from the prior distribution instead of the posterior. In this case, the state variance must be simulated first from the scaled inverse-χ^2-distribution and then the state mean must be simulated from the

normal distribution given the simulated variance.

$$\sigma_i^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2) \tag{5.65}$$

$$\mu_i|\sigma_i^2 \sim N(\mu_0, \sigma_i^2/\kappa_0) \tag{5.66}$$

The posterior distribution is obtained by multiplying the prior by the likelihood function in (5.63):

$$p(\mu_i, \sigma_i^2|\boldsymbol{Q}_i) \propto \sigma_i^{-1}(\sigma_i^2)^{-(\nu_{n_i}/2+1)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[\nu_{n_i}\sigma_{n_i}^2 + \kappa_{n_i}(\mu_{n_i} - \mu_i)^2\right]\right\} \tag{5.67}$$

where the parameters of the joint posterior distribution are

$$\kappa_{n_i} = \kappa_0 + n_i \tag{5.68}$$

$$\nu_{n_i} = \nu_0 + n_i \tag{5.69}$$

$$\mu_{n_i} = \frac{\kappa_0}{\kappa_0 + n_i}\mu_0 + \frac{n_i}{\kappa_0 + n_i}\bar{q}_i \tag{5.70}$$

$$\nu_{n_i}\sigma_{n_i}^2 = \nu_0\sigma_0^2 + (n_i - 1)s_i^2 + \frac{\kappa_0 n_i}{\kappa_0 + n_i}(\bar{q}_i - \mu_0)^2 \tag{5.71}$$

The parameters of the posterior distribution reflects the prior information and the information from the data. In the absence of any prior knowledge about the state distribution parameters, a diffuse prior will be used. When the diffuse prior distribution combines with the data likelihood function, the posterior distribution is dominated by data.

The posterior distribution may be written as

$$p(\mu_i, \sigma_i^2|\boldsymbol{Q}_i) = p(\mu_i|\sigma_i^2, \boldsymbol{Q}_i)p(\sigma_i^2|\boldsymbol{Q}_i) \tag{5.72}$$

so that sampling can be accomplished by first drawing σ_i^2 from its marginal distribution, (scaled inverse-χ^2)

$$\sigma_i^2|\boldsymbol{Q}_i \sim Inv - \chi^2(\nu_{n_i}, \sigma_{n_i}^2) \tag{5.73}$$

and then, given the simulated value of σ_i^2, simulating μ_i from its normal conditional posterior distribution,

$$\mu_i|\sigma_i^2, \boldsymbol{Q}_i \sim N(\mu_{n_i}, \sigma_i^2/\kappa_{n_i}) \tag{5.74}$$

Initializing the Parameter Vector in the Gibbs Sampler

The Gibbs sampler requires an initial set of parameters. The method given in Section 5.4 can be used for the Gibbs sampler as well. Alternatively, the parameters estimated by maximum likelihood can be used. The use of the estimated parameters as initial values will improve the convergence of the Gibbs sampler.

Assessing MCMC Convergence

When MCMC simulation is used instead of direct simulation from the posterior distribution, one must ensure that the Markov chain has converged to a stationary distribution. Hence, after selecting the initial values, the MCMC algorithm is allowed to sample continuously for a certain number of iterations until the effect of the starting

values has vanished. This initial sampling is known as *warm-up* or *burn-in.* Once converged, the samples from the MCMC algorithm can be considered as samples drawn from $p(\boldsymbol{\theta}|\boldsymbol{Q}_T)$. The required number of simulations for warm-up can be decided using various *convergence diagnostic tools.*

Cowles and Carlin [1996] reviewed 13 convergence diagnostic tools and recommended using a variety of methods and multiple independent parallel Markov chains. *Kass et al.* [1998] discussed several topics such as confidence in simulation results, methods for speeding and assessing convergence, and the estimation of standard errors of MCMC methods.

In this project, multiple independent parallel Markov chains have been used and the R statistic was employed to assess the convergence of these multiple chains [*Gelman and Rubin*, 1992]. The R statistic is a measure of the variance between-chain and within-chain. The R statistic is relatively high if multiple Markov chains are not mixing properly in the parameter space.

The method can be summarized as follows:

Step 1: Using the conditional distributions in the Gibbs sampler, independently simulate $J \geq 2$ parallel sequences with different starting points, each of length $2L$. To diminish the effect of the starting values, use only the second L iterations of each sequence.

Step 2: For each scalar parameter of interest, v, in the chain, calculate the between-

sequence variance B/L, and the within-sequence variance W, defined by

$$B/L = \frac{1}{J-1}\sum_{j=1}^{J}(\bar{v}_j - \bar{v})^2 \tag{5.75}$$

where

$$\bar{v}_j = \frac{1}{L}\sum_{l=1}^{L} v_{jl} \quad \text{and} \quad \bar{v} = \frac{1}{J}\sum_{j=1}^{J}\bar{v}_j \tag{5.76}$$

and

$$W = \frac{1}{J}\sum_{j=1}^{J} s_j^2 \tag{5.77}$$

where

$$s_j^2 = \frac{1}{L-1}\sum_{l=1}^{L}(v_{jl} - \bar{v}_j)^2 \tag{5.78}$$

Step 3: Estimate the marginal posterior variance of the estimand, v, by a weighted average of B and W,

$$\hat{V}(v) = \frac{L-1}{L}W + \frac{B}{L} \tag{5.79}$$

Step 4: Calculate the Gelman-Rubin statistic, $\sqrt{\hat{R}}$, the potential scale reduction, by

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{V}(v)}{W}} \tag{5.80}$$

As the individual sequences converge and range over the entire parameter space, $\hat{V}$ approaches W and the statistic $\sqrt{\hat{R}}$ decreases to 1.

Gelman and Rubin [1992] recommend to compute $\sqrt{\hat{R}}$ for all estimands of interest until it is near 1 for all of them. Once this is obtained, the simulated values from the

second halves of the sequences can be treated as samples from the posterior distribution of all parameters. Further details on this method can be found in *Gilks et al.* [1998] and *Gelman et al.* [1995].

Implementation Issue

There are several constraints that must be enforced in the implementation of the Gibbs sampler. In the MS model framework, the posterior distribution can have more than one mode. During the iteration of the Gibbs sampler, the chains can move from one mode to another. For example, in the 2-state MS model, state 1 parameters can become state 2 parameters and vice versa. This is called *aliasing* and it decreases the rate of convergence [*Gelman et al.*, 1995]. In order to prevent this, *Gelman et al.* [1995] recommended to use a constraint that $\mu_1 < \cdots < \mu_M$. In addition, for each state, the coefficient of variation, C_{v_i}, for $i = 1, 2, \ldots, M$, is bounded between 0.0001 and 2.0 [*McMahon and Mein*, 1986]. If the above constraints are not satisfied during simulation, the state mean and variance are resimulated.

Chapter 6

Multi-Site Markov-Switching Model

6.1 Introduction

As mentioned in Section 4.1, Manitoba Hydro's system requires multi-site modelling. In this chapter, the single-site MS model, proposed in the previous chapter, is extended to a multi-site model [*Akıntuğ and Rasmussen*, 2005a].

The chapter is organized as follows. The formulation of the multi-site MS model along with the model assumptions are presented in the following section. Maximum likelihood estimation, implemented using the EM algorithm, is then explained. In Section 6.4, the crosscorrelation structure of the MS model is derived. The incorporation of parameter uncertainty and missing data uncertainty into the multi-site MS model through Gibbs sampling is introduced in Section 6.5.

6.2 Formulation of the Multi-Site MS model

In the multi-site MS model, it is assumed that the hydrological process is stationary and switching between M unobserved climate states. As before, the state is described by a state variable $s_t = (1, 2, \ldots, M)$. The observed flows $\boldsymbol{q}_t = (q_t^{(1)}, \ldots, q_t^{(N)})'$ at N sites is assumed to have been drawn from an N-dimensional multivariate probability distribution whose parameters are conditional upon the particular state prevailing at time t. The use of multivariate distributions should make it possible to preserve the spatial correlation between sites, at least to a certain degree. It is assumed that the climate state is regional so that every site is in the same climate state at every point in time. It is also assumed that, given the state sequence $\boldsymbol{S}_T = (s_1, \ldots, s_T)'$, $\boldsymbol{q}_t$ is independent of previous observations. As in the single-site case, there is in principle no restriction on the type of state distributions, however, it is assumed here that $\boldsymbol{q}_t$ follows an N-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}_{s_t} = (\mu_{s_t}^{(1)}, \ldots, \mu_{s_t}^{(N)})'$ and covariance matrix $\boldsymbol{\Sigma}_{s_t}$ that are specific to the state. Hence, given s_t, one could generate $\boldsymbol{q}_t$ as

$$\boldsymbol{q}_t = \boldsymbol{\mu}_{s_t} + \boldsymbol{V}_{s_t}\boldsymbol{\epsilon}_t \tag{6.1}$$

where $\boldsymbol{\epsilon}_t$ is a vector of independent, standard normal variables and $\boldsymbol{V}_{s_t}$ is the decomposition of the state covariance matrix $\boldsymbol{\Sigma}_{s_t}$, satisfying $\boldsymbol{V}_{s_t}\boldsymbol{V}'_{s_t} = \boldsymbol{\Sigma}_{s_t}$. Finding $\boldsymbol{V}_{s_t}$ is a standard problem in multivariate time series analysis and several methods are available, see references in Section 4.3.4. In the above model, it is assumed that the state

variable s_t follows an M-state first order Markov chain with transition probability matrix $\boldsymbol{P}$, see (5.2).

6.3 Parameter Estimation of the Multi-Site MS Model

Parameter estimation for the multi-site MS model can be accomplished in a way very similar to the single-site model. In the multi-site MS model, it is assumed that the state distributions are N-dimensional multivariate normal. Thus, if s_t is known, the density of $\boldsymbol{q}_t$ conditional on the state variable s_t is

$$f(\boldsymbol{q}_t|s_t = i, \boldsymbol{\psi}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{q}_t - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{q}_t - \boldsymbol{\mu}_i)\right\}, \quad i = 1, \ldots, M \tag{6.2}$$

where $\boldsymbol{\psi} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M\}$ is the set of population parameters. However, s_t is unobserved (hidden). In order to proceed, one can consider the joint density of the observed variable $\boldsymbol{q}_t$ and the unobserved variable s_t which is given by

$$p(\boldsymbol{q}_t, s_t = i|\boldsymbol{\eta}) = f(\boldsymbol{q}_t|s_t = i, \boldsymbol{\psi}) \Pr\{s_t = i; \boldsymbol{P}\} \tag{6.3}$$

where $\boldsymbol{P}$ denotes the $(M \times M)$ matrix of Markov transition probabilities, the parameter set $\boldsymbol{\eta} = \{\boldsymbol{\psi}, \boldsymbol{P}\}$, and $\Pr\{s_t = i; \boldsymbol{P}\} = \pi_i$. By summing (6.3) over all possible states, the unconditional density of $\boldsymbol{q}_t$ can be obtained as

$$f(\boldsymbol{q}_t|\boldsymbol{\eta}) = \sum_{i=1}^{M} \pi_i \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{q}_t - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{q}_t - \boldsymbol{\mu}_i)\right\} \tag{6.4}$$

which is an M component N-dimensional multivariate normal density with mixing coefficients π_i.

Let $\boldsymbol{Q}_T = \{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_T\}$ be the set of all multi-site observations and $\boldsymbol{S}_T = (s_1, s_2, \ldots, s_T)'$ be the associated state variables for the basin. The multi-site MS model in (6.1) is described by the parameter set $\boldsymbol{\theta} = \{\boldsymbol{\psi}, \boldsymbol{P}, \boldsymbol{\varrho}\}$ where $\boldsymbol{\varrho} = (\varrho_1, \varrho_2, \ldots, \varrho_M)'$ is the probability distribution of the first state variable in the sequence. This distribution is needed in the formulation of the maximum likelihood estimation.

Given the model parameters $\boldsymbol{\theta}$, the joint density of $(\boldsymbol{Q}_T, \boldsymbol{S}_T)$ can be expressed as

$$\begin{aligned} p(\boldsymbol{Q}_T, \boldsymbol{S}_T|\boldsymbol{\theta}) &= \Pr\{\boldsymbol{S}_T|\boldsymbol{\theta}\} f(\boldsymbol{Q}_T|\boldsymbol{S}_T, \boldsymbol{\theta}) \\ &= \varrho_{s_1} \prod_{t=1}^{T-1} \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} \prod_{t=1}^{T} f(\boldsymbol{q}_t|\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t}) \end{aligned} \tag{6.5}$$

where it is assumed that $f(\boldsymbol{q}_t|\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t})$ is an N-dimensional multivariate normal distribution. This equation can also be considered as the complete-data likelihood function. In order to obtain the data likelihood function, $\boldsymbol{S}_T$ can be eliminated by summing over all possible state sequences:

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{Q}_T) &= p(\boldsymbol{Q}_T|\boldsymbol{\theta}) \\ &= \sum_{\text{All } \boldsymbol{S}_T} p(\boldsymbol{Q}_T, \boldsymbol{S}_T|\boldsymbol{\theta}) \\ &= \sum_{\text{All } \boldsymbol{S}_T} \varrho_{s_1} \prod_{t=1}^{T-1} \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} \prod_{t=1}^{T} f(\boldsymbol{q}_t|\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t}) \end{aligned} \tag{6.6}$$

which can be simplified as

$$l(\boldsymbol{\theta}|\boldsymbol{Q}_T) = \sum_{s_1=1}^{M} \varrho_{s_1} f(\boldsymbol{q}_1|\boldsymbol{\mu}_{s_1}, \boldsymbol{\Sigma}_{s_1}) \sum_{s_2=1}^{M} p_{s_1 s_2} f(\boldsymbol{q}_2|\boldsymbol{\mu}_{s_2}, \boldsymbol{\Sigma}_{s_2}) \cdots \sum_{s_T=1}^{M} p_{s_{T-1} s_T} f(\boldsymbol{q}_T|\boldsymbol{\mu}_{s_T}, \boldsymbol{\Sigma}_{s_T}) \tag{6.7}$$

In matrix form, (6.7) can be written as [*MacDonald and Zucchini*, 1997]

$$l(\boldsymbol{\theta}|\boldsymbol{Q}_T) = \boldsymbol{A}_1' \left(\prod_{t=2}^{T} \boldsymbol{A}_t \right) \mathbf{1}_M \tag{6.8}$$

where

$$\boldsymbol{A}_1 = \begin{bmatrix} \varrho_1 f(\boldsymbol{q}_1|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\ \vdots \\ \varrho_M f(\boldsymbol{q}_1|\boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M) \end{bmatrix},$$

$$\boldsymbol{A}_t = \begin{bmatrix} p_{11} f(\boldsymbol{q}_t|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \cdots & p_{1M} f(\boldsymbol{q}_t|\boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M) \\ \vdots & \ddots & \vdots \\ p_{M1} f(\boldsymbol{q}_t|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) & \cdots & p_{MM} f(\boldsymbol{q}_t|\boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M) \end{bmatrix},$$

and $\mathbf{1}_M$ is an M-dimensional column vector of 1s. The parameters of the multi-site MS model, $\boldsymbol{\theta}$, can be obtained by maximizing (6.8) using a numerical procedure.

As in the single-site MS model, the EM algorithm can alternatively be used in the estimation of the multi-site MS model. In the EM implementation, the log-likelihood

function is

$$\log l(\boldsymbol{\theta}|\boldsymbol{Q}_T, \boldsymbol{S}_T) = \log \varrho_{s_1} + \sum_{t=1}^{T-1} \log \Pr\{(s_{t+1}|s_t)|\boldsymbol{P}\} + \sum_{t=1}^{T} f(\boldsymbol{q}_t|\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t}) \tag{6.9}$$

In the first step of the EM algorithm, the Q function, which is the expected value of the log-likelihood function, is obtained as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= E[\log l(\boldsymbol{\theta}|\boldsymbol{Q}_T, \boldsymbol{S}_T)|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}] \\ &= \sum_{i=1}^{M} \Pr\{s_1 = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \log \varrho_i \\ &\quad + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{t=1}^{T-1} \Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \log p_{ij} \\ &\quad + \sum_{i=1}^{M}\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\} \log f(\boldsymbol{q}_t|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \end{aligned} \tag{6.10}$$

where $\boldsymbol{\theta}^{(k)}$ is the current estimate of $\boldsymbol{\theta}$, and $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ are estimated using $\boldsymbol{\theta}^{(k)}$ in place of $\boldsymbol{\theta}$.

In the second step, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ is maximized with respect to $\boldsymbol{\theta}$ to obtain a new estimate of the model parameters that improves the current estimate and becomes the next value of $\boldsymbol{\theta}$ in the iterative EM-algorithm:

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \tag{6.11}$$

The EM iteration steps are repeated until a local maximum of the log-likelihood function is reached.

The maximization of (6.10) yields:

$$\varrho_i^{(k+1)} = \Pr\{s_1 = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}\,, \quad i = 1, \ldots M. \tag{6.12}$$

$$p_{ij}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}{\sum_{t=1}^{T-1} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}\,, \quad i, j = 1, \ldots M. \tag{6.13}$$

$$\boldsymbol{\mu}_i^{(k+1)} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}\boldsymbol{q}_t}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}\,, \quad i = 1, \ldots M. \tag{6.14}$$

$$\boldsymbol{\Sigma}_i^{(k+1)} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}\left(\boldsymbol{q}_t - \boldsymbol{\mu}_i^{(k+1)}\right)\left(\boldsymbol{q}_t - \boldsymbol{\mu}_i^{(k+1)}\right)'}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}}\,, \quad i = 1, \ldots M. \tag{6.15}$$

The calculation of $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ and $\Pr\{s_t = i, s_{t+1} = j|\boldsymbol{Q}_T, \boldsymbol{\theta}^{(k)}\}$ are described in detail in Section 5.5. Note that in the multi-site case, the probability density of the univariate normal distribution used in the forward-backward algorithm or in Kim's algorithm must be replaced by the multivariate normal density.

In order to start the iterative estimation algorithm, initial parameter estimates are needed. For the initial parameter estimates, observed streamflow values of each site are standardized using their sample mean and standard deviation as

$$z_t^{(n)} = \frac{q_t^{(n)} - \bar{q}_t^{(n)}}{s_q^{(n)}} \tag{6.16}$$

Then a single sequence is obtained by adding the standardized values of N sites at each time step, $z_t = z_t^{(1)} + z_t^{(2)} + \cdots + z_t^{(N)}$. This aggregated sequence is used to determine M equiprobable ranges. According to these ranges the initial state sequence of the region is obtained. For example, for a 2-state multi-site MS model, if z_t is negative, the regional climate is in state 1, otherwise it is in state 2. Using the regional state sequence, data sets of each state at each site are obtained from observed streamflow time series of the region. The initial parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are then estimated using the data sets of each state. Similarly, transition probabilities may be obtained by counting the empirical frequency of transitions. The probability distribution of the first state may be set to $\varrho_i = 1/M$.

6.4 Crosscorrelation Structure of the MS Model

In the multi-site MS model, a key model property of interest is the crosscorrelation function of $\boldsymbol{q}_t$ which should be reasonably preserved by the model.

For site g and h, let $\gamma_0^{(g,h)} = E\left[(q_t^{(g)} - \mu^{(g)})(q_t^{(h)} - \mu^{(h)})\right]$ denote the lag-0 crosscovariance between $q_t^{(g)}$ and $q_t^{(h)}$. An analytical expression of the lag-0 crosscovariance

can be derived by substituting (6.1) into the definition of the crosscovariance function:

$$\begin{aligned}\gamma_0^{(g,h)} =& E\left[(q_t^{(g)} - \mu^{(g)})(q_t^{(h)} - \mu^{(h)})\right] \\ =& E\left[\left(\mu_{s_t}^{(g)} + v_{s_t}^{(g,g)}\epsilon_t^{(g)} + v_{s_t}^{(g,h)}\epsilon_t^{(h)} - \mu^{(g)}\right)\left(\mu_{s_t}^{(h)} + v_{s_t}^{(h,g)}\epsilon_t^{(g)} + v_{s_t}^{(h,h)}\epsilon_t^{(h)} - \mu^{(h)}\right)\right] \\ =& E\left[\mu_{s_t}^{(g)}\mu_{s_t}^{(h)}\right] + E\left[v_{s_t}^{(g,g)}v_{s_t}^{(h,g)}\left(\epsilon_t^{(g)}\right)^2\right] + E\left[v_{s_t}^{(g,h)}v_{s_t}^{(h,h)}\left(\epsilon_t^{(h)}\right)^2\right] - \mu^{(g)}\mu^{(h)}\end{aligned} \tag{6.17}$$

where $\mu^{(g)}$ and $\mu^{(h)}$ are the overall mean of q_t at site g and h, respectively, and $v^{(g,h)}$ is the (g,h)'th entry in $\boldsymbol{V}_{s_t}$. $\mu^{(g)}$ and $\mu^{(h)}$ can be obtained from (5.11) using the multi-site estimates of state means. The last expression is obtained by noting that all product terms involving only one $\epsilon_t^{(g)}$ or one $\epsilon_t^{(h)}$ or both vanish after taking expectation. Also, $E\left[\mu_{s_t}^{(g)}\mu^{(h)}\right] = E\left[\mu^{(g)}\mu_{s_t}^{(h)}\right] = \mu^{(g)}\mu^{(h)}$. The terms in the last line of (6.17) can be calculated as

$$E\left[\mu_{s_t}^{(g)}\mu_{s_t}^{(h)}\right] = \sum_{i=1}^{M}\pi_i\mu_i^{(g)}\mu_i^{(h)} \tag{6.18a}$$

$$E\left[v_{s_t}^{(g,g)}v_{s_t}^{(h,g)}\left(\epsilon_t^{(g)}\right)^2\right] = \sum_{i=1}^{M}\pi_i v_i^{(g,g)}v_i^{(h,g)} \tag{6.18b}$$

$$E\left[v_{s_t}^{(g,h)}v_{s_t}^{(h,h)}\left(\epsilon_t^{(h)}\right)^2\right] = \sum_{i=1}^{M}\pi_i v_i^{(g,h)}v_i^{(h,h)} \tag{6.18c}$$

and the lag-0 crosscovariance between site g and h can be summarized as

$$\gamma_0^{(g,h)} = \sum_{i=1}^{M}\pi_i\mu_i^{(g)}\mu_i^{(h)} + \sum_{i=1}^{M}\pi_i v_i^{(g,g)}v_i^{(h,g)} + \sum_{i=1}^{M}\pi_i v_i^{(g,h)}v_i^{(h,h)} - \mu^{(g)}\mu^{(h)} \tag{6.19}$$

The analytical expression of the lag-r crosscovariance function can be expressed as

$$\begin{aligned}\gamma_r^{(g,h)} &= E\left[(q_t^{(g)} - \mu^{(g)})(q_{t-r}^{(h)} - \mu^{(h)})\right] \\ &= E\left[\left(\mu_{s_t}^{(g)} + v_{s_t}^{(g,g)}\epsilon_t^{(g)} + v_{s_t}^{(g,h)}\epsilon_t^{(h)} - \mu^{(g)}\right)\left(\mu_{s_{t-r}}^{(h)} + v_{s_{t-r}}^{(h,g)}\epsilon_t^{(g)} + v_{s_{t-r}}^{(h,h)}\epsilon_t^{(h)} - \mu^{(h)}\right)\right] \\ &= E\left[\mu_{s_t}^{(g)}\mu_{s_{t-r}}^{(h)}\right] - \mu^{(g)}\mu^{(h)}, \quad r = 1, 2, \ldots \end{aligned} \tag{6.20}$$

As in the derivation of the lag-r autocovariance function in Section 5.3.4, the cross-covariance function above can be written concisely as

$$\gamma_r^{(g,h)} = \left(\boldsymbol{\mu}^{(g)}\right)' \boldsymbol{\Pi} \boldsymbol{P}^r (\boldsymbol{\mu}^{(h)}) - \mu^{(g)}\mu^{(h)}, \quad r = 1, 2, \ldots \tag{6.21}$$

where $\boldsymbol{\mu}^{(g)} = \left(\mu_1^{(g)}, \mu_2^{(g)}, \ldots, \mu_M^{(g)}\right)'$, $\boldsymbol{\mu}^{(h)} = \left(\mu_1^{(h)}, \mu_2^{(h)}, \ldots, \mu_M^{(h)}\right)'$, $\boldsymbol{\Pi}$ is a $(M \times M)$ diagonal matrix with the stationary probabilities π_i on the diagonal, and $\boldsymbol{P}$ is the transition probability matrix.

The cross-covariance matrix of several sites can be constructed by calculating the covariance between all pairs of sites. In order to obtain the lag-r crosscorrelation matrix, the elements of the covariance matrix can be divided by the standard deviations of the corresponding sites calculated from (5.12):

$$\rho_r^{(g,h)} = \frac{\gamma_r^{(g,h)}}{\sigma^{(g)}\sigma^{(h)}}, \quad r = 1, 2, \ldots \tag{6.22}$$

6.5 Incorporating Parameter and Data Uncertainty into Multi-Site MS Model

The methodology that was used to incorporate parameter uncertainty into the single-site MS model is extended to the multi-site MS model in this section. In addition, since the framework of the multi-site MS model enables one to handle the missing data problem, the uncertainty associated with missing data is here incorporated into the multi-site MS model as well.

6.5.1 Application of the Gibbs Sampler to the Multi-Site MS Model

In the multi-site analysis, the length of the observed streamflow record at the different sites must be the same. In traditional methods, one either truncates the different time series to the period of concomitant record which implies a loss of valuable information, or extends the shorter records using for example linear regression. By estimating the missing data, one ensures that all the available streamflow information in this basin is used in the analysis. Several techniques have been developed to estimate missing data in hydrology. For detailed explanations of some of these techniques, refer to *Hirsch* [1982].

In the Gibbs sampler, the missing data are treated as parameters and simulated from their conditional distribution along with the usual model parameters [*Gelman et al.*, 1995]. The sampled missing data are combined with the observed data in order

to form the complete data matrix $\boldsymbol{Q}_T$ which is a matrix of T vectors of the data at N sites expressed as

$$\boldsymbol{Q}_T = \{\boldsymbol{Q}_{mis}, \boldsymbol{Q}_{obs}\} \tag{6.23}$$

where $\boldsymbol{Q}_{mis}$ and $\boldsymbol{Q}_{obs}$ are the simulated missing data and observed data, respectively.

Using arbitrary starting parameter values $\boldsymbol{\theta}^{(0)}$, the following Gibbs sampler steps can be repeated to simulate the parameters from their full conditional posteriors.

$$\boldsymbol{S}_T^{(k)} \sim p\left(\boldsymbol{S}_T \mid \boldsymbol{\mu}_{\text{All}}^{(k-1)}, \boldsymbol{\Sigma}_{\text{All}}^{(k-1)}, \boldsymbol{P}^{(k-1)}, \boldsymbol{Q}_T^{(k-1)}\right) \tag{6.24}$$

$$\boldsymbol{P}^{(k)} \sim p\left(\boldsymbol{P} \mid \boldsymbol{S}_T^{(k)}\right) \tag{6.25}$$

$$\boldsymbol{Q}_{mis}^{(k)} \sim p\left(\boldsymbol{Q}_{mis} \mid \boldsymbol{S}_T^{(k)}, \boldsymbol{\mu}_{\text{All}}^{(k-1)}, \boldsymbol{\Sigma}_{\text{All}}^{(k-1)}, \boldsymbol{Q}_{obs}\right) \tag{6.26}$$

$$\boldsymbol{Q}_T^{(k)} = \left\{\boldsymbol{Q}_{mis}^{(k)}, \boldsymbol{Q}_{obs}\right\} \tag{6.27}$$

$$\boldsymbol{\mu}_i^{(k)}, \boldsymbol{\Sigma}_i^{(k)} \sim p\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i \mid \boldsymbol{S}_T^{(k)}, \boldsymbol{Q}_T^{(k)}\right) \tag{6.28}$$

where i refers to the state, $i = 1, 2, \ldots, M$, $\boldsymbol{\mu}_{\text{All}}^{(k-1)} = \left\{\boldsymbol{\mu}_1^{(k-1)}, \ldots, \boldsymbol{\mu}_M^{(k-1)}\right\}$ and $\boldsymbol{\Sigma}_{\text{All}}^{(k-1)} = \left\{\boldsymbol{\Sigma}_1^{(k-1)}, \ldots, \boldsymbol{\Sigma}_M^{(k-1)}\right\}$. The state time series $\boldsymbol{S}_T$ is simulated first because the simulation of the transition probabilities $\boldsymbol{P}$ and $\boldsymbol{Q}_{mis}$ are dependent upon knowledge of $\boldsymbol{S}_T$. As in the case of the univariate model (Section 5.7.2), Chip's procedure is employed to simulate $\boldsymbol{S}_T$ and $\boldsymbol{P}$. Note that in Chip's method, the probability density of the univariate normal distribution must be replaced by the multivariate normal probability density. The simulation of the missing data values, the state mean vector and covariance matrix is explained in the following.

Simulating the Missing Data

Since the distribution associated with each climate state is assumed to be multivariate normal, the missing data values are simulated from a multivariate normal distribution. Note that the state sequence is simulated first in each iteration of the Gibbs sampler so that the simulation of missing values can be conditioned on s_t.

In the simulation of missing data at time t, there must be at least one site with observed data. In other words, the number of missing data values m_d at time t must be less than number of the sites N. If this condition is satisfied, the vector of missing data values $\boldsymbol{q}_t^{mis}$ at time step t can be simulated from an m_d-dimensional normal distribution, conditional upon $\boldsymbol{q}_t^{obs}$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$, i.e.:

$$\boldsymbol{q}_t^{mis} | \boldsymbol{q}_t^{obs}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i \sim N_{m_d}(\boldsymbol{\mu}_{m_d}, \boldsymbol{\Sigma}_{m_d}) \qquad \text{if} \quad s_t = i \tag{6.29}$$

From standard normal theory $\boldsymbol{\mu}_{m_d}$ and $\boldsymbol{\Sigma}_{m_d}$ can be expressed in terms of $\boldsymbol{q}_t^{obs}$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ as [*Tong*, 1990]

$$\boldsymbol{\mu}_{m_d} = \boldsymbol{\mu}_i^{mis} + \boldsymbol{\Sigma}_i^{mis|obs} \left(\boldsymbol{\Sigma}_i^{obs}\right)^{-1} \left(\boldsymbol{q}_t^{obs} - \boldsymbol{\mu}_i^{obs}\right) \tag{6.30}$$

$$\boldsymbol{\Sigma}_{m_d} = \boldsymbol{\Sigma}_i^{mis} - \boldsymbol{\Sigma}_i^{mis|obs} \left(\boldsymbol{\Sigma}_i^{obs}\right)^{-1} \boldsymbol{\Sigma}_i^{obs|mis} \tag{6.31}$$

where $\boldsymbol{\mu}_i^{obs}$, $\boldsymbol{\mu}_i^{mis}$, $\boldsymbol{\Sigma}_i^{mis}$, $\boldsymbol{\Sigma}_i^{obs}$, $\boldsymbol{\Sigma}_i^{mis|obs}$, and $\boldsymbol{\Sigma}_i^{obs|mis}$ can be obtained as partitionings of

$\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$:

$$\boldsymbol{\mu}_i = \left[\begin{array}{c} \boldsymbol{\mu}_i^{obs} \\ \hline \boldsymbol{\mu}_i^{mis} \end{array}\right] \quad \boldsymbol{\Sigma}_i = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_i^{obs} & \boldsymbol{\Sigma}_i^{obs|mis} \\ \hline \boldsymbol{\Sigma}_i^{mis|obs} & \boldsymbol{\Sigma}_i^{mis} \end{array}\right] \tag{6.32}$$

Simulating the State Mean Vector and the Covariance Matrix

For a given simulated state sequence, consider the data (observed and simulated) associated with state i. Define $\boldsymbol{q}_\tau = \left(q_\tau^{(1)}, \ldots, q_\tau^{(N)}\right)$, $\tau = 1, \ldots, n_i$, to be the data coming from state i and let $\boldsymbol{Q}_i = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{n_i})'$. The joint likelihood function of $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ given $\boldsymbol{Q}_i$ can be written as

$$l(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i | \boldsymbol{Q}_i) \propto |\boldsymbol{\Sigma}_i|^{-n_i/2} \exp\left(-\frac{1}{2}\sum_{\tau=1}^{n_i}(\boldsymbol{q}_\tau - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{q}_\tau - \boldsymbol{\mu}_i)\right) \tag{6.33}$$

It can be shown [*DeGroot*, 1970] that

$$\sum_{\tau=1}^{n_i}(\boldsymbol{q}_\tau - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{q}_\tau - \boldsymbol{\mu}_i) = n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i) + \mathrm{tr}(\boldsymbol{S}_i\boldsymbol{\Sigma}_i^{-1}) \tag{6.34}$$

where $\bar{\boldsymbol{Q}}_i = \frac{1}{n_i}\sum_{\tau=1}^{n_i}\boldsymbol{q}_\tau$ is the vector of average observed data coming from state i. The notation $\mathrm{tr}(\boldsymbol{M})$ refers to the trace of the matrix $\boldsymbol{M}$, i.e. the sum of the diagonal elements of the matrix, and $\boldsymbol{S}_i$ is the sum of squares matrix relative to $\bar{\boldsymbol{Q}}_i$,

$$\boldsymbol{S}_i = \sum_{\tau=1}^{n_i}(\boldsymbol{q}_t - \bar{\boldsymbol{Q}}_i)(\boldsymbol{q}_t - \bar{\boldsymbol{Q}}_i)' \tag{6.35}$$

Substituting (6.34) into (6.33) yields

$$l(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i | \boldsymbol{Q}_i) \propto |\boldsymbol{\Sigma}_i|^{-n_i/2} \exp\left\{-\frac{1}{2}\left[n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i) + \mathrm{tr}(\boldsymbol{S}_i\boldsymbol{\Sigma}_i^{-1})\right]\right\} \quad (6.36)$$

which is the kernel of a multivariate normal-Wishart distribution.

The use of a non-informative (improper) prior distribution may cause some problems in the Gibbs sampler. As mentioned in Chapter 5, when no data is simulated in a particular state, the posterior becomes improper. However, in the multi-site case, an improper posterior is obtained when the number of data sampled in a particular state is less than the number of sites. Therefore, a proper prior distribution is used. Since there is no prior information about the parameters, a diffuse prior is used. In this way the data dominate the posterior distribution.

To develop a joint prior distribution, the relationship $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = p(\boldsymbol{\mu}_i|\boldsymbol{\Sigma}_i)\, p(\boldsymbol{\Sigma}_i)$ can be utilized. The use of the multivariate normal density for $p(\boldsymbol{\mu}_i|\boldsymbol{\Sigma}_i)$ and the inverse-Wishart density, which is the multivariate generalization of the scaled inverse-χ^2-distribution, for $p(\boldsymbol{\Sigma}_i)$ yields the conjugate joint prior density:

$$p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \propto |\boldsymbol{\Sigma}_i|^{-(\nu_0+N+2)/2} \exp\left\{-\frac{1}{2}\left[\kappa_0(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0) + \mathrm{tr}(\boldsymbol{\Lambda}_0\boldsymbol{\Sigma}_i^{-1})\right]\right\} \quad (6.37)$$

where ν_0, κ_0, $\boldsymbol{\mu}_0$, and $\boldsymbol{\Lambda}_0$ are the parameters of the joint prior distribution which has the same form as the likelihood function in (6.36). For the simulation of the prior parameters from (6.37), the prior state covariance matrix $\boldsymbol{\Sigma}_i$ must be simulated first from the inverse-Wishart distribution and then, given the simulated state covariance, the prior

state mean vector $\boldsymbol{\mu}_i$ must be simulated from the multivariate normal distribution:

$$\boldsymbol{\Sigma}_i \sim \text{Inv-Whishart}_N(\nu_0, \boldsymbol{\Lambda}_0^{-1}) \tag{6.38}$$

$$\boldsymbol{\mu}_i|\boldsymbol{\Sigma}_i \sim N_N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_i/\kappa_0) \tag{6.39}$$

where the parameters ν_0 and $\boldsymbol{\Lambda}_0$ are the degrees of freedom and the scale matrix of the prior inverse-Wishart, and $\boldsymbol{\mu}_0$ and κ_0 are the prior mean vector and the scale of $\boldsymbol{\Sigma}$, respectively.

Using the prior distribution in (6.37) and the likelihood function in (6.36), Bayes' theorem yields the following posterior distribution

$$\begin{aligned} p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i|\boldsymbol{Q}_i) \propto |\boldsymbol{\Sigma}_i|^{-(n_i+\nu_0+N+2)/2} \exp\{ -\frac{1}{2}[&n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i) \\ &+ \kappa_0(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0) \\ &+ \text{tr}((\boldsymbol{\Lambda}_0 + \boldsymbol{S}_i)\boldsymbol{\Sigma}_i^{-1})]\} \end{aligned} \tag{6.40}$$

According to *DeGroot* [1970]

$$\begin{aligned} &n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{Q}}_i) + \kappa_0(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0) \\ &\quad = (\kappa_0 + n_i)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_n)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_n) + \frac{\kappa_0 n_i}{\kappa_0 + n_i}(\boldsymbol{\mu}_0 - \bar{\boldsymbol{Q}}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_0 - \bar{\boldsymbol{Q}}_i) \end{aligned} \tag{6.41}$$

where $\boldsymbol{\mu}_n$ is given in (6.45). The rightmost term of (6.41) can also be expressed as

$$\frac{\kappa_0 n_i}{\kappa_0 + n_i}(\boldsymbol{\mu}_0 - \bar{\boldsymbol{Q}}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_0 - \bar{\boldsymbol{Q}}_i) = \text{tr}\left[\frac{\kappa_0 n_i}{\kappa_0 + n_i}(\boldsymbol{\mu}_0 - \bar{\boldsymbol{Q}}_i)'(\boldsymbol{\mu}_0 - \bar{\boldsymbol{Q}}_i)\boldsymbol{\Sigma}_i^{-1}\right] \tag{6.42}$$

Using the expression in (6.42), the joint posterior distribution in (6.40) can be rewritten as

$$p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i|\boldsymbol{Q}_i) \propto |\boldsymbol{\Sigma}_i|^{-(n_i+\nu_0+N+2)/2} \exp\{-\frac{1}{2}[(\kappa_0+n_i)(\boldsymbol{\mu}_i-\boldsymbol{\mu}_n)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_n) + \mathrm{tr}((\boldsymbol{\Lambda}_0+\boldsymbol{S}_i)+\frac{\kappa_0 n_i}{\kappa_0+n_i}(\boldsymbol{\mu}_0-\bar{\boldsymbol{Q}}_i)'(\boldsymbol{\mu}_0-\bar{\boldsymbol{Q}}_i)\boldsymbol{\Sigma}_i^{-1})]\} \tag{6.43}$$

which can be further modified to obtain the multivariate normal inverse-Wishart distribution as

$$p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i|\boldsymbol{Q}_i) \propto |\boldsymbol{\Sigma}_i|^{-(\nu_n+N+2)/2} \exp\{-\frac{1}{2}[(\kappa_{n_i})(\boldsymbol{\mu}_i-\boldsymbol{\mu}_{n_i})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_{n_i}) + \mathrm{tr}(\boldsymbol{\Lambda}_{n_i}\boldsymbol{\Sigma}_i^{-1})]\} \tag{6.44}$$

with posterior parameters

$$\begin{aligned} \nu_{n_i} &= \nu_0 + n_i, \qquad \boldsymbol{\mu}_{n_i} = \frac{\kappa_0\boldsymbol{\mu}_0 + n_i\bar{\boldsymbol{Q}}_i}{\kappa_0+n_i} \\ \kappa_{n_i} &= \kappa_0 + n_i, \qquad \boldsymbol{\Lambda}_{n_i} = \boldsymbol{\Lambda}_0 + \boldsymbol{S}_i + \frac{\kappa_0 n_i}{\kappa_0+n_i}(\boldsymbol{\mu}_0-\bar{\boldsymbol{Q}}_i)'(\boldsymbol{\mu}_0-\bar{\boldsymbol{Q}}_i) \end{aligned} \tag{6.45}$$

The simulation of the state mean vector and covariance matrix from (6.44) may be summarized as follows:

1. Assume prior parameters $\boldsymbol{\mu}_0$, κ_0, ν_0, and $\boldsymbol{\Lambda}_0$.

2. Calculate posterior parameters from (6.45).

3. Simulate $\boldsymbol{\Sigma}_i$ from an inverse-Wishart distribution

$$\boldsymbol{\Sigma}_i \sim \text{Inv-Whishart}(\nu_{n_i}, \boldsymbol{\Lambda}_{n_i}^{-1}) \tag{6.46}$$

4. Simulate $\boldsymbol{\mu}_i$, given $\boldsymbol{\Sigma}_i$, from a multivariate normal distribution

$$\boldsymbol{\mu}_i|\boldsymbol{\Sigma}_i \sim \mathrm{N}_N(\boldsymbol{\mu}_{n_i}, \boldsymbol{\Sigma}_i/\kappa_{n_i}) \tag{6.47}$$

In the incorporation of parameter uncertainty, one of the major challenges is the selection of appropriate prior parameters in order to get a suitable diffuse prior distribution. As a preliminary analysis, a comparison of the prior and posterior distributions must be included to ensure that the prior distribution is diffuse relative to the posterior distribution. The selection of prior parameters will be discussed later in connection with the application.

Initializing the Parameter Set in the Gibbs Sampler

The maximum likelihood estimates of the model parameters can be used as initial values in the Gibbs sampling. Use of ML-parameters will improve the convergence of the Gibbs sampler. The initial missing data values may be estimated based on correlations with the observed data from other sites in the basin.

Assessing MCMC Convergence

In the multi-site MS model, the same convergence diagnostics as in the single-site MS model are applied, see Section 5.7.2. In addition, diagrams showing the Gibbs samples for each individual parameter may be examined visually.

As in the single-site case, the constraint that $\boldsymbol{\mu}_1 < \cdots < \boldsymbol{\mu}_M$ is enforced in order to avoid aliasing.

Chapter 7

Disaggregation Using Markov-Switching Model

7.1 Introduction

As explained in Chapters 5 and 6, non-linear Markov-Switching models provide an interesting alternative for modeling climate regimes present in aggregated annual flows. In typical water resources system analyses, generation of sub-annual flows such as monthly flows may be needed. In this case, temporal or spatial disaggregation models may be employed to disaggregate generated higher-level flows into lower-level flows.

As mentioned in Section 4.4, traditional linear disaggregation models require normally distributed data. When data are not normally distributed, an appropriate transformation must be applied prior to the use of the model but in this case, the additivity property is lost and some form of adjustment is needed. However, the adjusted

variables may no longer represent the assumed marginal distributions and important statistics such as the mean and the variance of the observed lower level variables may not be well reproduced.

In this chapter, a Markov Switching disaggregation (MSD) model is proposed to disaggregate higher-level flows generated using the MS model into lower-level flows [*Akıntuğ and Rasmussen*, 2005c]. With this modeling approach, it is possible to preserve the non-linear structure of both the higher-level and lower-level flows, while their important statistics are implicitly preserved. The assumptions and the formulation of the MSD model are presented in the following section. In the subsequent section, the estimation of model parameters is outlined. In Section 7.4, the performance of the proposed model is investigated by a comparison with the classical Valencia-Schaake disaggregation (VSD) model [*Valencia and Schaake*, 1973] described in Section 4.4.1. In the last section of this chapter, the potential advantages and disadvantages of the proposed MSD model are discussed.

7.2 Model Formulation

The proposed MSD model can be used for both temporal and spatial disaggregation. In temporal disaggregation, annual flows q_t are generated using an M-state MS model as described in Chapter 5. It is assumed that monthly flows, q_{tm}, $m = 1, \ldots, 12$, are also described by M climate regimes and have the same state probability sequence as annual flows, i.e. $\Pr\{s_t = i|\boldsymbol{Q}_{Tm}\} = \Pr\{s_t = i|\boldsymbol{Q}_T\}$ where $\boldsymbol{Q}_{Tm} = (q_{1m}, \ldots, q_{Tm})'$ is the month m flows. In this way, every month is in the same climate state as the year

it belongs to. In simulations, the climate state s_t is known from the annual model. It is further assumed that, given the climate state, q_{tm} is normally distributed with mean and variance that are specific to the state. The MSD model is obtained by using the above assumptions with the VSD modeling structure given in (4.51). In the MSD model, the zero mean annual flows x_t may be disaggregated into zero mean monthly flows $\boldsymbol{y}_t = (y_{t,1}, \ldots, y_{t,12})'$ in year t as

$$\boldsymbol{y}_t = \boldsymbol{A}_{s_t} x_t + \boldsymbol{B}_{s_t} \epsilon_t \tag{7.1}$$

where the state dependent parameter matrices $\boldsymbol{A}_{s_t}$ and $\boldsymbol{B}_{s_t}$ have dimensions (12×1) and (12×12), respectively, and $\boldsymbol{\epsilon}_t$ is an (12×1) column matrix of independent standard normal variables with zero mean and unit variance.

7.3 Parameter Estimation

With the MSD model described in the previous section, the marginal distributions of monthly flows become mixtures of M-normal distributions. Since the state probability sequences are fixed and known from the annual MS model, the parameters of the state distributions for each month can be estimated using the maximum likelihood estimators given in (6.14) and (6.15) with $\boldsymbol{q}_t$ replaced by $\boldsymbol{q}_{tm} = (q_{t,1}, q_{t,2}, \ldots, q_{t,12})'$:

$$\boldsymbol{\mu}_i = \frac{\sum_{t=1}^{T} \Pr\{s_t = i | \boldsymbol{Q}_T, \boldsymbol{\theta}\} \boldsymbol{q}_{tm}}{\sum_{t=1}^{T} \Pr\{s_t = i | \boldsymbol{Q}_T, \boldsymbol{\theta}\}}, \quad i = 1, \ldots M. \tag{7.2}$$

$$\boldsymbol{\Sigma}_i = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\} (\boldsymbol{q}_{tm} - \boldsymbol{\mu}_i) (\boldsymbol{q}_{tm} - \boldsymbol{\mu}_i)'}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}}, \quad i = 1, \ldots M. \tag{7.3}$$

It should be emphasized again that the state probabilities $\Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}$ are inherited from the annual model. The diagonal of the state covariance matrix in (7.3) gives the state variances of each month. Alternatively the state mean and variance of each month m can be estimated individually using the single-site estimators of mean and variance given in (5.32) and (5.33).

In (7.1), the parameter matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ for state $i = 1, \ldots, M$, may be estimated using the method of moments as described in (4.52) and (4.53):

$$\hat{\boldsymbol{A}}_i = \boldsymbol{S}_{Y_i X_i} \boldsymbol{S}_{X_i X_i}^{-1} \tag{7.4}$$

$$\hat{\boldsymbol{B}}_i \hat{\boldsymbol{B}}_i' = \boldsymbol{S}_{Y_i Y_i} - \boldsymbol{A}_i \boldsymbol{S}_{X_i Y_i} \tag{7.5}$$

where the sample covariance matrices may be calculated as

$$\boldsymbol{S}_{Y_i Y_i} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\} (\boldsymbol{q}_{tm} - \boldsymbol{\mu}_i) (\boldsymbol{q}_{tm} - \boldsymbol{\mu}_i)'}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}} \tag{7.6}$$

$$\boldsymbol{S}_{Y_i X_i} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\} (\boldsymbol{q}_{tm} - \boldsymbol{\mu}_i) (q_t - \mu_i)'}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}} \tag{7.7}$$

$$\boldsymbol{S}_{X_i Y_i} = \frac{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\} (q_t - \mu_i) (\boldsymbol{q}_{tm} - \boldsymbol{\mu}_i)'}{\sum_{t=1}^{T} \Pr\{s_t = i|\boldsymbol{Q}_T, \boldsymbol{\theta}\}} \tag{7.8}$$

where μ_i is the state i mean of the higher-level variable known from the annual MS model. The sample covariance matrix $\boldsymbol{S}_{X_i X_i}$ is equal to the state i variance σ_i^2 of the higher-level variables and is known from annual MS model.

7.4 Preliminary Analysis

Before using the proposed disaggregation model for drought frequency analysis, the performance of the 2-state MSD model was explored by comparing it with the classical VSD model in (4.51).

The MSD model and VSD model were applied to 138 years (1860-1997) of monthly flows of the Niagara River at Queenston, Ontario. The data were obtained from the HYDAT CD-ROM provided by Environment Canada. Five hundred sequences of 138 years annual flows were generated using the 2-state MS model given in (5.1) and generated annual flows were disaggregated with the two models. The MSD model was applied according to the principles described in the previous section. In the case of the VSD model, observed annual flows and monthly flows were transformed to obtain a normal marginal distribution using the transformation models in Section 4.2, with the best model selected according to the Filliben statistic. For the Niagara River, Box-Cox transformations with different λ values were selected for observed monthly and annual flows. In the simulations, annual flows generated by the 2-state annual MS model have a normal mixture marginal distribution, so annual flows were subjected to a normalizing transformation before use of the VSD model. After application of the VSD model, the disaggregated monthly values were first inverse-transformed and then adjusted using the proportional adjustment procedure to sum up to annual values, see Section 4.4.4. Transformation and adjustment are not required for the MSD model. Both models were tested for their ability to reproduce the following important statistics of the observed flows: (1) Mean; (2) Standard deviation; (3) Cross-correlation; and

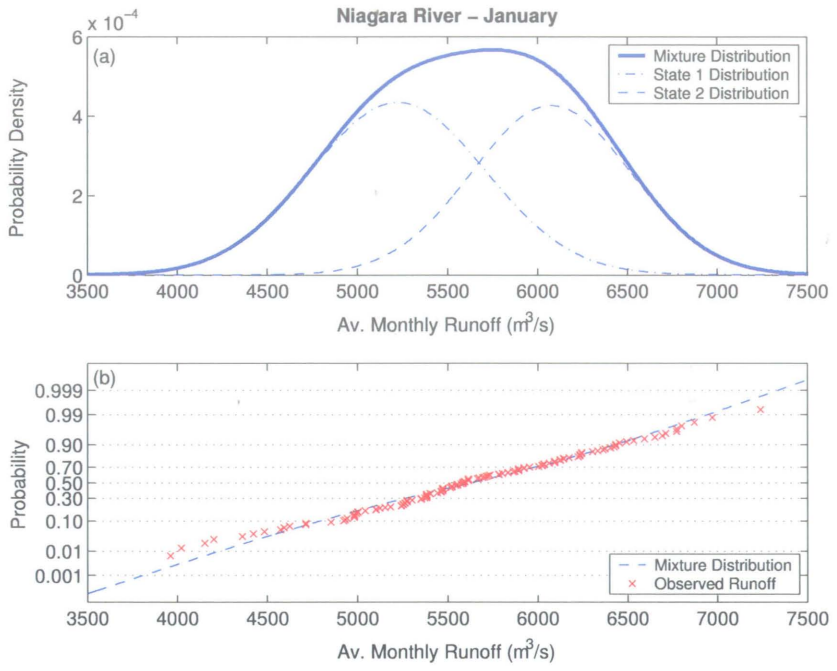


Figure 7.1: January flows of the Niagara River at Queenston, Ontario, simulated with the MSD model. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution

(4) Marginal distribution.

As a first step, the goodness-of-fit of the marginal distributions of the monthly flows is examined. Figure 7.1 illustrates the fitted marginal probability density function for the MSD model (mixture of two normals) and the normal probability plot of observed January flows of the Niagara River. The marginal distribution of January flows is very well preserved by the MSD model. The probability density and normal probability plot of the other months are given in Appendix B. Generally, the fit for the other months is very good.

To examine the sampling variability of various statistics, box plots were used. When an observed statistic does not fall within the range of the box, the model does not reproduce that statistic well.

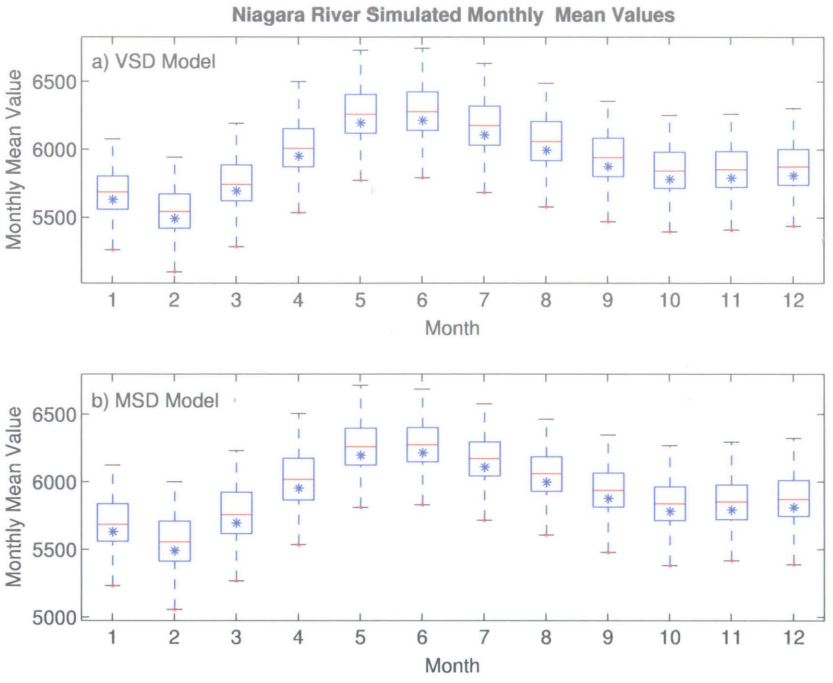


Figure 7.2: Niagara River simulated monthly mean values using a) VSD model. b) MSD model. The asterisk (∗) represents the observed statistic.

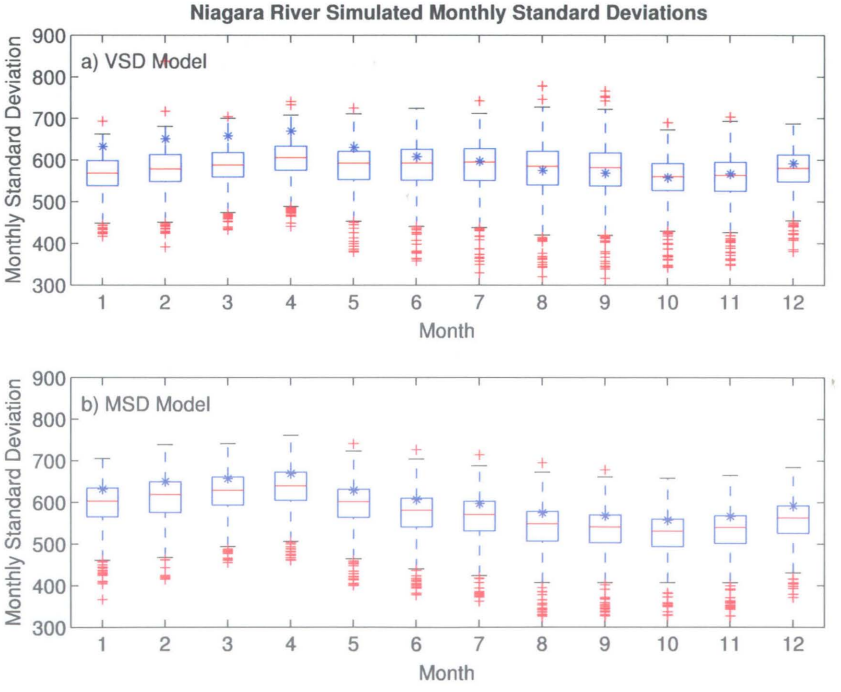


Figure 7.3: Niagara River simulated monthly standard deviations using a) VSD model. b) MSD model. The asterisk (∗) represents the observed statistic.

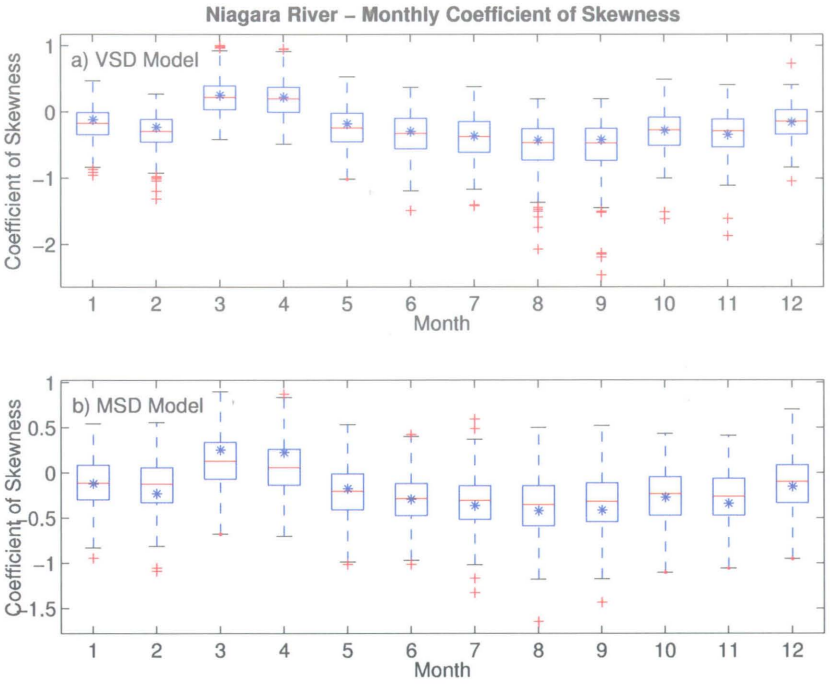


Figure 7.4: Niagara River simulated monthly coefficient of skewness using a) VSD model. b) MSD model. The asterisk (∗) represents the observed statistic.

The comparison of simulated and observed monthly means, standard deviations, and coefficients of skewness are given in Figure 7.2 to Figure 7.4 for both VSD and MDS models. The labels on the x-axis (1,2,...,12) indicates the calender month (January=1).

As illustrated in Figure 7.2 and Figure 7.4, the means and the coefficients of skewness of observed monthly flows for each month are well reproduced by both models. However, Figure 7.3 suggests the simulated standard deviation of the first five months are not well preserved by the VSD model. The MSD model reproduces the standard deviations relatively better. This bias in the VSD model is due to the inverse transformation and the required adjustment of the variables that distort the distribution of simulated monthly flows.

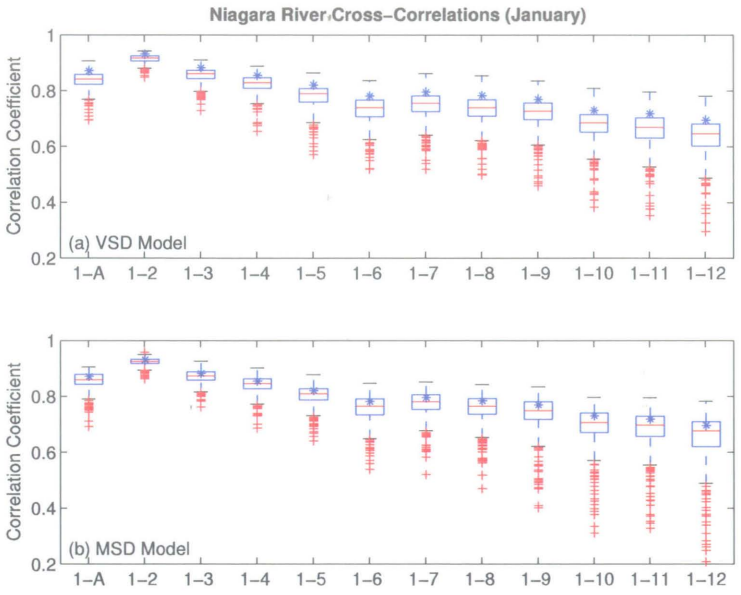


Figure 7.5: Simulated and observed cross-correlation pairs of January flows of the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk ($*$) represents the observed statistic. In the x-axis, 1-A indicates the correlation between January flows and annual flows, 1-2 indicates the correlation between January and February flows, and so on.

The two models were also compared in terms of their ability to preserve observed flow correlations. Figure 7.5 shows the correlation between January flows and annual flows, and between January flows and the remaining 11 months' flow of the same year. The comparison for other months is given in Appendix C. For all months, the MSD model is able to reproduce the dependence between monthly variables and between monthly and annual variables. However, the correlation between certain pairs of months is not well reproduced by the VSD model. This is due to the bias caused by the inverse-transformation and adjustment procedure when the observed data are not normally distributed.

To further evaluate the preservation of the assumed marginal distributions by the two models, the probability plot of 10,000 simulated flows are examined. In order

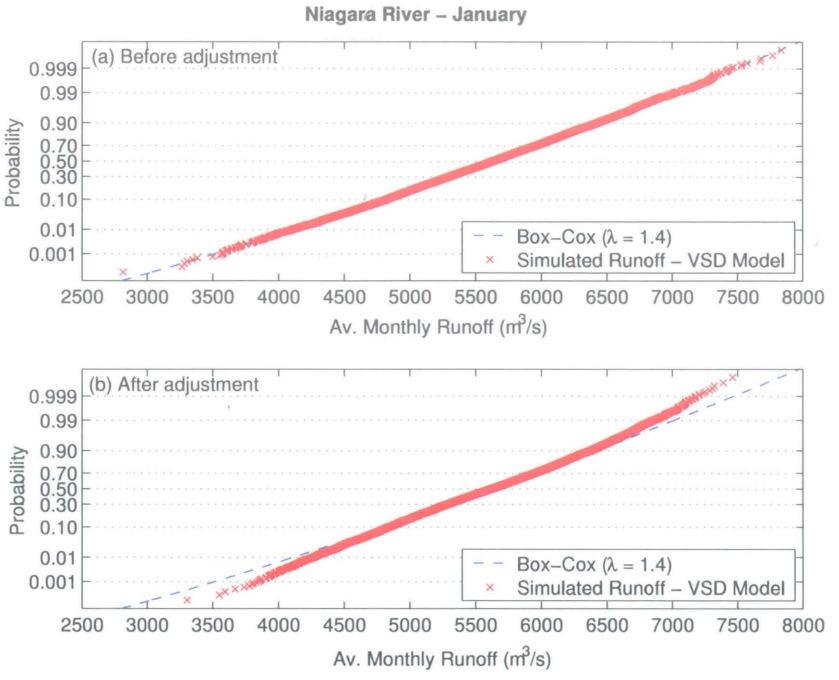


Figure 7.6: Simulated Niagara River January flows disaggregated by VSD model. a)Before adjustment procedure b) After adjustment procedure

to identify the possible source of distortion in the marginal distribution, the normal probability plot of January flows simulated from both the VSD and the MSD model were examined before and after the adjustment procedure. Since no transformation and adjustment are required for the MSD model, the marginal distributions, which are mixtures of two normals, are very well preserved with 10,000 simulated flows for all months. In the VSD model, the Box-Cox distribution with $\lambda = 1.4$ was suggested by the Filliben statistic for January flows and Figure 7.6 illustrates how the tails of January flows are distorted after application of the adjustment procedure. Similar distortions were identified for most of the other months, see Appendix D.

The application of the proposed MSD model to the Manitoba Hydro's system is described in Section 8.8.

7.5 Discussion

One of the potential disadvantages of the MSD model is that some flexibility in the marginal distributions of the lower-level variables is lost due to the fact that the mixing probabilities are imposed by the annual model. Therefore, it is important to verify that the marginal distributions of the lower-level variables are well preserved.

The MSD model preserves the additivity property because for a given year both the higher-level and lower-level variables are generated from normal distributions. Despite the fact that monthly means, variances, and cross-correlations are not modeled directly, they appear to be generally well preserved by the model.

The MSD model is not parsimonious in terms of parameters because a separate set of model parameters is required for each climate state. Hence it is not recommended to use the MSD model for multi-site disaggregation when the length of the observed record is short. The MSD model may be used for univariate temporal or spatial disaggregation. A reduction in parameters can be obtained using a staged disaggregation procedure [*Santos and Salas*, 1992] which disaggregate higher-level variables to lower-level variables in two or more steps, or Lane's condensed disaggregation model, given in (4.59a) and (4.59b). However, if such methods are used, the additivity property is lost.

Chapter 8

Application to Manitoba Hydro's System

8.1 Introduction

One of the objectives of this project is to quantify model, parameter, and data uncertainties in the drought frequency analysis of Manitoba Hydro's system. In order to achieve this objective, the proposed stochastic models given in Chapters 4, 5, 6, and 7 are applied to Manitoba Hydro's system. After the generation of synthetic flows, drought events are extracted from each scenario according to the definition of drought adopted in Chapter 3, and the drought frequency analysis is performed.

In this chapter, the details of the application of stochastic models are reported and the results of the energy drought frequency analysis are presented.

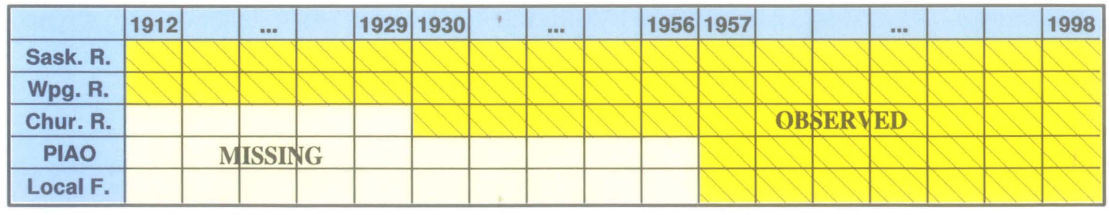


Figure 8.1: Observed and missing Manitoba Hydro streaflow data.

8.2 Historical Data Sets

The streamflow data used in the project were supplied by the Resource Planning and Market Analysis Department of Manitoba Hydro. As mentioned in Chapter 2, Manitoba Hydro's system is divided into five hydrological components. The flow records of the Winnipeg River at Slave Falls and the Saskatchewan River at Grand Rapids are available from 1912 to 1998 without any missing flows. The flow records of the Churchill River are available from 1930 to 1998. The Lake Winnipeg PIAO represents all Lake Winnipeg inflows except the Winnipeg and Saskatchewan River inflows. The inflow records of the tributaries on the western and eastern sides of Lake Winnipeg are available from 1957 to 1998. The local flows of the Burntwood and the Nelson River are also available from 1957 to 1998. Manitoba Hydro uses data for the period 1912-98 in their deterministic SPLASH model for planning and evaluation of their system. To obtain a full 1912-98 data record for all sites, data extension for the above hydrological components was performed by Manitoba Hydro [*Girling*, 1988, 1990].

In the stochastic time series models, the accuracy of estimated parameters depends on the record length and on the accuracy of the data. Therefore, initially three sets of data covering the periods 1912-98, 1930-98, and 1957-98 were intended for use

in the application. These periods represent different degrees of extension (see Table 8.1). However, eventually the 1957-98 data set was not used in the drought frequency analysis because together with gauged flows of PIAO more than 70% of the system's inflow is known for the period 1930-1956 which covers the critical drought period as well. The statistical data analysis and the drought frequency analysis are performed considering the 1912-98 and 1930-98 data sets.

8.3 Selection of Modeling Sites and Water Year

In Manitoba Hydro's system, there are 15 streamflow sites of interest and Lake Winnipeg PIAO (see Section 2.2). In the stochastic modeling, it was found to be beneficial to combine sites in order to decrease the dimension of the framework. The aggregation of the streamflow sites significantly reduces the dimension of the models and the number of parameters to be estimated. The aggregation is expected to improve the performance of the stochastic models.

In this project, the disaggregation approach was employed. Annual flows at one or more aggregated basin sites are generated first and then disaggregated into monthly flows, which can then be spatially disaggregated into monthly flows at sub-basin sites. If it is required for the analysis of the system, monthly flows at any sub-basin site can be further spatially disaggregated into monthly flows at key sites.

8.3.1 Basin Sites

The Winnipeg River, the Churchill River, and the Saskatchewan River are considered as basin sites in the stochastic modeling framework because they are the major rivers that contribute on average 28%, 27%, and 18%, respectively, to the total inflow into Manitoba Hydro's system. As discussed in Chapter 2, PIAO is an important component of Manitoba Hydro's system with a 17% contribution to total inflow and it is also selected as a basin site. The total contribution of the local flows in the Burntwood and the Nelson River basins is about 16% (Table 2.2). Site-to site correlation coefficients of local flows are high so these local flows are aggregated and considered as a basin site as well.

In summary, the following basin sites are selected:

1. **Churchill River**: The Churchill River flows upstream of Southern Indian Lake.
2. **Saskatchewan River**: The Saskatchewan River flows at Grand Rapids.
3. **Lake Winnipeg PIAO**: The partial inflow available as outflow values of Lake Winnipeg.
4. **Winnipeg River**: The Winnipeg River flows at Slave Falls
5. **Local Flows**: Aggregated local flows at the Burntwood and the Nelson Rivers.

In the stochastic models, the flows of the five basin sites given above were also aggregated to one artificial site called the Aggregated Basin which represents the total inflow to Manitoba Hydro's system.

Table 8.1: Mean, standard deviation, and coefficient of skewness of observed annual flows for five basins and Aggregated Basin.

	1912-98			1930-98		
Basins	μ	σ	C_s	μ	σ	C_s
Churchill River	35554	6655	-0.093	35488	7466	-0.057
Saskatchewan River	20101	5968	0.806	19201	5451	1.012
Lake Wpg PIAO	18504	9740	0.411	19452	10114	0.297
Winnipeg River	31651	8924	0.412	32302	9422	0.212
Local Flows	13142	3316	0.289	13486	3429	0.275
Aggregated Basin	118952	23589	0.306	119928	25069	0.152

The mean values, the standard deviations, and the skewness coefficients of the annual flows of the five basin sites as well as the Aggregated Basin Flows are given in Table 8.1.

8.3.2 Sub-basin Sites

As seen from Figure 2.3, six generation stations are located on the Winnipeg River. It is assumed that the flow at Slave Falls is representative for all other stations. The hydrometric station on the Saskatchewan River is located at the Grand Rapids generating station which means that the flow passes through the generating station directly. There are no generating stations on the Churchill River but a significant portion of this river is diverted through the Notigi Control structure into the Nelson River upstream of the Kettle generating station. This diverted flow will be taken into consideration for Kettle, Long Spruce, and Limestone generating stations. Therefore, there is no need to spatially disaggregate the Winnipeg River, the Saskatchewan River, and the Churchill River flows into sub-basin sites.

The PIAO flows are used as input to Lake Winnipeg in order to calculate the

Lake Winnipeg outflows which are used for energy generation on the Nelson River. Therefore, there is also no need to spatially disaggregate the Lake Winnipeg PIAO flows.

The 12 local flows, given in Table 2.2, are modeled as a single artificial basin site. The aggregated Local Flows are disaggregated into artificial sub-basin sites and key sites according to the contribution (physical location) of these local flows to the generating stations located on the Nelson River. As shown in Table 3.1, the Upper Nelson River local flows NR0, NR1, and NR2 contribute to all generating stations on the Nelson River. The Burntwood River local flows BR1, BR2, BR3, BR4A, and BR4B, and the Nelson River local flows NR3 and NR4 contribute to Kettle, Long Spruce, and Limestone generating stations. At the lower Nelson River, NR5 has contribution to Long Spruce and Limestone, and NR6 to Limestone only. Because of the very high correlation between NR5 and NR6, they are aggregated and considered as a sub-basin site in the modeling framework. In summary, the aggregated key sites are

1. **NR012**: NR0+NR1+NR2
2. **NR34BRAll**: NR3+NR4+BR1+BR2+BR3+BR4A+BR4B
3. **NR56**: NR5+NR6

Table 8.2: Coefficients of NR5 and NR6 Local Flows for the linear deterministic disaggregation model.

1912-98 Data Set:

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
NR5	0.74	0.68	0.66	0.62	0.70	0.74	0.71	0.68	0.71	0.75	0.78	0.76
NR6	0.26	0.32	0.34	0.38	0.30	0.26	0.29	0.32	0.29	0.25	0.22	0.24

8.3.3 Key Sites

If it is necessary, each sub-basin site flows can be further disaggregated to flows at control points according to their contribution to the generating stations. Because of the contribution of key sites BRAll, NR012, and NR34 to the same set of generating stations, there is no need to disaggregate these into flows at individual control points. However, key site NR56 must be disaggregated into two control points NR5 and NR6, because the local flow NR5 contributes to both Long Spruce and Limestone while NR6 contributes only to Long Spruce. Because of the high correlation between NR5 and NR6 (0.92), there is no need to use a stochastic model to disaggregate monthly NR56 flows to NR5 and NR6 flows. Instead a deterministic linear disaggregation procedure is used. The historical flows at NR5 and NR6 are used to calculate the proportions of these flows at the artificial key site NR56. The linear coefficients for the 1912-98 data set are given in Table 8.2.

8.4 Stochastic Modeling Frameworks

The first step in the modeling of the multi-site streamflow time series is the selection of an appropriate modeling framework defining the components to be modeled and

the associated stochastic models. In the selection of the frameworks, an attempt was made to minimize the required number of modeling steps. The two main modeling frameworks used in this research project are outlined in Figures 8.2 and 8.3. These two frameworks will be compared in terms of their ability to preserve the statistical properties of observed flows.

In Modeling Framework-1 given in Figure 8.2, a single-site annual model is used to generate annual flows for the entire basin. A spatial disaggregation model is then used to disaggregate generated annual flows into annual flows at the five basin sites. The annual flows at the five basin sites are subsequently disaggregated into monthly flows using a multi-site annual-to-monthly disaggregation model. The local flows in the Burntwood and Nelson Rivers must be further disaggregated into monthly flows at the three key sites. The last step of the framework is the deterministic disaggregation of monthly flows at key site NR56 into NR5 and NR6 flows using a deterministic linear disaggregation procedure (Section 8.3.3).

In Modeling Framework-2 given in Figure 8.3, annual flows at the five major basin sites are generated first using a multi-site annual model. The remaining steps of the framework are the same as in Modeling Framework-1.

As mentioned earlier, in this research project, the emphasis is placed on the modeling of the annual flows. Hence, the two frameworks differ only in the modeling of the annual flows. In the annual flow generation, parameter uncertainty is considered for both frameworks. In Framework-2, the data uncertainty in the multi-site MS model is also considered. Using AR(1) and MS models, a total of nine modeling frameworks

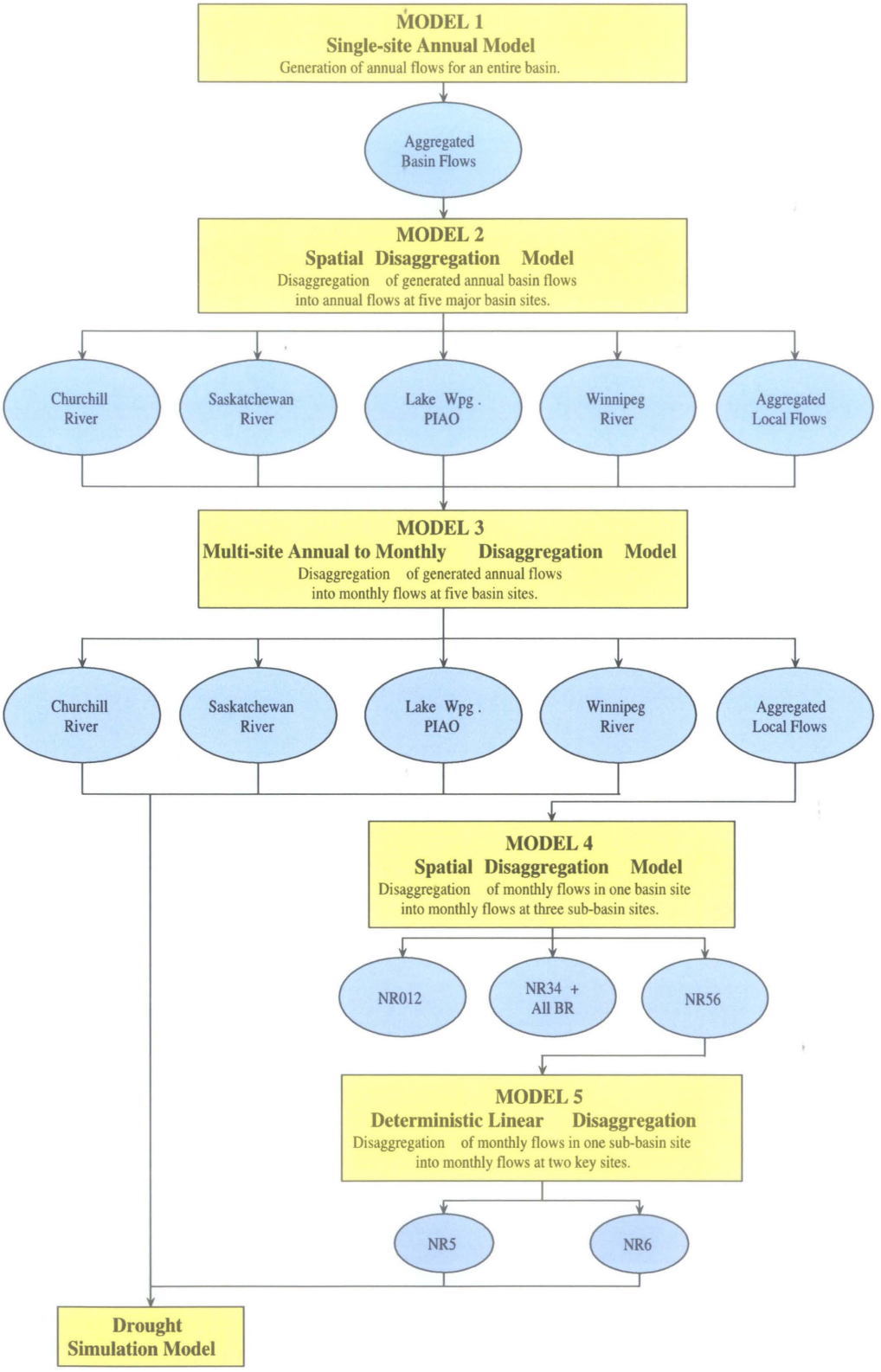


Figure 8.2: Stochastic Modeling Framework-1

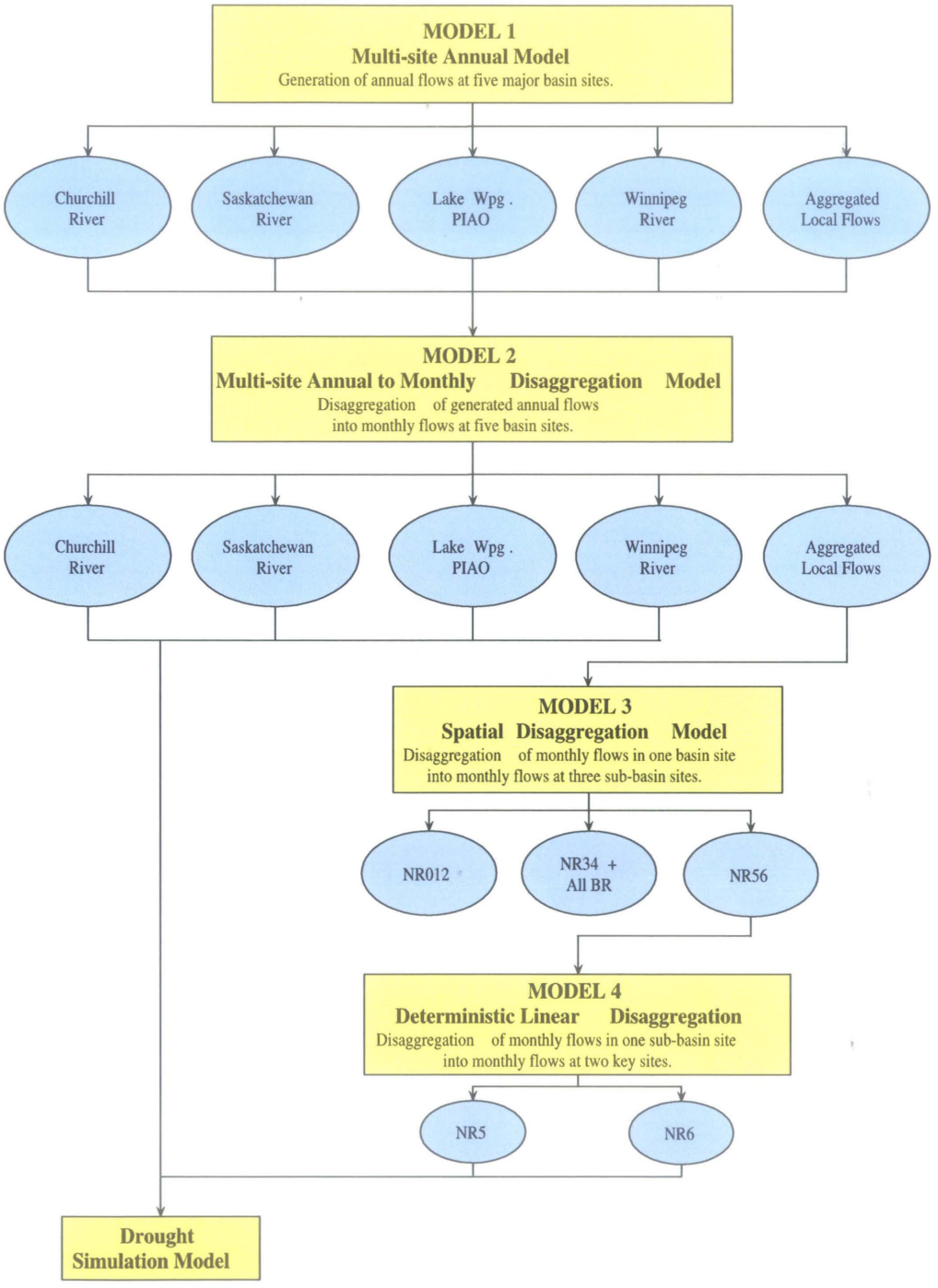


Figure 8.3: Stochastic Modeling Framework-2

Table 8.3: Selected models in Framework-1.

Modeling Framework-1:

Model	Model-1	Model-2	Model-3	Model-4	Model-5
MF-1.1	Single-site AR(1) Model	MR Spatial Disagg. Model	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disagg. Model
MF-1.2	Single-site AR(1) Model with par. uncertainty	MR Spatial Disagg. Model	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disagg. Model
MF-1.3	Single-site Markov-Switching Model	MS Spatial Disagg. Model	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disagg. Model
MF-1.4	Single-site MS Model with par. uncertainty	MS Spatial Disagg. Model	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disagg. Model

are employed (see Tables 8.3 and 8.4).

8.5 Application of the Traditional Models

In the drought frequency analysis, the performance of the proposed MS model was compared with the traditional AR(1) model. Details about the application of the AR(1) model are given in the following.

8.5.1 Marginal Distributions and Data Transformation

In the traditional first order autoregressive model and disaggregation models, described in Chapter 4, it is assumed that flows are normally distributed. The marginal distributions of the observed annual and monthly flows of the five basin sites and three key sites were investigated in terms of normality. Normal, 2-parameter lognormal, 3-parameter lognormal, and Box Cox distributions were fitted to observed flows and

Table 8.4: Selected models in Framework-2.

Modeling Framework-2:

Model	Model-1	Model-2	Model-3	Model-4
MF-2.1	Multi-site AR(1) Model	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disaggregation Model
MF-2.2	Multi-site AR(1) Model with parameter uncertainty	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disaggregation Model
MF-2.3	Multi-site Markov-Switching Model	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disaggregation Model
MF-2.4	Multi-site MS Model with parameter uncertainty	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disaggregation Model
MF-2.5	Multi-site MS Model with parameter and data uncertainty	Lane's Temporal Disaggregation Model	Lane's Spatial Disaggregation Model	Deterministic Disaggregation Model

the Filliben correlation coefficient statistic given in Section 4.2.4 was used to select the best distribution. For a given series, the distribution with the highest Filliben statistic is usually selected as transformation. However, in most cases, Filliben statistics for two or more distributions were very close to each other. Having similar Filliben statistics does not mean that two distributions fit data equally well in all regions of the distribution. Hence, beside the Filliben statistic, probability plots were visually investigated one by one and compared with other candidate distributions. The ones having a better fit in the lower tail (low flows) of the plot which is of most interest in drought studies were selected. The selected marginal distributions for the 1912-98 data set are given in Table 8.5. The marginal distribution of monthly flows of the Aggregated Basin Flows site is not included in Table 8.5 because they were not modeled in the frameworks in Figure 8.2 and 8.3.

Table 8.5: Selected marginal distributions for basin sites and key sites.

1912-98 Data Set:

Site	Ann	J	F	M	A	M	J	J	A	S	O	N	D
Agg. Basin F.	3	-	-	-	-	-	-	-	-	-	-	-	-
Churchill R.	1	1	1	4	1	1	1	4	1	1	4	2	2
Saskatchewan R.	3	1	1	3	3	2	3	4	4	4	4	2	1
Lake Wpg. PIAO	3	1	1	1	3	3	4	3	3	1	1	3	1
Winnipeg R.	4	4	3	4	3	3	3	3	4	4	3	2	2
Agg. Local F.	3	3	3	1	3	1	3	1	1	3	2	2	3
NR012	3	3	3	3	4	4	4	3	3	2	2	2	3
NR34BRAll	1	3	3	1	2	1	3	1	1	1	3	3	3
NR56	3	3	3	3	4	3	2	3	3	3	4	3	3

1: Normal, 2: 2-Par. Lognormal, 3: 3-Par-Lognormal, 4: Box-Cox

8.5.2 ARMA-Type Model Selection

The selection of the order of ARMA models is commonly based on the Akaike Information Criterion (AIC) [*Akaike*, 1974]:

$$AIC = -2\ln[\max(L_k)] + 2(k) \tag{8.1}$$

where $[\max(L_k)]$ is the maximum value of the likelihood function for a candidate model with k parameters. Hence, the AIC takes into account the parsimony of a model. For Autoregressive Moving Average [ARMA(p,q)] models, minimizing (8.1) is equivalent to minimizing [*Salas et al.*, 1988]

$$AIC(p,q) = T\ln(\hat{\sigma}_\varepsilon^2) + 2(p+q) \tag{8.2}$$

where T is the length of observed record and $\hat{\sigma}_\varepsilon^2$ is the estimate of the residual variance.

In this project, the best ARMA-type model was selected from among ARMA(1,1),

Table 8.6: AIC values for annual ARMA-type model identification.

1912-98 Data Set:

	AIC					
	Basin F.	Chur. R.	Sask. R.	PIAO	Wpg. R.	Local F.
ARMA(1,1)	-18.78	-67.01	-13.18	-11.97	-8.31	-8.36
AR(2)	-18.74	-67.00	-13.18	-11.97	-8.01	-8.27
AR(1)	-19.84	-68.63	-14.84	-13.92	-8.47	-9.35
AR(0)	2184.3	1964.1	1945.1	2030.4	2015.2	1842.9

AR(2), AR(1), and AR(0) models. The model that gave the minimum AIC according to (8.2) was selected. As seen from Table 8.6, the AR(1) model is the best choice between competing models for all sites.

8.6 Application of the 2-state Single-site MS Model to Manitoba Hydro Data

In the drought frequency analysis, Framework-1 generates aggregated annual flows for the entire Manitoba Hydro basin using a single site annual stochastic time series model. In this section, the application of the single site MS model for Manitoba Hydro along with some implementation issues of this model is presented. The MS model identification and parameter uncertainty are also explained in this section.

To determine the order of the MS model for the Aggregated Basin Flows and the five basin sites, the BIC described in Section 5.6 was employed. More specifically, models of order 2, 3, and 4 were considered. Models with five states or more were considered of little interest because of the high number of parameters. The corresponding BIC values for each data set are given in Table 8.7. In the case of the Churchill River 1912-98 data,

Table 8.7: BIC values for MS Model identification.

1912-98 Data Set:

		$\mathbf{BIC_M}$					
States (M)	d_M	Basin F.	Chur. R.	Sask. R.	PIAO	Wpg. R.	Local F.
2	6	590.2	605.2	602.3	644.5	638.9	565.1
3	12	594.7	603.9	609.3	651.9	645.3	571.6
4	20	604.2	612.3	620.6	664.3	649.8	580.5

1930-98 Data Set:

		$\mathbf{BIC_M}$					
States (M)	d_M	Basin F.	Chur. R.	Sask. R.	PIAO	Wpg. R.	Local F.
2	6	471.4	390.6	377.8	418.9	415.9	354.5
3	12	477.4	395.5	384.2	426.6	417.1	362.0
4	20	490.6	403.5	393.2	438.1	423.3	372.7

a 3-state model appears to be the best choice, while for the other sites, the 2-state process wins the model comparison for all data sets. However, when the 1930-98 data set is considered, the 2-state process appears to be the best choice for all sites. Since only the 1930-98 period of the Churchill River is actually observed, a 2-state model was selected for all basin sites for all data sets. In the case of annual runoff, $s_t = 1$ represents a *dry* state (relative to long-term annual mean) and $s_t = 2$ a *wet* state. For the 2-state MS model, the unknown parameter vector is $\boldsymbol{\theta} = (\varrho_1, p_{12}, p_{21}, \mu_1, \mu_2, \sigma_1, \sigma_2)$.

In order to start the iteration for parameter estimation, initial parameters are needed. For the 2-state MS model, the observed data $\mathbf{Q}_T$ were compared with the long term annual mean $\bar{Q}$ to obtain the two data sets, $\mathbf{Q}_{S_1}$ for state 1 and $\mathbf{Q}_{S_2}$ for state 2. Specifically if the observed data in year t is less than $\bar{Q}$, it is considered as belonging to $\mathbf{Q}_{S_1}$, otherwise to $\mathbf{Q}_{S_2}$. The initial population parameters μ_i and σ_i, for $i = 1, 2$, are then estimated from these data sets.

The initial elements of the transition probability matrix can also be obtained from

$\mathbf{Q}_T$. Let n_{ij} denote the number of moves from state i to state j, and n_i denote the length of $\mathbf{Q}_{S_i}$. In this case, the empirical transition probability is $p_{ij} = n_{ij}/n_i$. Once p_{12} and p_{21} are known, the remaining elements of the transition probability matrix, $\mathbf{P}$, can be calculated as $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$.

For a 2-state MS model, there are two initial probabilities, ϱ_1 and ϱ_2, but only one of them needs to be estimated because $\varrho_1 = 1 - \varrho_2$. Since this parameter converges very quickly, it can be set to $\varrho_1 = 1/2$.

Once the initial parameters are selected, the ML-estimates of the model parameters can be calculated using the EM algorithm as explained in Section 5.4.

After estimation of the parameter vector $\boldsymbol{\theta}$, the streamflow sequences can be simulated in two steps. First, the state time series is simulated by a Markovian process

$$s_t | s_{t-1} \sim \text{Markovian}(\boldsymbol{\varrho}, \mathbf{P}) \tag{8.3}$$

Once the state sequence is known, the streamflow variable can be simulated from

$$q_t = \begin{cases} N(\mu_1, \sigma_1) & \text{if} \quad s_t = 1 \\ N(\mu_2, \sigma_2) & \text{if} \quad s_t = 2 \end{cases} \tag{8.4}$$

The single-site 2-state MS model is applied to the five basin sites in Manitoba Hydro's system as well as to the Aggregated Basin Flows site. The objective is to investigate the evidence of a 2-state persistence structure in the observed annual runoff of Manitoba Hydro.

Table 8.8: Estimated ML-parameters and associated model properties of the single-site 2-state MS model (Annual runoff in cfs).

1912-98 Data Set:

	Churchill River	Saskatchewan River	Lake Wpg. PIAO	Winnipeg River	Aggregated Local Flows	Aggregated Basin Flows
μ_1	25,817	17,716	13,836	25,977	11,402	107,090
μ_2	37,874	28,319	25,858	38,411	16,404	137,760
σ_1	3,193	3,615	6,843	5,462	2,206	17,027
σ_2	4,879	4,934	8,907	7,274	2,412	19,632
p_{12}	0.1274	0.0898	0.0934	0.1876	0.1726	0.1160
p_{21}	0.0308	0.3551	0.1213	0.2164	0.3010	0.1628
μ	35,526 (35,554)	19,857 (20,101)	19,065 (18,504)	31,751 (31,651)	13,225 (13,142)	119,850 (118,950)
σ	6,630 (6,555)	5,784 (5,968)	9,823 (9,740)	8,889 (8,924)	3,318 (3,316)	23,627 (23,589)
c_s	-0.3111 (-0.0935)	0.8902 (0.8058)	0.3627 (0.4110)	0.3536 (0.4122)	0.3056 (0.2891)	0.2510 (0.3076)
ρ_1	0.4367 (0.7456)	0.3005 (0.4195)	0.2891 (0.4089)	0.2901 (0.3367)	0.2771 (0.3497)	0.2953 (0.4712)
ρ_2	0.3677 (0.5269)	0.1668 (0.1251)	0.2270 (0.1472)	0.1729 (-0.0043)	0.1459 (0.0322)	0.2130 (0.1431)

The observed values are given in parentheses

Maximum likelihood estimates of the MS model for the 1912-98 record are obtained by means of the EM-algorithm and are reported in Table 8.8, along with selected statistics of the observed data.

A number of characteristics of the MS model are of interest. Many of these can be calculated analytically using the various formulas provided in Section 5.3. First of all, the model should adequately reproduce the marginal distribution of runoff. In the 2-state MS model, the marginal distribution is a mixture of two normal distributions. The resulting pdf and probability plot of the annual Aggregated Basin Flows are shown in Figure 8.4. As seen from the probability plot, the marginal distribution of the Aggregated Basin site is very well preserved. Appendix E contains probability plots of the five basin sites.

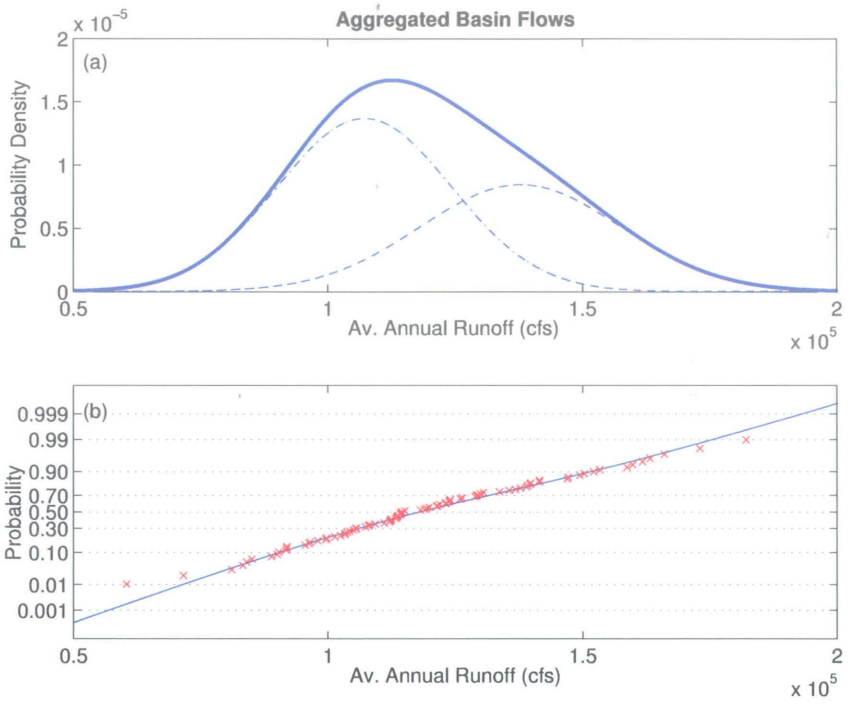


Figure 8.4: Annual Aggregated Basin Flows of Manitoba Hydro's system. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

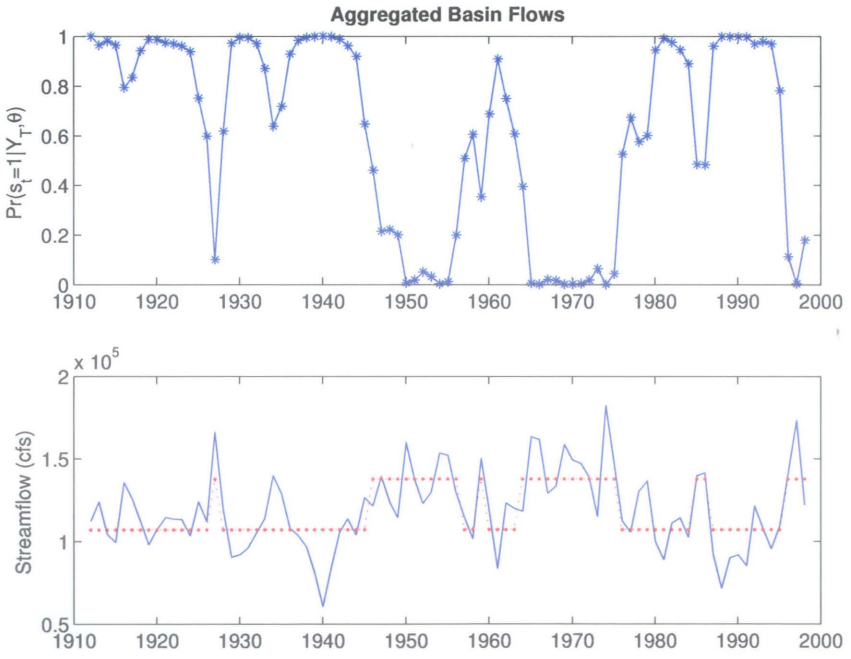


Figure 8.5: State probabilities and regime shifts for the mean annual Aggregated Basin Flows.

First, second, and third order moments are reported in Table 8.8 for all sites. There are some minor discrepancies between observed and modeled moments, but overall the reproduction of moments is good, as one would expect based on the visually good reproduction of the marginal distribution.

It is of interest to look at the estimated state probability sequences, $\Pr\{s_t = i|\mathbf{Q}_T\}$, for $i = 1, 2$, and $t = 1, 2, \ldots, T$. Figure 8.5 shows the state probabilities and the mean level associated with the most probable state for the annual Aggregated Basin Flows. At all sites, the state probabilities are very well defined, that is, most of the state probabilities are either close to zero or one (see Appendix F).

Based on the transition probabilities given in Table 8.8, it is possible to determine the average duration of each regime. For example, for the Aggregated Basin Flows, the dry regime has an expected duration of $1/p_{12} \cong 9$ years, and the wet regime has an expected duration of $1/p_{21} \cong 6$ years.

Figure 8.6 shows the autocorrelation function of a 2-state MS model and the conventional AR(1) fitted directly to the data. In this particular example, the observed short-term (lag-1) correlation is not well captured by the MS model. On the other hand, an AR(1) model preserves very well the first autocorrelation coefficient. For higher lags, the scenario is reversed. Beyond lag-2, the AR(1) model considerably underestimates the observed autocorrelation whereas the MS model does a more reasonable job in preserving the observed autocorrelation. If the interest is multi-year droughts with typical durations greater than two years, then the MS-model may seem a more prudent (conservative) choice.

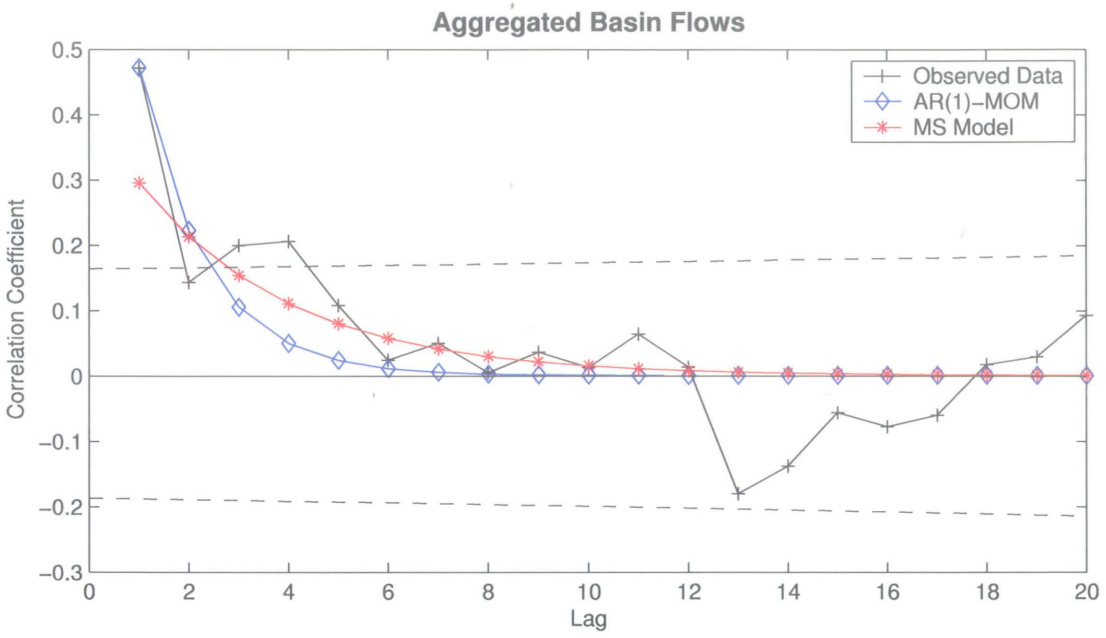


Figure 8.6: Aggregated Annual Basin Flows. Observed and modeled autocorrelation function.

As mentioned in Section 5.3.4, an M-state MS model has the same autocorrelation structure as an ARMA$(M-1, M-1)$ model. Therefore, the functional form of autocorrelation of the 2-state MS model considered here is equivalent to that of an ARMA(1,1) model. In Appendix G, the autocorrelation function for the 2-state MS model and AR(1) and ARMA(1,1) models are compared for all sites.

8.6.1 Incorporating Parameter Uncertainty

In Section 5.7, the incorporation of parameter uncertainty was developed for the M-state MS model. In this section, the specific case of $M = 2$ case will be considered. Given initial parameters, the entire state sequence will be simulated first using Chib's method.

For $M = 2$, the Multinomial likelihood function of the transition probabilities in (5.59) reduces to the Binomial function

$$l(\mathbf{S}_T|p_{ij}) \propto (p_{ij})^{(n_{ij})}(1-p_{ij})^{n_i-n_{ij}} \tag{8.5}$$

and the Dirichlet prior distribution in (5.60) becomes the Beta distribution

$$p(p_{ij}) \propto (p_{ij})^{(\alpha-1)}(1-p_{ij})^{(\beta-1)} \tag{8.6}$$

$$p_{ij} \sim Beta\,(\alpha, \beta) \tag{8.7}$$

where α and β are parameters of the prior distribution. For a diffuse prior distribution, $\alpha = \beta = 1$. Using (8.5) and (8.6), the updated posterior distribution becomes a Beta distribution

$$p(p_{ij}|\mathbf{S}_T) \propto (p_{ij})^{(n_{ij}+\alpha-1)}(1-p_{ij})^{(n_i-n_{ij}+\beta-1)} \tag{8.8}$$

$$p_{ij}|\mathbf{S}_T \sim Beta\,(\alpha + n_{ij}, \beta + n_i - n_{ij}) \tag{8.9}$$

For the 2-state single-site MS model, the k'th iteration of the Gibbs sampler pro-

ceeds as follows:

$$\begin{aligned}
\mathbf{S}_T^{(k)} &\sim p\left[\mathbf{S}_T \mid \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\sigma}^{(k-1)}, \mathbf{P}^{(k-1)}, \mathbf{Q}_T\right] && (8.10)\\
p_{12}^{(k)}|\mathbf{S}_T^{(k)} &\sim Beta\left[1+n_{12}^{(k)}, 1+n_1^{(k)}-n_{12}^{(k)}\right] && (8.11)\\
p_{21}^{(k)}|\mathbf{S}_T^{(k)} &\sim Beta\left[1+n_{21}^{(k)}, 1+n_2^{(k)}-n_{21}^{(k)}\right] && (8.12)\\
(\sigma_1^2)^{(k)}|\mathbf{Q}_1 &\sim Inv-\chi^2\left[\nu_{n_1}^{(k)}, (\sigma_{n_1}^2)^{(k)}\right] && (8.13)\\
\mu_1^{(k)}|(\sigma_1^2)^{(k)}, \mathbf{Q}_1 &\sim N\left[\mu_{n_1}^{(k)}, (\sigma_1^2)^{(k)}/\kappa_{n_1}^{(k)}\right] && (8.14)\\
(\sigma_2^2)^{(k)}|\mathbf{Q}_2 &\sim Inv-\chi^2\left[\nu_{n_2}^{(k)}, (\sigma_{n_2}^2)^{(k)}\right] && (8.15)\\
\mu_1^{(k)}|(\sigma_2^2)^{(k)}, \mathbf{Q}_2 &\sim N\left[\mu_{n_2}^{(k)}, (\sigma_2^2)^{(k)}/\kappa_{n_2}^{(k)}\right] && (8.16)
\end{aligned}$$

where $n_{ij}^{(k)}$, $i,j=1,2$, is the number of times that $s_t=i$ and $s_{t+1}=j$, and n_i is the number of times that $s_t=i$ in the state sequence $\mathbf{S}_T^{(k)}$. The posterior parameters of the Normal and Inv-Chi-square distribution, $\kappa_{n_i}^{(k)}, \nu_{n_i}^{(k)}, \mu_{n_i}^{(k)}, (\sigma_{n_i}^2)^{(k)}$, are given in (5.68)-(5.71).

For the hyperparameters (parameters of the prior distribution) of μ_i and σ_i^2, the following parameters are selected in order to obtain diffuse prior distributions:

$$\begin{aligned}
\kappa_0 &= 1 \qquad \mu_{0_i} = \mu_{i_{MLE}}\\
\nu_0 &= 5 \qquad \sigma_{0_i}^2 = \sigma_{i_{MLE}}^2 && (8.17)
\end{aligned}$$

where $\mu_{i_{MLE}}$ and $\sigma_{i_{MLE}}^2$ are the maximum likelihood estimates of state i mean and variance, respectively.

As mentioned in Section 5.7, during the simulation of the state sequence $\mathbf{S}_T^{(k)}$ in the Gibbs sampler, it is possible to obtain less than two occurrences of one of the states. In this case, since there is not enough data to simulate the parameters from the posterior distribution of that state, they are directly simulated from their prior distributions. Specifically

$$\left(\sigma_i^2\right)^{(k)} \sim Inv - \chi^2(\nu_0, \sigma_{0_i}^2) \tag{8.18}$$

$$\mu_i^{(k)} | \left(\sigma_i^2\right)^{(k)} \sim N(\mu_{0_i}, \left(\sigma_i^2\right)^{(k)} / \kappa_0) \tag{8.19}$$

However, since the prior distributions are diffuse, some constraints are applied in order to minimize the simulation of highly unlikely parameters. For the state mean and variance, the prior distribution is bounded as

$$\left(\sigma_i^2\right)^{(k)} = \left[E\left[\left(\sigma_i^2\right)^{(k)}\right] \pm \sqrt{Var\left[\left(\sigma_i^2\right)^{(k)}\right]}\right] \tag{8.20}$$

$$\mu_i^{(k)} = \left[\mu_{0_i} \pm \left(\sigma_i^2\right)^{(k)} / \sqrt{\kappa_0}\right] \tag{8.21}$$

where for the scaled inverse-χ^2-distribution

$$E\left[\left(\sigma_i^2\right)^{(k)}\right] = \frac{\nu_0}{\nu_0 - 2}\sigma_{0_i}^2 \tag{8.22}$$

$$Var\left[\left(\sigma_i^2\right)^{(k)}\right] = \frac{2\nu_0}{(\nu_0 - 2)^2(\nu_0 - 4)}\sigma_{0_i}^2 \tag{8.23}$$

When a heuristic method is used for the initial parameter vector in the Gibbs sampler, the algorithm is allowed to sample for a certain number of iterations (warm

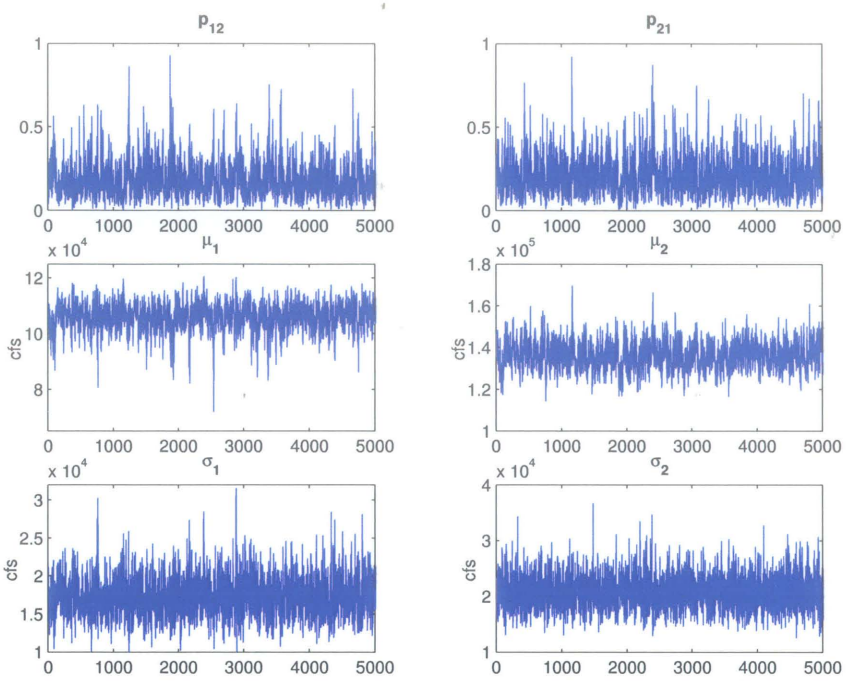


Figure 8.7: 2-State MS model parameters simulated using Gibbs sampling of the posterior distributions for Aggregated Basin Flows.

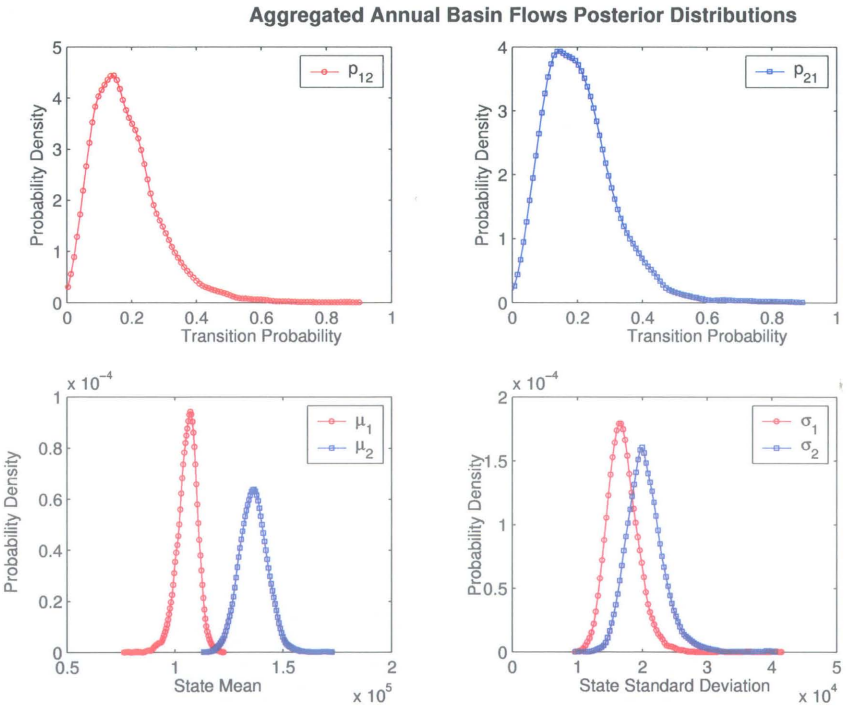


Figure 8.8: Posterior distribution of 2-State MS model parameters for Aggregated Basin Flows.

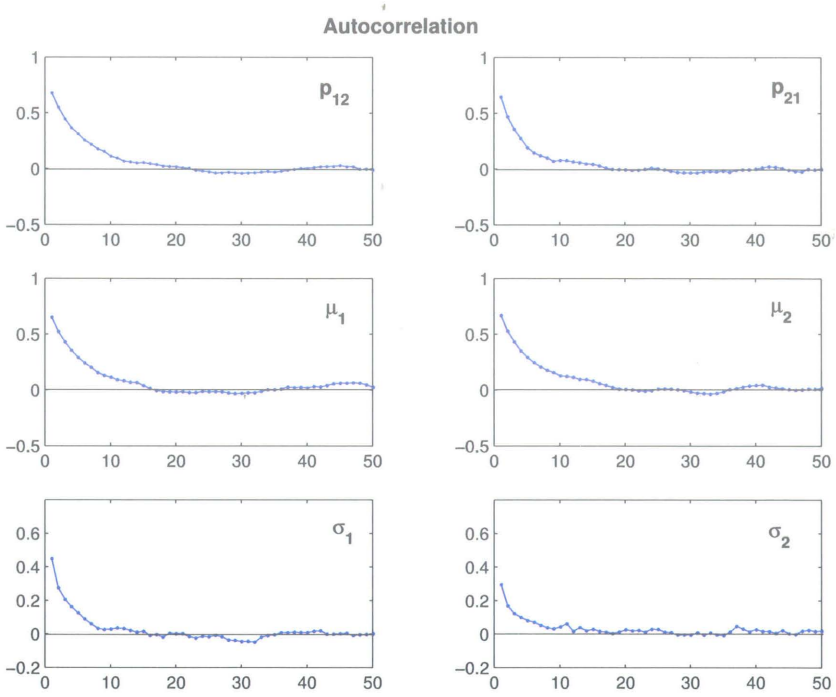


Figure 8.9: Correlograms for the parameters of the 2-state MS model simulated from Gibbs sampler for Aggregated Basin Flows.

up) in order to minimize the impact of the initial parameters. The convergence of the chain can be tested using different methods. However, when the MLEs are used as initial parameters, it was observed that according to the R statistic (Section 5.7), this warm-up period is not necessary because the MLEs are close to the mode of the marginal posterior distributions. Five thousand simulated model parameters and their marginal posterior distributions for Aggregated Basin Flows are given in Figures 8.7 and 8.8, respectively.

As mentioned in Section 4.3.3, when parameter uncertainty is considered, each streamflow sequence is simulated using a different set of model parameters. In MCMC simulations, a significant autocorrelation can be observed in the chain. To avoid this correlation, the chain can be thinned by taking every n'th valued in the chain. This

is a common practice in MCMC simulations. In order to decide the thinning interval, the correlograms of the 5000 simulated parameters for Aggregated Basin Flows are examined. As seen from Figure 8.9, the autocorrelation is close to zero after lag 20 for all parameters. After examining several chains, it was decided to use every 30th parameter sets for the simulation of synthetic flows.

8.7 Application of the 2-State Multi-site MS Model to Manitoba Hydro Data

In Framework-2, annual flows at the five hydrological components of Manitoba Hydro's system are generated using a multi-site stochastic time series models. The multi-site extension of the MS model was described in Chapter 6 for the general M-state case. In this section, the application of the 2-state multi-site MS model and the corresponding parameter and data uncertainty procedures are presented for Manitoba Hydro's system.

The basic assumption of the multi-site model is that a single state variable describes the flows at all sites. In other words, the model does not allow one site to be in dry state and another site in wet state in the same year. If the watersheds are geographically close, this should not be an overly restrictive assumption since it is reasonable to assume that the climate regime is regional in scope. For watersheds that do not belong to the same climate region, this could be a somewhat restrictive assumption. However, if individual state variables are applied to individual sites, it becomes impossible to

preserve the cross-correlation between sites.

The multi-site application of the 2-state MS model for Manitoba Hydro's system considers 5-dimensional vectors of observations, $\boldsymbol{q}_t = \left(q_t^{(1)}, \ldots, q_t^{(5)}\right)'$, for $t = 1, \ldots, T$. In order to model the correlation between the five sites, it is assumed that the five annual flows are generated from a multivariate normal distribution. After estimating the model parameter set $\boldsymbol{\theta} = \{\varrho_1, p_{12}, p_{21}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}$, the state sequence can be simulated as

$$s_t | s_{t-1} \sim \text{Markovian}(\boldsymbol{\varrho}, \mathbf{P}) \tag{8.24}$$

Given the state sequence at time t, the streamflow variable can be simulated from

$$\boldsymbol{q}_t = \begin{cases} N\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\right) & \text{if} \quad s_t = 1 \\ N\left(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\right) & \text{if} \quad s_t = 2 \end{cases} \tag{8.25}$$

The parameter estimation for the multi-site MS model was presented in Section 6.3. Table 8.9 gives the state means and standard deviations of the 2-state multi-site MS model for the five sites along with first, second, and third order moments for the 1912-98 data set.

Figure 8.10 shows the associated marginal distribution for the Saskatchewan River using the multi-site model. The fit is good, particularly in the left tail which is of most interest in drought studies. Generally, it is possible to obtain a good distribution fit with the multi-site MS model. The fit tends to be better if sites are considered individually. However, even with the 5-site model employed here it is possible to

Table 8.9: ML-parameters and moments of the multi-site MS model of Manitoba Hydro annual runoff (cfs) with transition probabilities $p_{12} = 0.0506$ and $p_{21} = 0.0820$.

1912-98 Data Set:

	Churchill River	Saskatchewan River	Lake Wpg. PIAO	Winnipeg River	Aggregated Local Flows
μ_1	34130	17119	16720	29347	12761
μ_2	37218	23588	20589	34345	13588
σ_1	8159	3556	9548	8398	3466
σ_2	3464	6246	9422	8651	3028
μ	35309 (35554)	19588 (20101)	18197 (18504)	31255 (31651)	13077 (13142)
σ	6927 (6555)	5709 (5968)	9684 (9740)	8836 (8924)	3330 (3316)
c_s	-0.3540 (-0.0935)	0.7305 (0.8058)	-0.0037 (0.4110)	0.0322 (0.4122)	-0.0443 (0.2891)
ρ_1	0.0407 (0.7456)	0.2629 (0.4195)	0.0327 (0.4089)	0.0655 (0.3367)	0.0126 (0.3497)
ρ_2	0.0353 (0.5269)	0.2280 (0.1251)	0.0283 (0.1472)	0.0568 (-0.0043)	0.0110 (0.0322)

The observed values are given in parentheses

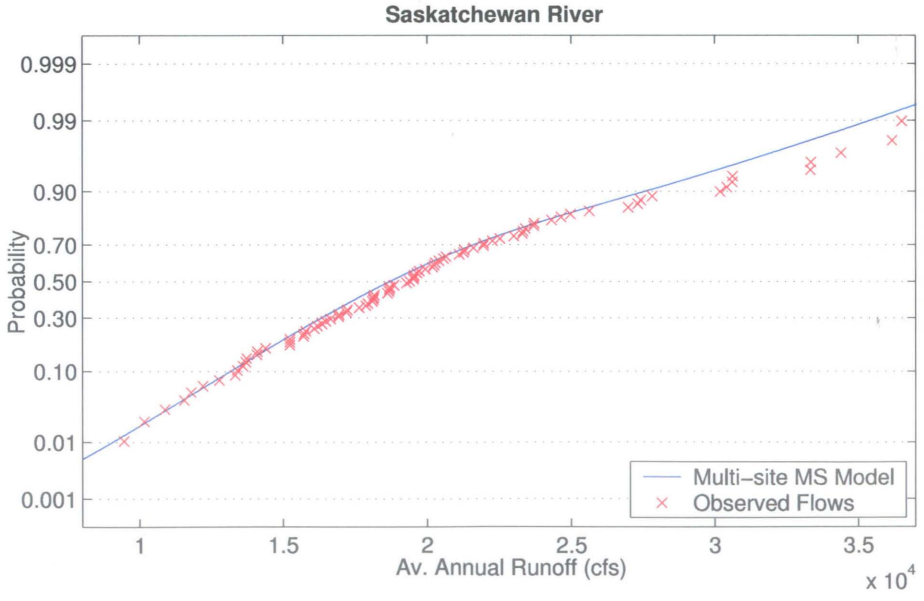


Figure 8.10: Saskatchewan River. Normal probability plot of observed annual flows and fitted multi-site MS model.

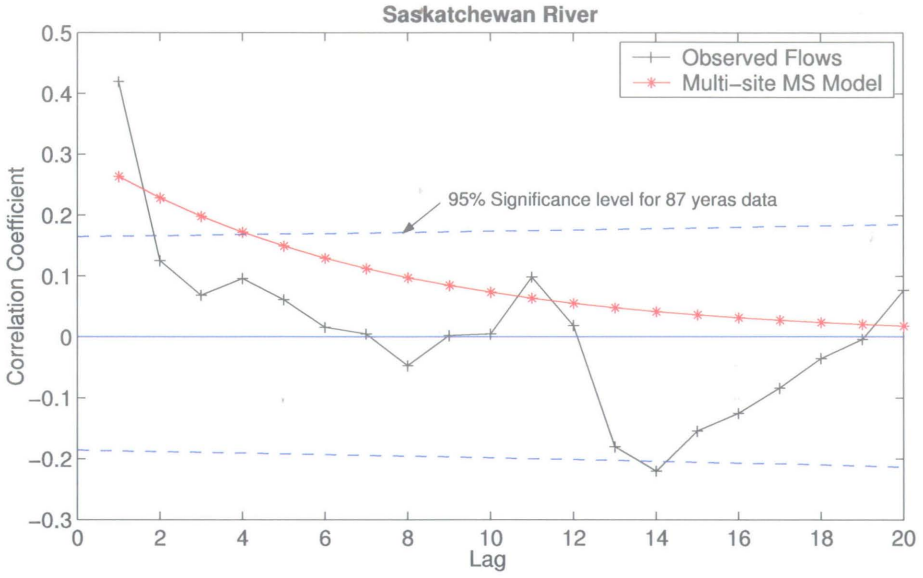


Figure 8.11: Saskatchewan River. Observed and multi-site MS model autocorrelation.

obtain good fits. The probability plots for all five sites of the multi-site MS model are given in Appendix H.

Serial correlation plays a key role in determining drought frequencies and is one of the main properties that a model should preserve. The autocorrelation function for the Saskatchewan River mean annual flow is shown in Figure 8.11 along with the observed autocorrelation function. Except for lag 1, the modeled autocorrelation somewhat overestimates the observed autocorrelation. This again appears to be a consequence of the simultaneous fitting of five sites, because if only the Saskatchewan River is considered, the fit is much better. It is worth noting the substantial autocorrelation of the MS model even at very high lags. This could have considerable impact on the characteristics of droughts. The long-term autocorrelation is partly due to the considerable persistence of the Markov Chain and partly due to the difference in mean levels. For some of the other sites where the difference in the estimated mean levels

Table 8.10: Modeled and observed cross-correlation of Manitoba Hydro annual runoff (cfs). Parameters are estimated using the multi-site MS model. The observed cross-correlations are given in parenthesis.

1912-98 Data Set:

	Churchill River	Saskatchewan River	Lake Wpg. PIAO	Winnipeg River	Aggregated Local Flows
Churchill River	1	0.2239 (0.2290)	0.1866 (0.1829)	-0.0592 (-0.0217)	0.1714 (0.1586)
Saskatchewan River		1	0.3315 (0.2948)	0.3505 (0.3523)	0.1820 (0.1597)
Lake Wpg. PIAO			1	0.5376 (0.5394)	0.6981 (0.6962)
Winnipeg River				1	0.4604 (0.4699)
Aggregated Local Flows					1

The observed values are given in parentheses

is modest, the autocorrelation function tends to die out much faster. As seen from Table 8.9, the simultaneous fitting of the five sites considerably underestimates the observed lag-1 and lag-2 autocorrelations for some of the sites. The comparison of the autocorrelation functions for the single-site MS model, the multi-site MS model, and the MS disaggregation model for the five sites are given in Appendix I.

It is of practical interest to be able to calculate the overall covariance of $\boldsymbol{q}_t$, given information about the covariance matrix for each state. The calculation of the covariance matrix is described in Section 6.4. The crosscovariance matrix of several sites can be constructed by calculating the covariance between all pairs of sites and can be standardized to give the crosscorrelation matrix. The crosscorrelations of the multi-site MS model are given in Table 8.10. It can be seen that the crosscorrelations are very well preserved by the model. The preservation of crosscorrelation is an important model property in drought studies.

8.7.1 Incorporating Parameter and Data Uncertainty

For the purpose of incorporating parameter and data uncertainty in the drought frequency analysis of Manitoba Hydro's system, the Gibbs sampler is applied to the 2-state, multi-site MS model. The methodology for the M-state MS model was described in Section 6.5.

For the 2-state 5-site MS model, the k'th iteration of the Gibbs sampler proceeds as follows:

$$\mathbf{S}_T^{(k)} \sim p\left[\mathbf{S}_T \mid \boldsymbol{\mu}_1^{(k-1)}, \boldsymbol{\mu}_2^{(k-1)}, \boldsymbol{\Sigma}_1^{(k-1)}, \boldsymbol{\Sigma}_2^{(k-1)}, \mathbf{P}^{(k-1)}, \mathbf{Q}_T^{(k-1)}\right] \tag{8.26}$$

$$p_{12}^{(k)}|\mathbf{S}_T^{(k)} \sim Beta\left[1+n_{12}^{(k)}, 1+n_1^{(k)}-n_{12}^{(k)}\right] \tag{8.27}$$

$$p_{21}^{(k)}|\mathbf{S}_T^{(k)} \sim Beta\left[1+n_{21}^{(k)}, 1+n_2^{(k)}-n_{21}^{(k)}\right] \tag{8.28}$$

$$\boldsymbol{\Sigma}_1^{(k)}|\mathbf{Q}_1 \sim \text{Inv-Whishart}\left[\nu_{n_1}^{(k)}, (\boldsymbol{\Lambda}_{n_1}^{-1})^{(k)}\right] \tag{8.29}$$

$$\boldsymbol{\mu}_1^{(k)}|\boldsymbol{\Sigma}_1^{(k)}, \mathbf{Q}_1 \sim N\left[\boldsymbol{\mu}_{n_1}^{(k)}, \boldsymbol{\Sigma}_1^{(k)}/\kappa_{n_1}^{(k)}\right] \tag{8.30}$$

$$\boldsymbol{\Sigma}_2^{(k)}|\mathbf{Q}_2 \sim \text{Inv-Whishart}\left[\nu_{n_2}^{(k)}, (\boldsymbol{\Lambda}_{n_2}^{-1})^{(k)}\right] \tag{8.31}$$

$$\boldsymbol{\mu}_1^{(k)}|\boldsymbol{\Sigma}_2^{(k)}, \mathbf{Q}_2 \sim N\left[\boldsymbol{\mu}_{n_2}^{(k)}, \boldsymbol{\Sigma}_2^{(k)}/\kappa_{n_2}^{(k)}\right] \tag{8.32}$$

where $n_{ij}^{(k)}$, $i,j = 1,2$, is the number of times that $s_t = i$ and $s_{t+1} = j$, and n_i is the number of times that $s_t = i$ in the state sequence $\mathbf{S}_T^{(k)}$. As in the single-site case, a state sequence is first simulated using Chip's method. The posterior parameters of the multivariate Normal and Inv-Whishart distributions, $\kappa_{n_i}^{(k)}, \nu_{n_i}^{(k)}, \boldsymbol{\mu}_{n_i}^{(k)}, \boldsymbol{\Lambda}_{n_i}^{(k)}$, are given in (6.45).

In the application of the Gibbs sampler for a multi-site MS model, it was found

that if the prior parameters for the state distributions are not carefully chosen, the Gibbs sampler may not achieve convergence. Therefore, the selection of suitable prior parameters was a challenge in the application.

As in the single-site case, the state mean vector and covariance matrix are simulated from their prior distributions when the number of the data in a particular state in the simulated state sequence $\boldsymbol{S}_T$ is less than or equal to the number of the sites. However, in the multi-site application, sampling from prior distributions rarely occurs.

For the prior parameter, *Thyer and Kuczera* [2003a] suggested to use the data-based empirical Bayes estimates defined as

$$\boldsymbol{\mu}_0 = \bar{\boldsymbol{y}} \tag{8.33}$$

$$\boldsymbol{\Lambda}_0 = \frac{\boldsymbol{S}_{obs}^{-1}}{\left(\nu_0 - \frac{N+1}{2}\right)} \tag{8.34}$$

where $\bar{\boldsymbol{y}}$ and $\boldsymbol{S}_{obs}$ are the empirical mean vector and covariance matrix of the entire observed data set $\boldsymbol{Q}_T$. *Thyer and Kuczera* [2003a] used the same prior parameters for both states. In this project, the above hyperparameter values were initially employed. However, using these prior parameters it was found that certain posterior distributions became bimodal. Therefore, the multivariate extension of the single-site prior parameters given in (8.17) were used for each state. The multivariate parameters are given by

$$\kappa_0 = 1 \qquad \boldsymbol{\mu}_{0_i} = \boldsymbol{\mu}_{i_{MLE}}$$
$$\nu_0 = 7 \qquad \boldsymbol{\Lambda}_{0_i} = \frac{\boldsymbol{\Sigma}^{-1}_{i_{MLE}}}{\left(\nu_0 - \frac{N+1}{2}\right)} \tag{8.35}$$

where $\boldsymbol{\mu}_{i_{MLE}}$ and $\boldsymbol{\Sigma}_{i_{MLE}}$ are the maximum likelihood estimates of the state i mean vector and covariance matrix, respectively. The constants κ_0 and ν_0 were selected low enough to obtain a diffuse proper prior distribution.

To consider data uncertainty in the Gibbs sampler, missing data values are simulated from a multivariate normal distribution and used to augment the observed data to obtain the complete data series, $\boldsymbol{Q}_T^{(k)} = \left\{\boldsymbol{Q}_{mis}^{(k)}, \boldsymbol{Q}_{obs}\right\}$. The updated complete data series is then used to obtain data series from state 1 and 2, $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, that are used in the simulation of the state mean vector and covariance matrix. It was found that in the Gibbs sampler, the number of data in a particular state usually was greater than the number of sites which means that the model parameters can be simulated from posterior distribution in the large majority of iterations.

As in the single-site MS model, a thinning interval of 30 was used. Examination of the simulated parameter chains revealed that a warm up period is not necessary when the MLEs are used as initial parameter values in the Gibbs sampler.

The simulated posterior distributions are given in Appendix J. There is a good agreement between the maximum likelihood estimates and posterior distributions of the model parameters.

8.8 Application of the 2-State MS Disaggregation Model

In the application of the spatial MSD model, the annual Aggregated Basin Flows q_t generated using the single-site 2-state MS model (see Section 8.6) are disaggregated into annual flows in five sites $\boldsymbol{q}_t = (q_t^{(1)}, \ldots, q_t^{(5)})'$. It is assumed that lower level variables are described by two climate regimes and have the same state probability sequence as annual flows. This means that every site is in the same climate state in a given year. In the disaggregation, the climate state s_t is known from the single-site 2-state MS model. In the MSD model, the zero mean annual Aggregated Basin Flows x_t may be disaggregated into zero mean annual flows $\boldsymbol{y}_t = (y_t^{(1)}, \ldots, y_t^{(5)})'$ at the five sites as

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{A}_1 x_t + \boldsymbol{B}_1 \epsilon_t & \text{if} \quad s_t = 1 \\ \boldsymbol{A}_2 x_t + \boldsymbol{B}_2 \epsilon_t & \text{if} \quad s_t = 2 \end{cases} \tag{8.36}$$

Despite the loss of some flexibility, it was found that the marginal distributions are well preserved for all five sites (see Appendix K). However, because of the constraint that all sites must be at the same state in the same year, the mean levels are closer to each other at all sites compared to the single-site models (see Appendices F and L). Estimated state means and standard deviations and overall means, standard deviations, and coefficients of skewness are given in Table 8.11. Although there are some minor discrepancies between observed and modeled moments, in most cases, the MSD model preserves moments better than the multi-site and single-site MS model (Table 8.8 and

Table 8.11: ML-parameters and moments of the MS spatial disaggregation model of Manitoba Hydro annual runoff (cfs) estimated using Aggregated Basin Flows' state sequences $\Pr\{s_t = i|\boldsymbol{Q_T}\}$ for $i = 1, 2$ and transition probabilities $p_{12} = 0.1160$ and $p_{21} = 0.1628$.

1912-98 Data Set:

	Churchill River	Saskatchewan River	Lake Wpg. PIAO	Winnipeg River	Aggregated Local Flows	Aggregated Basin Flows
μ_1	34070	18458	14158	28478	11920	107090
μ_2	37904	22705	25391	36679	15078	137760
σ_1	7046	5282	7063	7359	2750	17027
σ_2	5046	5977	9273	8742	3163	19632
μ	35666 (35554)	20225 (20101)	18832 (18504)	31891 (31651)	13234 (13142)	119850 (118950)
σ	6569 (6555)	5961 (5968)	9776 (9740)	8931 (8924)	3317 (3316)	23627 (23589)
c_s	-0.2303 (-0.0935)	0.1290 (0.8058)	0.3784 (0.4110)	0.2183 (0.4122)	0.1892 (0.2891)	0.2510 (0.3076)
ρ_1	0.0597 (0.7456)	0.0889 (0.4195)	0.2314 (0.4089)	0.1478 (0.3367)	0.1588 (0.3497)	0.2953 (0.4712)
ρ_2	0.0430 (0.5269)	0.0641 (0.1251)	0.1669 (0.1472)	0.1066 (-0.0043)	0.1146 (0.0322)	0.2130 (0.1431)

The observed values are given in parentheses

8.9). However, it is worth noting that the problems with the fitting of autocorrelation functions are not overcome with the MSD model. As in the multi-site MS model, for some of the sites, the MSD model considerably underestimates the observed lag-1 and lag-2 autocorrelations. The autocorrelation functions for the MSD model for the five sites are given in Appendix I.

Using information about the covariance matrix for each state, the crosscorrelations produced by the MSD model are given in Table 8.12. Although the multi-site MS model preserves crosscorrelation quite well, the comparison of Tables 8.10 and 8.12 reveals that the MSD model is modeling crosscorrelation even better than the multi-site MS model.

Table 8.12: Modeled and observed cross-correlation of Manitoba Hydro annual runoff (cfs). Parameters are estimated by disaggregating Aggregated Basin Flows using the MS disaggregation model. The observed cross-correlations are given in parentheses.

1912-98 Data Set:

	Churchill River	Saskatchewan River	Lake Wpg. PIAO	Winnipeg River	Aggregated Local Flows
Churchill River	1	0.2278 (0.2290)	0.1829 (0.1829)	-0.0126 (-0.0217)	0.1587 (0.1586)
Saskatchewan River		1	0.2947 (0.2948)	0.3571 (0.3523)	0.1638 (0.1597)
Lake Wpg. PIAO			1	0.5407 (0.5394)	0.6955 (0.6962)
Winnipeg River				1	0.4744 (0.4699)
Aggregated Local Flows					1

8.9 Verification of Annual Models

In stochastic modeling applications, it is important to verify that the computer implementation of the model is done correctly. The verification of a stochastic streamflow model typically involves a demonstration that statistics that should be preserved by the model are indeed reproduced. In the present application, 100,000 years of flows were generated using four modeling frameworks (MF-1.1, MF-1.3, MF-2.1, and MF-2.3 in Tables 8.3 and 8.4) and generated statistics were compared with corresponding observed statistics for all sites. Only models without parameter uncertainty were considered for validation. Since normally transformed data are used in the traditional AR(1) model, the means, the standard deviations, the lag-1 auto-correlations, and the lag-0 cross-correlations of observed and generated flows were compared before and after inverse transformation. As seen from Sections M.1 and M.3 in Appendix M, there is a good agreement between observed and simulated annual flows in both frameworks

involving AR-models. For the verification of the MS-type models (Model MF-1.3 and MF-2.3), key statistics are compared with corresponding theoretical statistics given in Tables 8.9, 8.10, 8.11, and 8.12. As in the case of the traditional AR(1) models, all selected statistics are well preserved.

In the stochastic modeling, preservation of the marginal distributions is of particular interest as well. When traditional disaggregation models are used, the marginal distributions may be distorted because of the application of the adjustment procedure. The analysis of simulated annual flows reveals that the marginal distributions are well preserved in all modeling frameworks. However, it is worth mentioning that the spatial disaggregation employed in the model MF-1.1 introduces a minor distortion in the upper tail of the annual Churchill River, Winnipeg River, and Local Flows which may not be critical for the drought analysis.

One of the general criteria for the validation of models with parameter uncertainty is to check if the estimated model parameters are located close to the mode of the corresponding posterior distributions. The posterior distributions of the single-site MS model parameters that is used in model MF-1.4 for the Aggregated Basin Flows are given in Figure 8.8. The posterior distributions of the multi-site MS model that are used in model MF-2.4 can be found in Appendix J. In all posterior distributions, maximum likelihood estimates are close to the mode of the corresponding posterior distribution. This indicates that the models have been correctly implemented.

8.10 Energy Drought Frequency Analysis for Manitoba Hydro's System

The drought frequency analysis can be performed using any of the three drought variables discussed in Chapter 3, namely duration, severity, and magnitude. However, the magnitude of a drought event is a poor indicator of a drought for a hydropower system because two drought events could have the same magnitude but their impact could be very different in terms of power production capabilities. To describe the impact of drought events, the duration and severity are used in the frequency analysis. After the generation of synthetic monthly flows for Manitoba Hydro's system using the nine stochastic modeling frameworks, drought severities and associated durations were extracted from each synthetic series according to the definition of drought for Manitoba Hydro (see Section 3.5).

In the design of engineering structures, the return period of extreme events such as floods and droughts has been widely used. The return period can be defined in different ways. In this study, the return period is defined as the average elapsed time (years) between the occurrence of specified drought events. In the case of annual floods, the maximum flood in any given year can be assumed to be independent and identically distributed with a known exceedance probability p_{exc}, and the return period T can be calculated in a straightforward manner as $T = 1/p_{exc}$. However, since droughts can span several years, the return period of a given drought event cannot be computed in the same way as for floods. In the case of droughts, the return period can be obtained

as follows [*Sadeghipour and Dracup*, 1985; *DeWit*, 1995]

$$T = \frac{\tau}{p_{exc}} \tag{8.37}$$

where τ is the average cycle length of drought events which can be calculated as the total number of flow years divided by the total number of drought events extracted.

The drought frequency analysis was performed using synthetic flows generated by the nine modeling frameworks. A total of 1000 sequences of 1000 years synthetic flows were generated using each modeling framework, and drought events (durations and severities) were extracted from each sequence according to the definition of drought (see Chapter 3). The extracted drought events obtained from the 1000 sequences were then combined and sorted from the highest (most severe) event to the lowest (least severe) event. The exceedance probabilities were calculated using the Weibull plotting position formula, $p_{exc} = m/(n+1)$, where m is the order number of the event and n is the total number of events. Linear interpolation is employed for the calculation of the exceedance probability of a given drought severity. For each modeling framework, the average cycle length τ in years is calculated as the total number of years (1,000,000) divided by the total number of drought events obtained.

8.10.1 Frequency Analysis Results and Discussion

The result of the stochastic drought analysis is presented in the form of frequency curves. The simulated energy drought frequency curves based on the 1912-98 data set

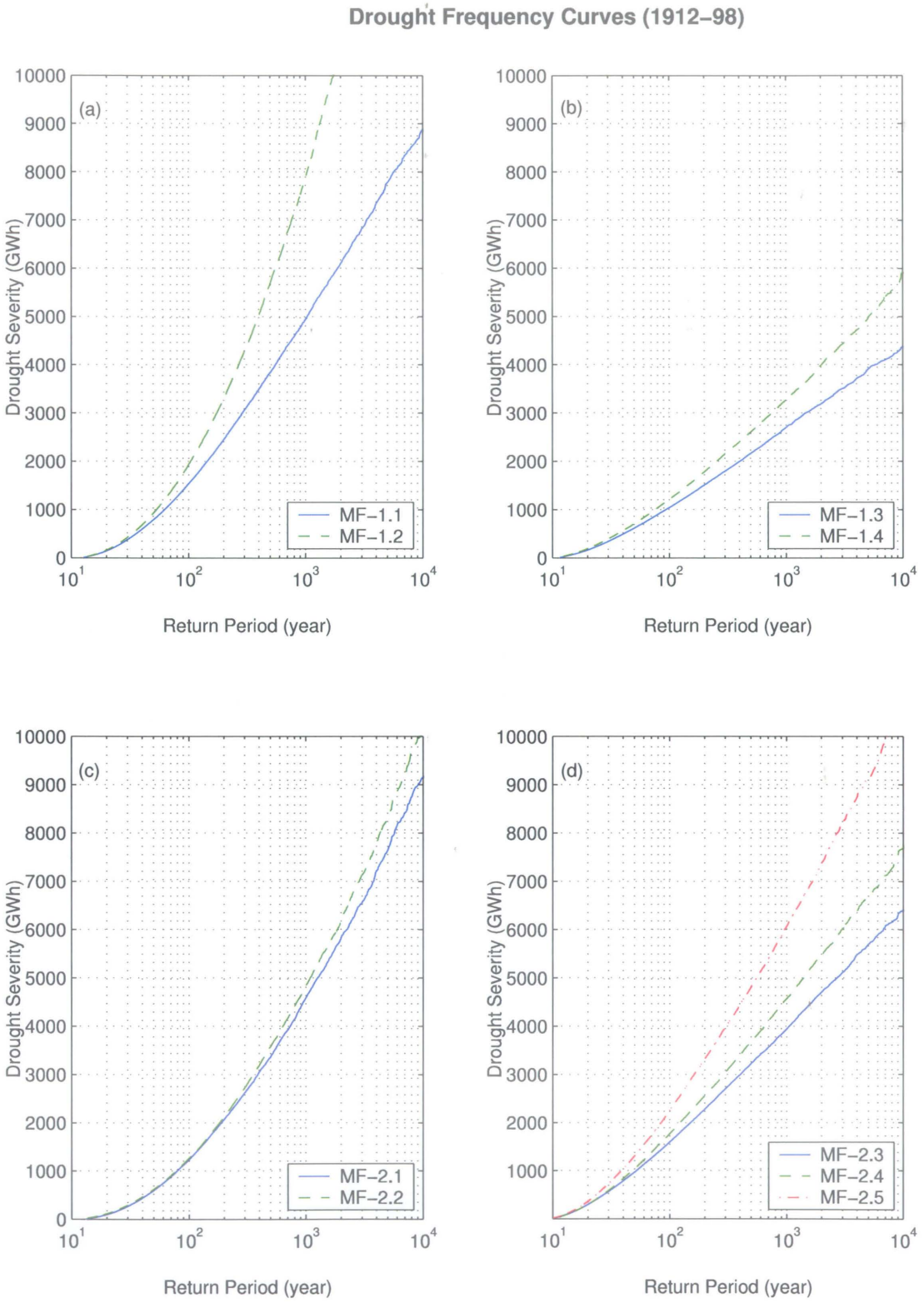


Figure 8.12: Drought frequency curves obtained from all modeling frameworks (1912-98).

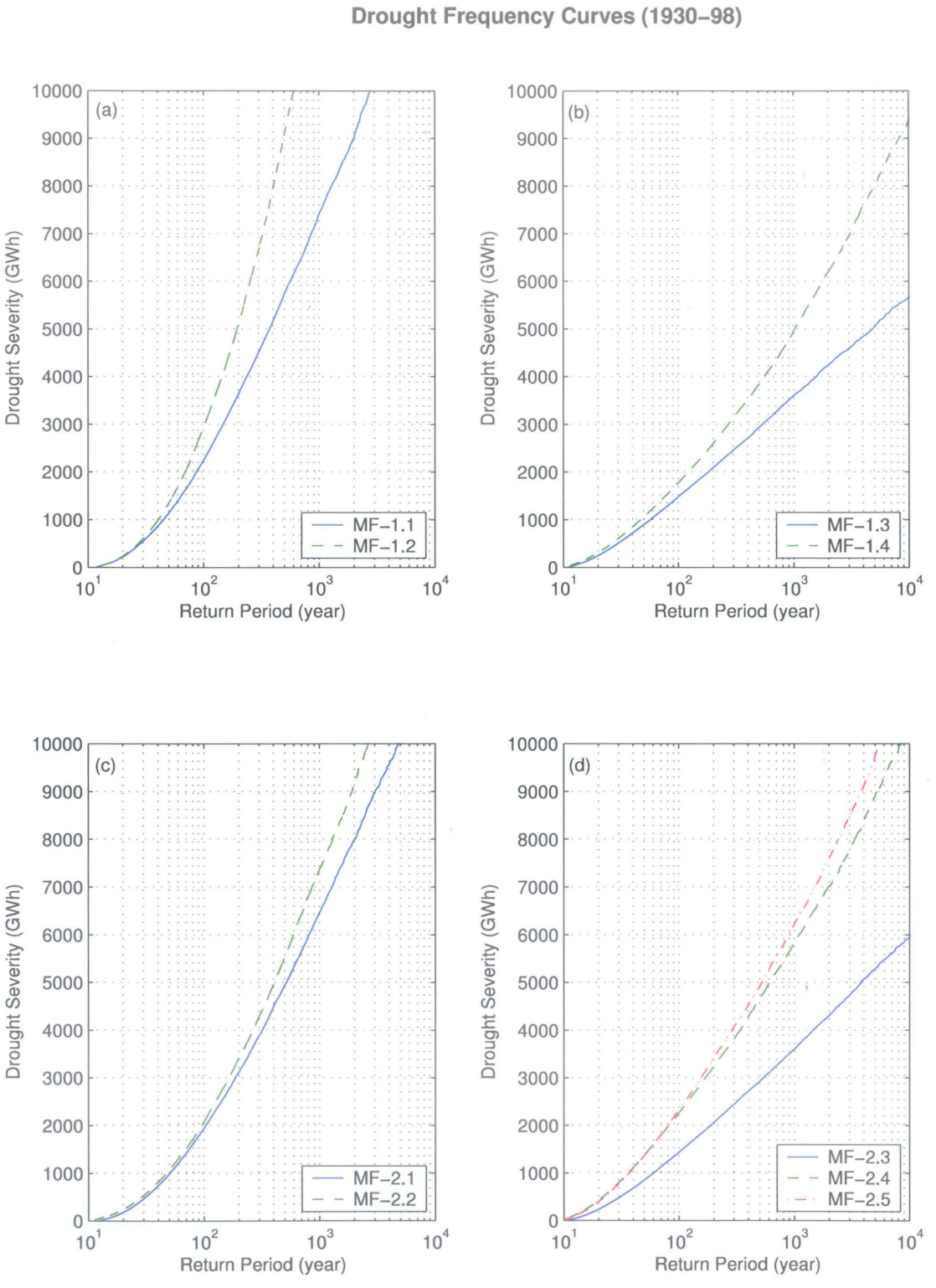


Figure 8.13: Drought frequency curves obtained from all modeling frameworks (1930-98).

Table 8.13: Statistics of calculated return periods of given drought severities using model MF-1.1. The statistics are base on hundred repeated analysis under identical conditions.

	Drought Severity (GWh)			
	1000	3000	5000	7000
Mean	62.8	296.4	1029.8	3218.7
Standard Deviation	0.53	4.6	29.7	175.6
Median	63	297	1031	3213.5
Maximum	64	305	1104	3739
Minimum	62	286	955	2842

are given in Figure 8.12. It should be noted that minor variations in the estimated frequency curves are observed from one simulation to another. In order to quantify this variation, 100 frequency curves were simulated using the MF-1.1 model (1912-98 data set) and the return period of certain drought severities was calculated from each frequency curve. The statistics of the calculated return periods, given in Table 8.13, show that there is no significant variation between simulations. Although model MF-1.1 is the simplest and therefore fastest model among others, it takes approximately 10 days to simulate 100 frequency curves using a 2.8 GHz Pentium 4 processor computer. Therefore the quantification of the variation in frequency curves is not done for the other models. It is assumed that the variation obtained from model MF-1.1 is representative of the variation for all models.

Figure 8.12 illustrates that in all cases, the inclusion of parameter uncertainty in the annual model yields smaller return period for a given drought severity than when using the corresponding annual model without parameter uncertainty. This is to be expected. The effect of including parameter uncertainty is relatively more significant when short records are used. Although the 1912-98 data set is relatively long, Figure

8.12 shows that when extreme drought events with return periods in excess of 100 years are considered, there is a significant impact of parameter uncertainty.

In addition to parameter uncertainty, Figure 8.12d illustrates the consideration of missing data uncertainty in the multi-site MS model. It can be seen from this figure that the consideration of missing data uncertainty significantly affects the results of the frequency analysis. For example, in Figure 8.12d, the return period of a drought event with a severity of 4000 GWh is approximately 600 years when model MF-2.4 is used. This model considers only parameter uncertainty in the multi-site MS model. However, the return period of the same event is 300 years according to model MF-2.5 which considers both parameter and data uncertainty in the same model.

The same drought frequency analysis was repeated using historical data from the period 1930-98. The frequency curves for this data set are given in Figure 8.13. The comparison of frequency curves for the periods 1912-98 and 1930-98 reveals that the return period of a given drought severity decreases when the 1930-98 data set is used. Since the percentage of missing data is smaller for the 1930-98 data set (compared to 1912-98), the impact of missing data uncertainty on the frequency analysis is also less in the 1930-98 data set (Figure 8.13d).

The frequency curves for drought severity conditional upon drought duration are analyzed for all modeling frameworks and data sets. The conditional frequency curves that are obtained using model MF-2.5 are given in Figure 8.14a. For example, the return period of a 2000 GWh drought severity with duration less than or equal to 2 years can be obtained from the figure as approximately 300 years. The conditional

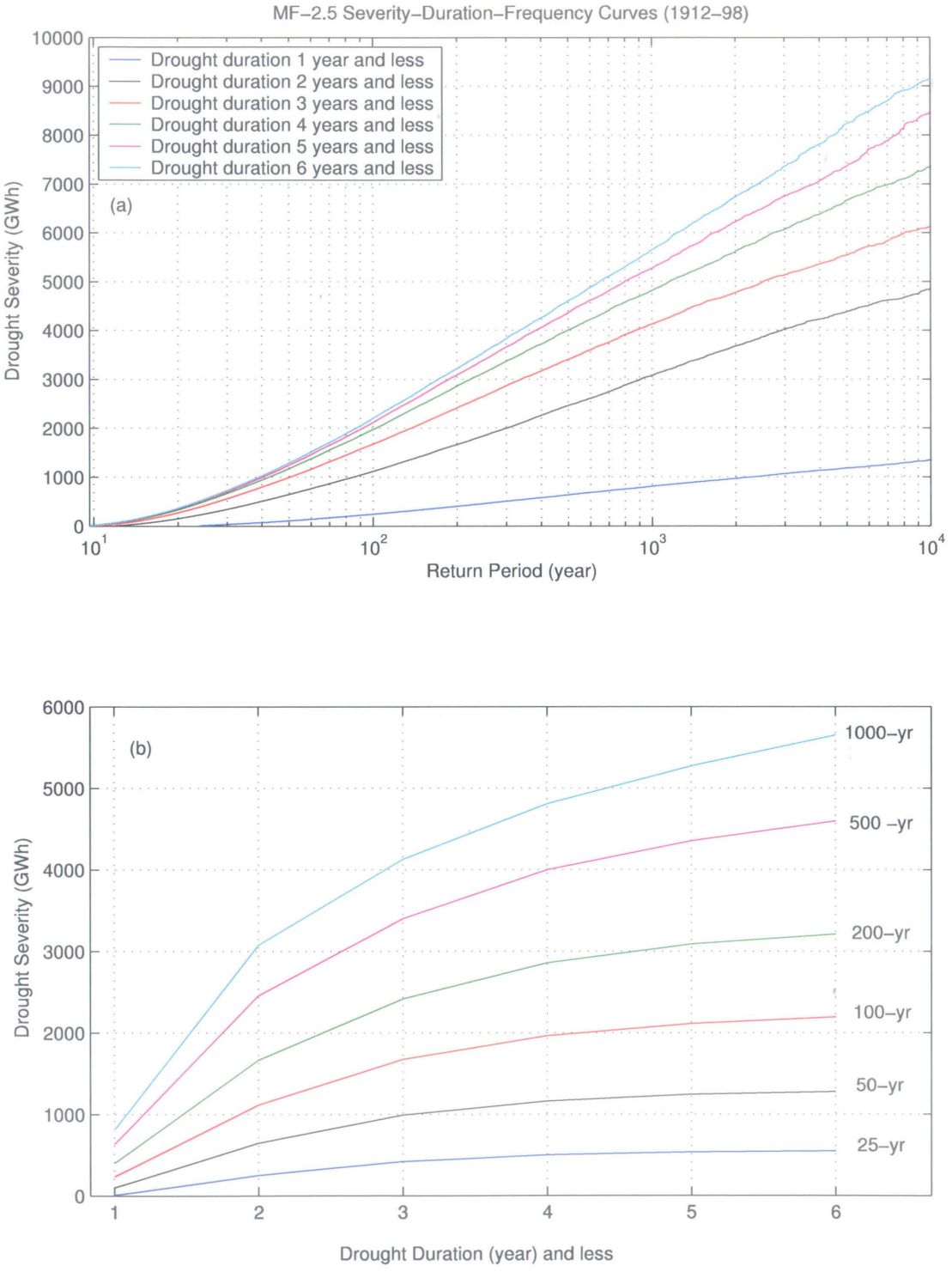


Figure 8.14: Conditional drought frequency curves obtained from MF-2.5 (1912-98).

Table 8.14: Return Period (T) of critical drought (3309 GWh) using all modeling frameworks and data sets.

	Framework-1				Framework-2				
	MF-1.1	MF-1.2	MF-1.3	MF-1.4	MF-2.1	MF-2.2	MF-2.3	MF-2.4	MF-2.5
$T_{cr_{1912-98}}$	360	200	2380	1040	480	420	540	370	200
$T_{cr_{1930-98}}$	170	120	740	340	230	190	430	210	190

frequency curves can also be plotted as Severity-Duration-Frequency curves as in Figure 8.14b. The conditional frequency curves for other modeling frameworks are given in Appendix N for both data sets.

Manitoba Hydro has an interest in determining the frequency of the critical drought of 3309 GWh (see Table 3.3). The return period of the critical drought, calculated using each modeling framework and each data set, is given in Table 8.14. Several observations can be made. Models MF-1.1 and MF-1.2, which differ only in the consideration of parameter uncertainty in the single-site AR(1) model, suggest that the consideration of parameter uncertainty in the modeling of Aggregated Basin Flows has significant impact on the estimation of the return period of the critical drought. When the uncertainty in the parameters of the single-site AR(1) model is considered, the return period changes from 360 to 200 years for the 1912-98 data set, and from 170 to 120 years for the 1930-98 data set. In the MF-2.1 and MF-2.2 models, which differ only in the consideration of parameter uncertainty in the multi-site AR(1) model, the impact of parameter uncertainty is relatively small.

In the application of the MS model, there is a significant difference between the results of the two frameworks. Model MF-1.3 generates annual Aggregated Basin Flows

using a single-site MS model and disaggregates them into annual flows at five sites. Model MF-2.3 generates annual flows at five sites using the multi-site MS model. As seen from Tables 8.9 and 8.11, there are significant differences between model parameters in the two models MF-1.3 and MF-2.3. Although consideration of parameter uncertainty significantly reduces the calculated return period in model MF-1.4, it is still not a realistic value in comparison with the other modeling frameworks. In Framework-1, the modeling of Aggregated Basin Flows influences the results of the frequency analysis. The difference between models MF-1.3 and MF-2.3 may be explained by their persistence structure that can be obtained using the estimated transition probabilities. In the MF-1.3 model, the probability of moving from state 1 (dry) to state 2 (wet) is estimated as $p_{12} = 0.12$, and moving from state 2 to state 1 is estimated as $p_{21} = 0.16$. These estimates suggest that the dry regime has an expected duration of around eight years and the wet regime has an expected duration of around six years. In the MF-2.3 model, the persistence structure is stronger with expected durations of around 20 years for the dry regime and 13 years for the wet regime.

In order to identify the source of the difference in the frequency analysis results between the AR(1) model and the MS model in Framework-1, the single-site MS model and the AR(1) model for Aggregated Basin Flows is further investigated by a drought frequency of annual streamflow (as opposed to energy flows). In streamflow analysis, a multi-year drought may be defined as a period of consecutive years during which the annual streamflow is continuously below the long-term mean annual runoff. The period of time for which the annual flow is below the long-term mean annual runoff

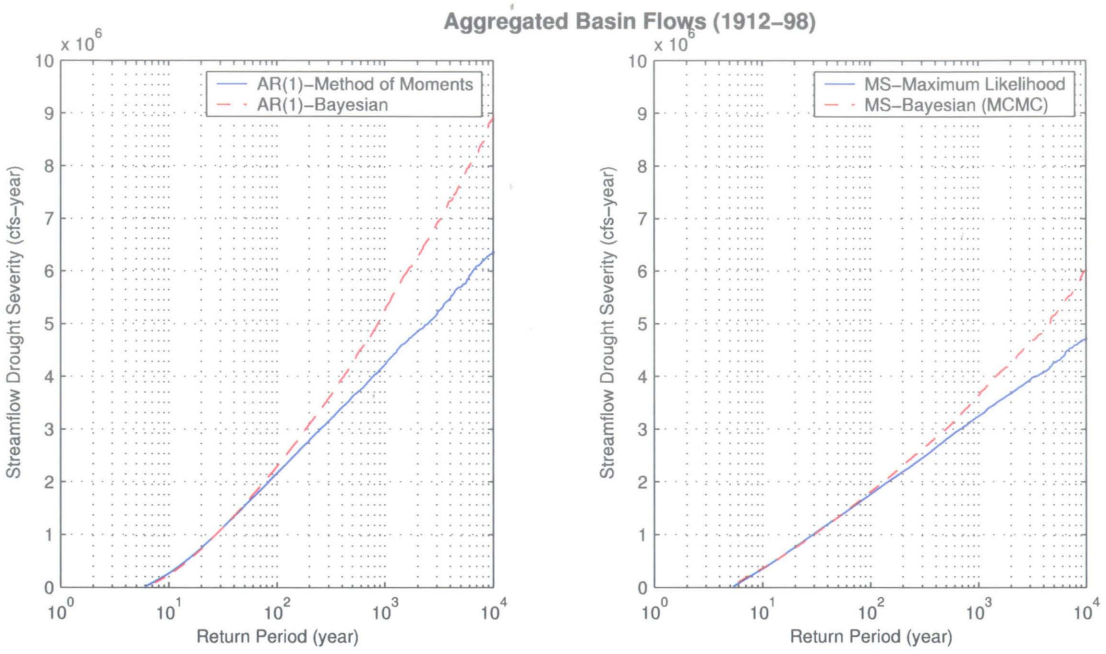


Figure 8.15: Aggregated Basin Flows (1912-98). Streamflow drought frequency curves obtained using the Modeling Framework-1.

is defined as the drought duration and the cumulative deficit of streamflow during a drought is defined as the drought severity.

According to above definition of streamflow drought, 15 drought events were extracted from the 1912-98 record of Aggregated Basin Flows (see Figure 5.2). The critical drought is observed from 1936 to 1944 with a severity of 2,581,000 cfs-years. This critical streamflow drought corresponds to the critical energy drought for Manitoba Hydro's system observed between 1938 and 1942 (see Table 3.3). The frequency curves obtained using 500 sequences of 1000-yr synthetic flows generated using the AR(1) and the MS models with and without parameter uncertainty are given in Figure 8.15. Considering parameter uncertainty, Figure 8.15 suggests that the return period of the critical drought is approximately 130 years with the AR(1) model and

approximately 290 years with the MS model. The difference in results increases with increasing drought severities. For example, the return period of a severity of 4,000,000 cfs-years is approximately 400 years with the AR(1) model and approximately 1300 years with the MS model. In this particular application, although the MS model preserves all important statistics well, it generates less severe extended low flows than the AR(1) model for Aggregated Basin Flows.

In the energy drought frequency analysis, it was found that the modeling of Aggregated Basin Flows dominates the results of the drought frequency analysis in Framework-1 for both the AR(1) and the MS models. Therefore, for energy drought frequency analysis Framework-1 may not be the best choice of framework. In addition, the structure of Framework-1 does not allow one to incorporate missing data uncertainty.

Chapter 9

Conclusions

The overall objective of this project was to develop a comprehensive probabilistic framework for energy drought frequency analysis by considering model, parameter, and missing data uncertainty. In this last chapter, several conclusions of the project are presented.

The MS models are applied to Manitoba Hydro's system along with traditional models in the assessment of energy drought frequency. The importance of model, parameter, and missing data uncertainties was quantified by comparing the results of the comprehensive energy drought frequency analysis with and without considering uncertainties. Two modeling frameworks were employed in the generation of synthetic flows. In one framework, a single series of all inflow sources is modeled first and then disaggregated spatially to the five basin sites representing the major inflow components of Manitoba Hydro's system. In the second framework, the five basin sites are simulated simultaneously with a multivariate stochastic model. Although both frame-

works preserve the important statistical properties of observed flows well, the results of the drought frequency analysis differ significantly from each other. Since the modeling of Aggregated Basin Flows, which is an artificial site, dominates the results of the drought frequency analysis in Framework-1, Framework-2 may be the better choice of framework for Manitoba Hydro's system.

The impact of parameter uncertainty may in some cases be more important than the choice of model [*Stedinger and Taylor*, 1982b]. This is especially true when the observed record is short and/or the number of parameters to be estimated is high. In such circumstances, it becomes important to quantify parameter uncertainty. Although, the 1912-98 data set is relatively long, the frequency analysis shows a significant impact of parameter and missing data uncertainty, especially when considering extreme drought events which was one of the main objectives of this project.

The structure of the multi-site MS model allows one to incorporate the uncertainty associated with missing data to ensure all information in the basin is utilized. The drought frequency analysis results reveal that different conclusions are obtained when missing data uncertainty is introduced in the multi-site MS model.

From a physical point of view, the regime switching property of MS models may have some justification. A number of well documented hydrologic time series exhibit regime-like behavior that in many cases can be associated with oceanic circulation patterns.

A major componenent of the research has been to study the performance of MS models when applied to hydrological data series. The structure of the MS model is

simple and easy to understand. Low order models are also relatively parsimonious in terms of parameters. A 2-state single-site model has six independent parameters to be estimated. Model parameters can be estimated efficiently by the method of maximum likelihood.

All basic statistical properties of the model can be derived analytically from the estimated parameters. This is useful for understanding model characteristics and for quick model evaluation.

The marginal distribution of data generated from an M-state MS model is a mixture of M normal distributions. Although mixture densities are known to be quite flexible, the marginal distribution associated with MS models may give a poor fit to the observed data compared to conventional transformation such as Box-Cox. It was found that if the sole objective is to fit a mixture density to the observed data, then a good agreement can usually be achieved. However, when the estimation includes the Markov chain parameters, some of the flexibility is lost and a good fit apparently is not always possible. This may be a limitation of MS models. It should be emphasized however that in many cases the MS model provides a good representation of the marginal distribution of data.

As it is the case with most model comparisons, the relative merits of different models depend on the data used in the study. For the data used in this study, the marginal distributions were well represented by the MS model. However, the multi-site version of the MS model produced significantly weaker autocorrelation than the single-site model. A general finding is that single-site MS models often result in stronger

autocorrelations at higher lags than traditional low-order ARMA-type models. This may be important in drought studies where one would expect autocorrelation to have impact on simulated statistics and where a model with strong autocorrelation would be considered a conservative choice. In multi-site applications, the MS model should be used with caution especially when the basin is large and not homogeneous. In such cases, the preservation of the autocorrelation may be seriously compromised.

Bayesian inference was used to incorporate parameter uncertainty into the single-site and multi-site MS models. Since analytical expressions for posterior distributions are not available for the MS model, a Markov chain Monte Carlo method known as the Gibbs sampler is used to simulate model parameters from the posterior distributions. In all applications, a good agreement was observed between the posterior distribution of a given parameter and its corresponding maximum likelihood estimate.

In the MS model, the assumption that all sites are in the same state at the same time may not be justified when the basin is large and consists of nonhomogeneous sub-basins. This is indeed a questionable assumption for Manitoba Hydro's system which is large and diverse. In the single-site MS model estimation (see Table 8.8), transition probabilities suggest average duration of dry spells from 5 to 11 years for the five sites. However, wet spell durations for the Churchill River differ from the other four sites with an average of 33 years compared to 3-8 years for the remaining sites.

While there is no basis for concluding that the MS model is generally superior to the AR(1) model, there are cases where the model is quite appropriate and in fact may

have some physical justification, for example, when a particular hydrologic variable is known to be influenced by low-frequency oceanic circulations such as ENSO, PDO, and NAO.

The MS disaggregation (MSD) model proposed in Chapter 7 is a new contribution to the hydrology literature. The main advantage of the MSD model is that it preserves the additivity property without distorting the marginal distributions of lower level flows by adjustments. This is a common problem in the application of traditional disaggregation models.

There are several possible extensions to the MS model considered in this project. In cases where the autocorrelation structure of the data is not well represented by the simple MS model, one could assume that the flows within a given state spell are autocorrelated, perhaps in the form of a simple first-order autoregression. This will add considerable complexity to the model, but also improve the preservation of autocorrelation.

Although the marginal distributions associated with the MS model give good fits to Manitoba Hydro's data, it may be possible to improve the fit further by considering other distributions than the normal as components of the model. For example, it might be possible to obtain a better fit if a combinations of lognormal distributions were used. Although not a problem in this study, the simulation of negative flows could happen in other situations. In this case, truncated normal distributions could be a solution.

In large basins, the assumption of a common state sequence for all sites is too

restrictive. An alternative modeling approach could be developed for spatial variations in the state. However, this would significantly increase the number of parameters.

The author believes that this thesis has made an original contribution to the hydrology literature. The MS model that has recently gained popularity in stochastic hydrology has been investigated further by defining its theoretical properties for an arbitrary number of states, by finding solutions for the implementation of the estimation procedures, by developing the MS disaggregation model, and by applying the model to Manitoba Hydro's system in the assessment of energy drought frequency analysis along with traditional models.

References

Akaike, H., A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, *19*, 716–723, 1974.

Akıntuğ, B., Energy drought frequency analysis of Manitoba Hydro, *Tech. rep.*, Manitoba Hydro, Winnipeg, Manitoba, Canada, 2002.

Akıntuğ, B., and P. F. Rasmussen, Analysis of energy drought using a Markov-Switching model, Canadian Society for Civil Engineers (CSCE), 17th Canadian Hydrotechnical Conference, Edmonton, Alberta, Canada, 17-19 August, 2005a.

Akıntuğ, B., and P. F. Rasmussen, A Markov-Switching model for annual hydrologic time series, *Water Resources Research*, *41, W09424, doi:10.1029/2004WR003605*, 2005b.

Akıntuğ, B., and P. F. Rasmussen, Simulation of monthly flows using a Markov-Switching model, American Geophysical Union (AGU) Meeting, The 2005 Joint Assembly, New Orleans, Louisiana, USA, 23-27 May, 2005c.

Andel, J., A time series model with suddenly changing parameters, *Journal of Time Series Analysis*, *14*, 111–123, 1993.

Bates, B. C., A Markov Chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resources Engineering*, *37*, 937–947, 2001.

Bell, L. S. J., P. J. Binning, G. Kuczera, and P. M. H. Kau, Rigorous uncertainty assessment in contaminant transport inverse modelling: A case study of fluoride diffusion through clay liners, *Journal of Contaminant Hydrology*, *57*, 1–20, 2002.

Bilmes, J. A., A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, *Technical report TR-97-021*, International Computer Science Institue, University of Berkeley, CA, USA, 1998.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley and Sons, Inc., New York, USA, 1992.

Bras, R. L., and I. Rodríguez-Iturbe, *Random Functions and Hydrology*, Dover Publications, Inc., New York, 1993.

Campbell, E. P., and B. C. Bates, Regionalization of rainfall-runoff model parameters using Markov chain Monte Carlo samples, *Water Resources Research*, *37*, 731–739, 2001.

Casella, G., and E. I. George, Explaining the Gibbs sampler, *The American Statistician*, *46*, 167–174, 1992.

Chander, S., N. S. Kambo, S. K. Spolia, and A. Kumar, Analysis of surplus - and deficit - using runs, *Journal of Hydrology*, *49*, 193–208, 1981.

Chang, T. J., and J. R. Stenson, Is it realistic to refine a 100-year drought for water management, *Water Resources Bulletin*, *26*, 823–829, 1990.

Chen, M., Q. Shao, and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springler-Verlag New York Inc., USA, 2000.

Chib, S., Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics*, *75*, 79–97, 1996.

Cowles, M. K., and B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: A comparative review, *Journal of the American Statistical Association*, *91*, 883–904, 1996.

DeGroot, M. H., *Optimal Statistical Decisions*, McGraw-Hill, 1970.

Dempster, A. P., N. M. Laird, and D. B. Rubin, Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society. Series B*, *39*, 1–38, 1977.

DeWit, W., Probabilistic drought analysis, M.Sc. Thesis, University of Manitoba, Winnipeg, Manitoba, Canada, 1995.

Dracup, J. A., K. S. Lee, and E. G. Paulson, On the definition of drought, *Water Resources Research*, *16*, 297–302, 1980.

Filliben, J. J., The probability plot correlation coefficient test, *Technometrics*, *17*, 111–117, 1975.

Fortin, V., L. Perreault, and J. D. Salas, Retrospective analysis and forecasting of streamflows using a shifting level model, *Journal of Hydrology*, *296*, 135–163, 2004.

Fraley, C., and A. E. Raftery, How many clusters? Which clustering method? Answers via model based cluster analysis, *Computer Journal*, *41*, 578–588, 1998.

Gelman, A., and D. B. Rubin, Inference from iterative simulation using multiple sequence, *Statistical Science*, *7*, 457–511, 1992.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall, Suffolk, Great Britain, 1995.

Gemerman, D., *Markov Chain Monte Carlo, Stochastic Simulation for Bayesian Inference*, Chapman and Hall, Suffolk, Great Britain, 1997.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall / CRC, New York, USA, 1998.

Girling, W. C., Long term streamflow data, phase I, investigative studies, *Tech. rep.*, Manitoba Hydro, Winnipeg, Manitoba, Canada, 1988.

Girling, W. C., Long term streamflow data, phase II studies, stage I: Streamflow data extention 1968-1988 period investigative studies, *Tech. rep.*, Manitoba Hydro, Winnipeg, Manitoba, Canada, 1990.

Grygier, J. C., and J. R. Stedinger, Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resources Research*, *24*, 1574–1584, 1988.

Grygier, J. C., and J. R. Stedinger, *SPIGOT Technical Description, Version 2.6*, School of Civil and Environmental Eng., Cornell University, Ithaca, New York, 1990.

Grygier, J. C., and J. R. Stedinger, *SPIGOT User Manual, Version 2.7*, School of Civil and Environmental Eng., Cornell University, Ithaca, New York, 2001.

Hamilton, J. D., Analysis of time series subject to changes in regime, *Journal of Econometrics*, *45*, 39–70, 1990.

Hansen, B. E., The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP, *Journal of Applied Econometrics*, *7*, S61–S82, 1992.

Hirsch, R. M., A comparison of four streamflow record extension techniques, *Water Resources Research*, *18*, 1081–1088, 1982.

Hoshi, K., and S. J. Burges, Disaggregation of streamflow volumes, *Journal of Hydraulics Division*, *105*, 27–41, 1979.

Hsieh, W. W., J. Li, A. Shabbar, and S. Smith, Seasonal prediction with error estimation of Columbia River streamflow in British Columbia, *Journal of Water Resources Planning and Management*, *129*, 146–149, 2003.

Hughes, J. P., P. Guttorp, and S. P. Charles, A non-homogeneous hidden Markov model for precipitation occurrence, *Applied Statistics*, *48*, 15–30, 1999.

Jackson, B. B., Markov mixture models for drought lengths, *Water Resources Research*, *11*, 64–74, 1975.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal, Markov chain Monte Carlo in practice: A roundtable discussion, *The American Statistician*, *52*, 93–100, 1998.

Kehagias, A., A hidden Markov model segmentation procedure for hydrological and environmental time series, *Stochastic Environmental Research*, *18*, 117–130, 2004.

Kim, C., Dynamic linear models with Markov-switching, *Journal of Econometrics*, *60*, 1–22, 1994.

Koutsoyiannis, D., A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series, *Water Resources Research*, *28*, 3175–3191, 1992.

Koutsoyiannis, D., and A. Manetas, Simple disaggregation by accurate adjusting procedures, *Water Resources Research*, *32*, 2105–2117, 1996.

Kuczera, G., and E. Parent, Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *Journal of Hydrology*, *211*, 69–85, 1998.

Lane, W. L., *Applied Stochastic Techniques (LAST Computer Package); User Manual*, Division of Planning Technical Services, Bureau of Reclamation, Denver, Colorado, 1979.

Lane, W. L., Corrected parameter estimates for disaggregation schemes, International Symposium On Raninfall Runoff Modeling, Mississippi State University, USA., 1981.

Leroux, B. G., and M. L. Puterman, Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models, *Biometrics*, *48*, 545–558, 1992.

Loucks, D. P., J. R. Stedinger, and D. A. Haith, *Water Resources Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1981.

Lu, Z.-Q., and L. M. Berliner, Markov switching time series models with application to a daily runoff series, *Water Resources Research*, *35*, 523–534, 1999.

MacDonald, I. L., and W. Zucchini, *Hidden Markov and Other Models for Discrete-Valued Time Series*, Chapman and Hall, London, 1997.

Matalas, N. C., Mathematical assessment of synthetic hydrology, *Water Resources Research*, *3*, 937–945, 1967.

Mathier, L., L. Perreault, B. Bobee, and F. Ashkar, Frequency analysis of water deficit duration and severity, *Proceedings of the Second International Conference on Computer Methods and Water Resources II, Marrakech, Morocco*, pp. 141–156, 1991.

Mathier, L., L. Perreault, B. Bobee, and F. Ashkar, The use of geometric and gamma-related distributions for frequency analysis of water deficit, *Stochastic Hydrology and Hydraulics*, *6*, 239–254, 1992.

McLachlan, G., and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics, USA, 2000.

McMahon, T. A., and R. G. Mein, *River and Reservoir Yield*, Water Resources Publications, Littleton, Colorado, USA., 1986.

Mejia, J. M., and J. Rousselle, Disaggregation models in hydrology revisited, *Water Resources Research*, *12*, 185–186, 1976.

Neal, E. G., M. T. Walter, and C. Coffeen, Linking the Pacific Decadal Oscillation to seasonal stream discharge patterns in southeast Alaska, *Journal of Hydrology*, *263*, 188–197, 2002.

Poskitt, D. S., and S. Chung, Markov chain models, time series analysis and extreme value theory, *Advances in Applied Probability*, *28*, 405–425, 1996.

Rabiner, L. R., A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, *77*, 257–286, 1989.

Rabiner, L. R., and B. H. Juang, An introduction to hidden Markov models, *IEEE Acoustics, Speech, and Signal Processing Society*, *3*, 4–16, 1986.

Rangarajan, S., Drought flow analysis, *Tech. rep.*, Manitoba Hydro, Manitoba, Canada, 1998.

Sadeghipour, J., and J. A. Dracup, Regional frequency analysis of hydrologic multiyear droughts, *Water Resources Bulletin*, *21*, 481–487, 1985.

Salas, J. D., *Analysis and Modeling of Hydrologic Time Series*, chap. 19, pp. 19.1–19.72, In: Maidment, D., (Ed.), Handbook of Hydrology, McGraw-Hill, Inc., New York, USA, 1993.

Salas, J. D., and D. C. Boes, Shifting level modeling of hydrologic series, *Advances in Water Resources*, *3*, 59–63, 1980.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane, *Applied Modeling of Hydrologic Time Series*, 3rd ed., Water Resources Publication, Littleton, Colorado, 1988.

Salas, J. D., N. Saada, C. H. Chung, W. L. Lane, and D. K. Frevert, Stochastic analysis, modeling, and simulation (SAMS) version 2000 - user's manual, *Tech. rep.*, Colorado State University, USA, 2000.

Sanso, B., and L. Guenni, A stochastic model for tropical rainfall at a single location, *Journal of Hydrology*, *214*, 64–73, 1999.

Santos, E. G., and J. D. Salas, Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering*, *118*, 765–784, 1992.

Schwarz, G., Estimating the dimension of a model, *The Annals of Statistics*, *6*, 461–464, 1978.

Sen, Z., Run-sum of annual flow series, *Journal of Hydrology*, *35*, 311–324, 1977.

Sen, Z., Statistical analysis of hydrologic critical drought, *Journal of the Hydraulics Division*, *106*, 99–115, 1980a.

Sen, Z., Critical drought analysis of periodic-stochastic processes, *Journal of Hydrology*, *46*, 251–263, 1980b.

Sharma, A., D. G. Tarboton, and U. Lall, Streamflow simulation: A nonparametric approach, *Water Resources Engineering*, *33*, 291–308, 1997.

Sharma, T. C., Drought parameters in relation to truncation level, *Hydrological Processes*, *14*, 1279–1288, 2000.

Shiau, J.-T., and H. W. Shen, Recurrence analysis of hydrologic droughts of differing severities, *Journal of Water Resources Planning and Management*, *127*, 30–40, 2001.

Smith, K., and M. B. Richman, Recent hydroclimatic fluctuations and their effects on water resources in Illinois, *Climatic Change*, *24*, 249–269, 1993.

Spence, C., Streamflow variability (1965 to 1998) in five Northwest Territories and Nunavut rivers, *Canadian Water Resources Journal*, *27*, 135–154, 2002.

Stedinger, J. R., Fitting log normal distributions to hydrologic data, *Water Resources Research*, *16*, 481–490, 1980.

Stedinger, J. R., and M. R. Taylor, Synthetic streamflow generation 1. Model verification and validation, *Water Resources Research*, *18*, 909–918, 1982a.

Stedinger, J. R., and M. R. Taylor, Synthetic streamflow generation 2. Effects of parameter uncertainty, *Water Resources Research*, *18*, 919–924, 1982b.

Stedinger, J. R., and R. M. Vogel, Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, *20*, 47–56, 1984.

Stedinger, J. R., D. Pei, and T. A. Cohn, A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulation, *Water Resources Research*, *21*, 665–675, 1985.

Tao, P. C., and J. W. Delleur, Multistation, multiyear synthesis of hydrologic time series by disaggregation, *Water Resources Research*, *12*, 1303–1312, 1976.

Tarboton, D. G., A. Sharma, and U. Lall, Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, *Water Resources Research*, *34*, 107–119, 1998.

Thyer, M., and G. Kuczera, Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model, *Water Resources Research*, *36*, 3301–3310, 2000.

Thyer, M., and G. Kuczera, A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach, *Journal of Hydrology*, *275*, 12–26, 2003a.

Thyer, M., and G. Kuczera, A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 2. Real data analysis, *Journal of Hydrology*, *275*, 27–48, 2003b.

Thyer, M., G. Kuczera, and Q. J. Wang, Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *Journal of Hydrology*, *265*, 246–257, 2002.

Tiao, G. C., and A. Zellner, On the Bayesian estimation of multivariate regression, *Journal of the Royal Statistical Society, Series B*, *26*, 277–285, 1964.

Timmermann, A., Moments of Markov switching models, *Journal of Econometrics*, *96*, 75–111, 2000.

Tong, Y. L., *The Multivariate Normal Distribution*, Springer-Verlag, New York, USA, 1990.

Valdes, J. B., and I. Rodriguez-Iturbe, Bayesian generation of synthetic stremflows 2. The multivariate case, *Water Resources Research*, *13*, 291–295, 1977.

Valencia, D., and J. C. Schaake, Disaggregation processes in stochastic hydrology, *Water Resources Research*, *9*, 580–585, 1973.

Vicens, G. J., I. Rodriguez-Iturbe, and J. C. Schaake, A Bayesian framework for the use of regional information in hydrology, *Water Resources Research*, *11*, 405–414, 1975a.

Vicens, G. J., I. Rodriguez-Iturbe, and J. C. Schaake, Bayesian generation of synthetic streamflows, *Water Resources Reserach*, *11*, 827–838, 1975b.

Vogel, R. M., The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distribution hypotheses, *Water Resources Research*, *22*, 587–590, 1986.

Wilks, D. S., Multisite generalization of a daily stochastic precipitation generation model, *Journal of Hydrology*, *210*, 178–191, 1998.

Yevjevich, V., An objective approach to definitions and investigations of continental hydrologic drought, *Hydrology Paper*, *23, Colorado State University, Fort Collins, Colo.*, 1967.

Zelenhasic, E., On the extreme streamflow drought analysis, *Water Resources Management*, *16*, 105–132, 2002.

Zelenhasic, E., and A. Salvai, A method of streamflow drought analysis, *Water Resources Reserach*, *23*, 156–168, 1987.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc., USA, 1971.

Zhang, J., and R. A. Stine, Autoregressive structure of Markov regime switching models and model selection, *Journal of Time Series Analysis*, *22*, 107–124, 2001.

Appendix A

Evidence of Long-term Cycles in Manitoba Hydro's Annual Streamflow

A.1 Introduction

During the preliminary analysis, Manitoba Hydro's annual streamflow data were smoothed using a 5-yr running average. As seen from following plots, there is evidence of extended long-term wet and dry cycles in all five sites and aggregated basin flows. This evidence motivated the choice of a 2-state Markov-Switching model in this project.

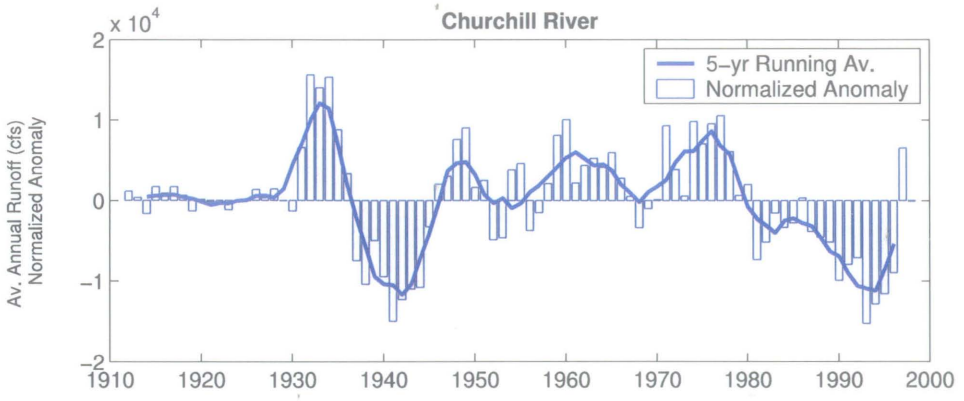


Figure A.1: Churchill River annual flows, 1912-98. Normalized anomalies and 5-yr running average.

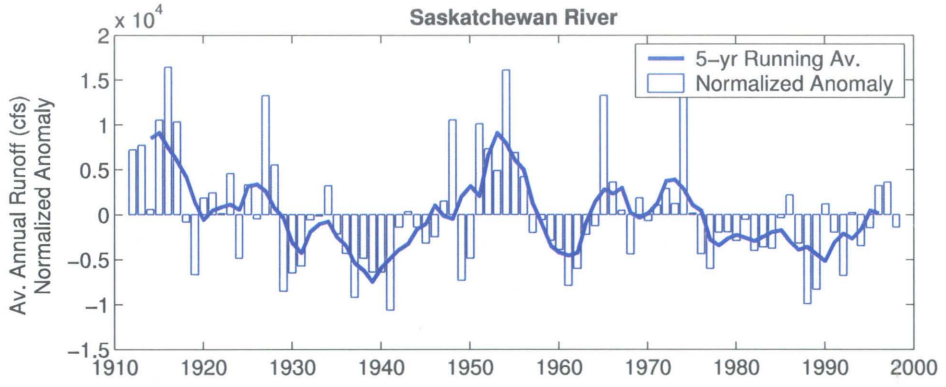


Figure A.2: Saskatchewan River annual flows, 1912-98. Normalized anomalies and 5-yr running average.

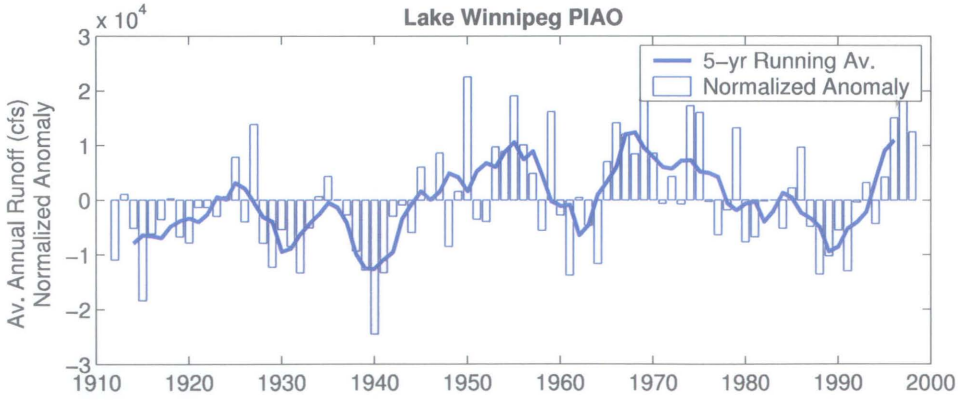


Figure A.3: Lake Winnipeg annual PIAO, 1912-98. Normalized anomalies and 5-yr running average.

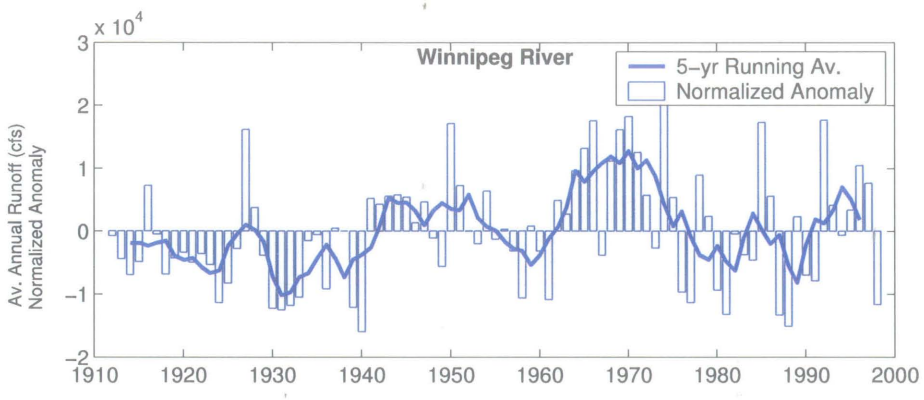


Figure A.4: Winnipeg River annual flows, 1912-98. Normalized anomalies and 5-yr running average.

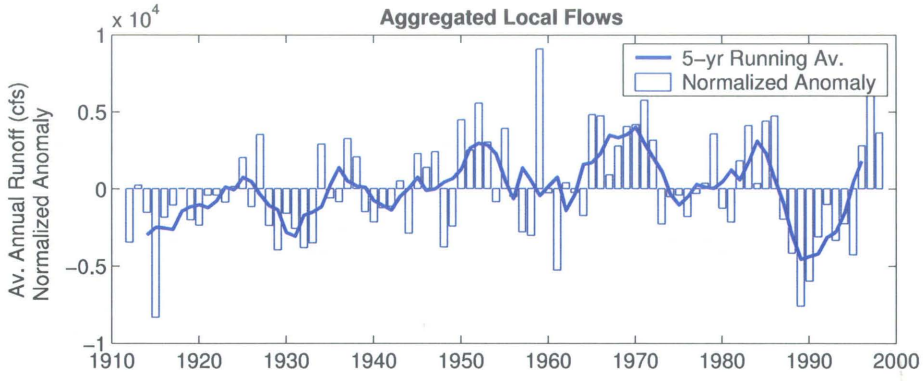


Figure A.5: Aggregated annual Local Flows, 1912-98. Normalized anomalies and 5-yr running average.

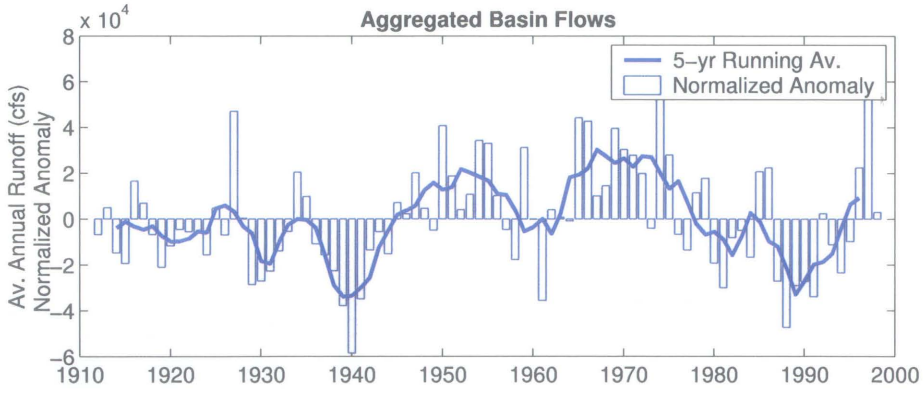


Figure A.6: Aggregated annual Basin Flows, 1912-98. Normalized anomalies and 5-yr running average.

Appendix B

Probability Density Functions and Normal Probability Plots for the Niagara River Monthly Flows

B.1 Introduction

In the MSD model, proposed in Chapter 7, the parameters of the marginal distributions of the lower-level variables are estimated conditional on the higher-level variable's state probability sequences. Therefore the goodness of fit of the marginal distributions of the monthly flows must be tested before application of the model. As illustrated in Figure B.1 to Figure B.12, good fits of the marginal distributions of all observed monthly flows of the Niagara River were obtained by the MSD model.

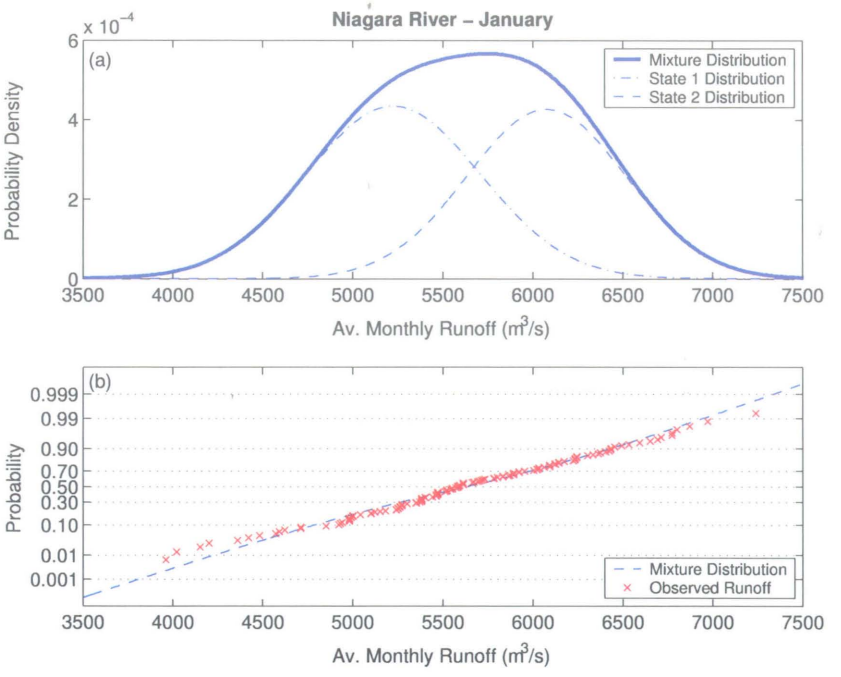


Figure B.1: January flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

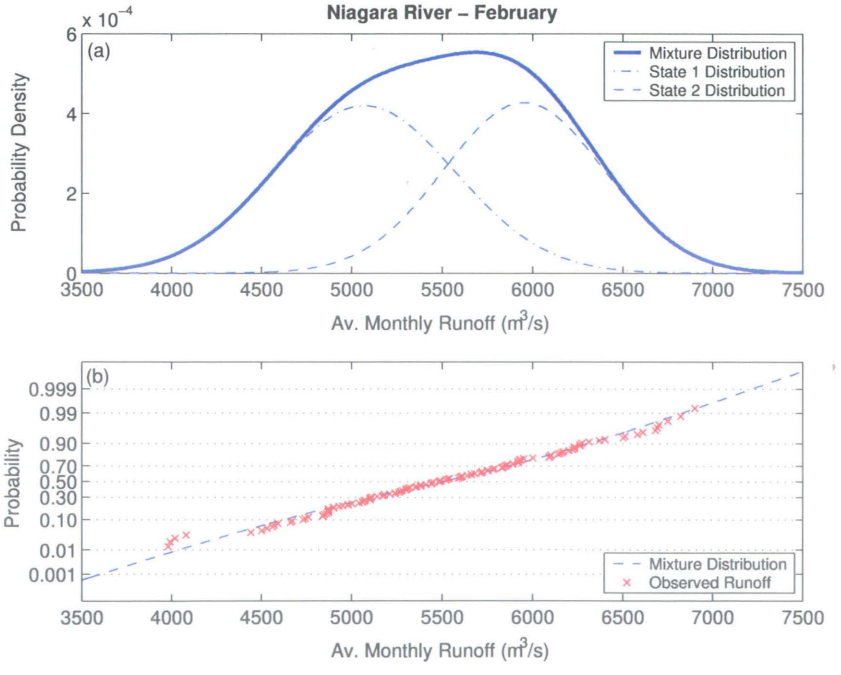


Figure B.2: February flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

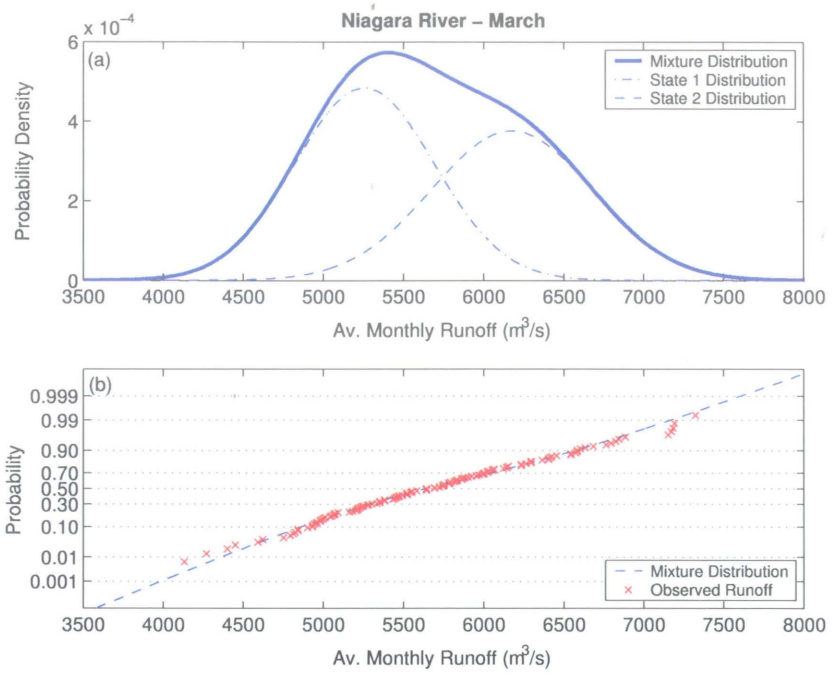


Figure B.3: March flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

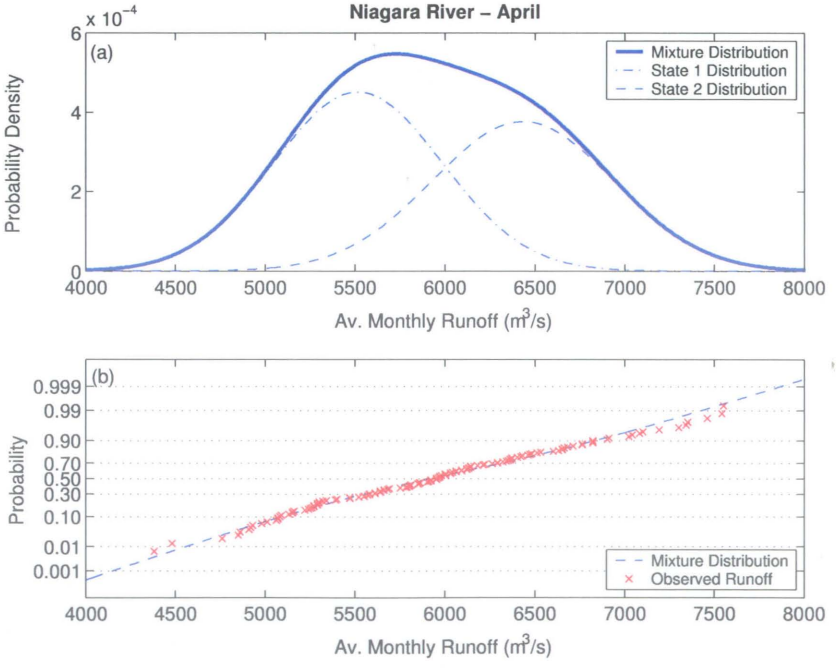


Figure B.4: April flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

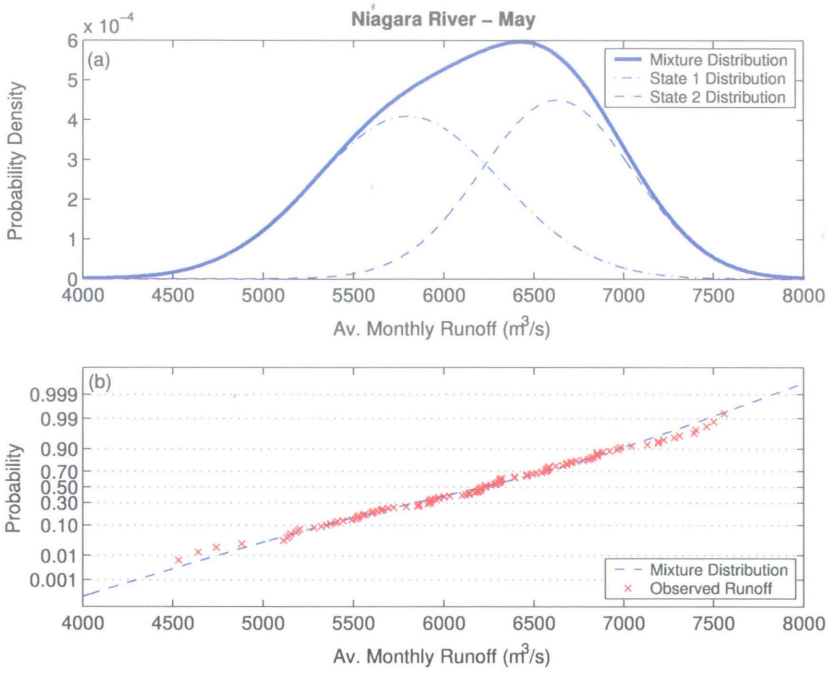


Figure B.5: May flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

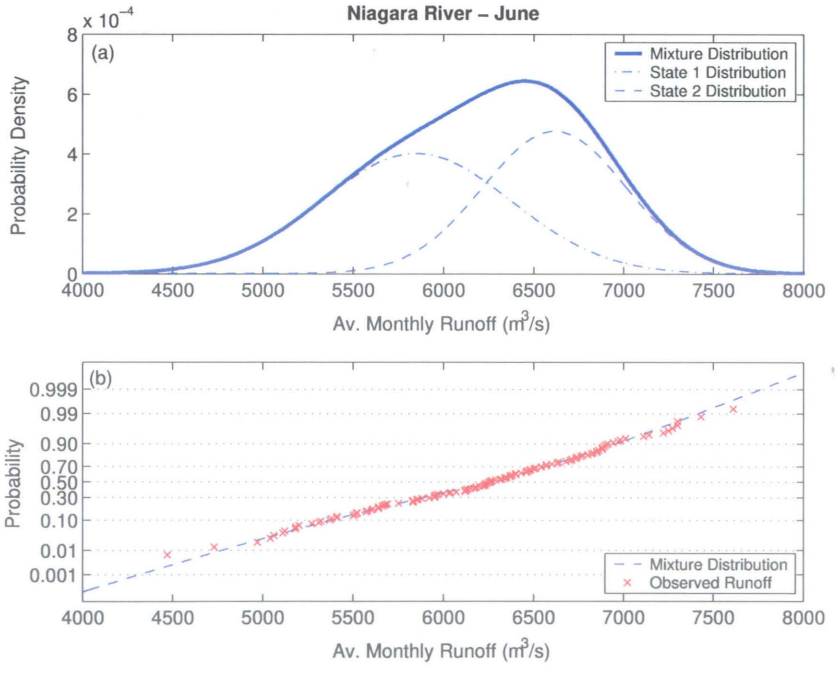


Figure B.6: June flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

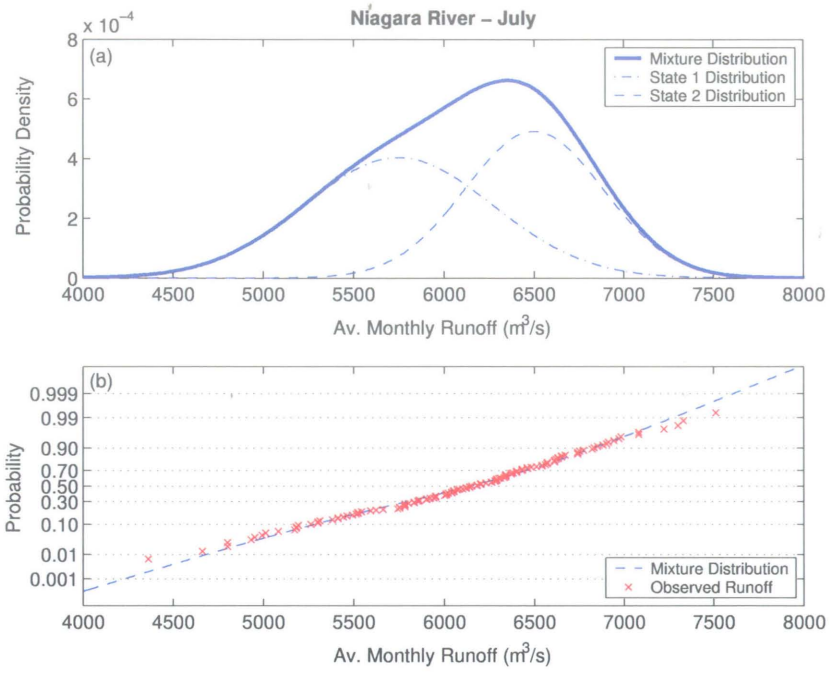


Figure B.7: July flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

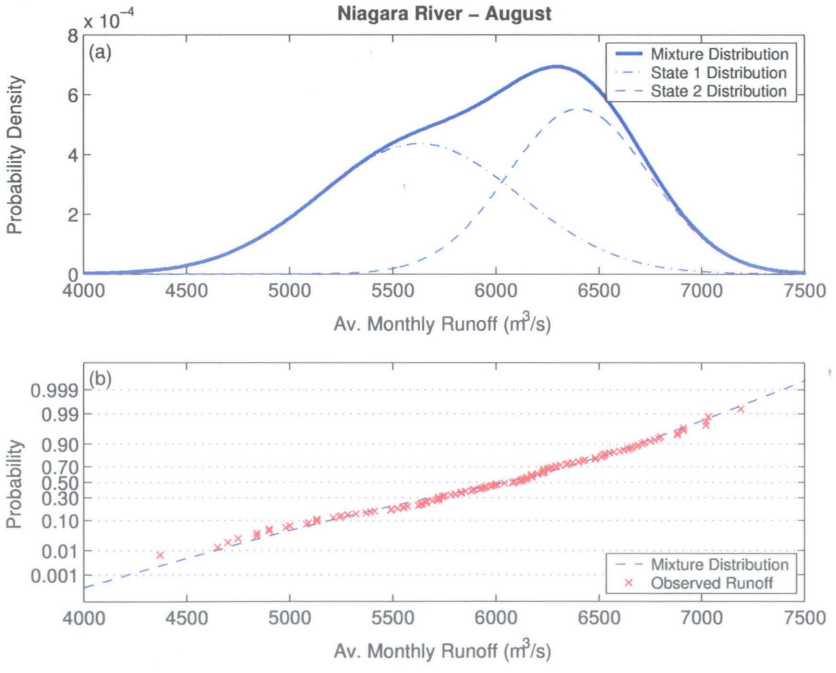


Figure B.8: August flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

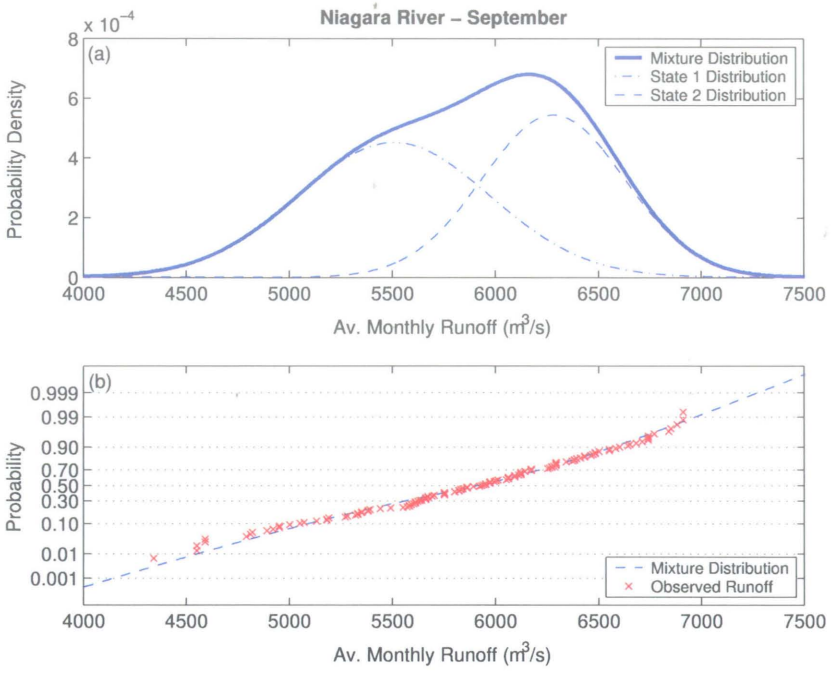


Figure B.9: September flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

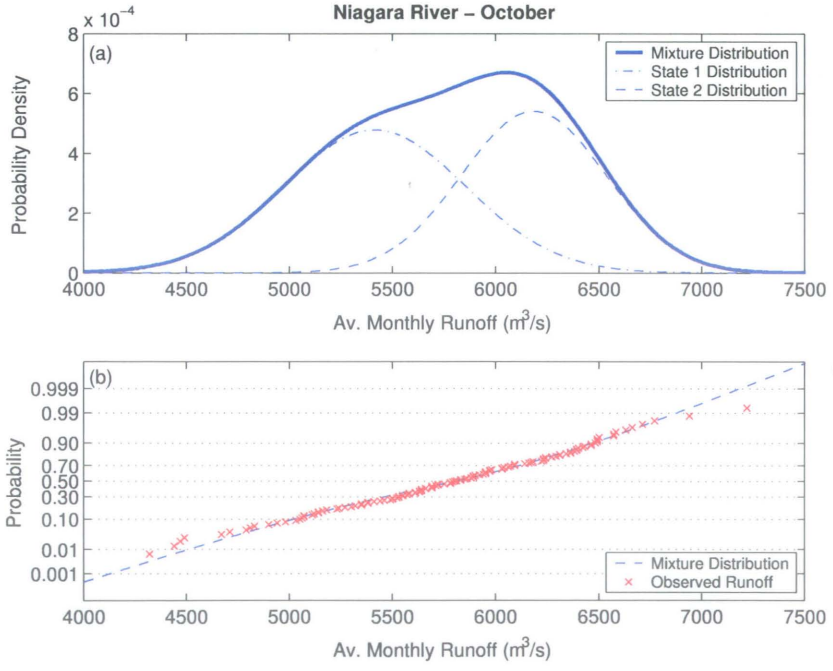


Figure B.10: October flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

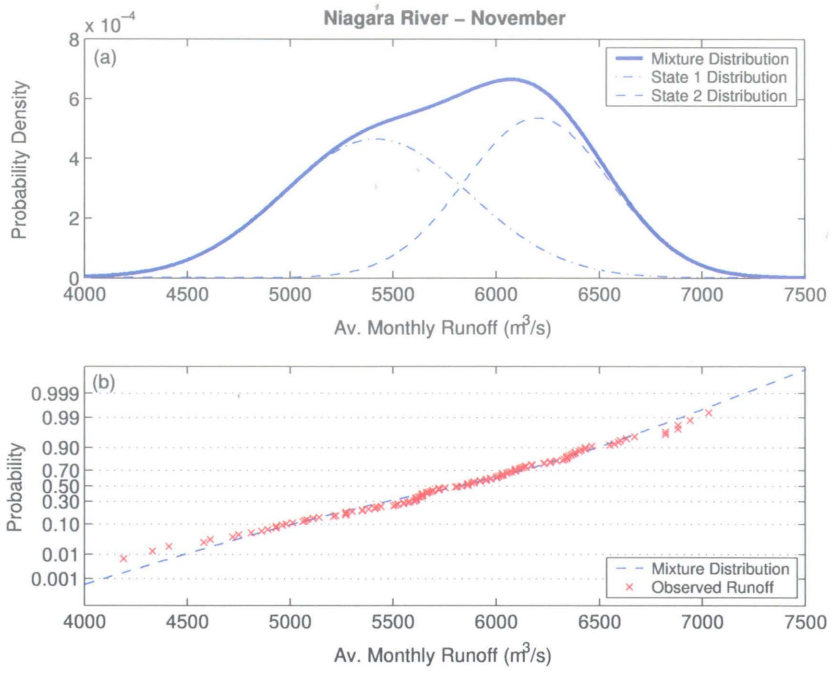


Figure B.11: November flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

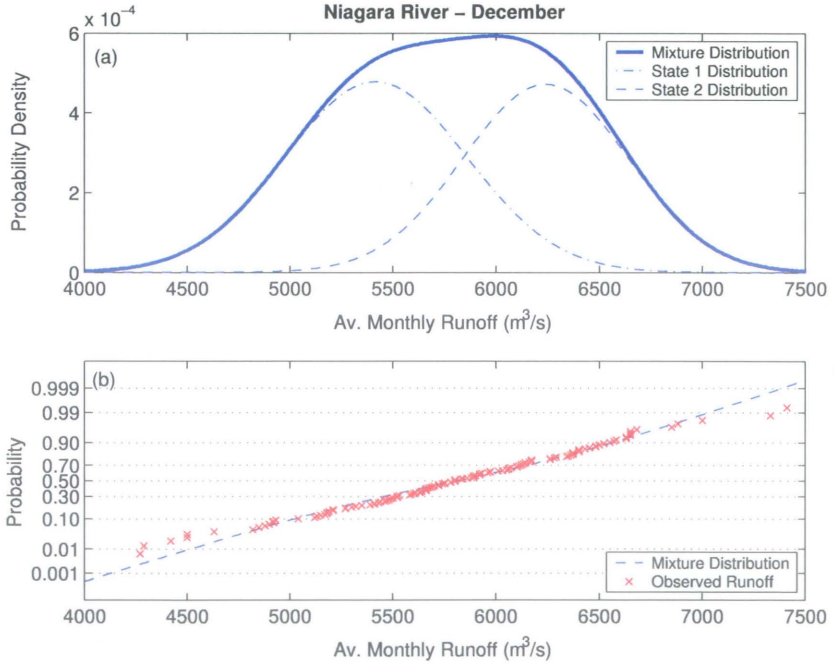


Figure B.12: December flows of the Niagara River. a) Marginal probability density function. b) Normal probability plot of observed flows and fitted mixture distribution.

Appendix C

Comparison of Simulated and Observed Cross-Correlations for the Niagara River Monthly Flows

C.1 Introduction

In Chapter 7, the MSD model and the VSD model are compared using simulated and observed lag-0 cross-correlation coefficients. Figures C.1 - C.12 illustrate the correlation between all monthly variables, and between all monthly and annual variables. On the x-axis, numbers $1, 2, \ldots, 12$ represent the months January (1), February (2), ..., December (12), and A represents annual flow. For example, 1-A indicates the correlation between January flows and annual flows, 1-2 indicates the correlation between January and February flows, and so on.

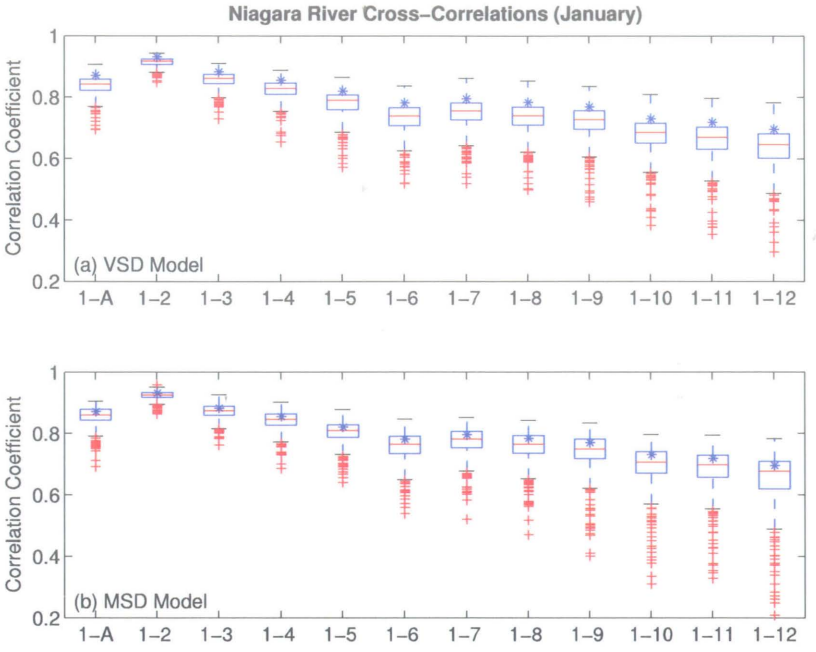


Figure C.1: Simulated and observed cross-correlations for January flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

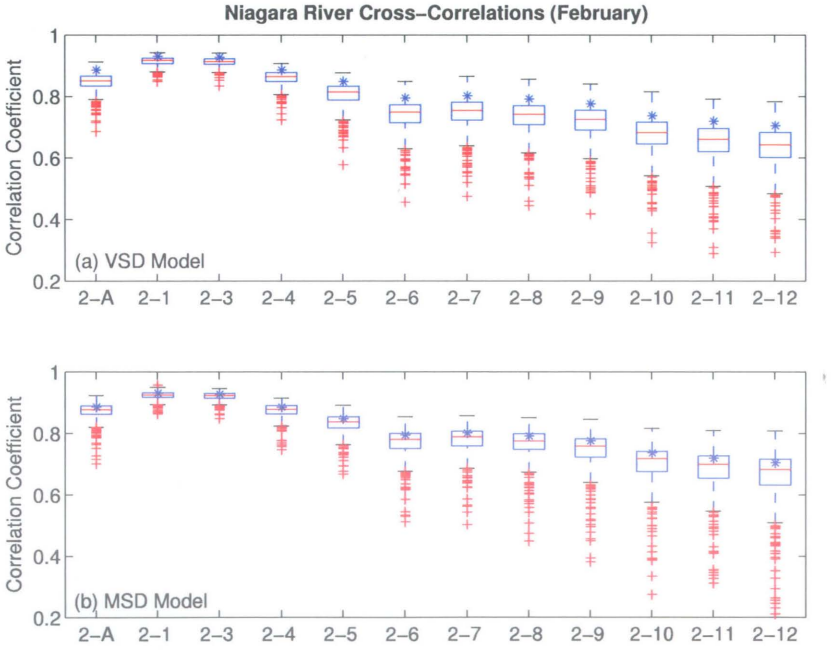


Figure C.2: Simulated and observed cross-correlations for February flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

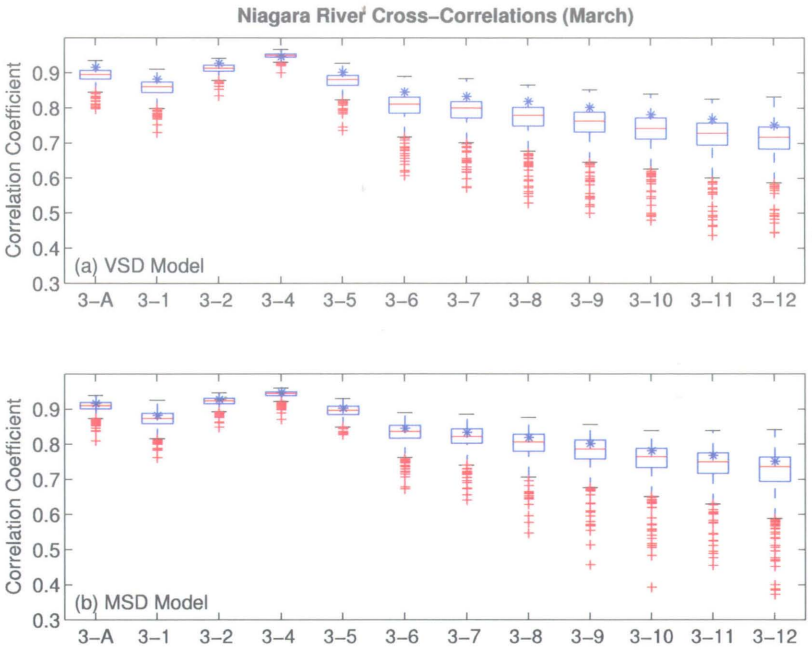


Figure C.3: Simulated and observed cross-correlations for March flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

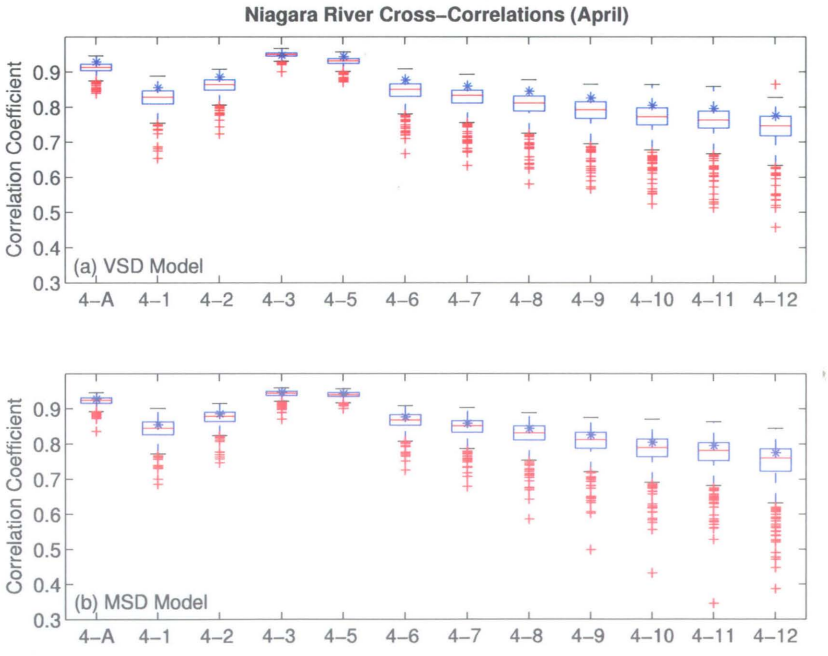


Figure C.4: Simulated and observed cross-correlations for April flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

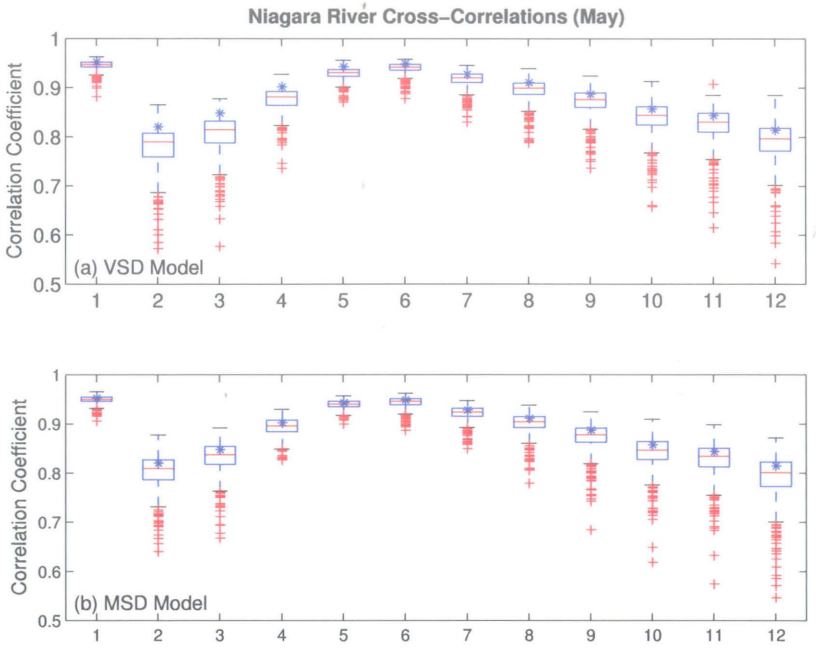


Figure C.5: Simulated and observed cross-correlations for May flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

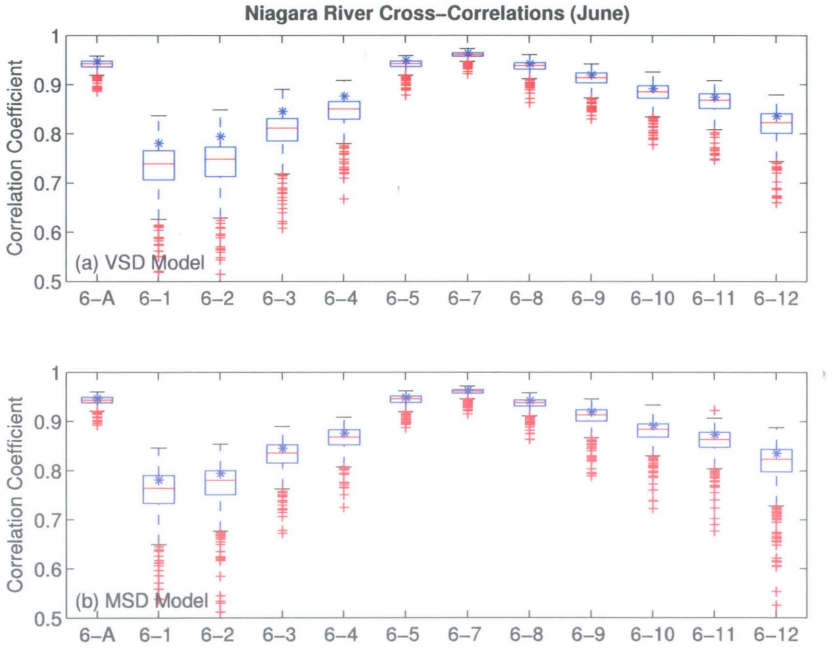


Figure C.6: Simulated and observed cross-correlations for June flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

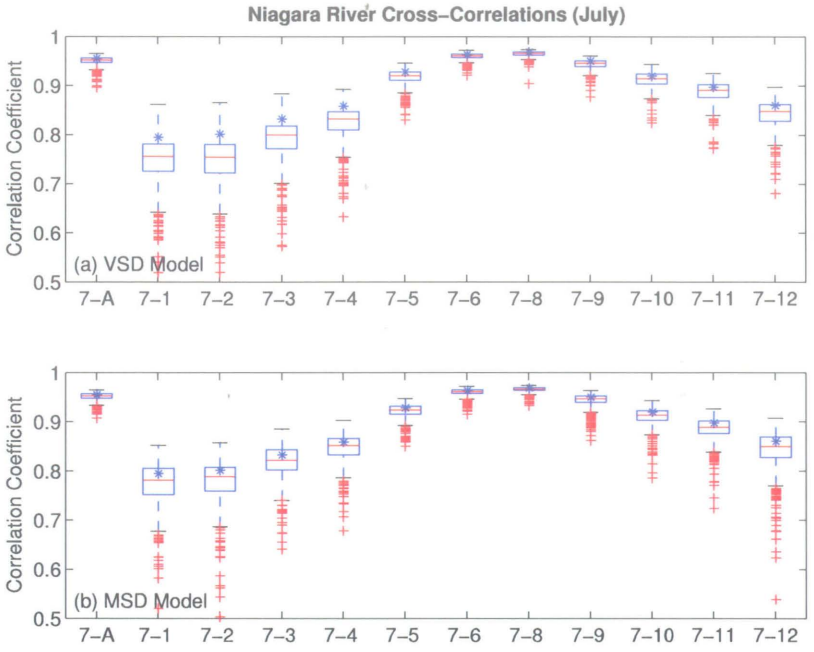


Figure C.7: Simulated and observed cross-correlations for July flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

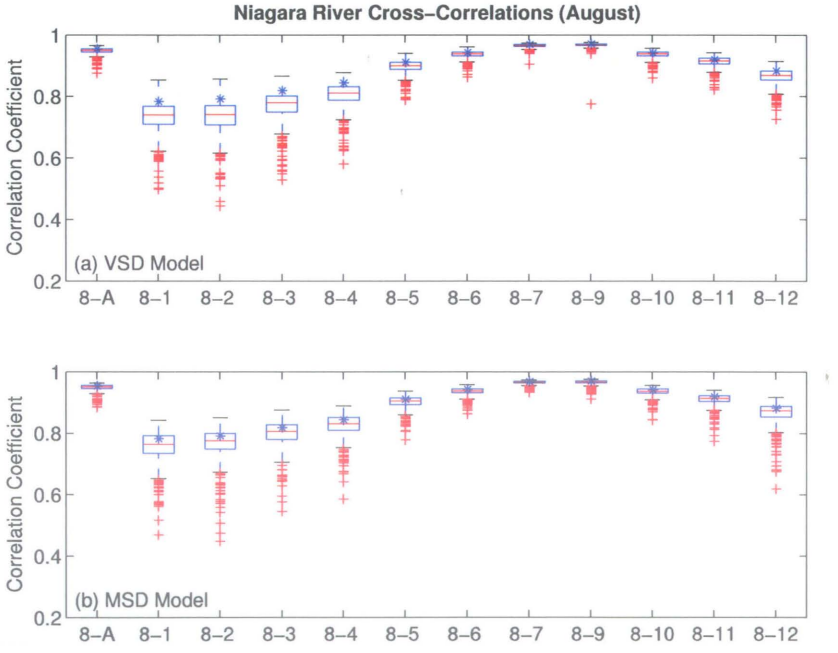


Figure C.8: Simulated and observed cross-correlations for August flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

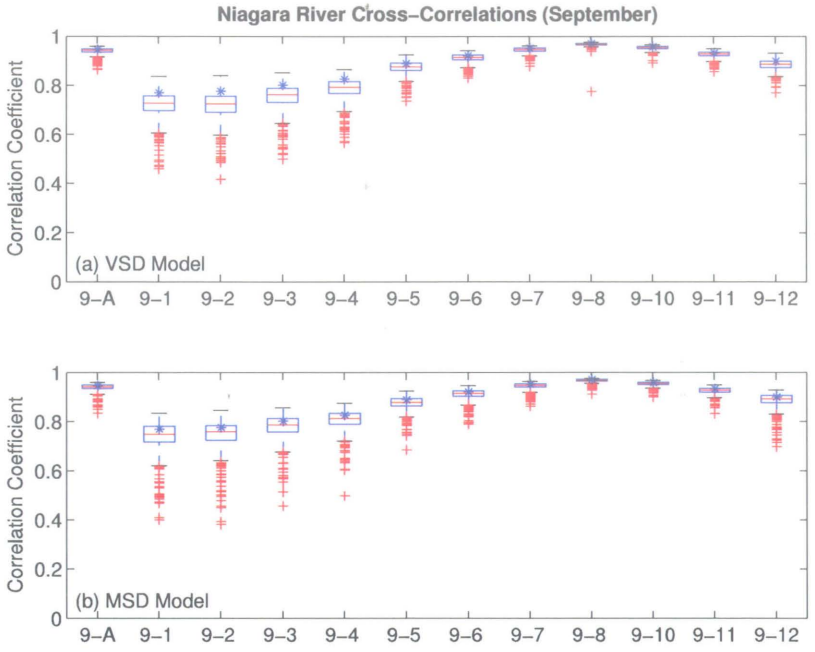


Figure C.9: Simulated and observed cross-correlations for September flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

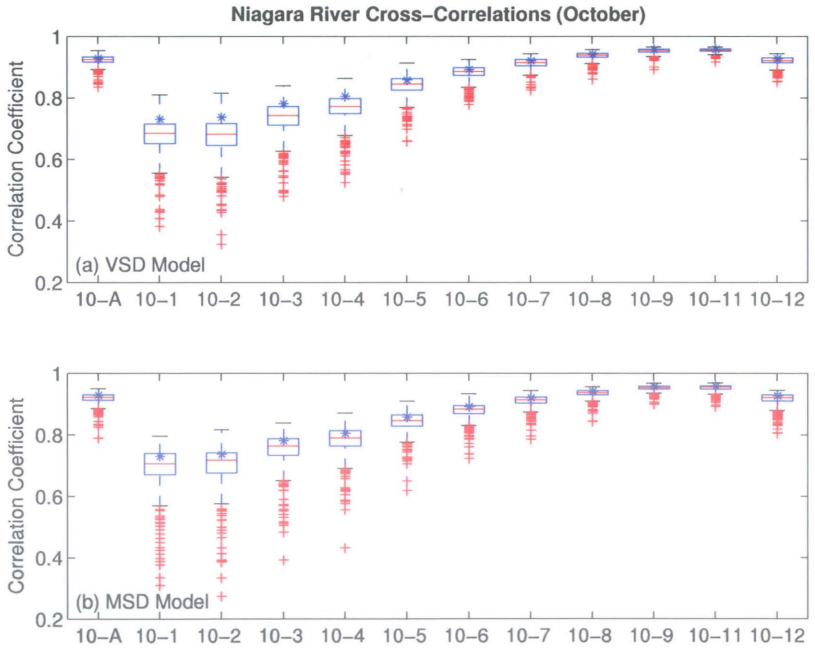


Figure C.10: Simulated and observed cross-correlations for October flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

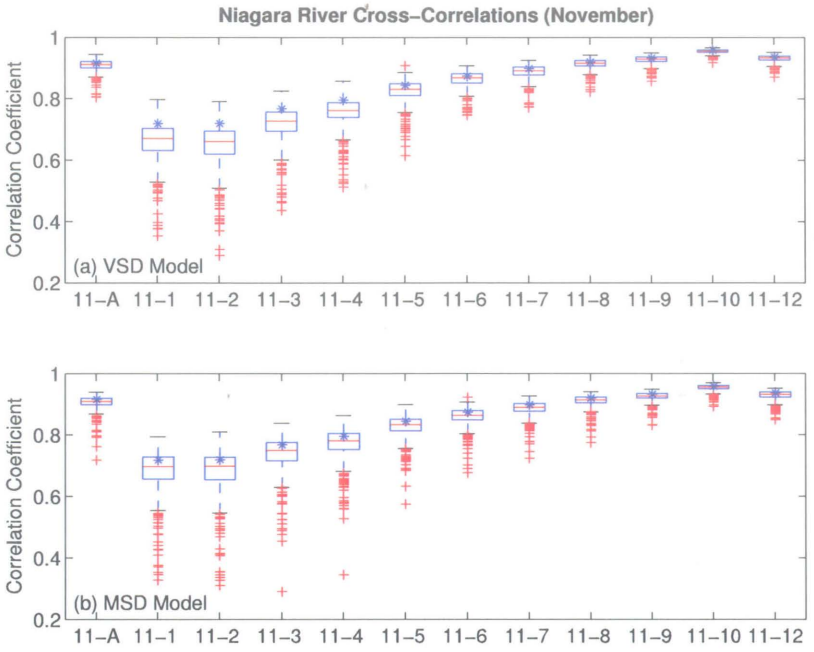


Figure C.11: Simulated and observed cross-correlations for November flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

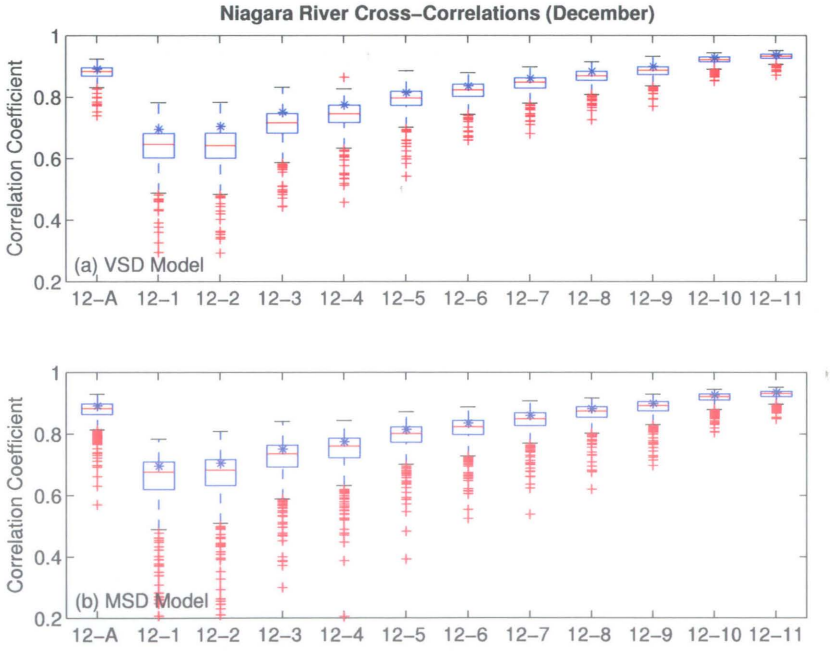


Figure C.12: Simulated and observed cross-correlations for December flows for the Niagara River. a) Simulated using VSD Model. b) Simulated using MSD Model. The asterisk (∗) represents the observed statistic.

Appendix D

Probability Plots of the Simulated Niagara River Monthly Flows

D.1 Introduction

To evaluate the modeling of the marginal distributions by the MSD and the VSD models, the probability plot of 10,000 simulated flows are examined. Figures D.1 - D.12 compare normal probability plots of simulated monthly flows using the two models. The marginal distribution of the MSD model is a mixture of two normals. In the VSD model, the Box-Cox distribution was suggested by the Filliben correlation coefficient statistic for all months. In most cases, the tails of the modeled marginal distribution were not well reproduced by the VSD model, whereas the MSD model performs relatively well.

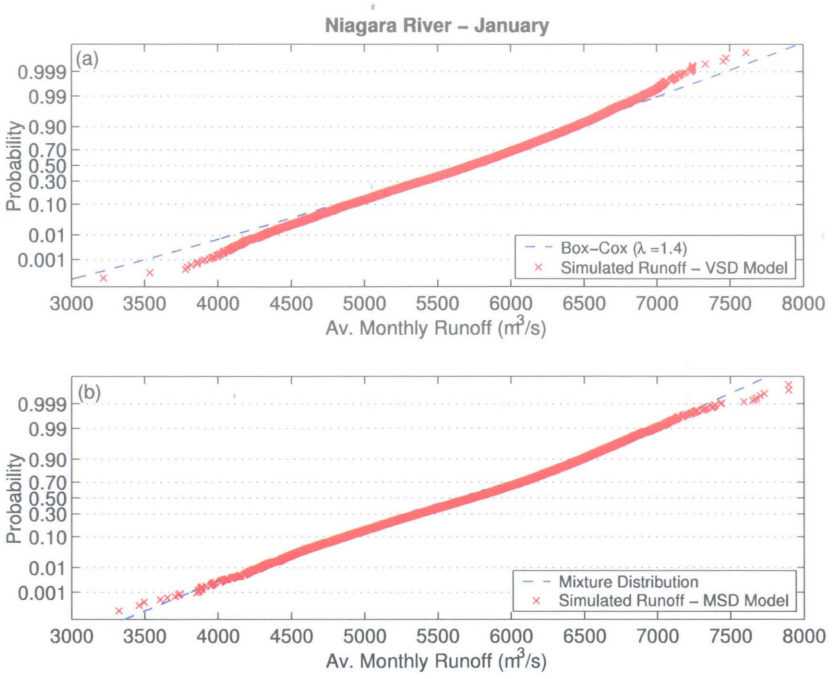


Figure D.1: Simulated January flows of the Niagara River using a) VSD Model. b) MSD Model.

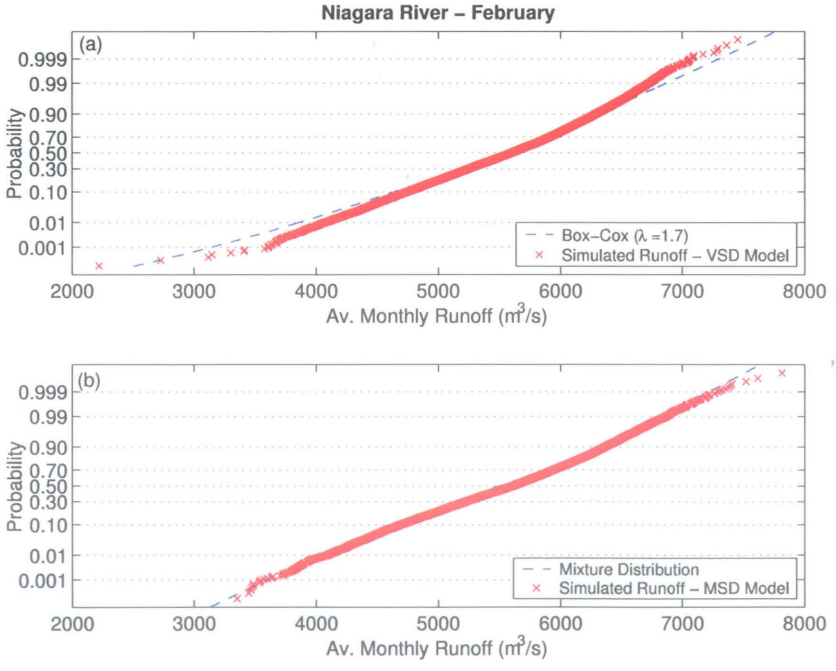


Figure D.2: Simulated February flows of the Niagara River using a) VSD Model. b) MSD Model.

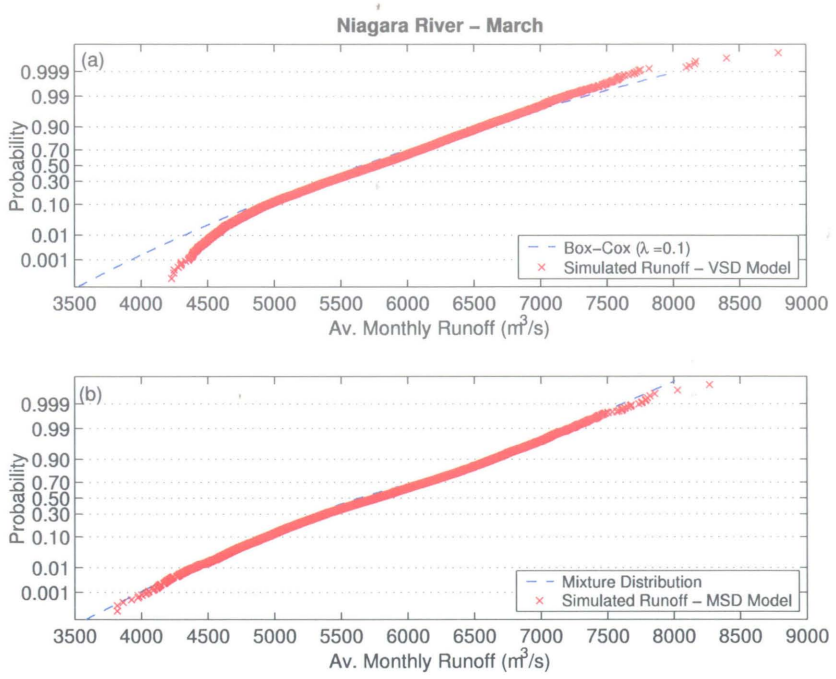


Figure D.3: Simulated March flows of the Niagara River using a) VSD Model. b) MSD Model.

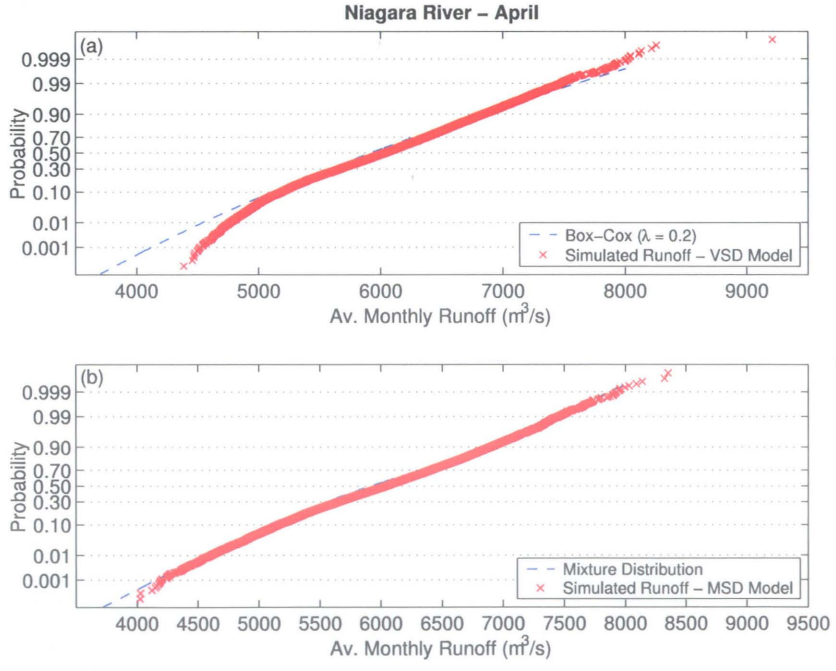


Figure D.4: Simulated April flows of the Niagara River using a) VSD Model. b) MSD Model.

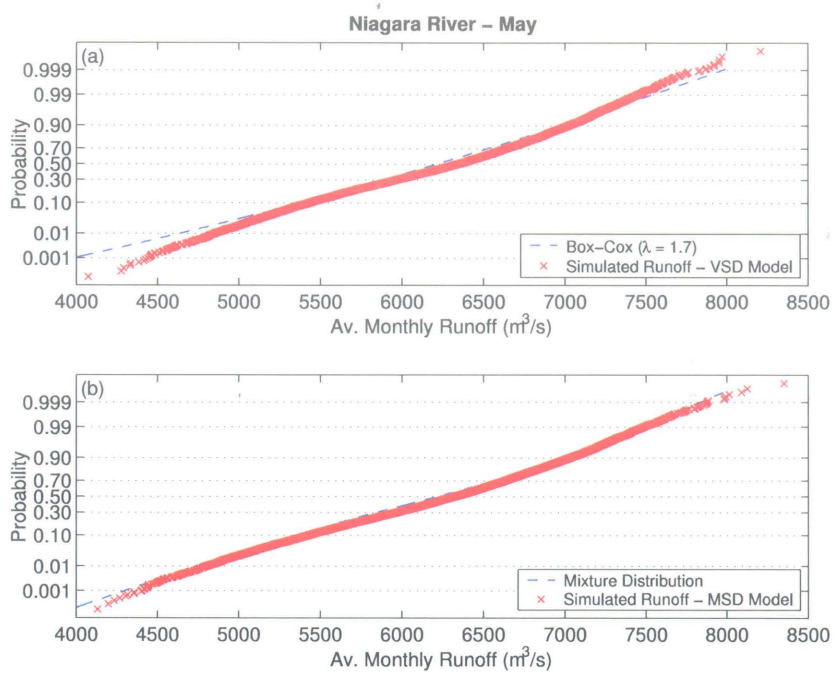


Figure D.5: Simulated May flows of the Niagara River using a) VSD Model. b) MSD Model.

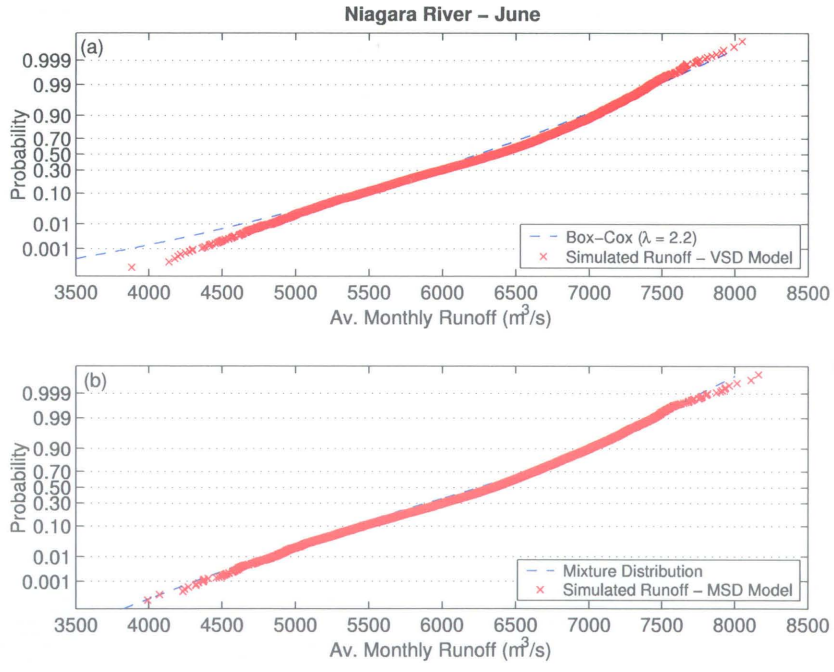


Figure D.6: Simulated June flows of the Niagara River using a) VSD Model. b) MSD Model.

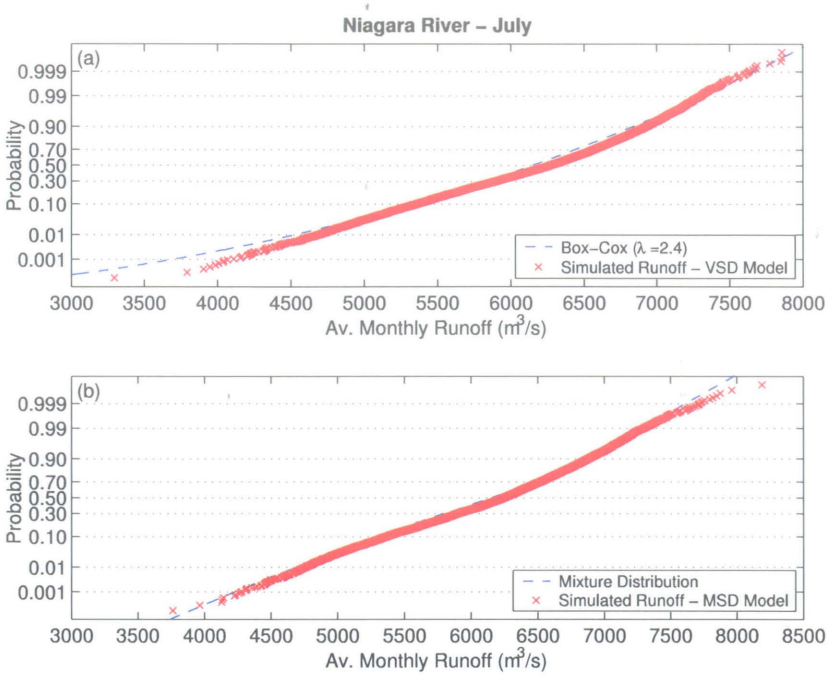


Figure D.7: Simulated July flows of the Niagara River using a) VSD Model. b) MSD Model.

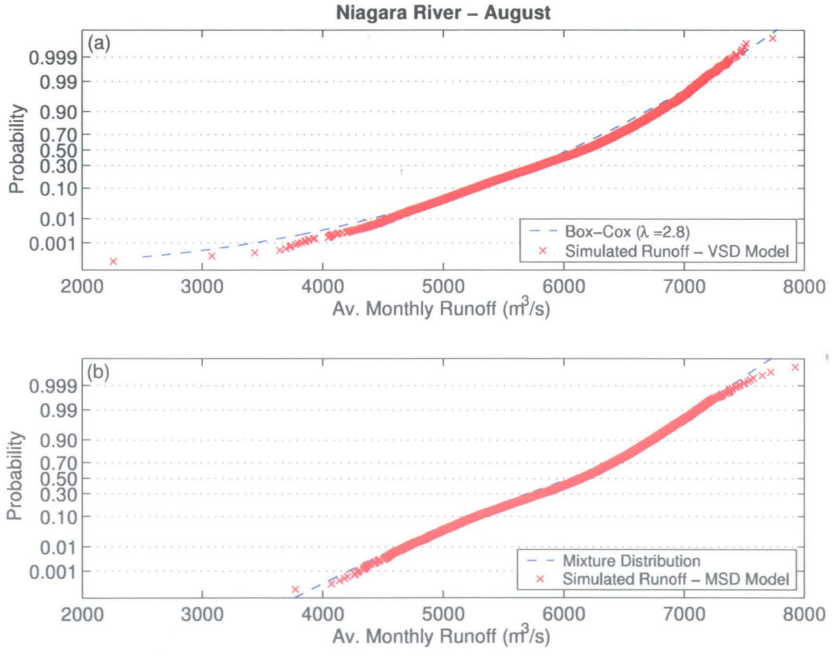


Figure D.8: Simulated August flows of the Niagara River using a) VSD Model. b) MSD Model.

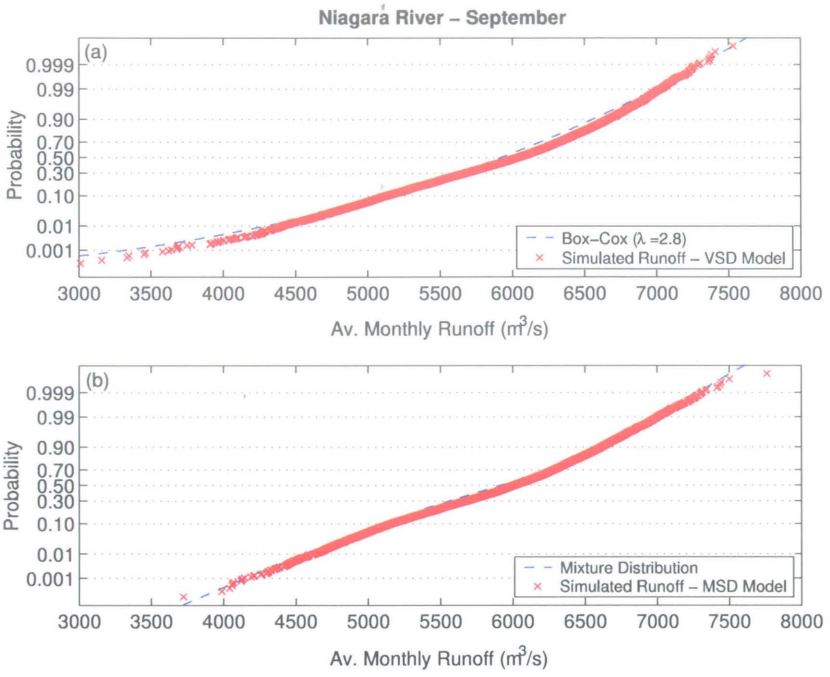


Figure D.9: Simulated September flows of the Niagara River using a) VSD Model. b) MSD Model.

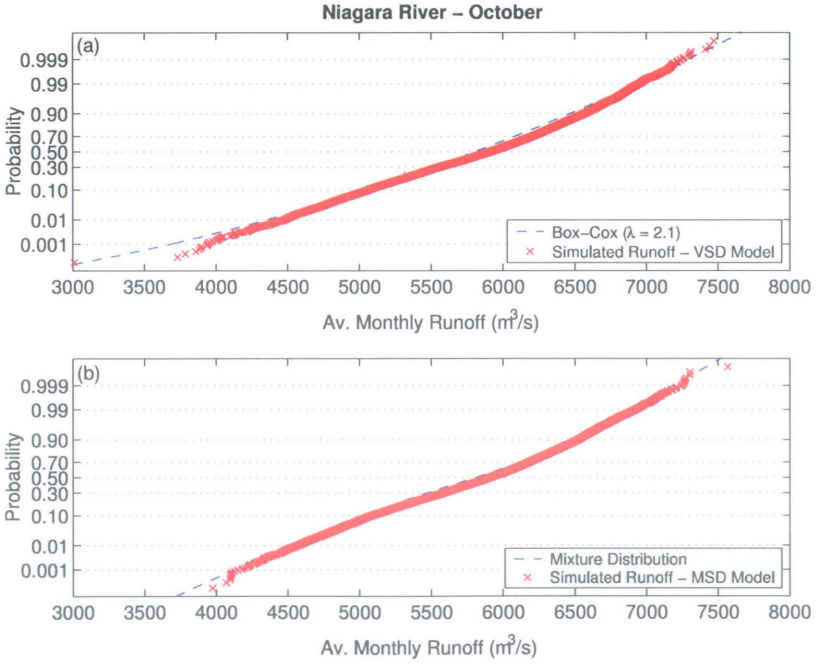


Figure D.10: Simulated October flows of the Niagara River using a) VSD Model. b) MSD Model.

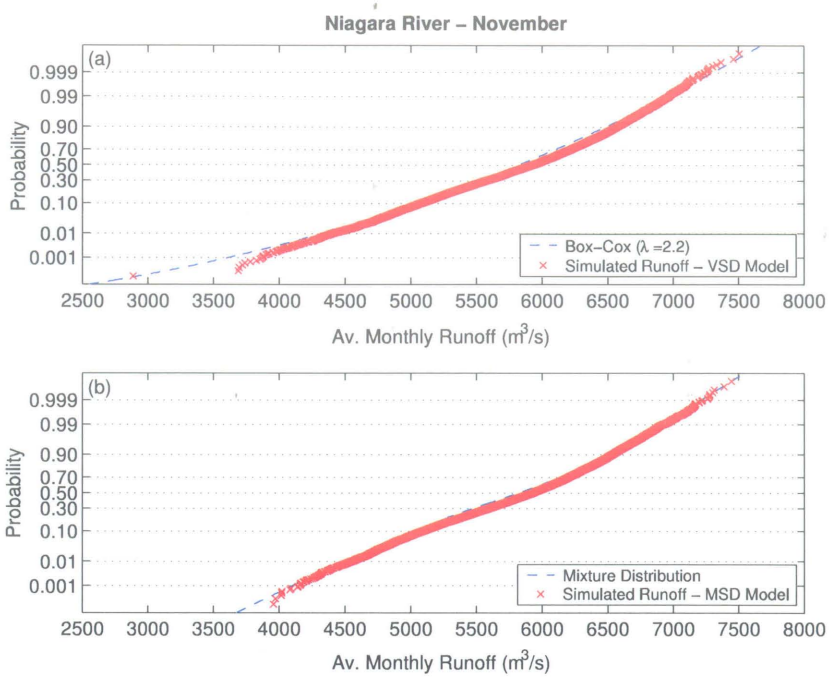


Figure D.11: Simulated November flows of the Niagara River using a) VSD Model. b) MSD Model.

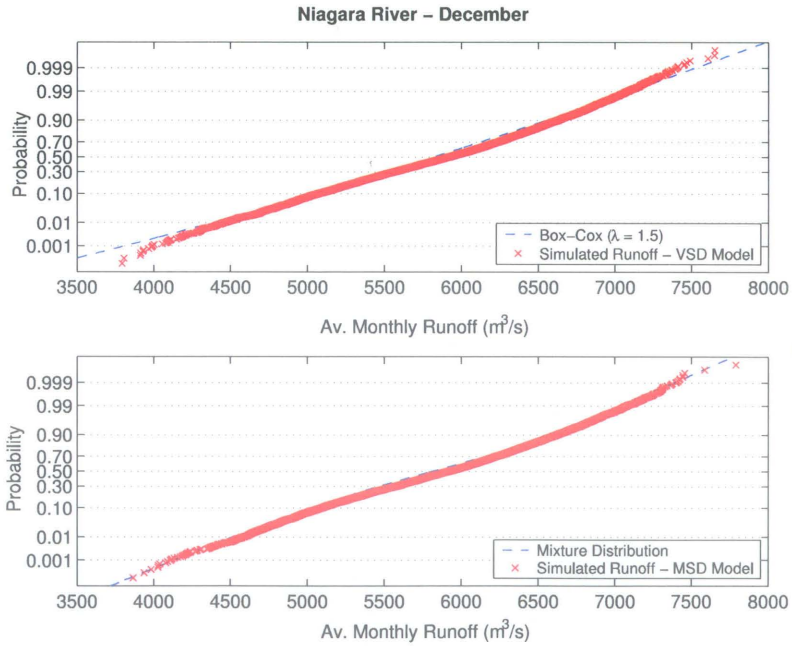


Figure D.12: Simulated December flows of the Niagara River using a) VSD Model. b) MSD Model.

Appendix E

The Single-site 2-state MS Model: Fitted Marginal Distributions

E.1 Introduction

After estimating the model parameters for the five sites of Manitoba Hydro's system individually, the goodness-of-fit of the marginal distributions produced by the single-site 2-state MS model can be assessed visually by probability plots (Figures E.1-E.6. Although the data are not transformed into Normal in the MS model, the model appears to fit the data very well at all sites.

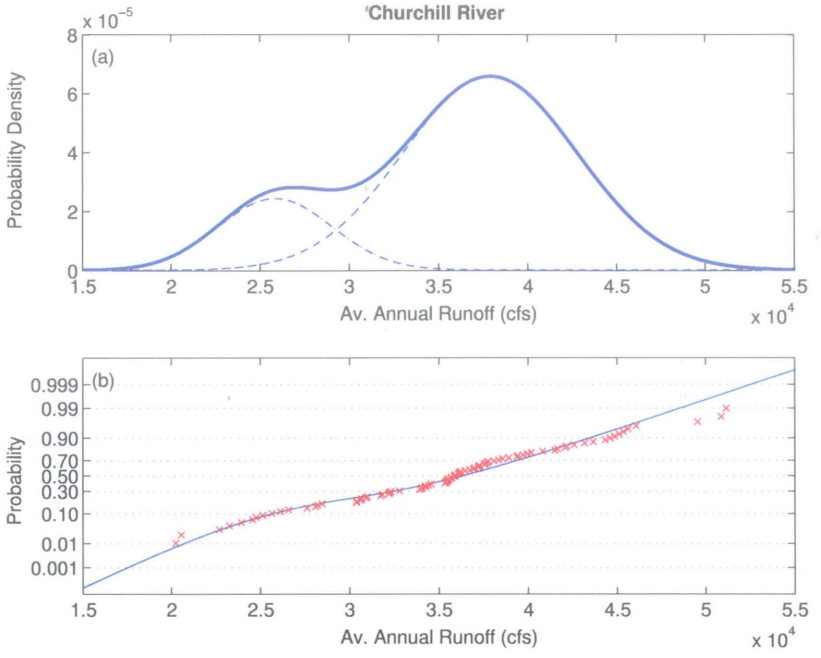


Figure E.1: Annual flows of the Churchill River at Southern Indian Lake. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

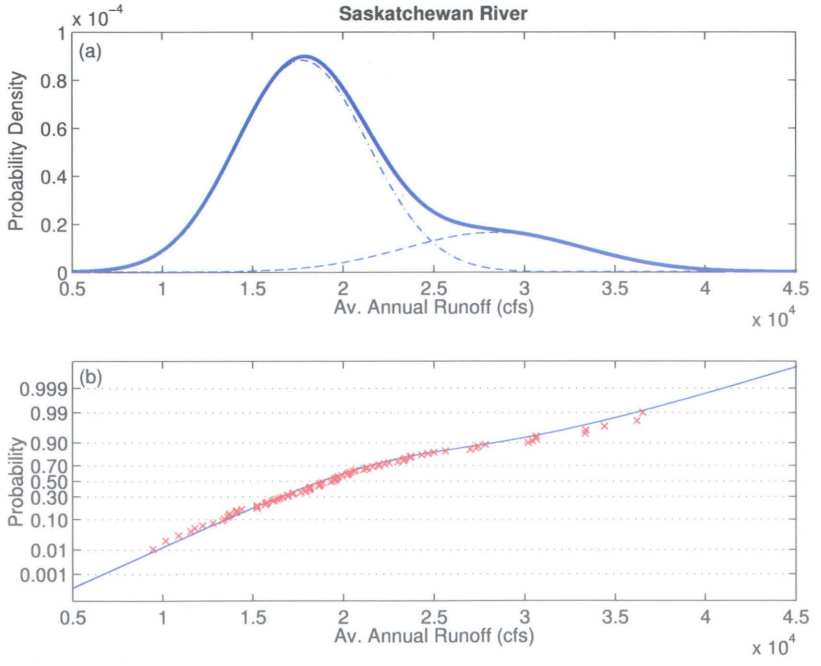


Figure E.2: Annual flows of the Saskatchewan River at Grand Rapids. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

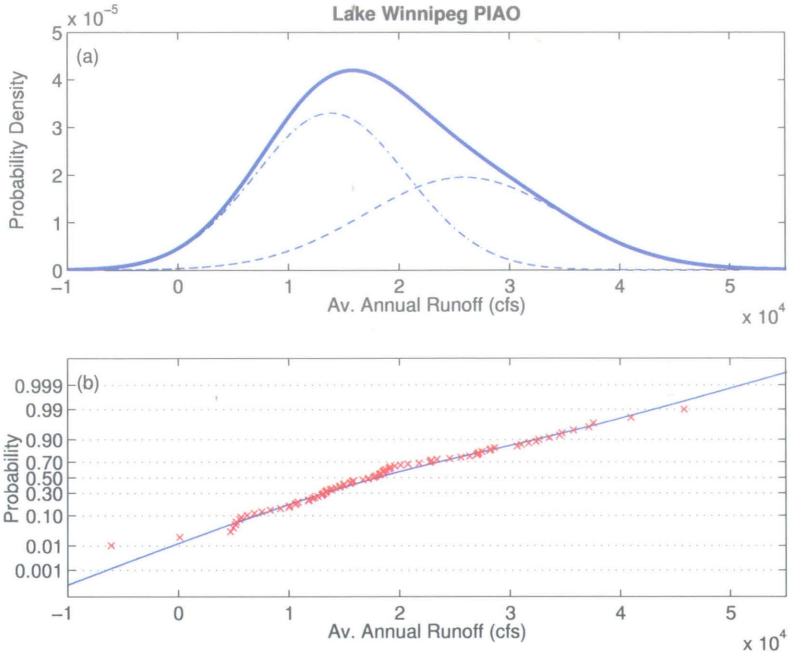


Figure E.3: Annual flows of the Lake Winnipeg PIAO. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

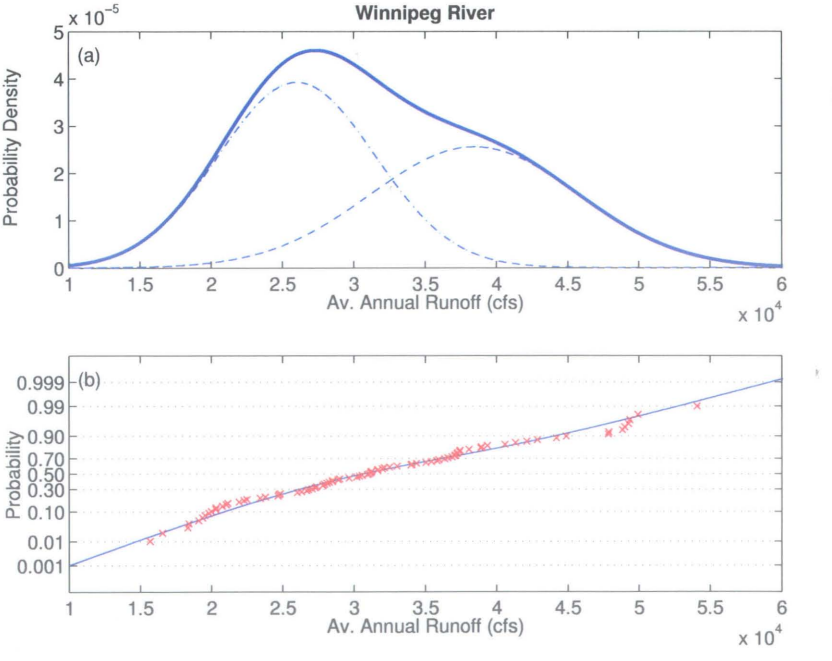


Figure E.4: Annual flows of Winnipeg River at Slave Falls. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

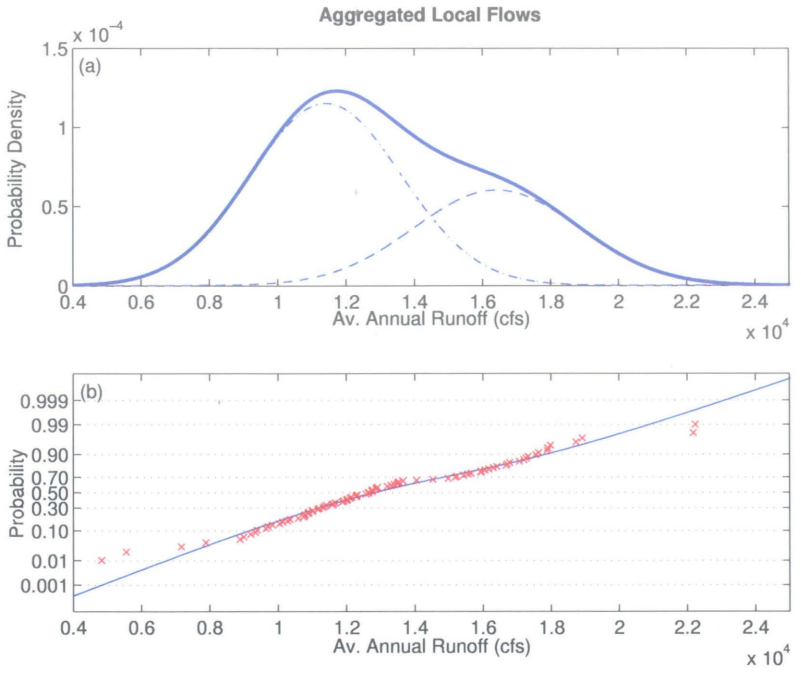


Figure E.5: Annual flows of Aggregated Local Flows in Burntwood and Nelson River. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

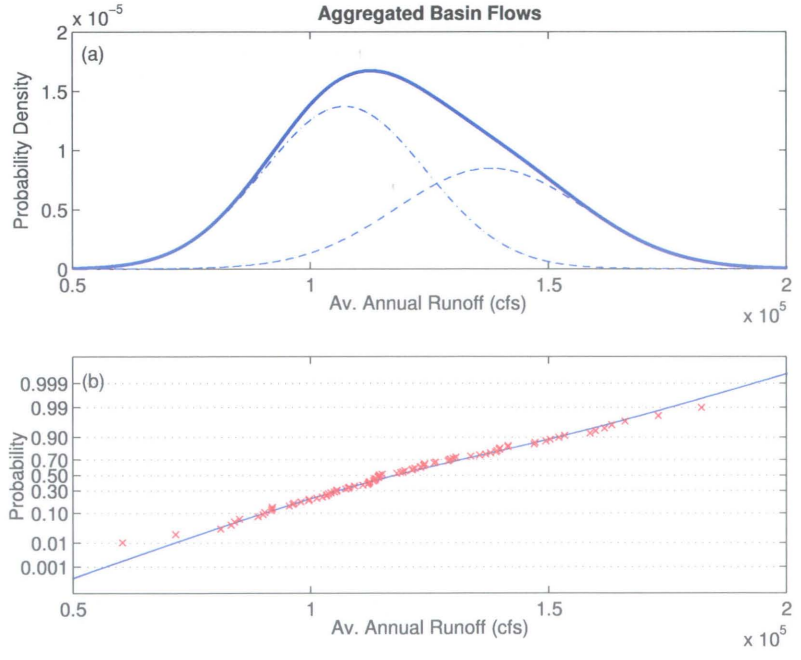


Figure E.6: Annual flows of Aggregated Basin Flows in Nelson and Churchill River Basin. a) Marginal probability density function of MS model. b) Normal probability plot of observed annual flows and fitted MS model

Appendix F

The Single-site 2-state MS Model: State Probabilities and Regime Changes in Observed Streamflow

F.1 Introduction

To appreciate the usefulness of a 2-state MS model for Manitoba Hydro's data, it is of interest to look at the estimated state probability sequences, $\Pr\{s_t = i|\mathbf{Q}_T\}$, for $i = 1, 2$, and $t = 1, 2, \ldots, T$. Figures F.1- F.5 illustrates the state probabilities and the mean level associated with the most probable states. At all sites, the state probabilities are very well defined, that is, most of the state probabilities are either close to zero or one.

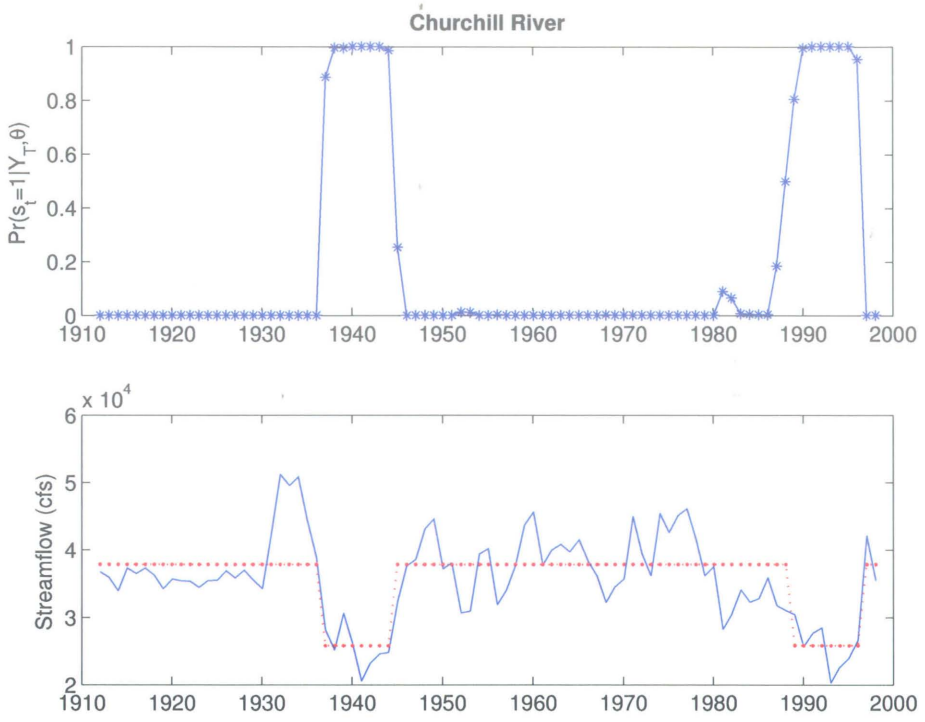


Figure F.1: State probabilities and regime shifts for the Churchill River annual runoff.

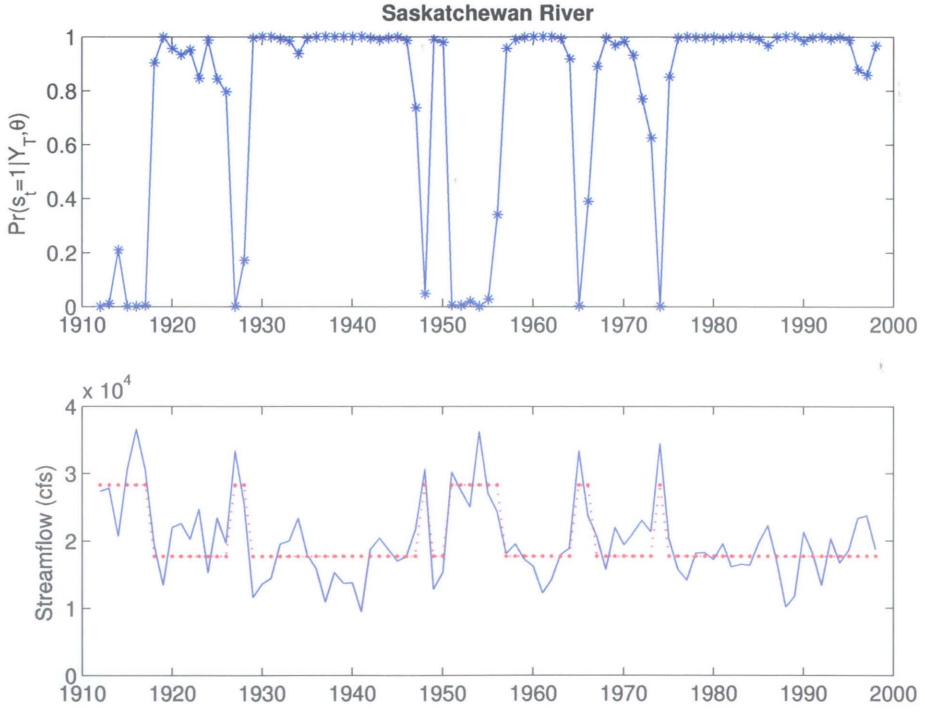


Figure F.2: State probabilities and regime shifts for the Saskatchewan River annual runoff.

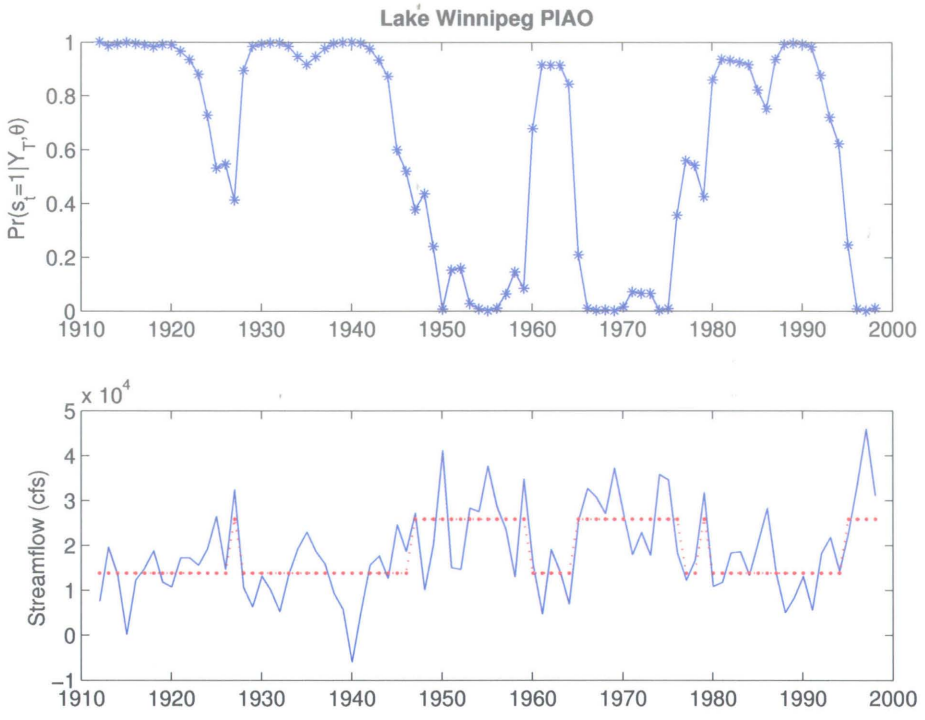


Figure F.3: State probabilities and regime shifts for the annual Lake Winnipeg PIAO.

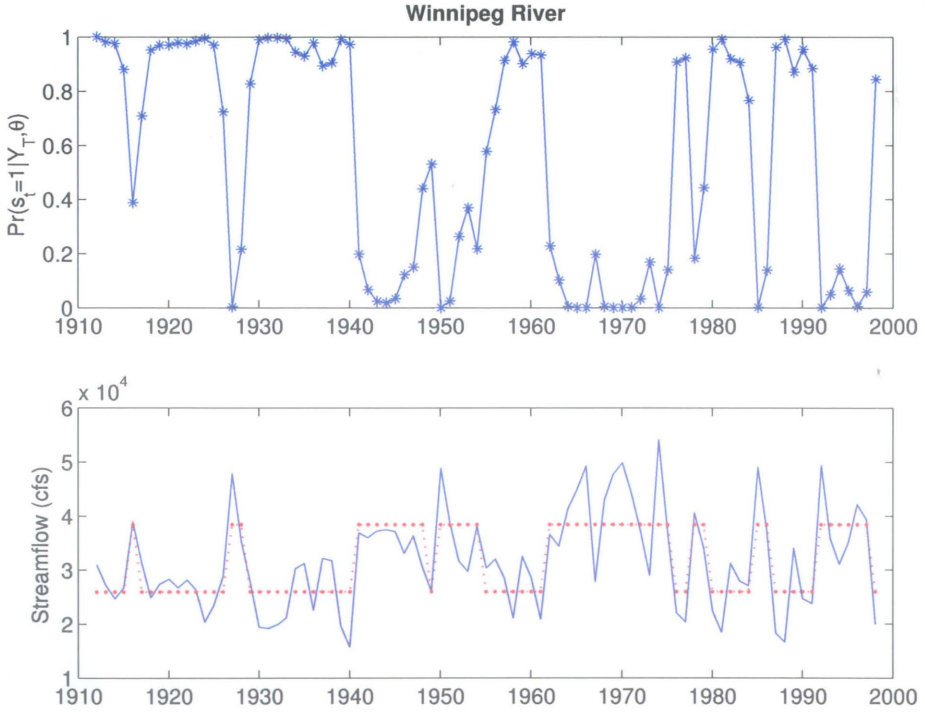


Figure F.4: State probabilities and regime shifts for the Winnipeg River annual runoff.

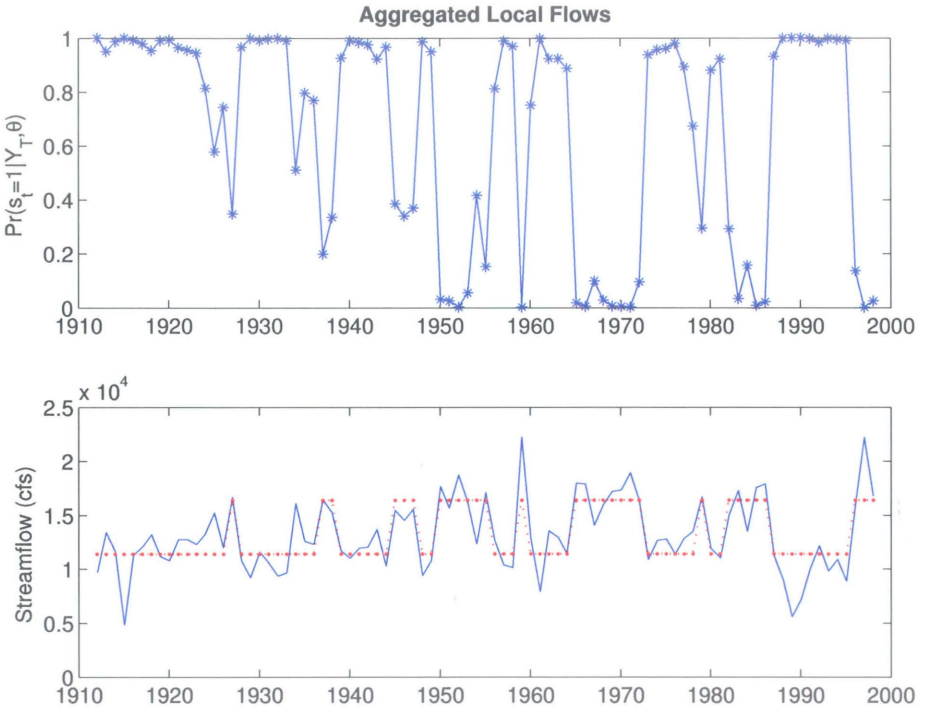


Figure F.5: State probabilities and regime shifts for the annual Aggregated Local Flows.

Appendix G

The Single-site 2-state MS Model: Autocorrelation Function Comparison

G.1 Introduction

It is of interest to examine and compare the autocorrelation function of the 2-state MS model and the conventional AR(1) and ARMA(1,1) models (for more details see *Akıntuğ and Rasmussen* [2005b]). Figures G.1-G.6 show the autocorrelation function for the equivalent ARMA(1,1) model denoted ARMA(1,1)-MS (see Section 5.3.4 for details). The result, obtained from the theoretical autocorrelation equations, confirms that the autocorrelation structure of a 2-state MS-model can be exactly reproduced by a properly selected ARMA(1,1) model. The following figures also show the autocor-

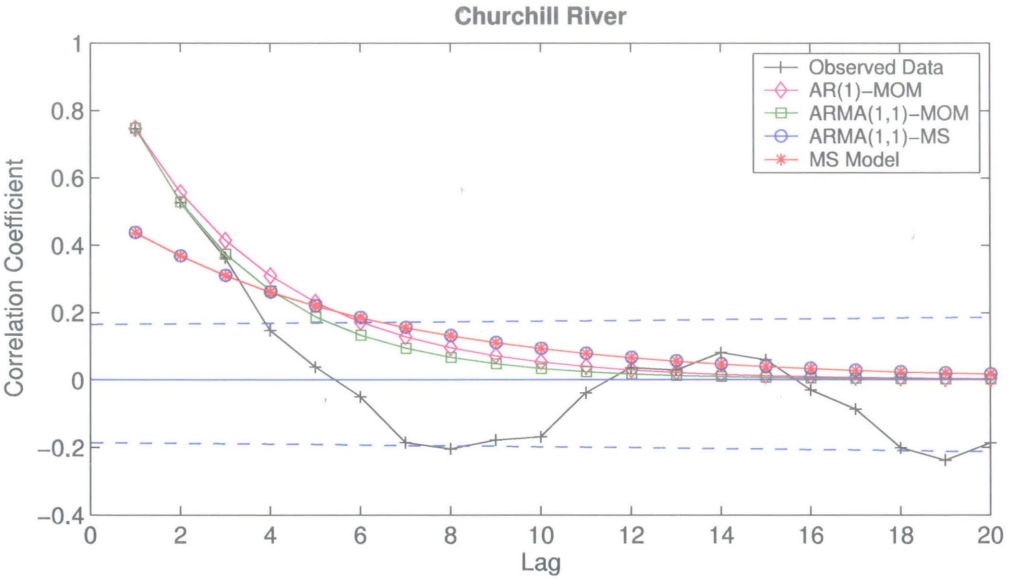


Figure G.1: Churchill River. Observed and modeled autocorrelation functions.

relation function of ARMA(1,1) and AR(1) models fitted directly to the data by the method of moments (ARMA(1,1)-MOM and AR(1)-MOM, respectively) and 95% significance levels. At all sites, the observed short-term (lag-1) correlation is not captured by the MS model. On the other hand, the AR(1)-MOM and ARMA(1,1)-MOM models preserve very well the first and the first two autocorrelation coefficients, respectively. For higher lags, the ARMA-type models in some cases considerably underestimate the observed autocorrelation whereas the MS model does a more reasonable job in preserving the observed autocorrelation.

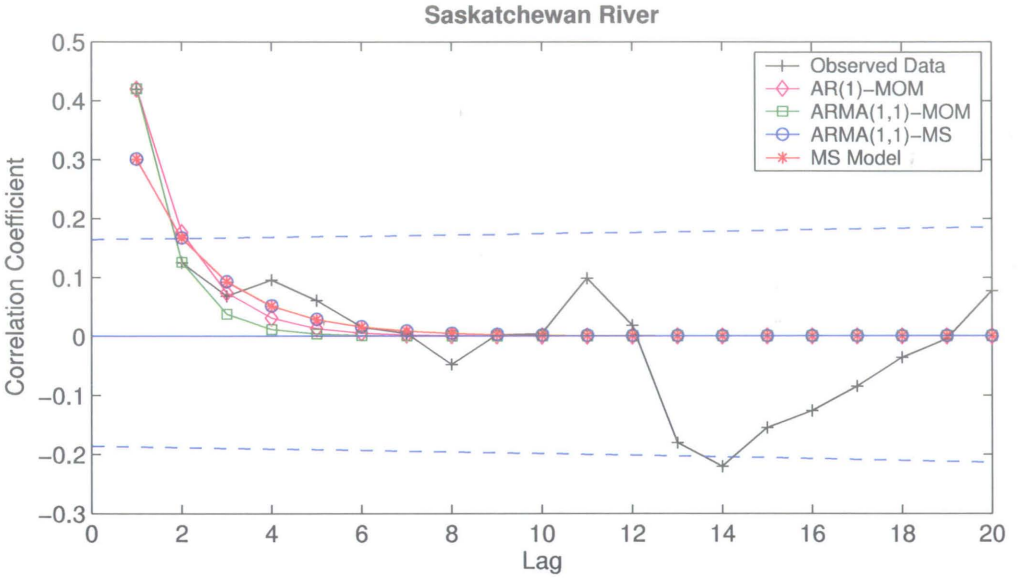


Figure G.2: Saskatchewan River. Observed and modeled autocorrelation functions.

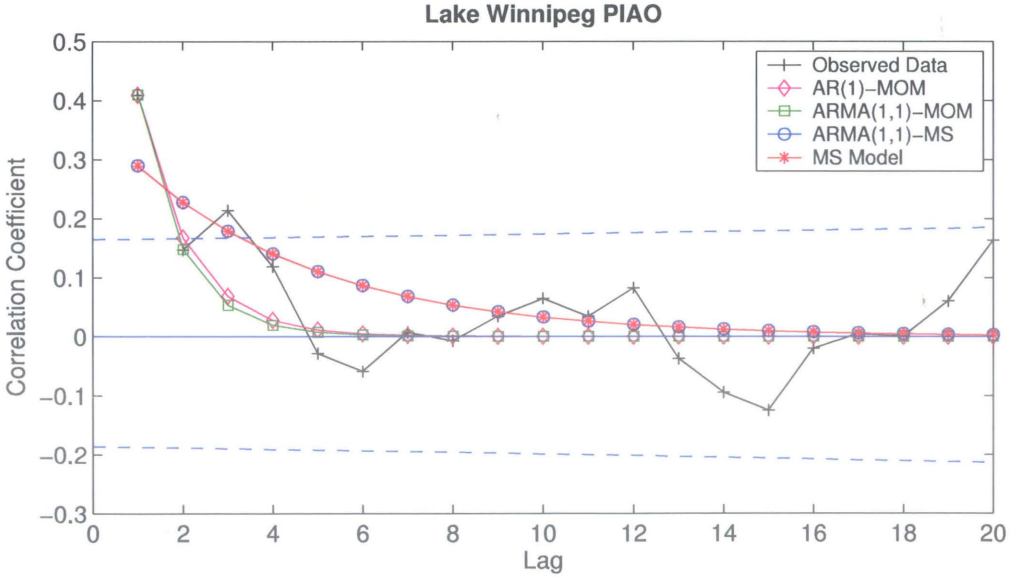


Figure G.3: Lake Winnipeg PIAO. Observed and modeled autocorrelation functions.

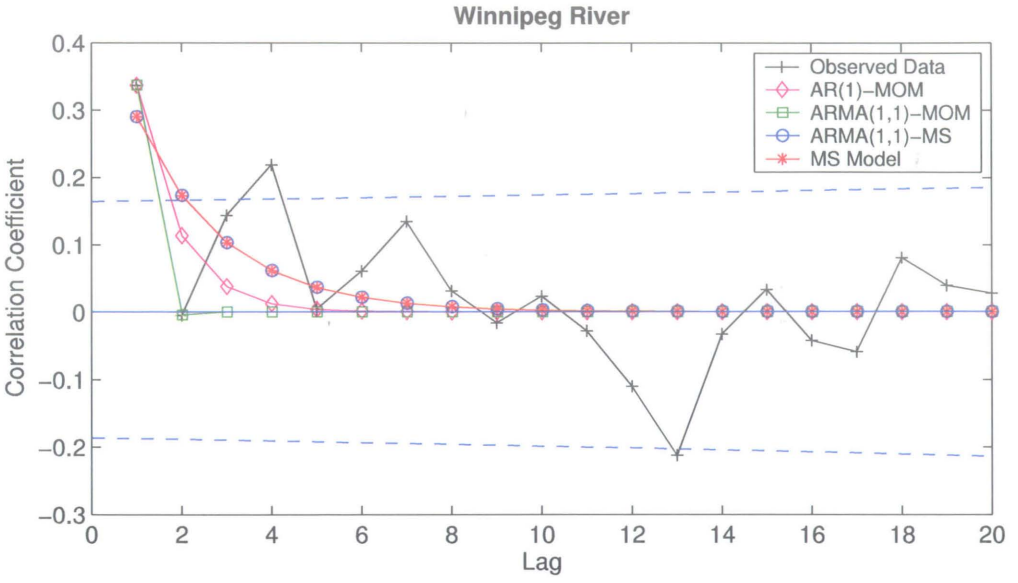


Figure G.4: Winnipeg River. Observed and modeled autocorrelation functions.

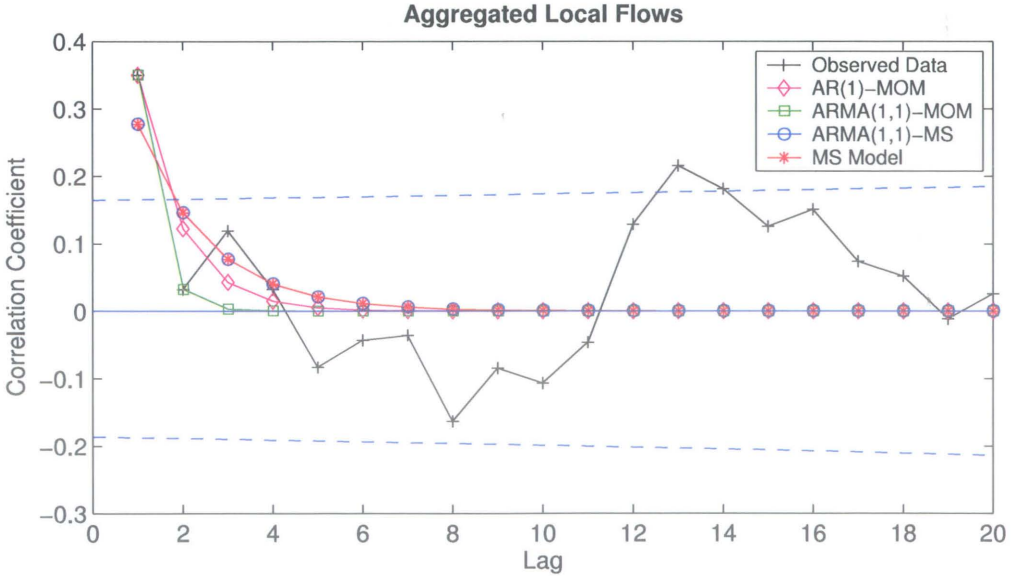


Figure G.5: Aggregated Local Flows. Observed and modeled autocorrelation functions.

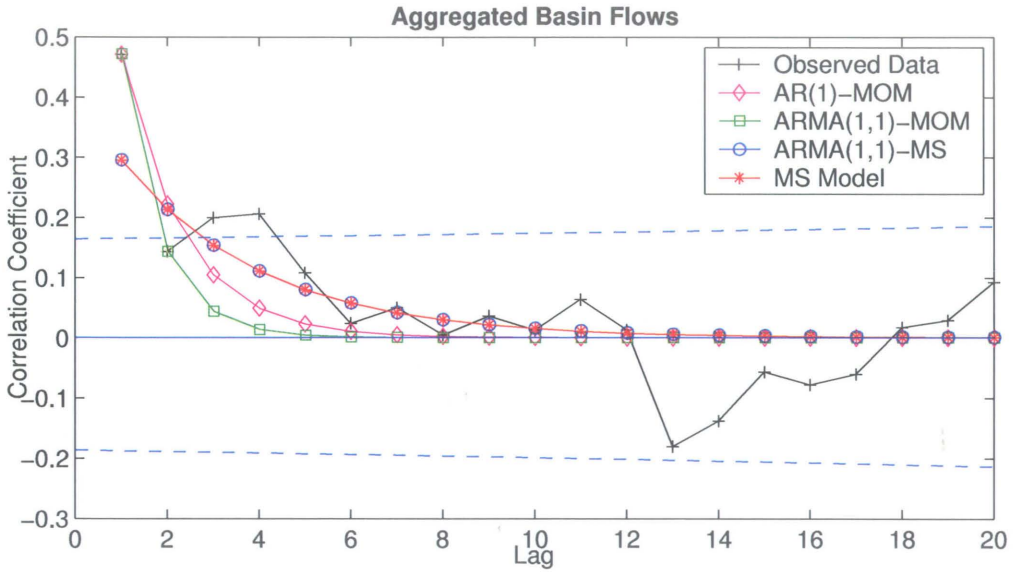


Figure G.6: Aggregated Basin Flows. Observed and modeled autocorrelation functions.

Appendix H

Multi-site MS Model Probability Plots

H.1 Introduction

In the MS model, the agreement between the observed data and the fitted distributions is assessed graphically using probability plots. Figures H.1-H.5 illustrate the associated marginal distributions for the five sites using the 2-state multi-site MS model. When sites are considered individually, the goodness-of-fit tends to be better, however, good fits are obtained with five sites as well.

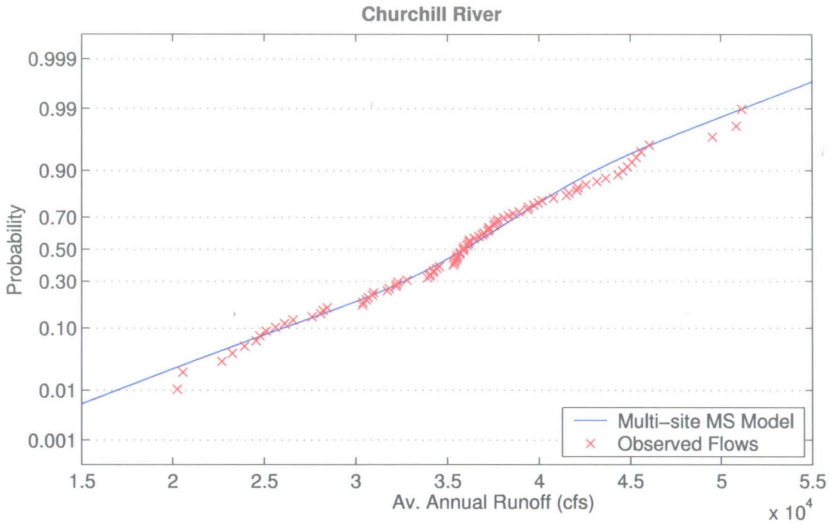


Figure H.1: Churchill River. Normal probability plot of observed annual flows and fitted multi-site MS model.

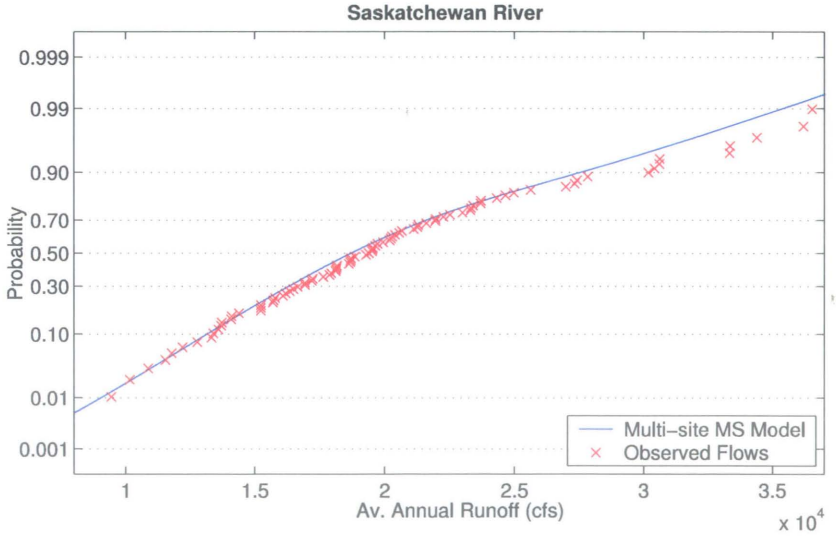


Figure H.2: Saskatchewan River. Normal probability plot of observed annual flows and fitted multi-site MS model.

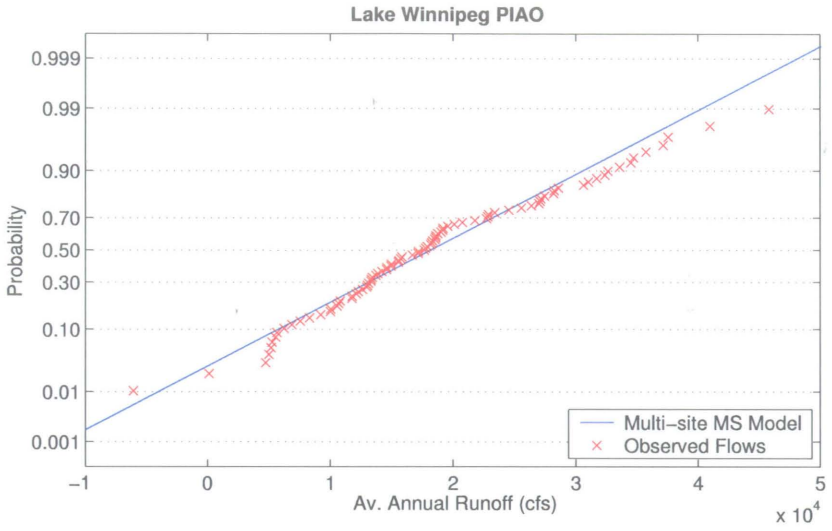


Figure H.3: Lake Winnipeg PIAO. Normal probability plot of observed annual flows and fitted multi-site MS model.

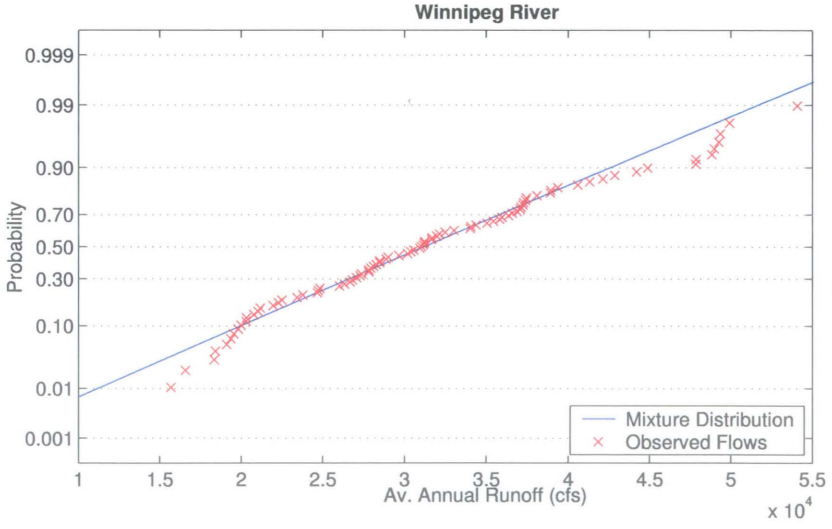


Figure H.4: Winnipeg River. Normal probability plot of observed annual flows and fitted multi-site MS model.

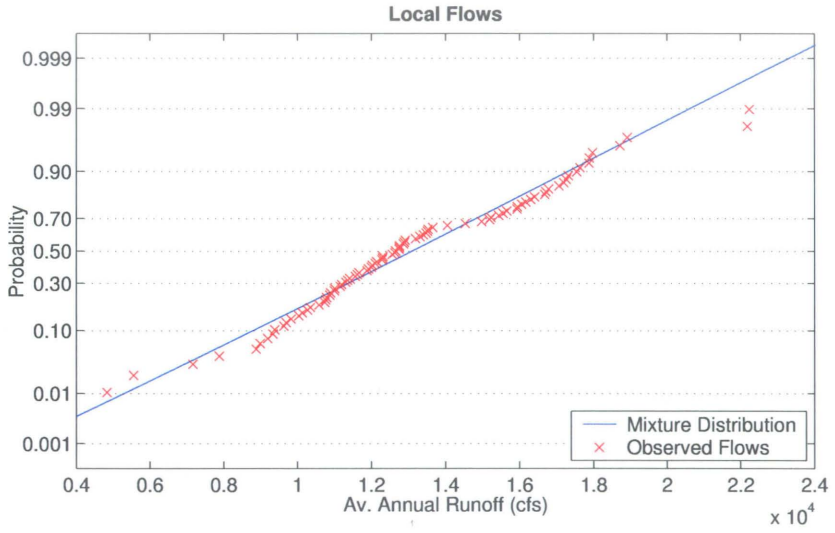


Figure H.5: Aggregated Local Flows. Normal probability plot of observed annual flows and fitted multi-site MS model.

Appendix I

Autocorrelation Function Comparison of MS Models

I.1 Introduction

Figures I.1-I.5 show the autocorrelation function for the single-site MS model, the multi-site MS model, and the MS Disaggregation model along with the observed autocorrelation function. The result, obtained from the theoretical autocorrelation equations, confirms that the fit is much better when sites are considered individually compared to the simultaneous fitting of the five sites. For some of the sites, the simultaneous fitting of the five sites considerably underestimates the observed autocorrelation.

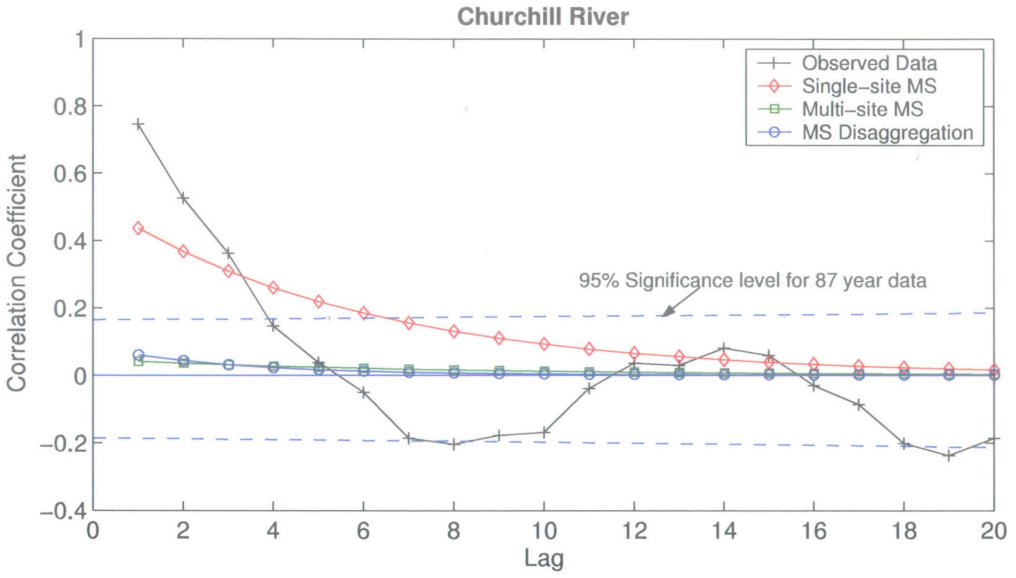


Figure I.1: Churchill River. Observed and modeled autocorrelation functions.

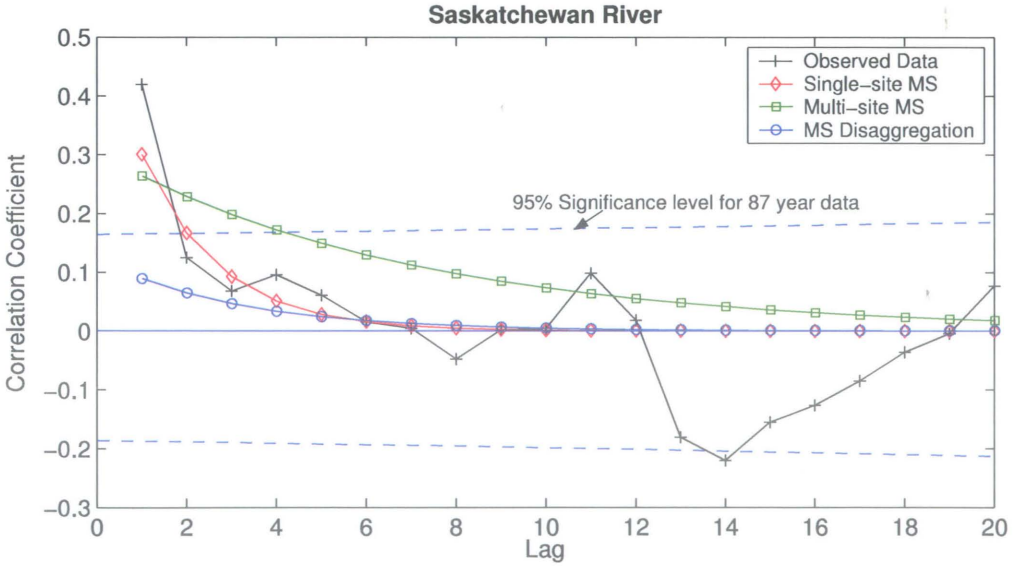


Figure I.2: Saskatchewan River. Observed and modeled autocorrelation functions.

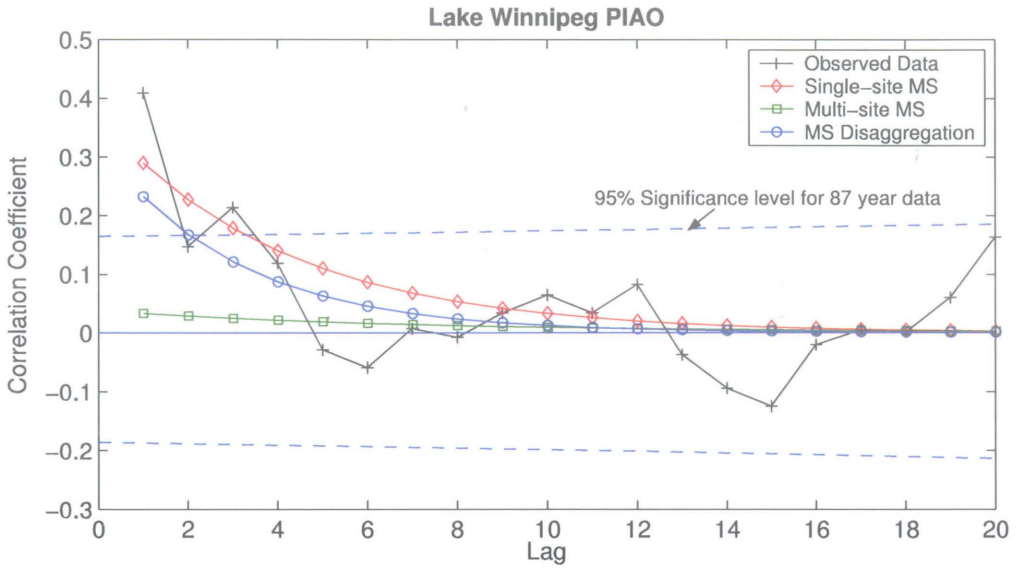


Figure I.3: Lake Winnipeg PIAO. Observed and modeled autocorrelation functions.

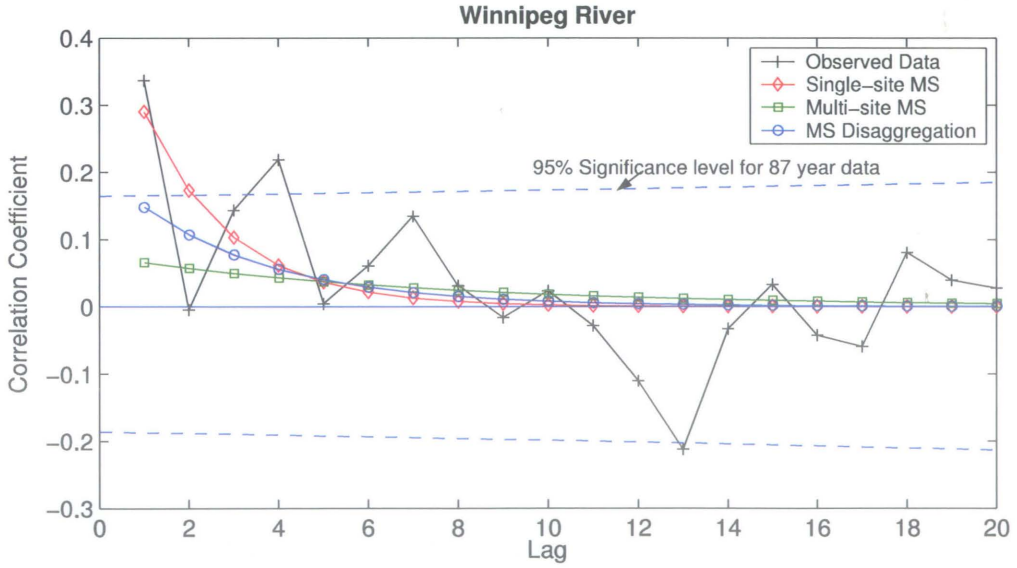


Figure I.4: Winnipeg River. Observed and modeled autocorrelation functions.

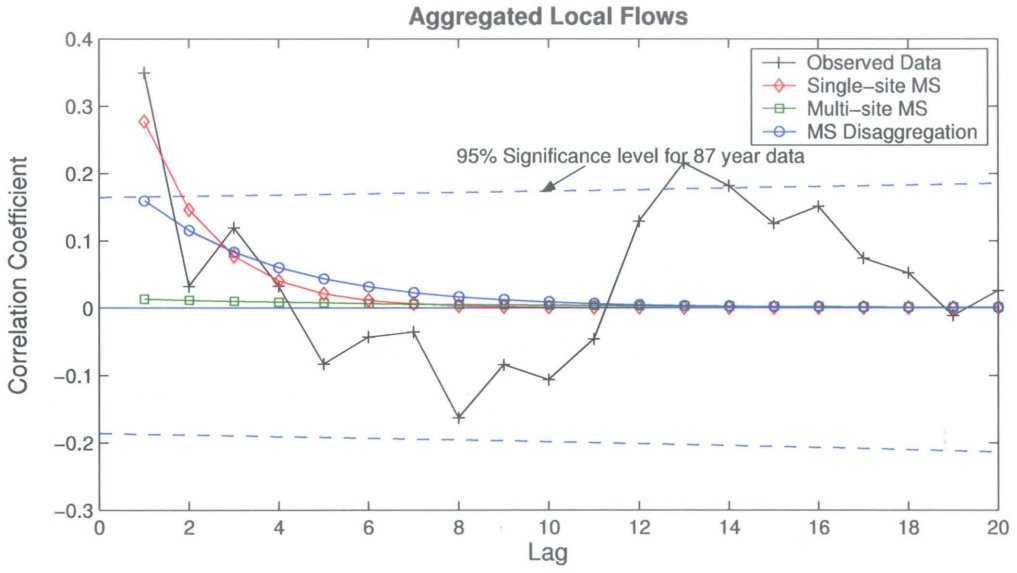


Figure I.5: Aggregated Local Flows. Observed and modeled autocorrelation functions.

Appendix J

The Multi-site 2-state MS Model: Posterior Distributions

J.1 Introduction

The simulated posterior distributions of multi-site MS model parameters are shown in the following sequence of figures. There is a good agreement between the maximum likelihood estimates and the posterior distributions of the model parameters. The ML-estimates are close to the mode of the posterior distributions.

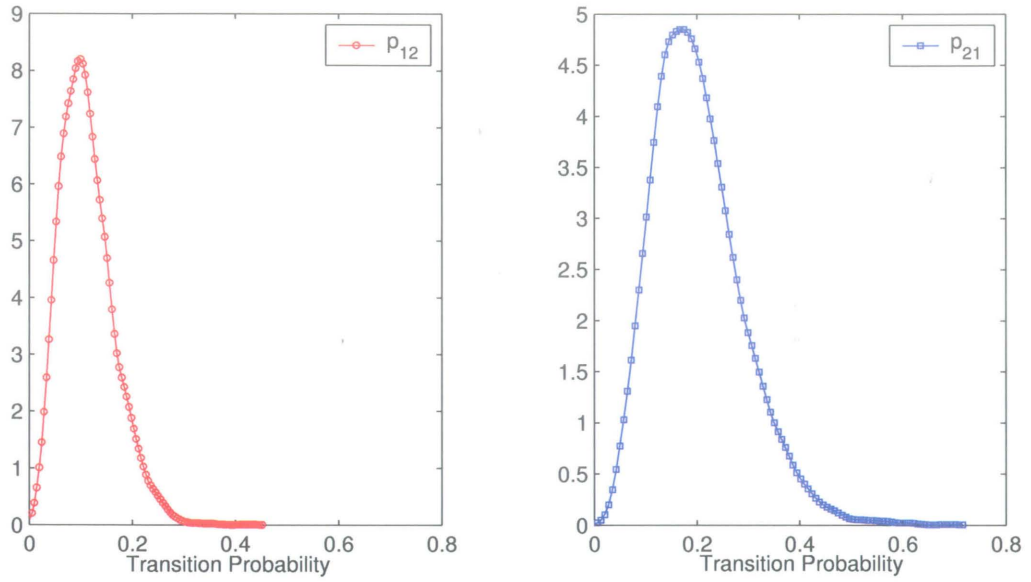


Figure J.1: Posterior distribution of transition probabilities for Multi-site MS model.

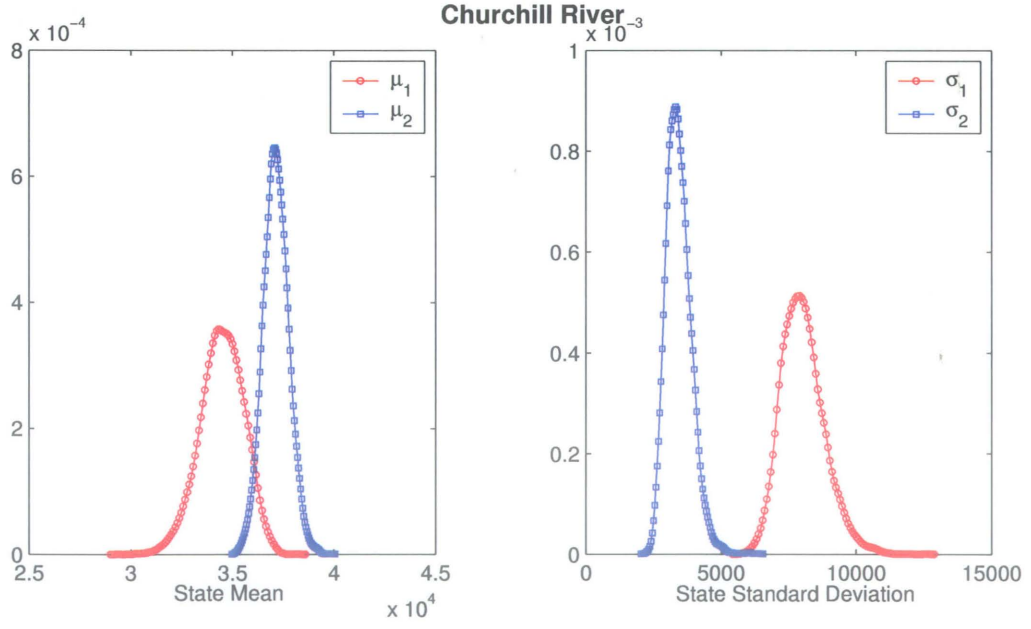


Figure J.2: Churchill River annual flows. Posterior distribution of state means and standard deviations for Multi-site MS model.

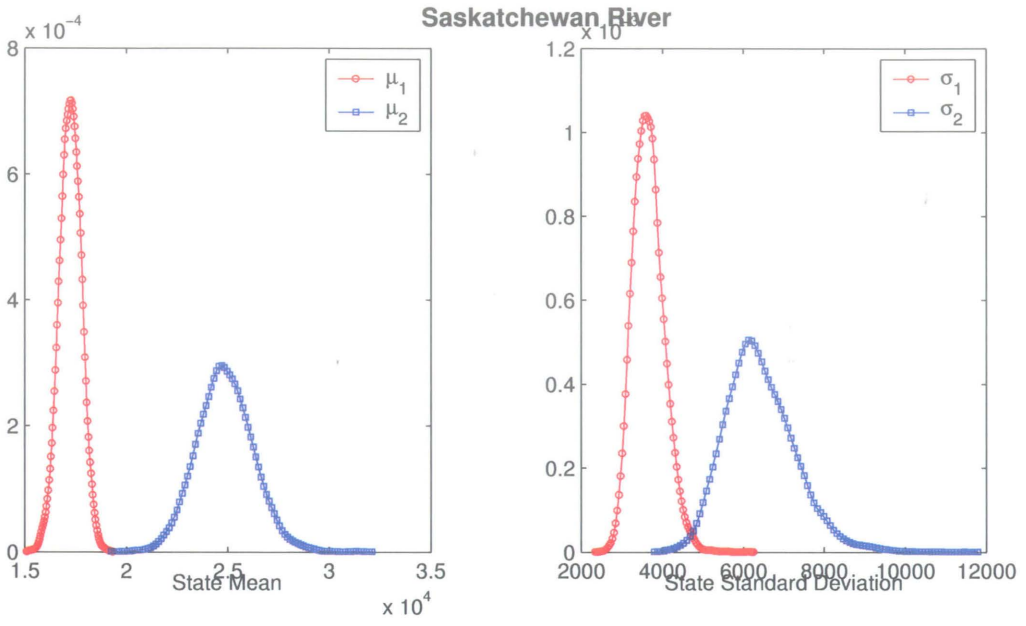


Figure J.3: Saskatchewan River annual flows. Posterior distribution of state means and standard deviations for Multi-site MS model.

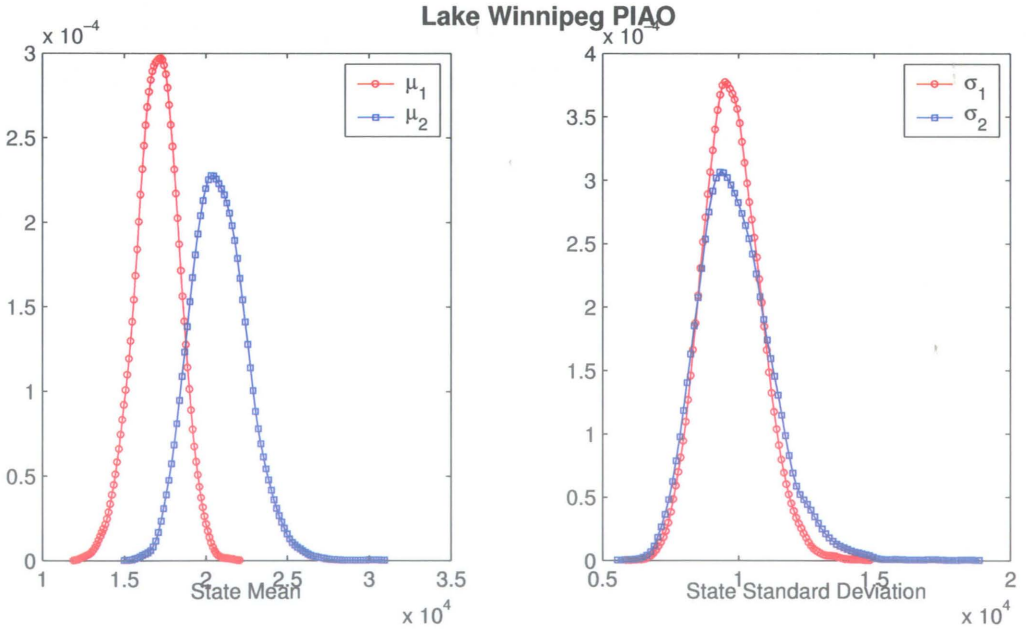


Figure J.4: Lake Winnipeg PIAO. Posterior distribution of state means and standard deviations for Multi-site MS model.

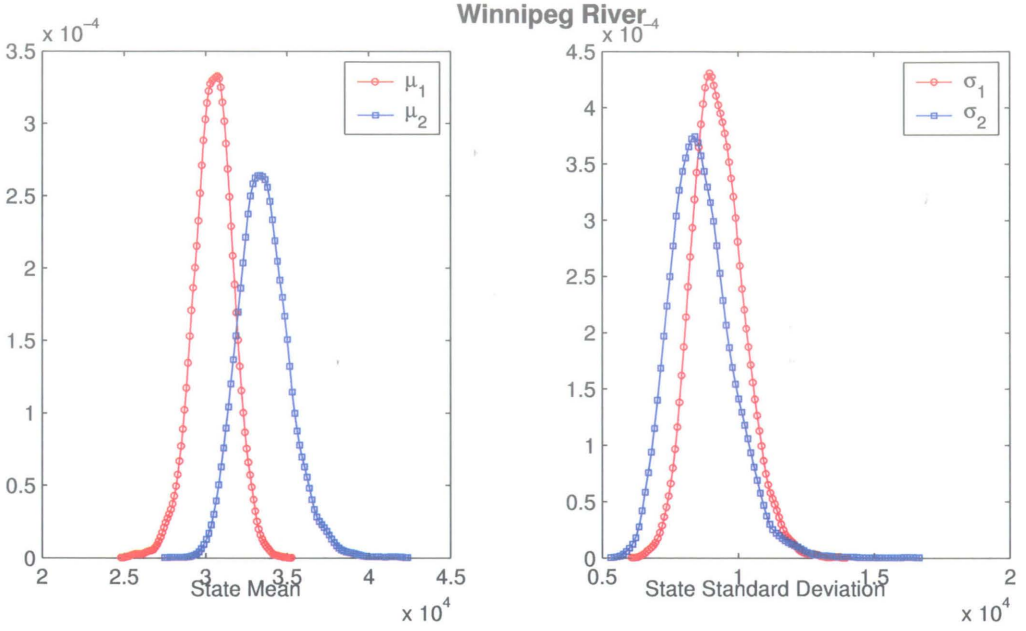


Figure J.5: Winnipeg River annual flows. Posterior distribution of state means and standard deviations for Multi-site MS model.

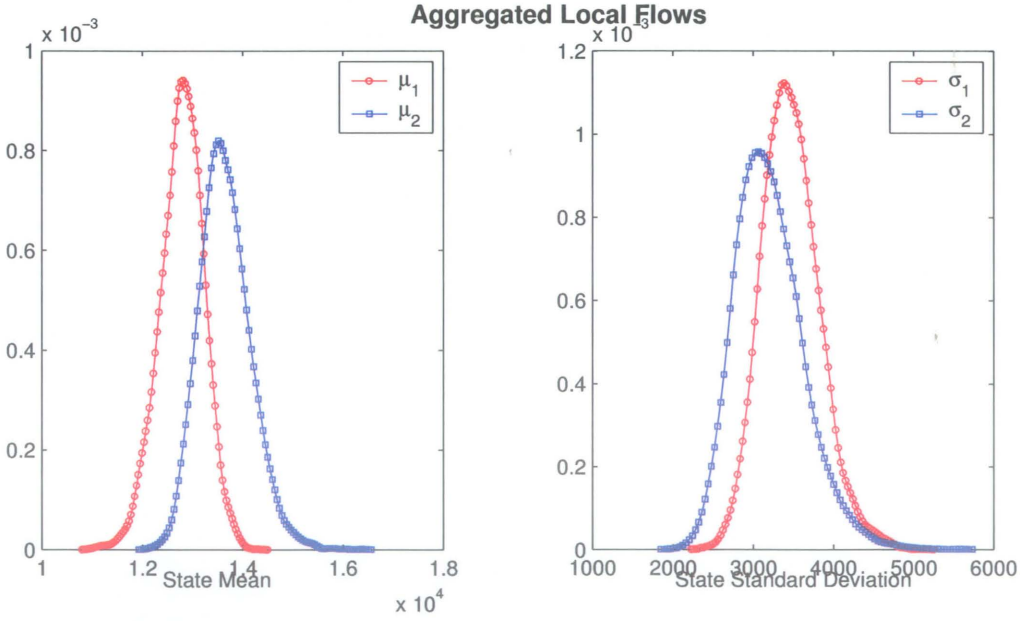


Figure J.6: Annual Aggregated Local Flows. Posterior distribution of state means and standard deviations for Multi-site MS model.

Appendix K

The Spatial 2-state MS Disaggregation Model: Fitted Marginal Distributions

K.1 Introduction

In the MSD spatial disaggregation model, Aggregated Basin Flows are first simulated with a single-site MS model and the MS disaggregation model parameters are estimated for the five sites. The goodness-of-fit of the marginal distribution of the MSD model is illustrated in the following figures. The model appears to fit the data well at all sites.

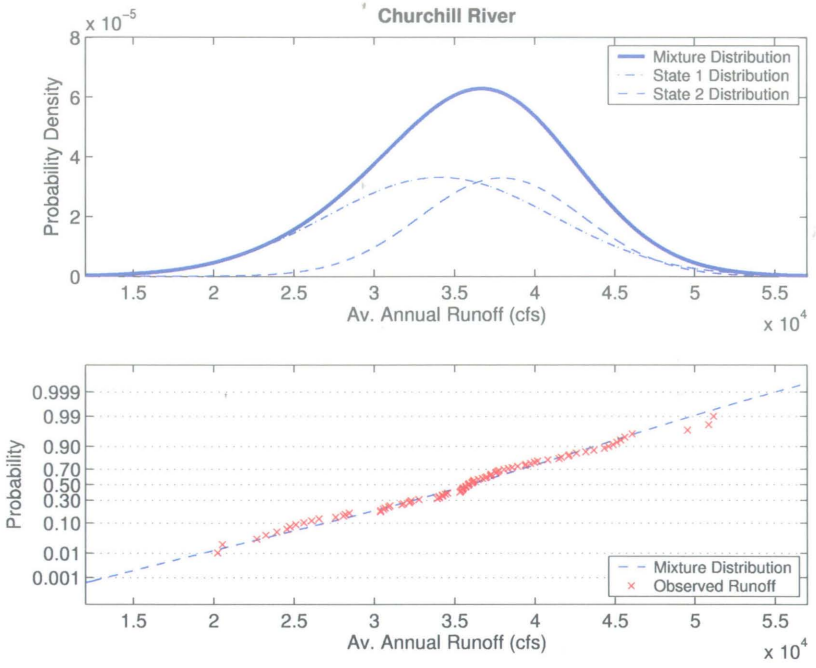


Figure K.1: Annual flows of the Churchill River at Southern Indian Lake. a) Marginal probability density function of MSD model. b) Normal probability plot of observed annual flows and fitted MSD model.

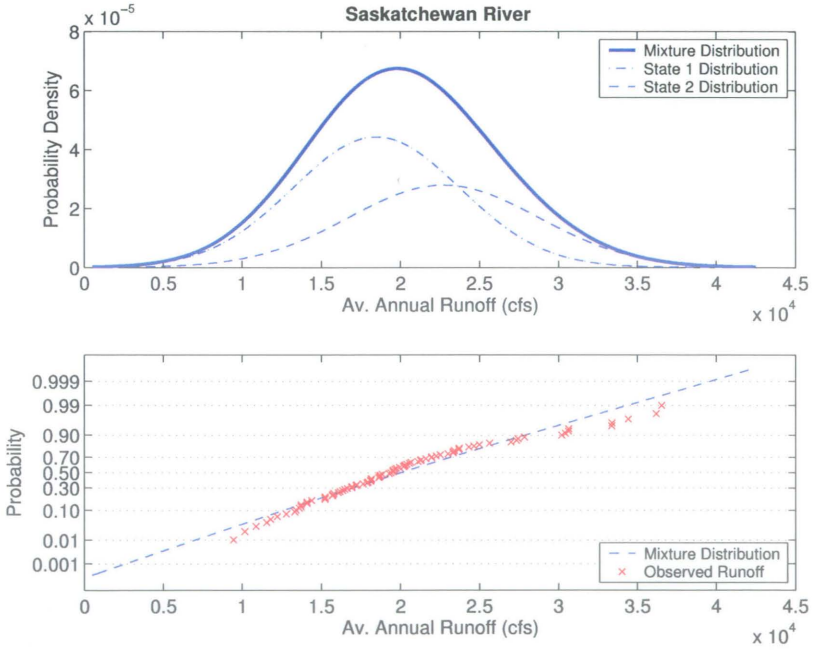


Figure K.2: Annual flows of the Saskatchewan River at Grand Rapids. a) Marginal probability density function of MSD model. b) Normal probability plot of observed annual flows and fitted MSD model.

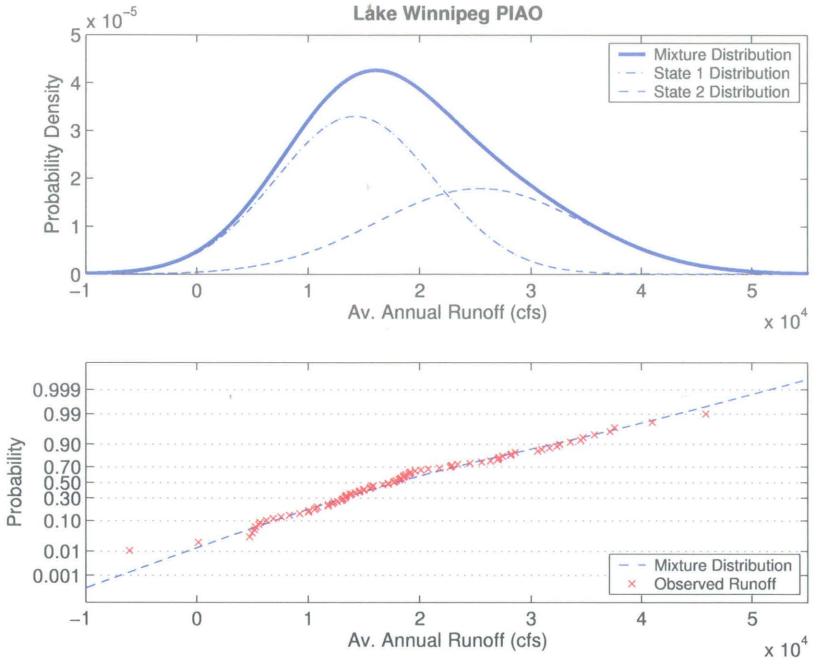


Figure K.3: Annual flows of the Lake Winnipeg PIAO. a) Marginal probability density function of MSD model. b) Normal probability plot of observed annual flows and fitted MSD model.

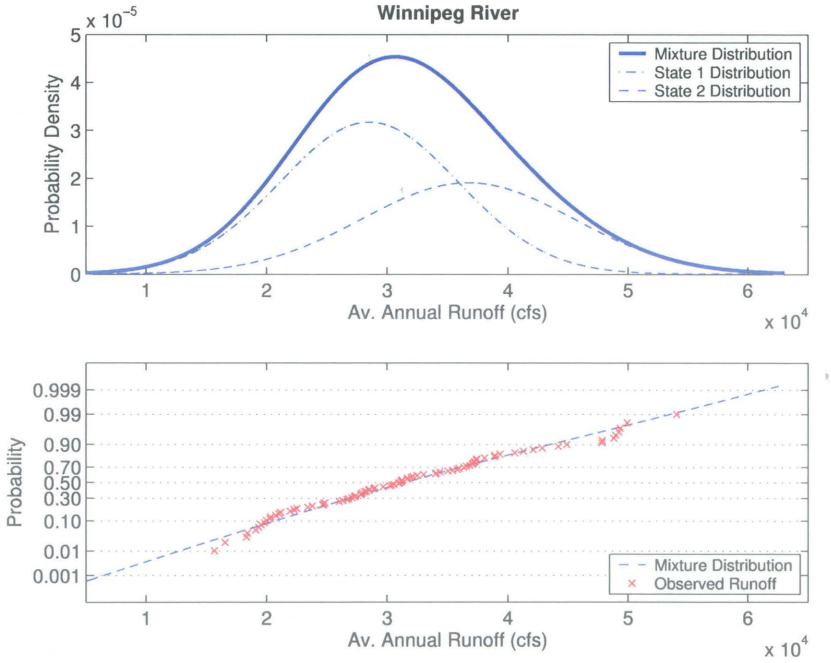


Figure K.4: Annual flows of Winnipeg River at Slave Falls. a) Marginal probability density function of MSD model. b) Normal probability plot of observed annual flows and fitted MSD model.

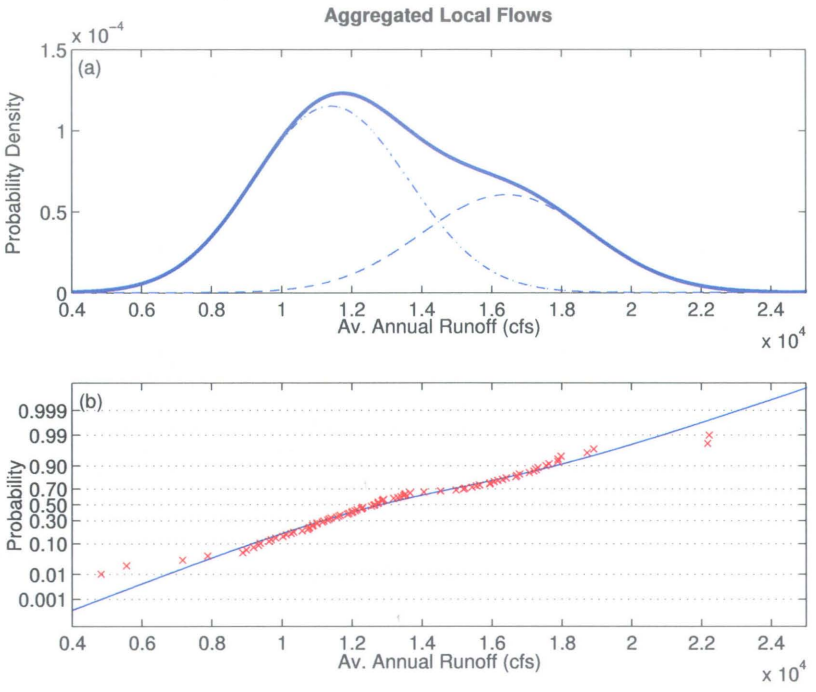


Figure K.5: Annual flows of Aggregated Local Flows in Burntwood and Nelson River. a) Marginal probability density function of MSD model. b) Normal probability plot of observed annual flows and fitted MSD model.

Appendix L

The Spatial 2-state MS Disaggregation Model: Regime Changes in Observed Streamflow

L.1 Introduction

Figure L.1 illustrates the state probabilities for the MSD model, estimated using the single-site MS model for Aggregated Basin Flows. The mean levels associated with the most probable states are given in Figures L.2-L.6. Because of the constraint that all sites must have the same state sequence, the mean levels are closer to each other at all sites compared to the single-site models.

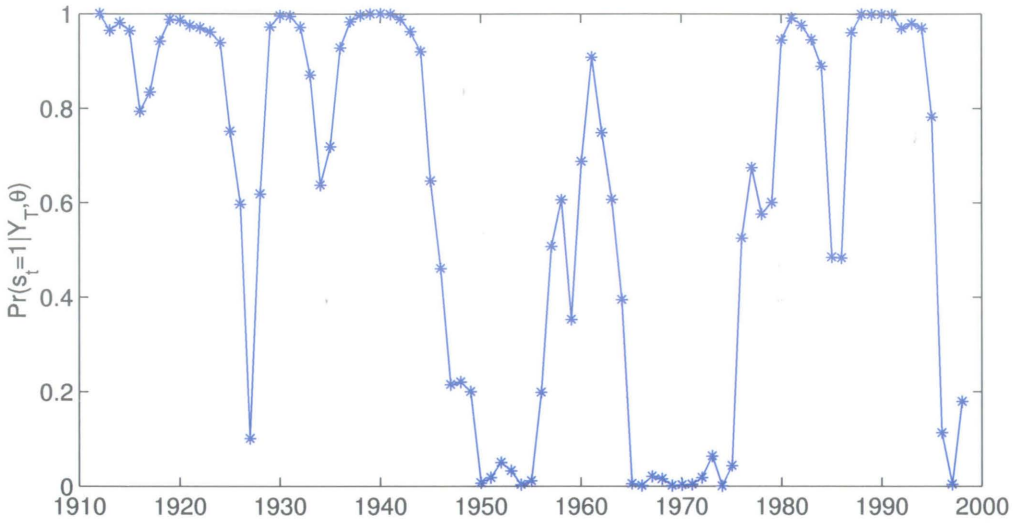


Figure L.1: The MSD model state probabilities for all five sites.

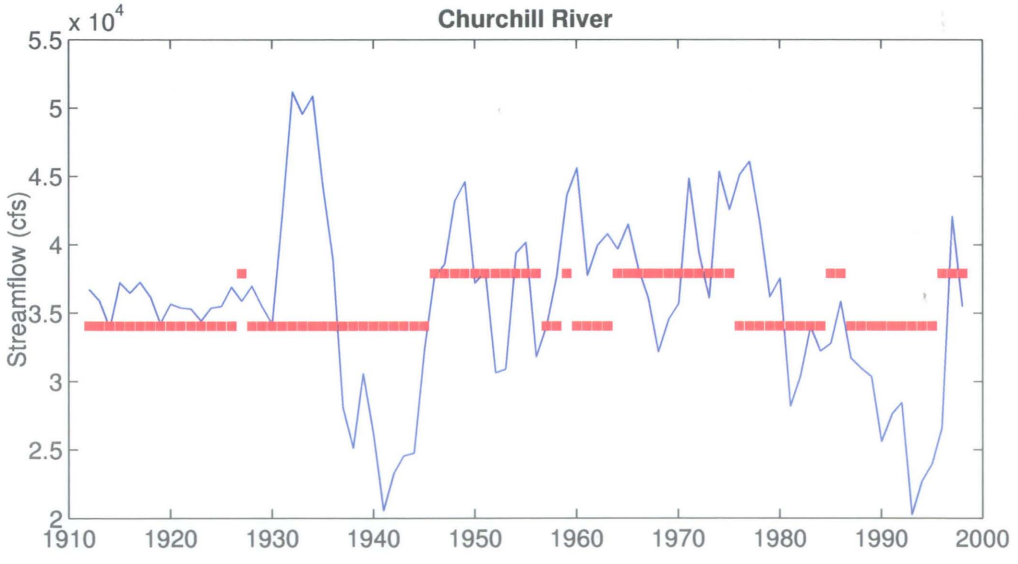


Figure L.2: The MSD model regime shifts for the Churchill River annual runoff.

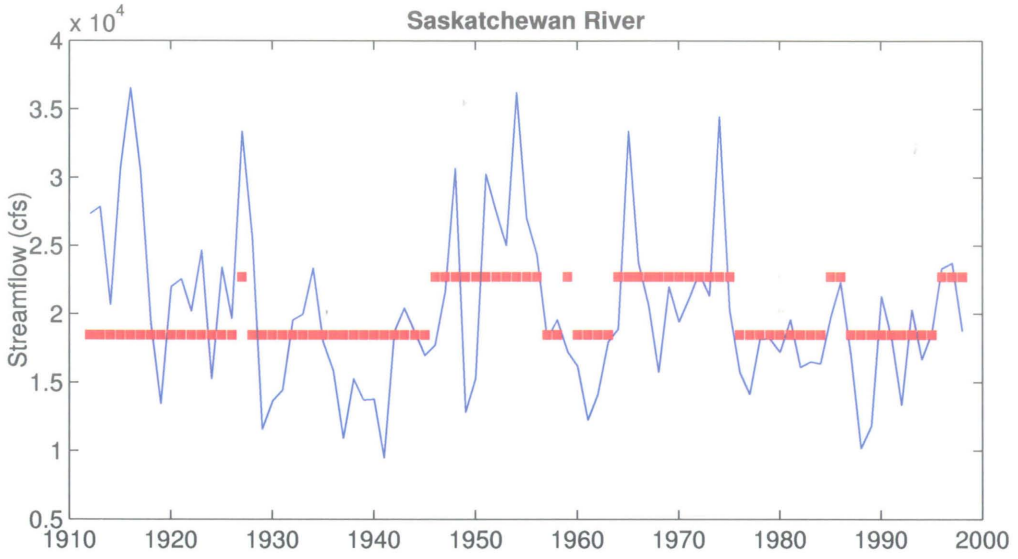


Figure L.3: The MSD model regime shifts for the Saskatchewan River annual runoff.

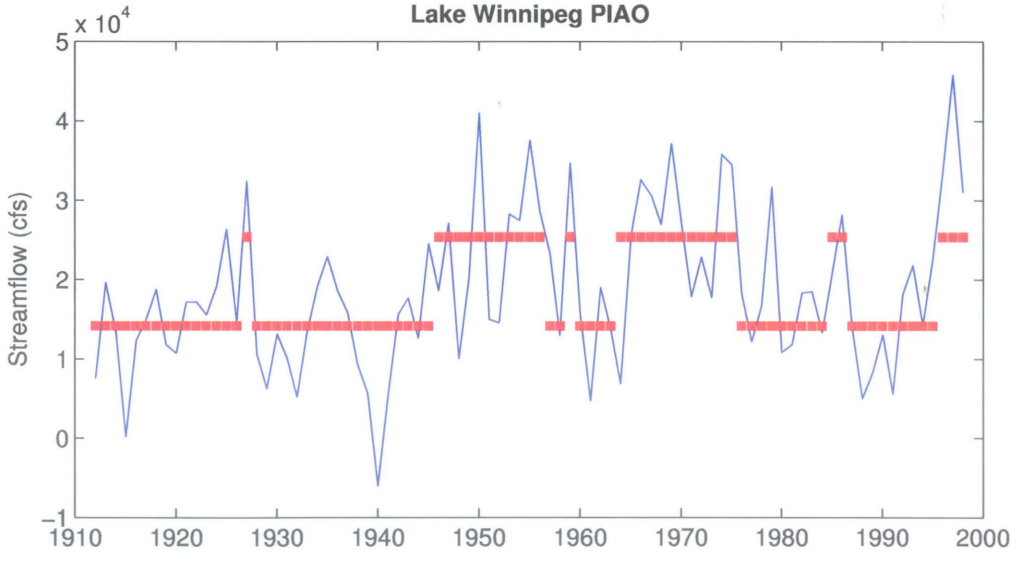


Figure L.4: The MSD model regime shifts for the annual Lake Winnipeg PIAO.

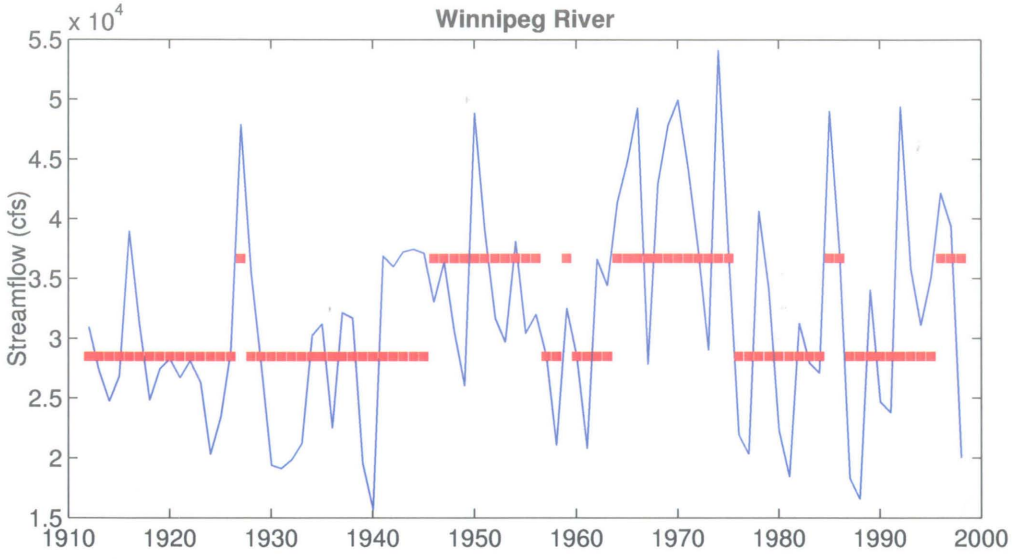


Figure L.5: The MSD model regime shifts for the Winnipeg River annual runoff.

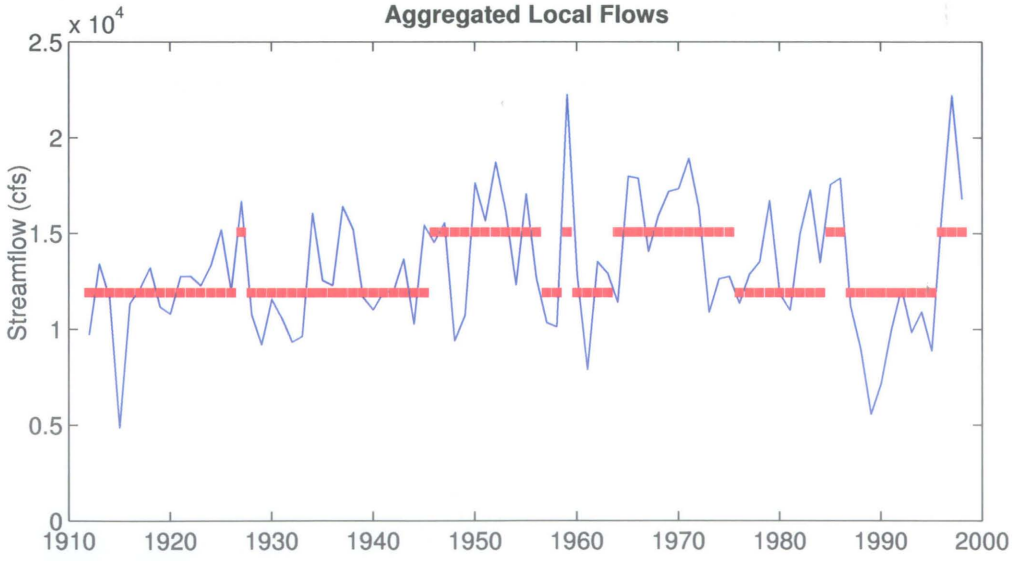


Figure L.6: The MSD model regime shifts for the Aggregated Local Flows annual runoff.

Appendix M

Verification of Stochastic Models

M.1 Modeling Framework 1.1

Simulated flows 100,000-yr

MODEL 1: SINGLE-SITE AR(1) ANNUAL MODEL

	MEAN	STANDARD DEVIATION	LAG-1 AUTOCORRELATION
Obs (Norm)	14.993	0.0865	0.489
Syn (Norm)	14.993	0.0862	0.488
Obs (Raw)	1427424	283063	0.471
Syn (Inv-T)	1425755	281094	0.487

MODEL 2: MEJIA AND ROUSSELLE SPATIAL DISAGGREGATION MODEL

MEAN

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	4772326	12.559	14.415	152.67	13.157
Syn (Norm)	4767817	12.559	14.415	152.70	13.157
Obs (Raw)	426644	241214	222048	379811	157705
Syn (Inv-T)	426295	241045	222121	380346	157733
Syn (Adj)	426389	241078	221294	378962	158032

STANDARD DEVIATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	1062897	0.237	0.063	13.352	0.076
Syn (Norm)	1058128	0.235	0.063	13.436	0.076
Obs (Raw)	79857	71610	116881	107091	39797
Syn (Inv-T)	79567	69346	115109	108414	39714
Syn (Adj)	82882	69905	114055	103815	41565

LAG-0 CROSS-CORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local	
Obs (Norm)	1	0.232	0.179	-0.029	0.151	Chur R
Syn (Norm)	1	0.226	0.172	-0.036	0.147	
Obs (Raw)	1	0.229	0.183	-0.022	0.159	
Syn (Inv-T)	1	0.223	0.171	-0.037	0.147	
Syn (Adj)	1	0.262	0.163	-0.049	0.226	
		1	0.333	0.356	0.189	Sask R
		1	0.355	0.361	0.208	
		1	0.295	0.352	0.166	
		1	0.351	0.354	0.204	
		1	0.342	0.346	0.249	
			1	0.541	0.693	PIAO
			1	0.548	0.693	
			1	0.539	0.696	
			1	0.543	0.692	
			1	0.532	0.671	
				1	0.463	Wpg R
				1	0.471	
				1	0.470	
				1	0.466	
				1	0.467	
					1	Local
					1	
					1	
					1	
					1	

LAG-1 AUTOCORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	0.743	0.428	0.413	0.343	0.358
Syn (Norm)	0.739	0.413	0.465	0.300	0.446
Obs (Raw)	0.746	0.420	0.409	0.337	0.350
Syn (Inv-T)	0.739	0.405	0.465	0.296	0.445
Syn (Adj)	0.730	0.401	0.470	0.304	0.450

M.2 Modeling Framework 1.3

Simulated flows 100,000-yr

MODEL 1: SINGLE-SITE MS(2) ANNUAL MODEL

	MEAN	STANDARD DEVIATION	LAG-1 AUTOCORRELATION
Obs	1427424	283063	0.471
Syn	1435796	282509	0.296

MODEL 2: MARKOV-SWITCHING SPATIAL DISAGGREGATION MODEL

MEAN

	Chur R	Sask R	PIAO	Wpg R	Local
Obs	426644	241214	222048	379811	157705
Syn	427870	242105	225375	381800	158652

STANDARD DEVIATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs	79857	71610	116881	107091	39797
Syn	78654	71288	117148	106892	39560

LAG-1 AUTOCORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs	0.746	0.420	0.409	0.337	0.350
Syn	0.059	0.084	0.233	0.146	0.155

LAG-0 CROSS-CORRELATION

Obs	1	0.229	0.183	-0.022	0.159	Chur R
Syn	1	0.227	0.183	-0.017	0.157	
		1	0.295	0.352	0.160	Sask R
		1	0.294	0.357	0.164	
			1	0.539	0.696	PIAO
			1	0.540	0.694	
				1	0.470	Wpg R
				1	0.472	
					1	Local
					1	

==========

TRANSFORMED (NORMAL) ANNUAL FLOW STATISTICS

==========

MEAN

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	4772327	12.559	14.415	152.668	13.157
Syn (Norm)	4770589	12.557	14.415	152.560	13.156

STANDARD DEVIATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	1062897	0.237	0.064	13.352	0.0764
Syn (Norm)	1060528	0.236	0.063	13.320	0.0760

LAG-1 AUTOCORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	0.743	0.428	0.413	0.343	0.356
Syn (Norm)	0.056	0.083	0.221	0.142	0.152

LAG-0 CROSS-CORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local	
Obs (Norm)	1	0.232	0.179	-0.029	0.151	Chur R
Syn (Norm)	1	0.224	0.179	-0.024	0.152	
		1	0.333	0.356	0.189	Sask R
		1	0.292	0.352	0.160	
			1	0.541	0.693	PIAO
			1	0.534	0.694	
				1	0.463	Wpg R
				1	0.466	
					1	Local
					1	

M.3 Modeling Framework 2.1

Simulated flows 100,000-yr

MODEL 1: MULTI-SITE AR(1) MODEL

MEAN

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	4772327	12.559	14.415	152.668	13.157
Syn (Norm)	4762787	12.560	14.416	152.729	13.157
Obs (Raw)	426644	241214	222048	379811	157705
Syn (Inv-T)	425943	241163	223074	380223	157767

STANDARD DEVIATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	1062897	0.237	0.064	13.352	0.076
Syn (Norm)	1048865	0.233	0.062	13.227	0.075
Obs (Raw)	79857	71611	116881	107091	39797
Syn (Inv-T)	78897	69348	113996	106714	39066

LAG-1 AUTOCORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	0.743	0.428	0.413	0.344	0.358
Syn (Norm)	0.760	0.445	0.425	0.348	0.366
Obs (Raw)	0.746	0.420	0.409	0.337	0.350
Syn (Inv-T)	0.760	0.438	0.424	0.345	0.365

LAG-0 CROSS-CORRELATION

Obs (Norm)	1	0.232	0.180	-0.029	0.151	Chur R
Syn (Norm)	1	0.238	0.184	-0.028	0.157	
Obs (Raw)	1	0.229	0.183	-0.022	0.159	
Syn (Inv-T)	1	0.235	0.184	-0.029	0.157	
		1	0.333	0.356	0.189	Sask R
		1	0.349	0.356	0.201	
		1	0.295	0.352	0.160	
		1	0.344	0.351	0.197	
			1	0.541	0.693	PIAO
			1	0.543	0.688	
			1	0.539	0.696	
			1	0.539	0.688	
				1	0.463	Wpg R
				1	0.461	
				1	0.470	
				1	0.457	
					1	Local
					1	
					1	
					1	

M.4 Modeling Framework 2.3

Simulated flows 100,000-yr

MODEL 1: MULTI-SITE MS(2) ANNUAL MODEL (Annual)

MEAN

	Chur R	Sask R	PIAO	Wpg R	Local
Obs	426644	241214	222048	379811	157705
Syn	423557	235067	217752	374542	156590

STANDARD DEVIATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs	79857	71610	116881	107091	39797
Syn	83140	68751	116253	105860	39683

LAG-1 AUTOCORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs	0.746	0.420	0.409	0.337	0.350
Syn	0.044	0.265	0.036	0.068	0.013

LAG-0 CROSS-CORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local	
Obs	1	0.229	0.183	-0.022	0.159	Chur R
Syn	1	0.226	0.185	-0.059	0.167	
		1	0.295	0.352	0.160	Sask R
		1	0.336	0.354	0.184	
			1	0.539	0.696	PIAO
			1	0.535	0.694	
				1	0.470	Wpg R
				1	0.460	
					1	Local
					1	

TRANSFORMED (NORMAL) ANNUAL FLOW STATISTICS

MEAN

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	4772327	12.559	14.415	152.668	13.157
Syn (Norm)	4770147	12.559	14.415	152.603	13.156

STANDARD DEVIATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	1062897	0.237	0.064	13.352	0.0764
Syn (Norm)	1062881	0.238	0.064	13.331	0.0759

LAG-1 AUTOCORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local
Obs (Norm)	0.743	0.428	0.413	0.343	0.356
Syn (Norm)	0.038	0.232	0.036	0.068	0.013

LAG-0 CROSS-CORRELATION

	Chur R	Sask R	PIAO	Wpg R	Local	
Obs (Norm)	1	0.232	0.179	-0.029	0.151	Chur R
Syn (Norm)	1	0.223	0.180	-0.050	0.158	
		1	0.333	0.356	0.189	Sask R
		1	0.377	0.353	0.208	
			1	0.541	0.693	PIAO
			1	0.535	0.694	
				1	0.463	Wpg R
				1	0.460	
					1	Local
					1	

Appendix N

Conditional Drought Frequency Curves

N.1 Introduction

As part of the drought frequency analysis, the frequency curves for drought severity conditional upon drought duration are plotted for all modeling frameworks and the two data sets. The conditional frequency curves are presented as Severity-Duration-Frequency curves in Figures N.1-N.18.

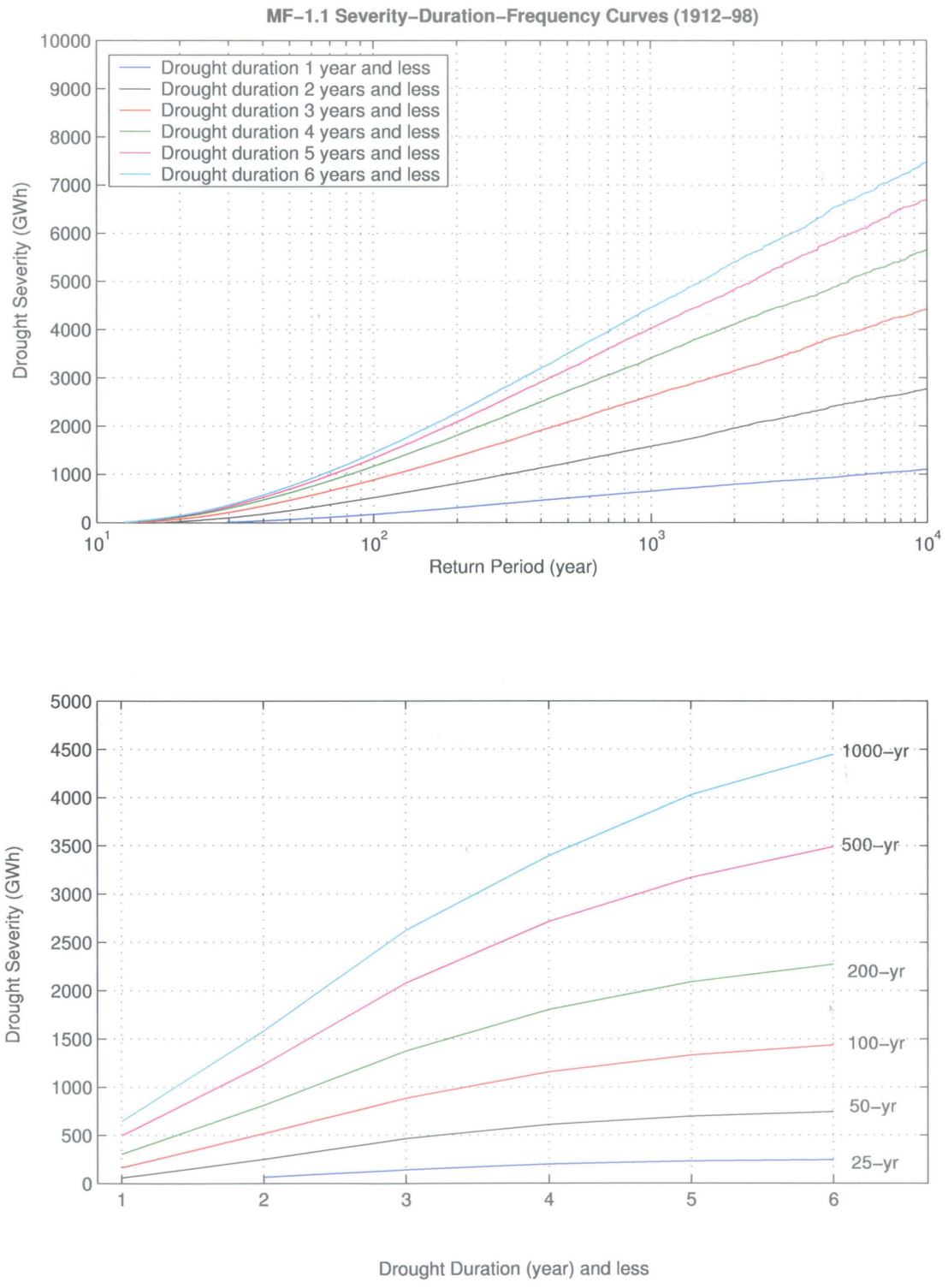


Figure N.1: Conditional Drought Frequency Curves obtained from MF-1.1 (1912-98).

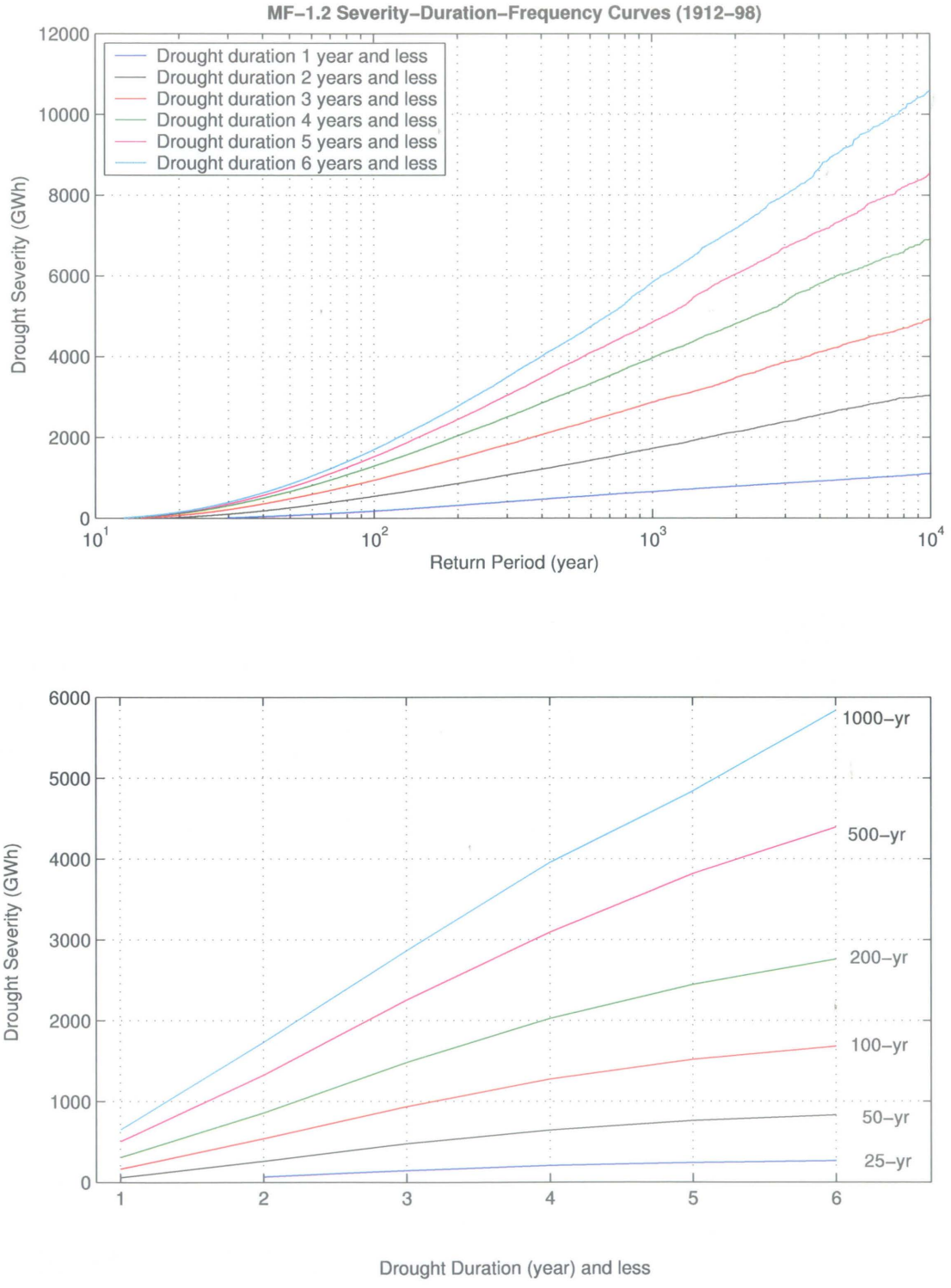


Figure N.2: Conditional Drought Frequency Curves obtained from MF-1.2 (1912-98).

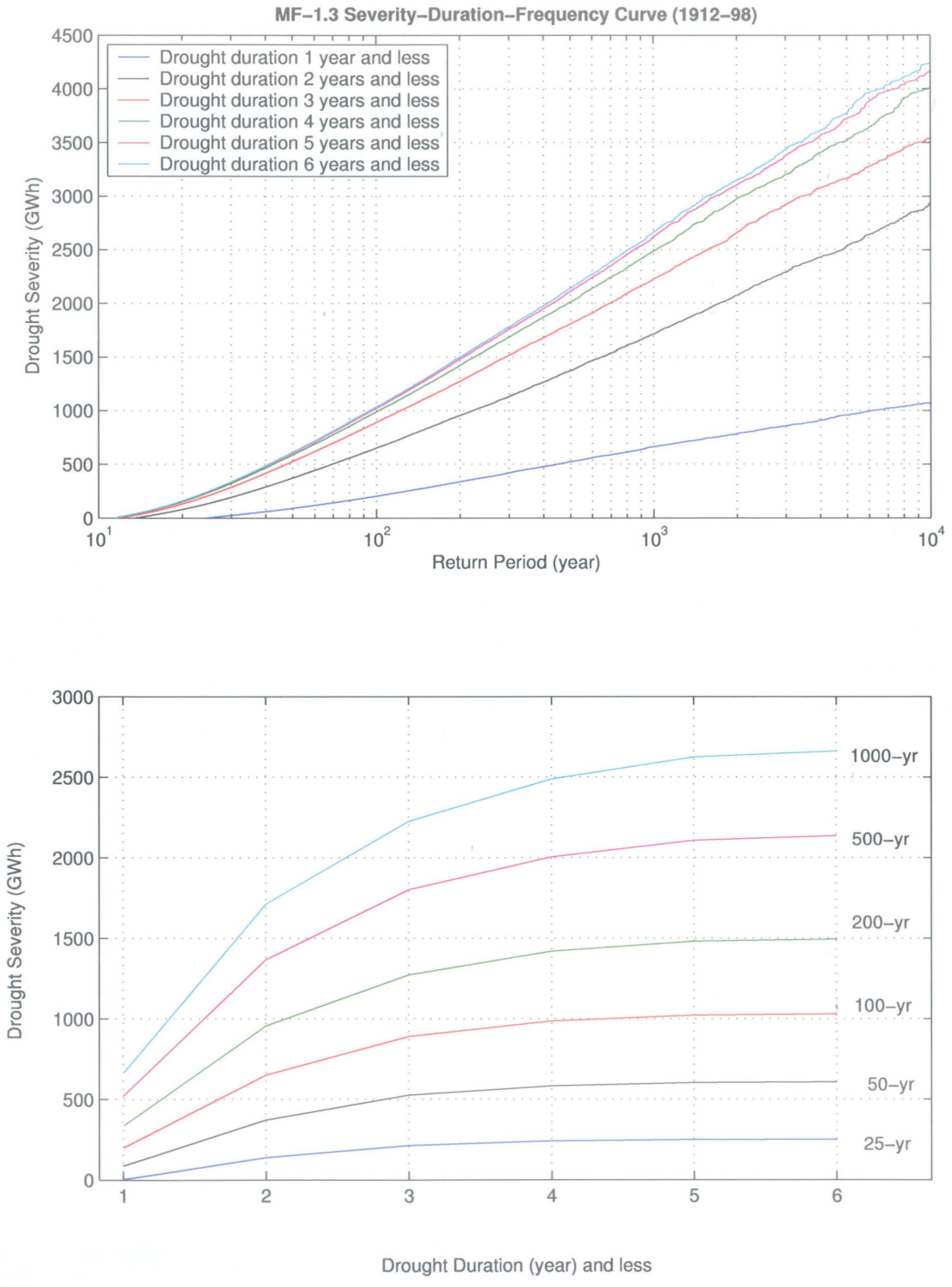


Figure N.3: Conditional Drought Frequency Curves obtained from MF-1.3 (1912-98).

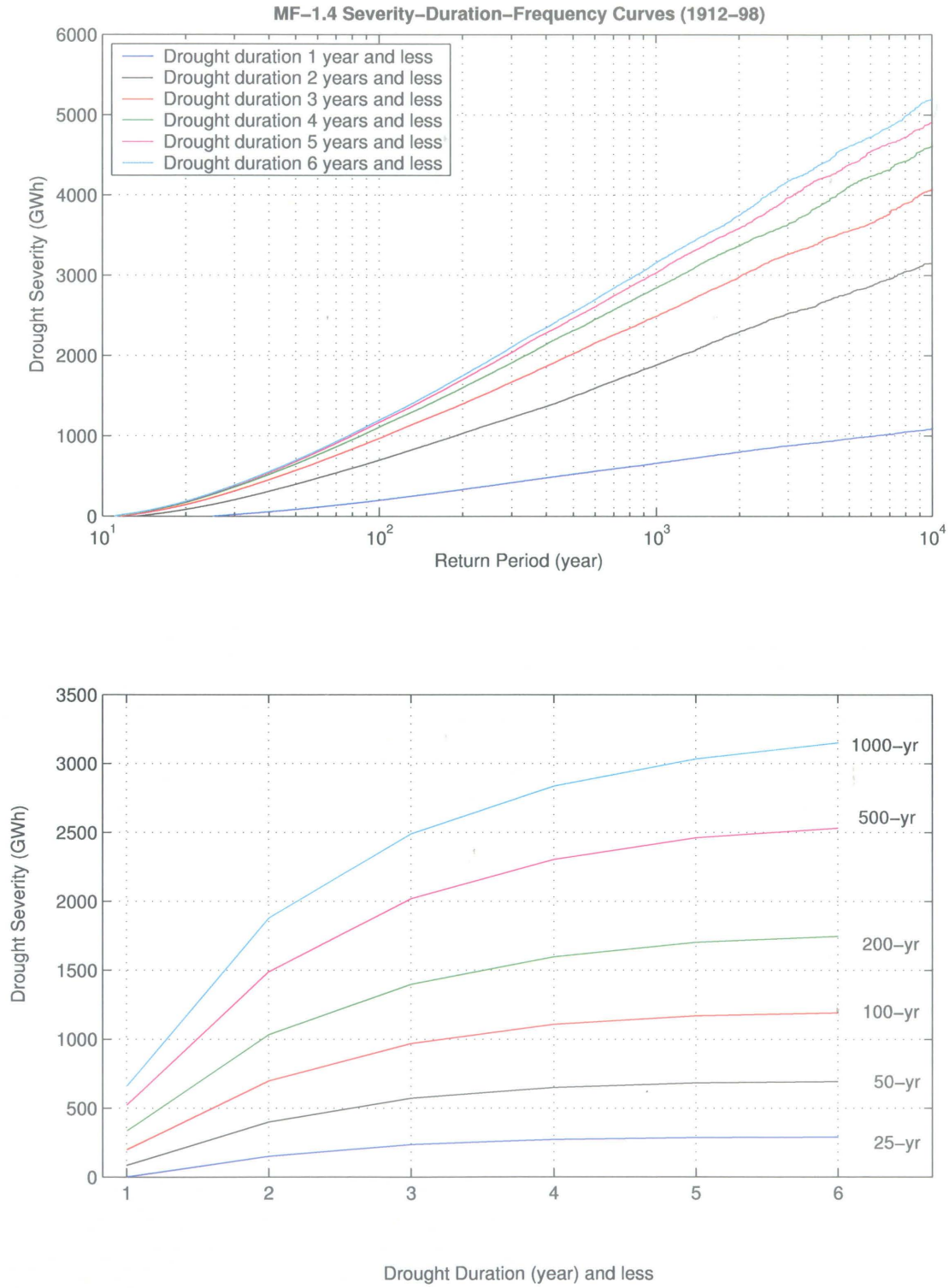


Figure N.4: Conditional Drought Frequency Curves obtained from MF-1.4 (1912-98).

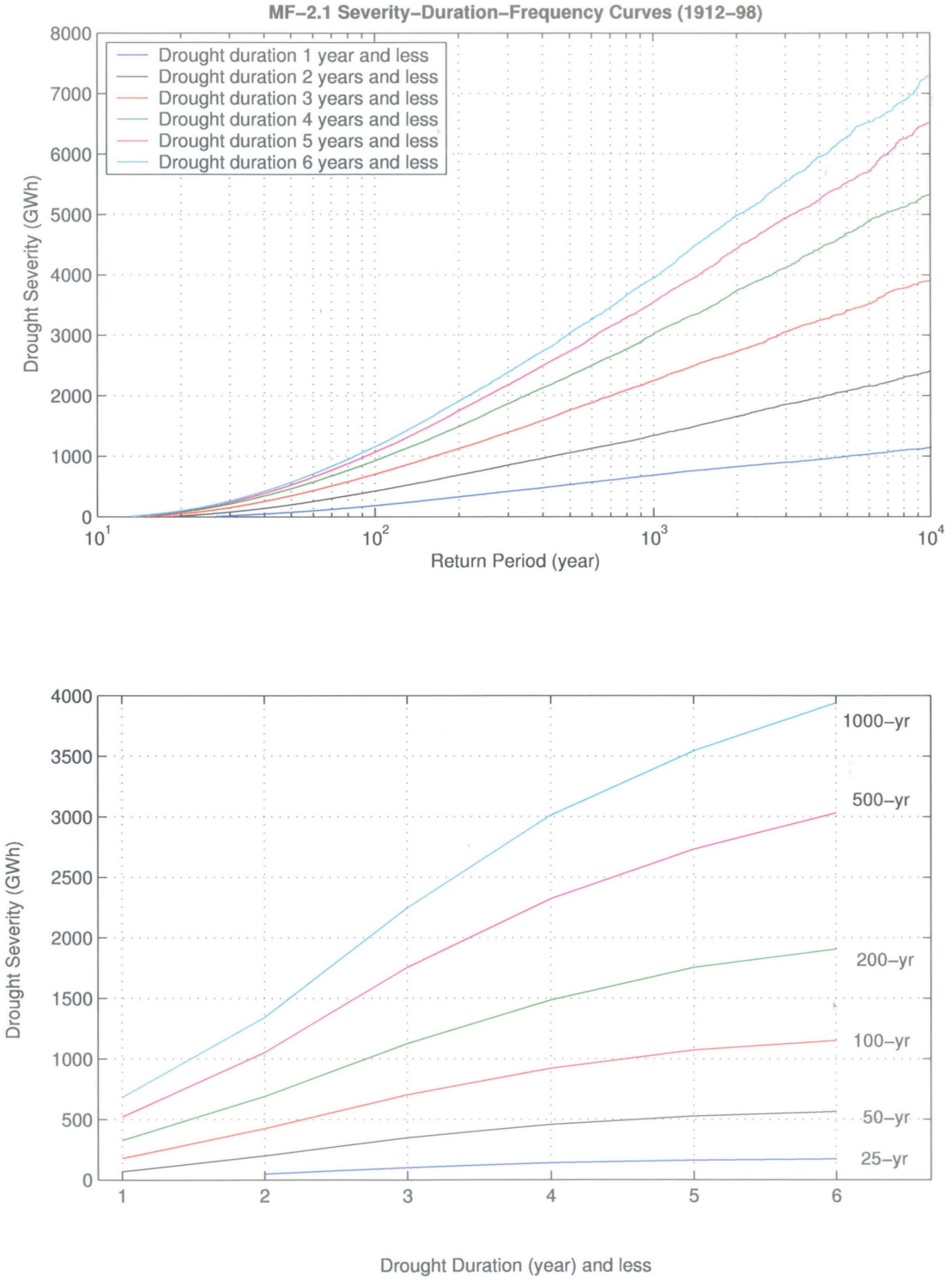


Figure N.5: Conditional Drought Frequency Curves obtained from MF-2.1 (1912-98).

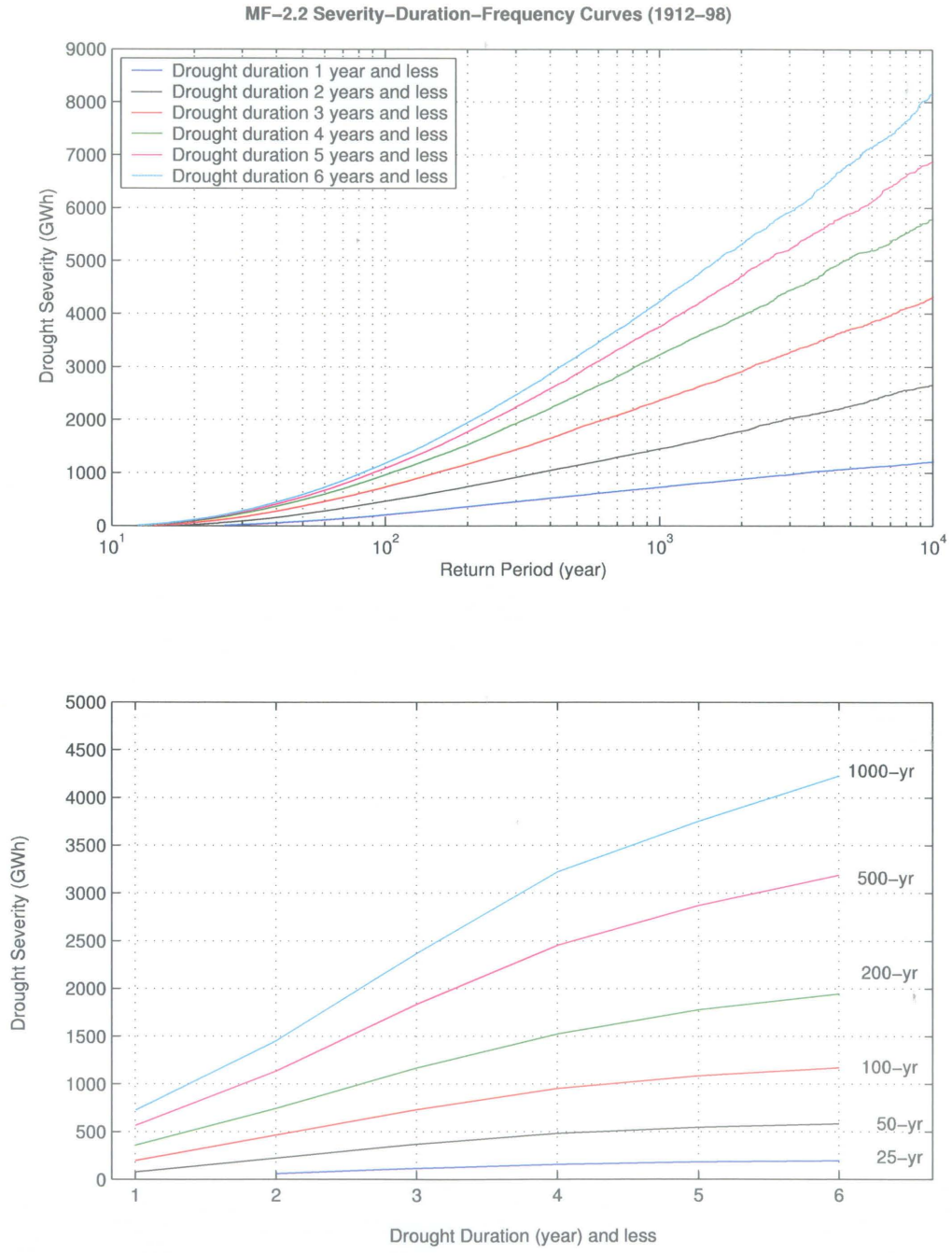


Figure N.6: Conditional Drought Frequency Curves obtained from MF-2.2 (1912-98).

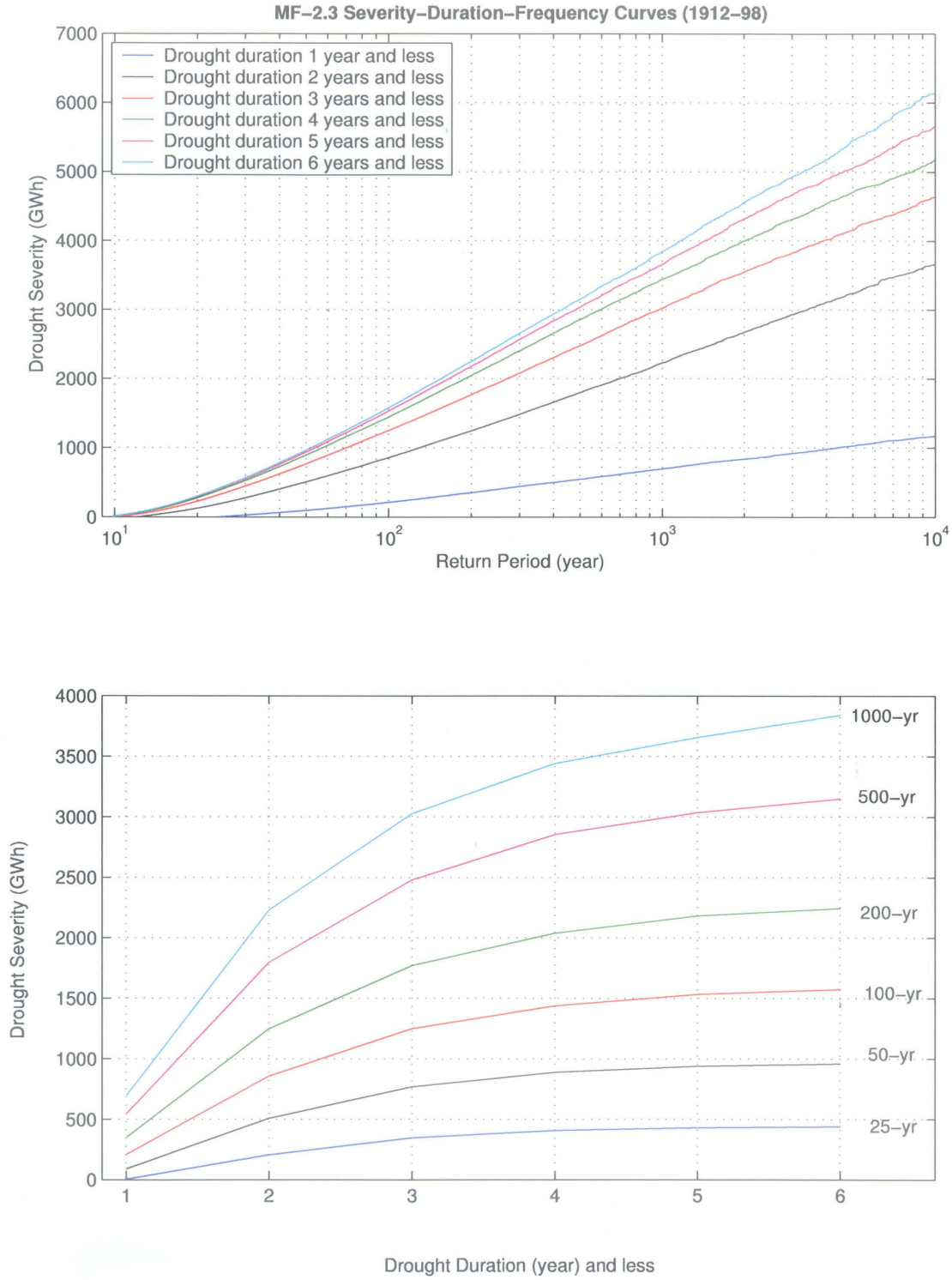


Figure N.7: Conditional Drought Frequency Curves obtained from MF-2.3 (1912-98).

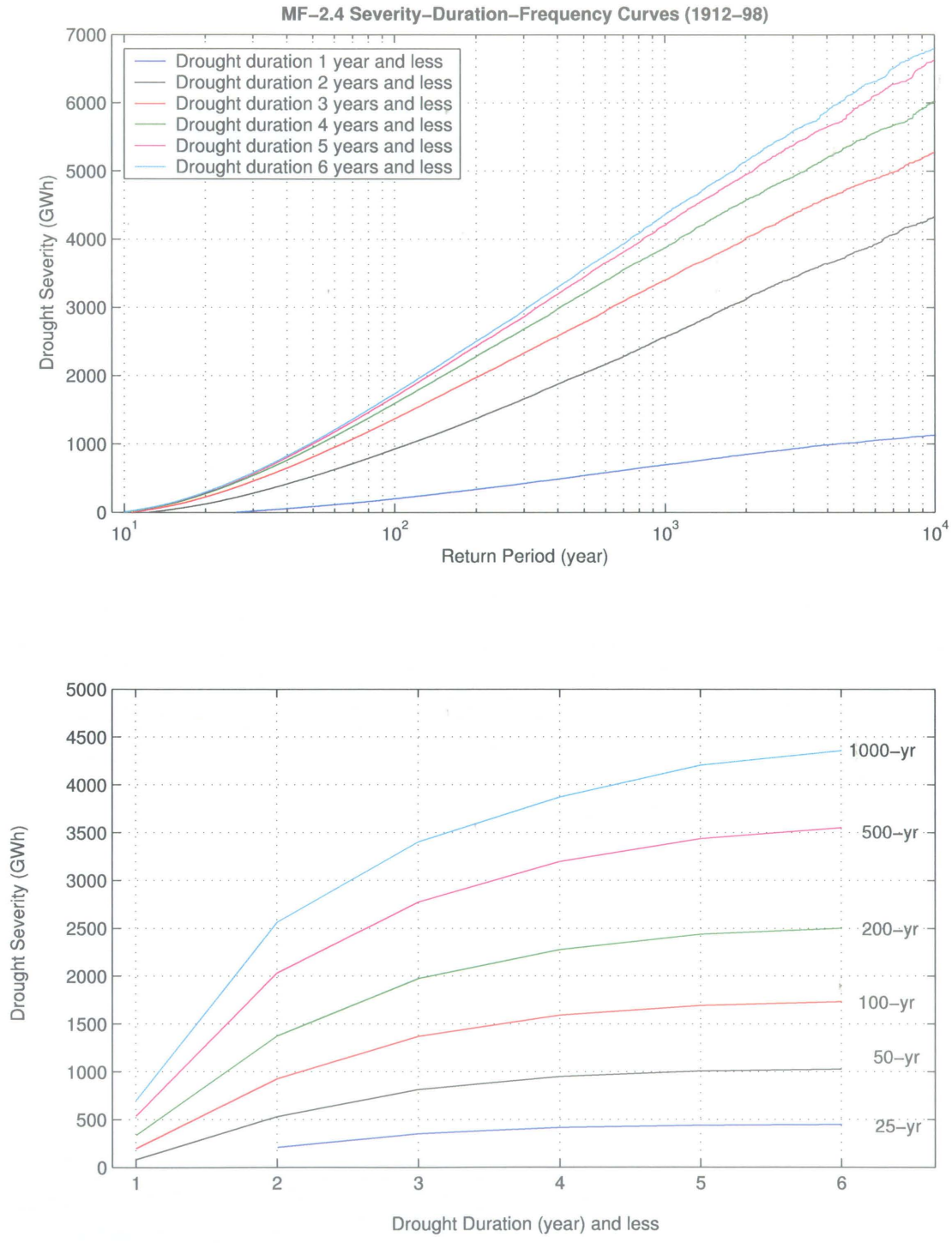


Figure N.8: Conditional Drought Frequency Curves obtained from MF-2.4 (1912-98).

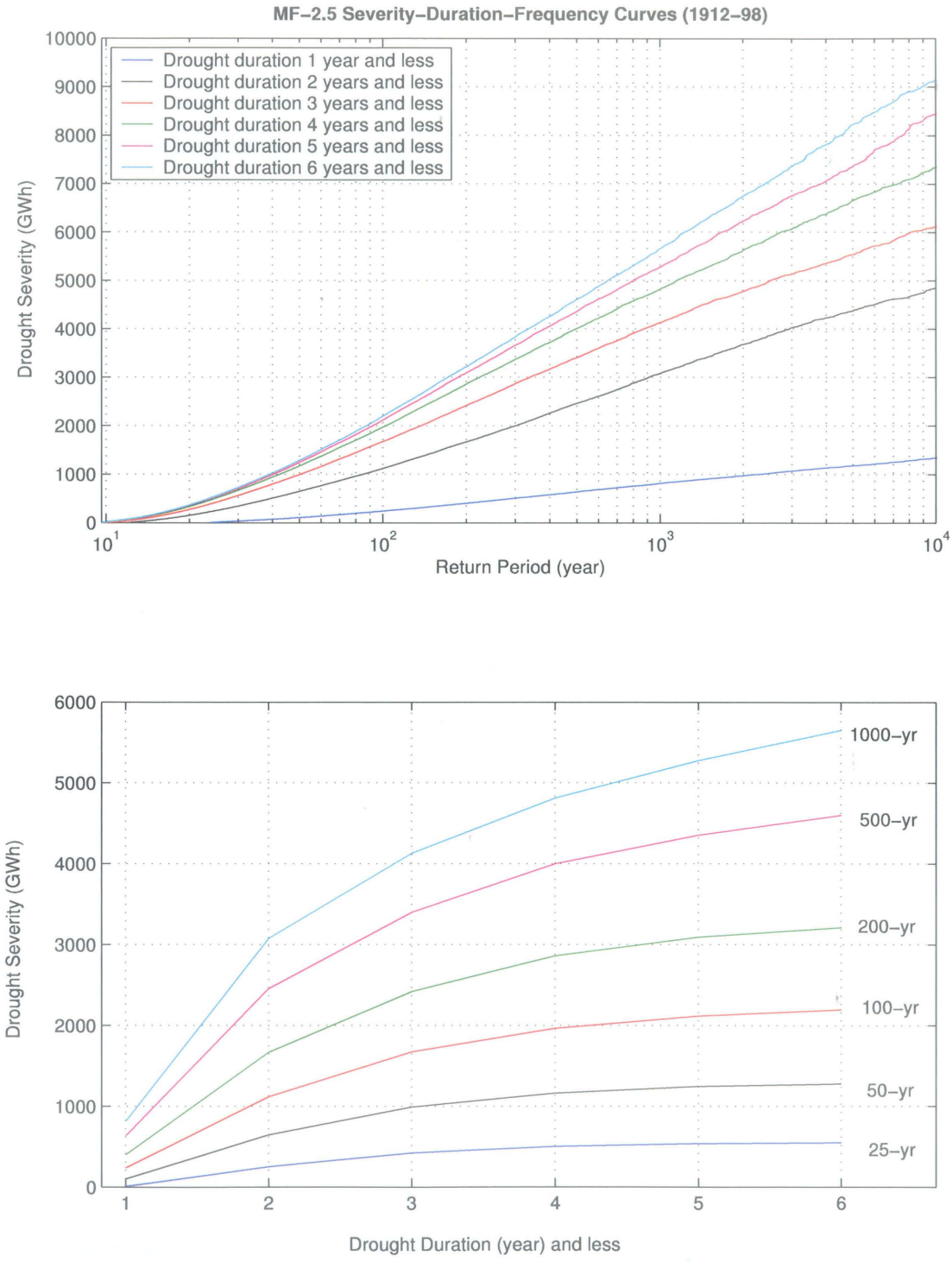


Figure N.9: Conditional Drought Frequency Curves obtained from MF-2.5 (1912-98).

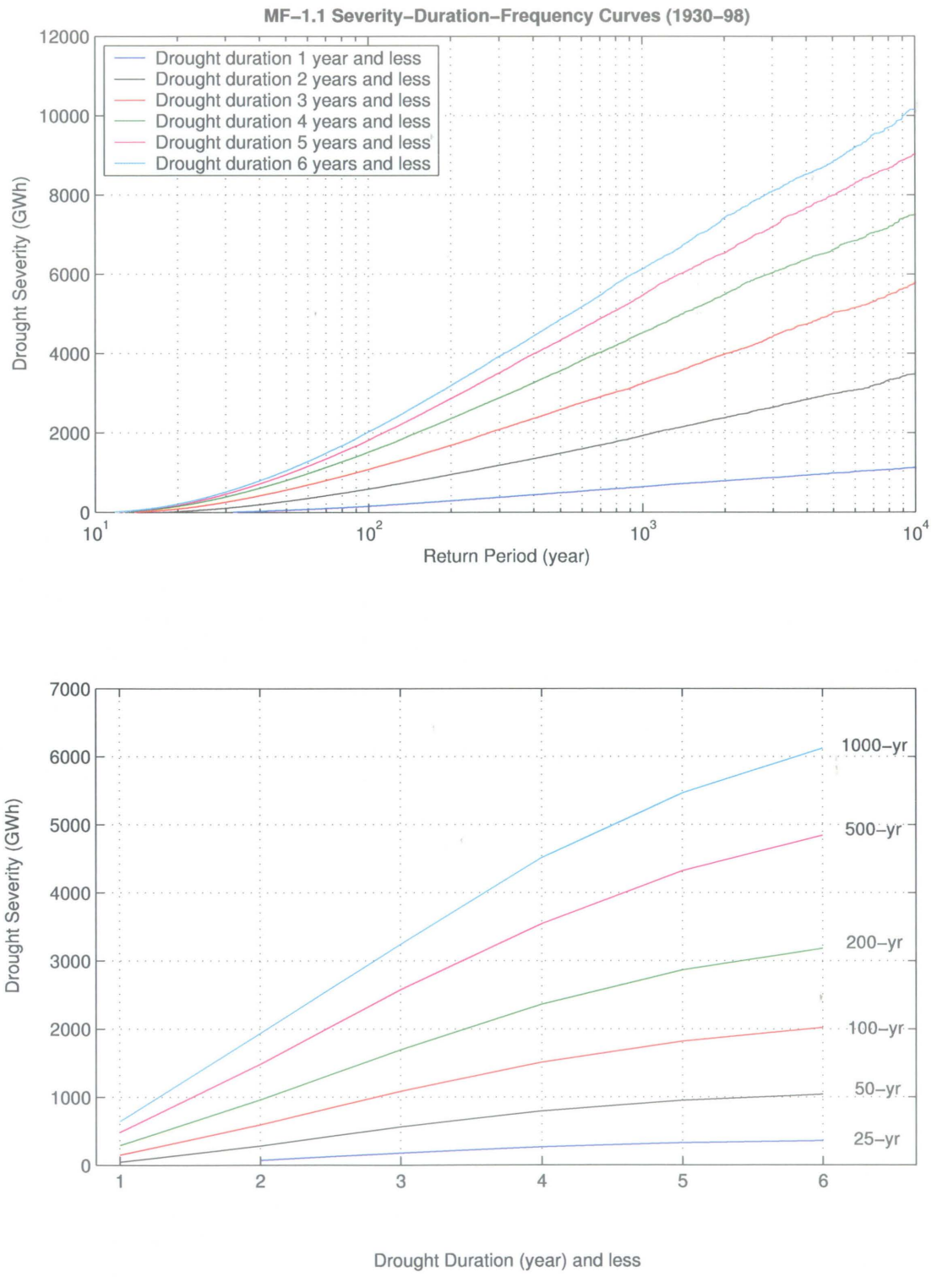


Figure N.10: Conditional Drought Frequency Curves obtained from MF-1.1 (1930-98).

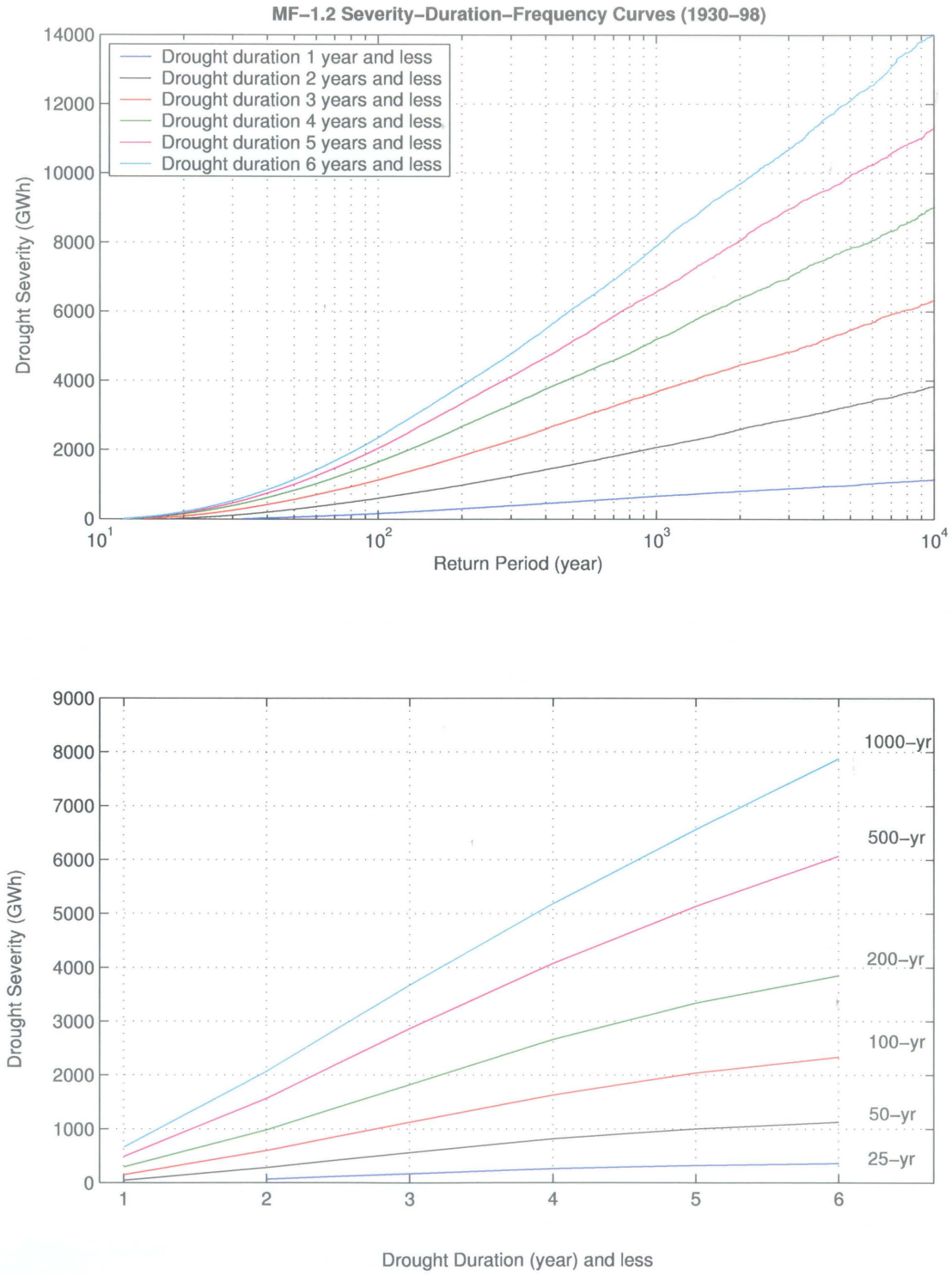


Figure N.11: Conditional Drought Frequency Curves obtained from MF-1.2 (1930-98).

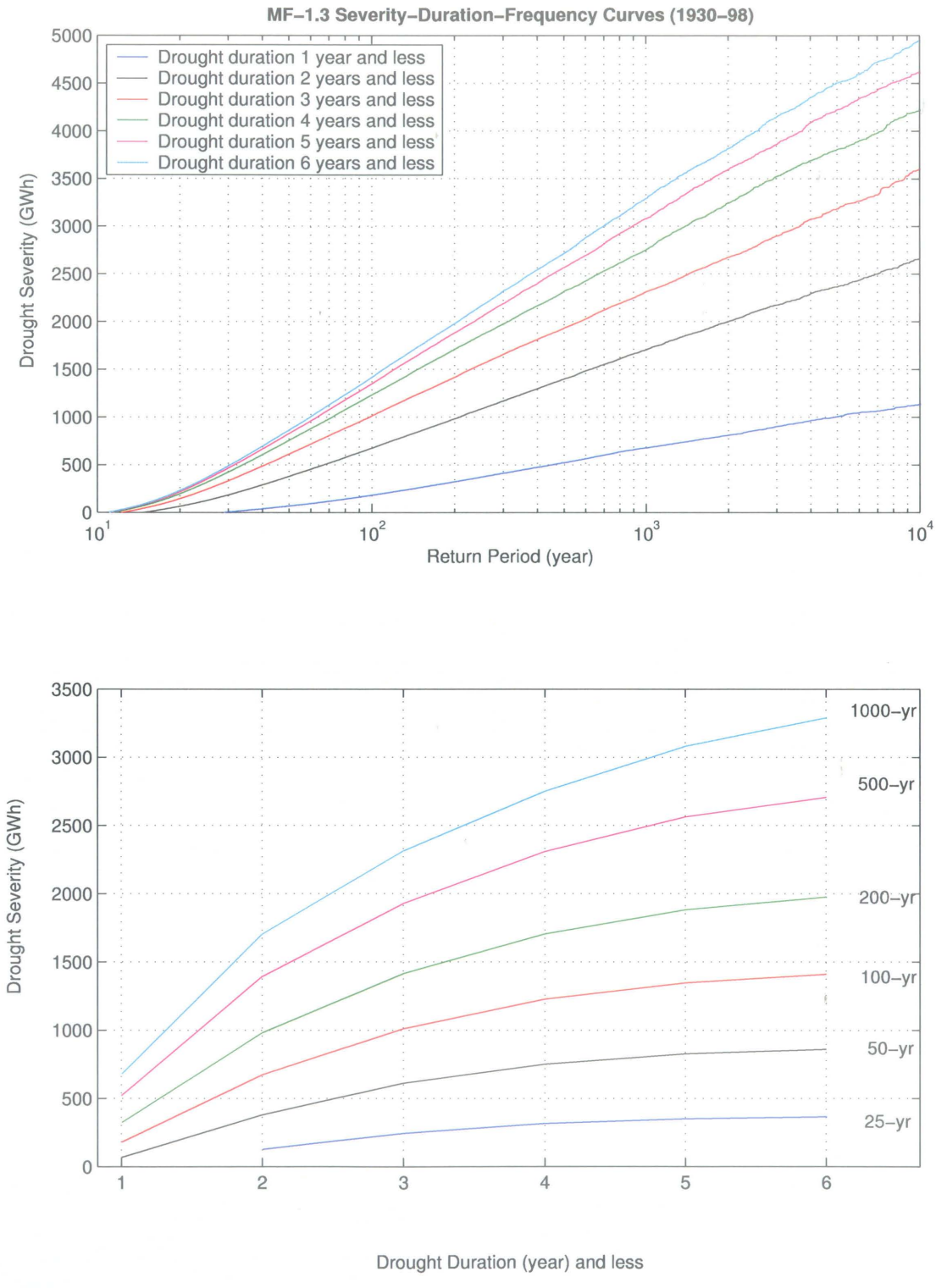


Figure N.12: Conditional Drought Frequency Curves obtained from MF-1.3 (1930-98).

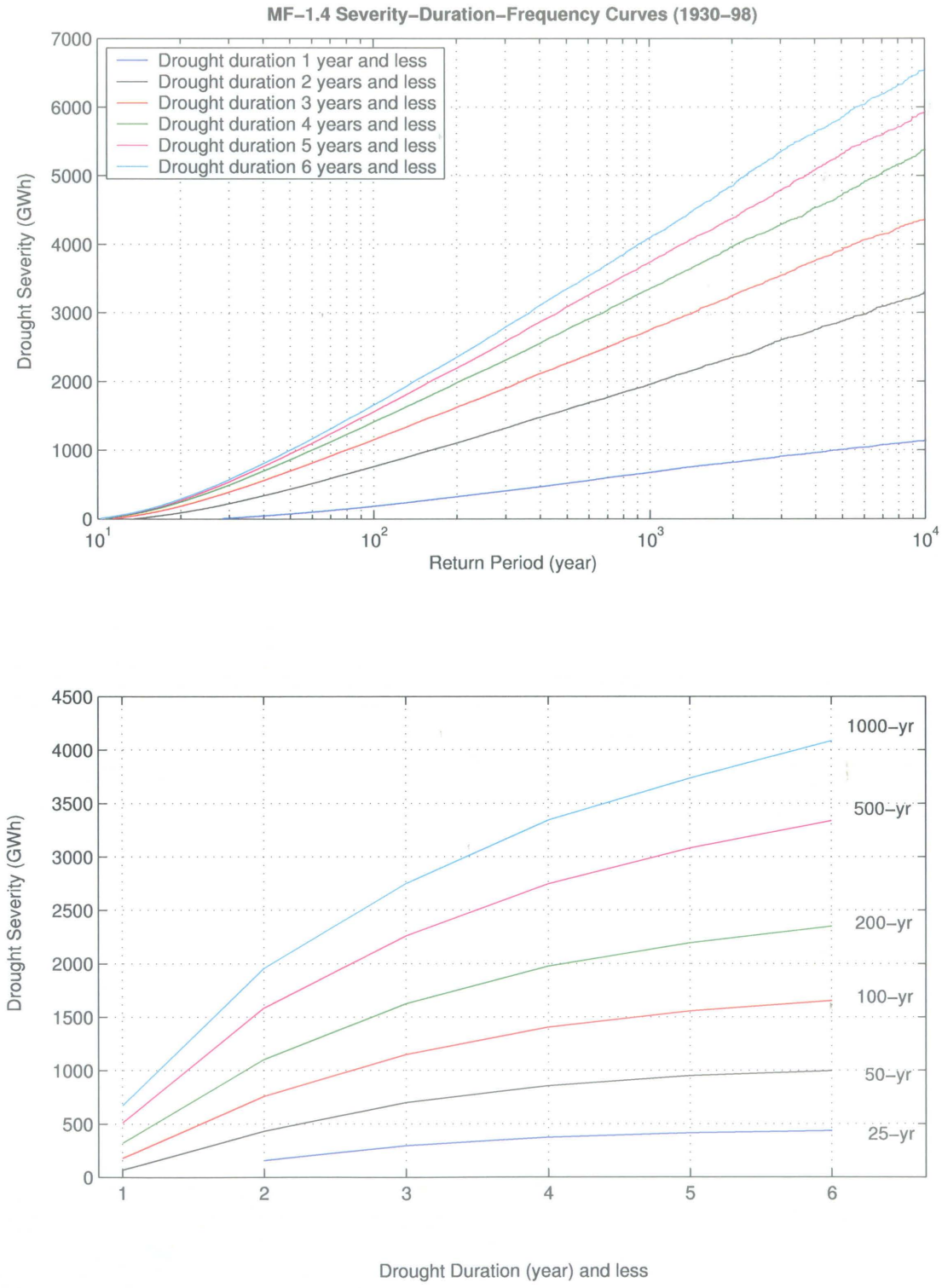


Figure N.13: Conditional Drought Frequency Curves obtained from MF-1.4 (1930-98).

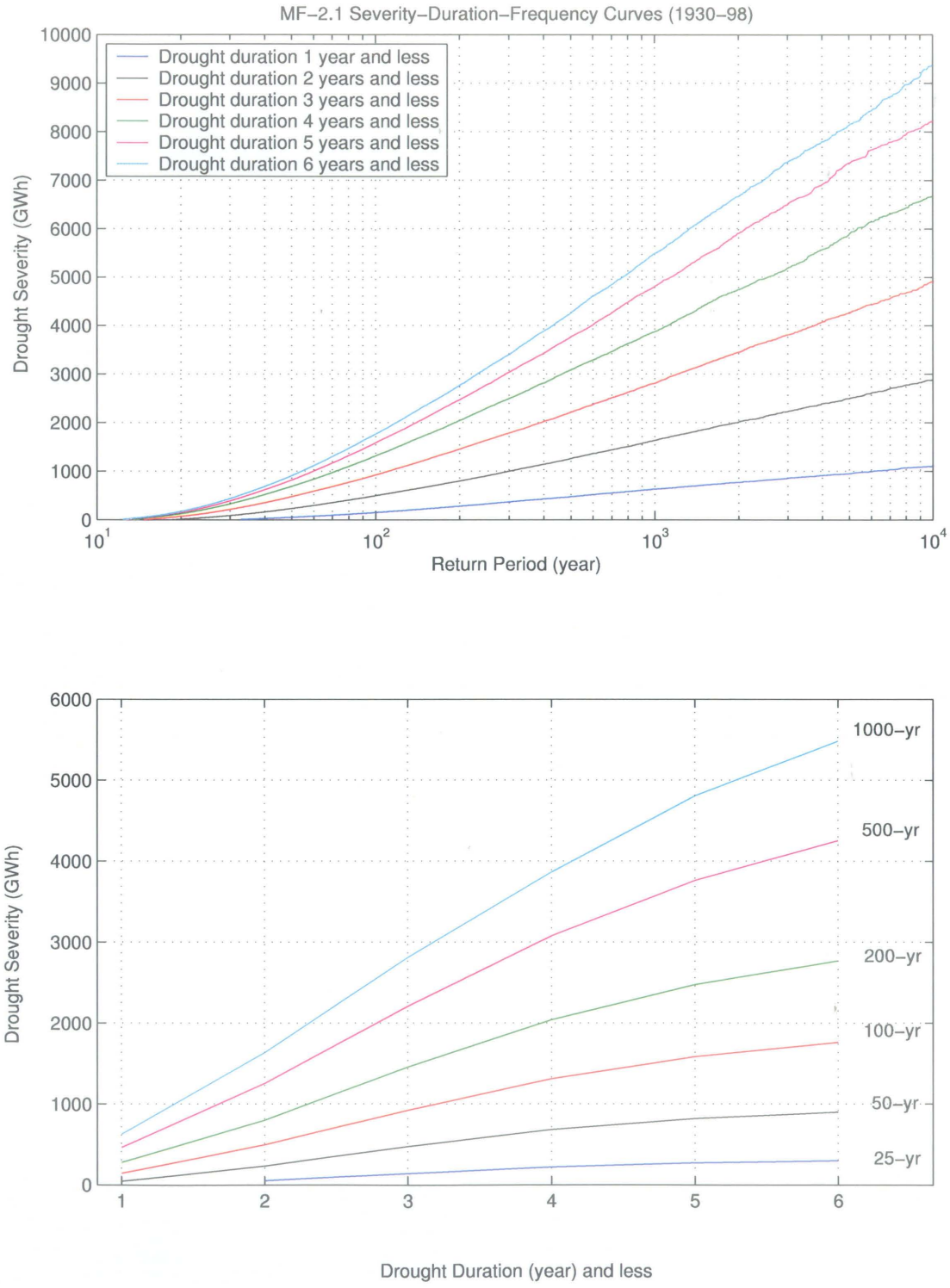


Figure N.14: Conditional Drought Frequency Curves obtained from MF-2.1 (1930-98).

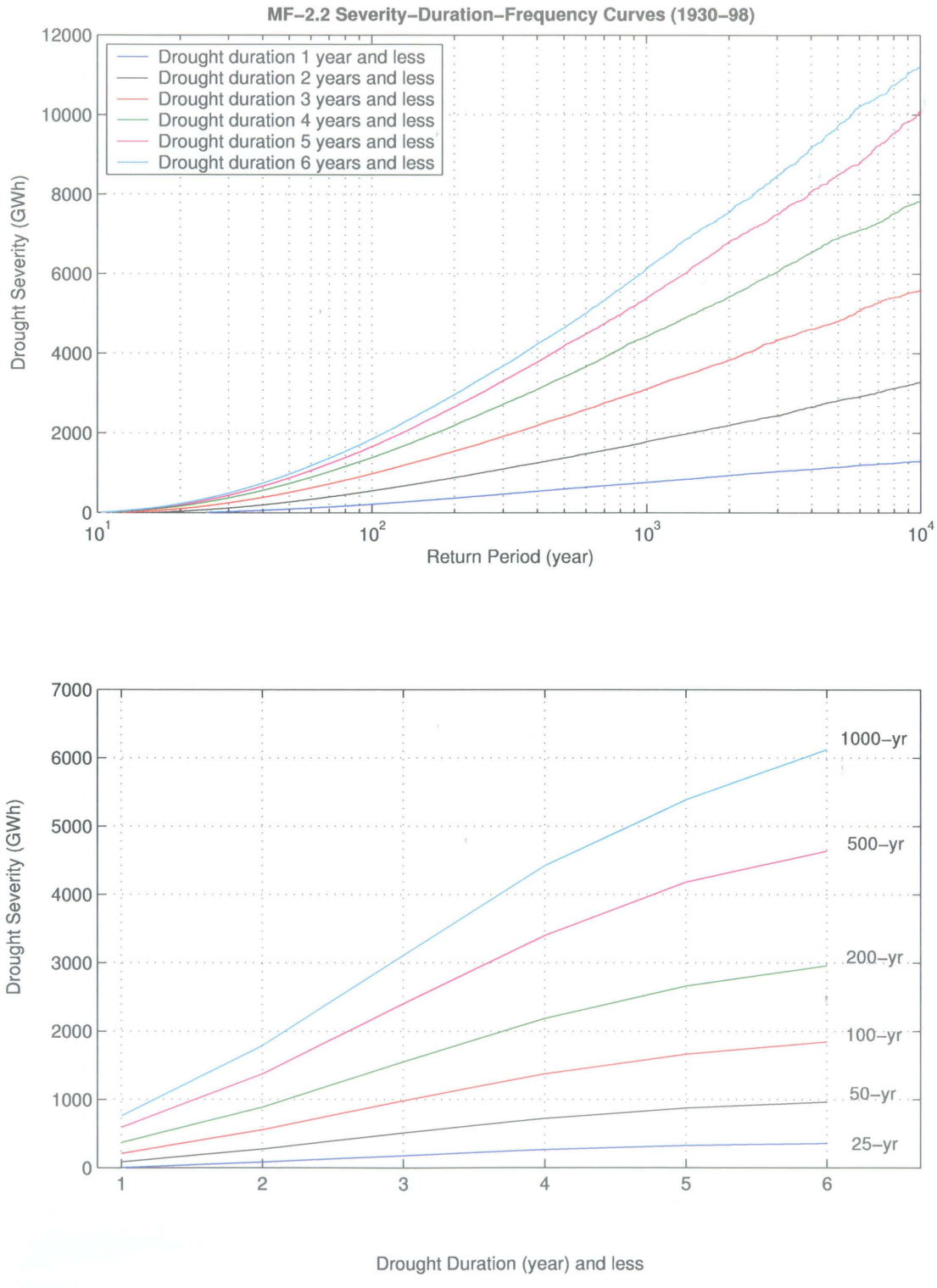


Figure N.15: Conditional Drought Frequency Curves obtained from MF-2.2 (1930-98).

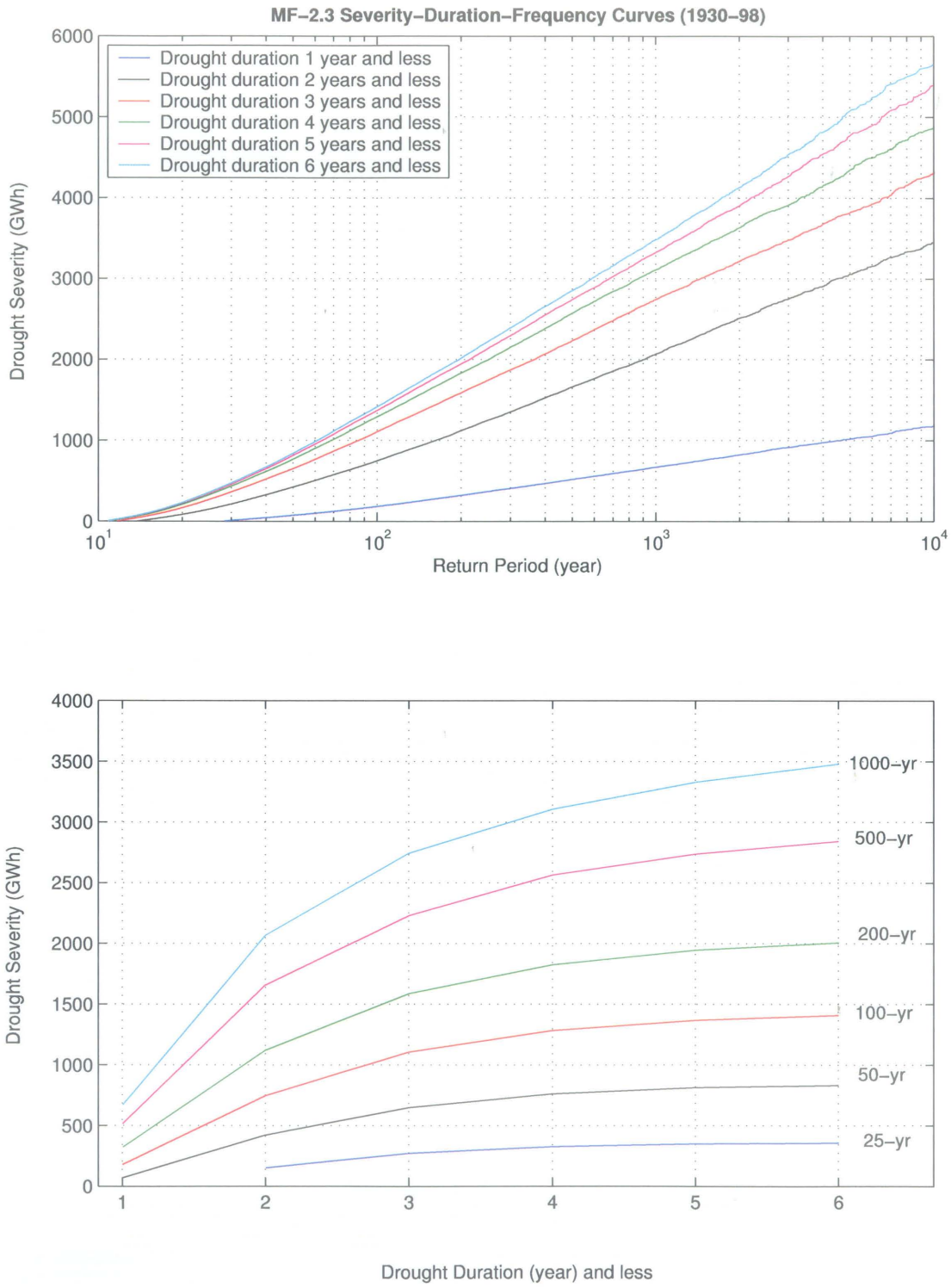


Figure N.16: Conditional Drought Frequency Curves obtained from MF-2.3 (1930-98).

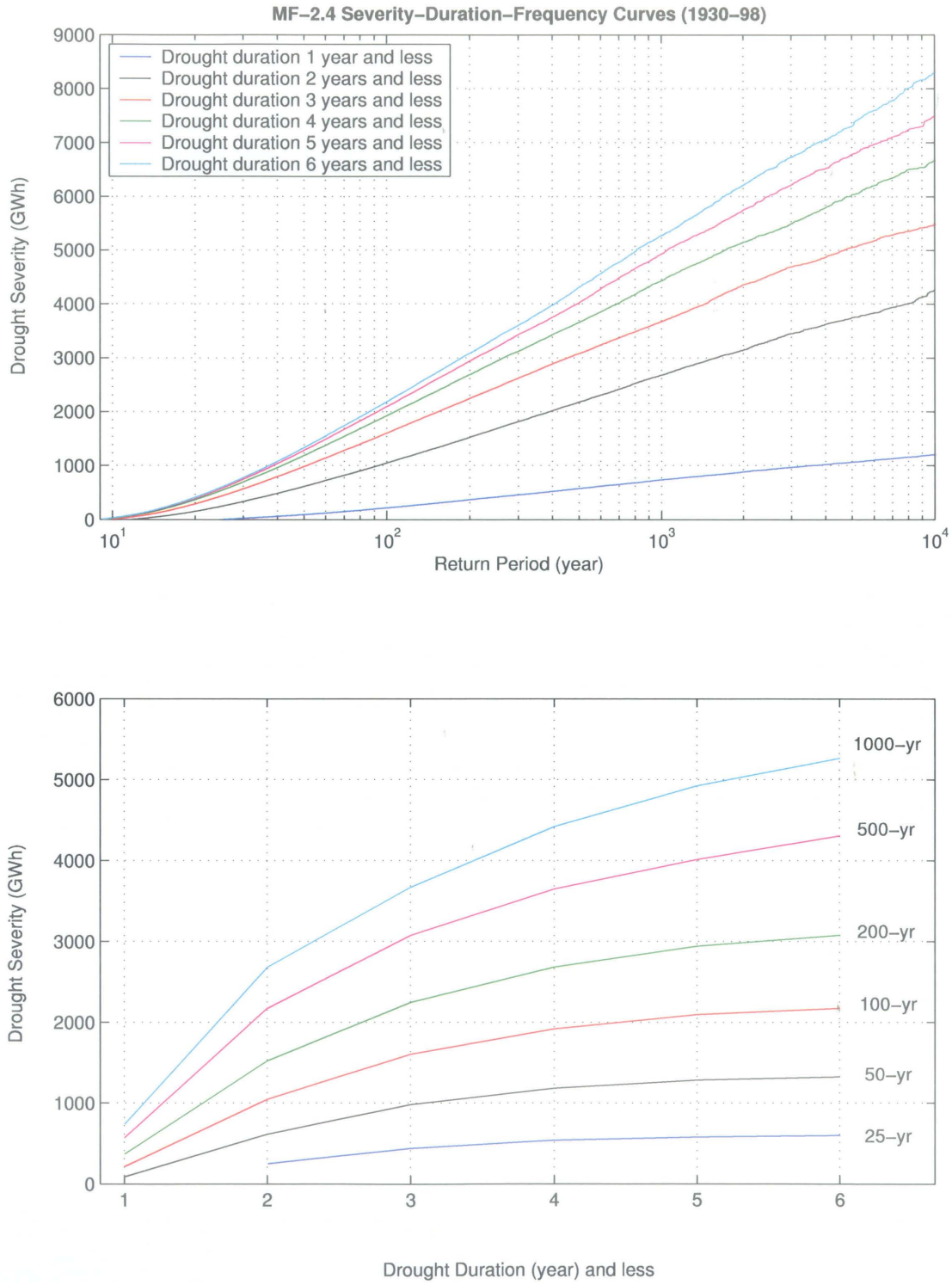


Figure N.17: Conditional Drought Frequency Curves obtained from MF-2.4 (1930-98).

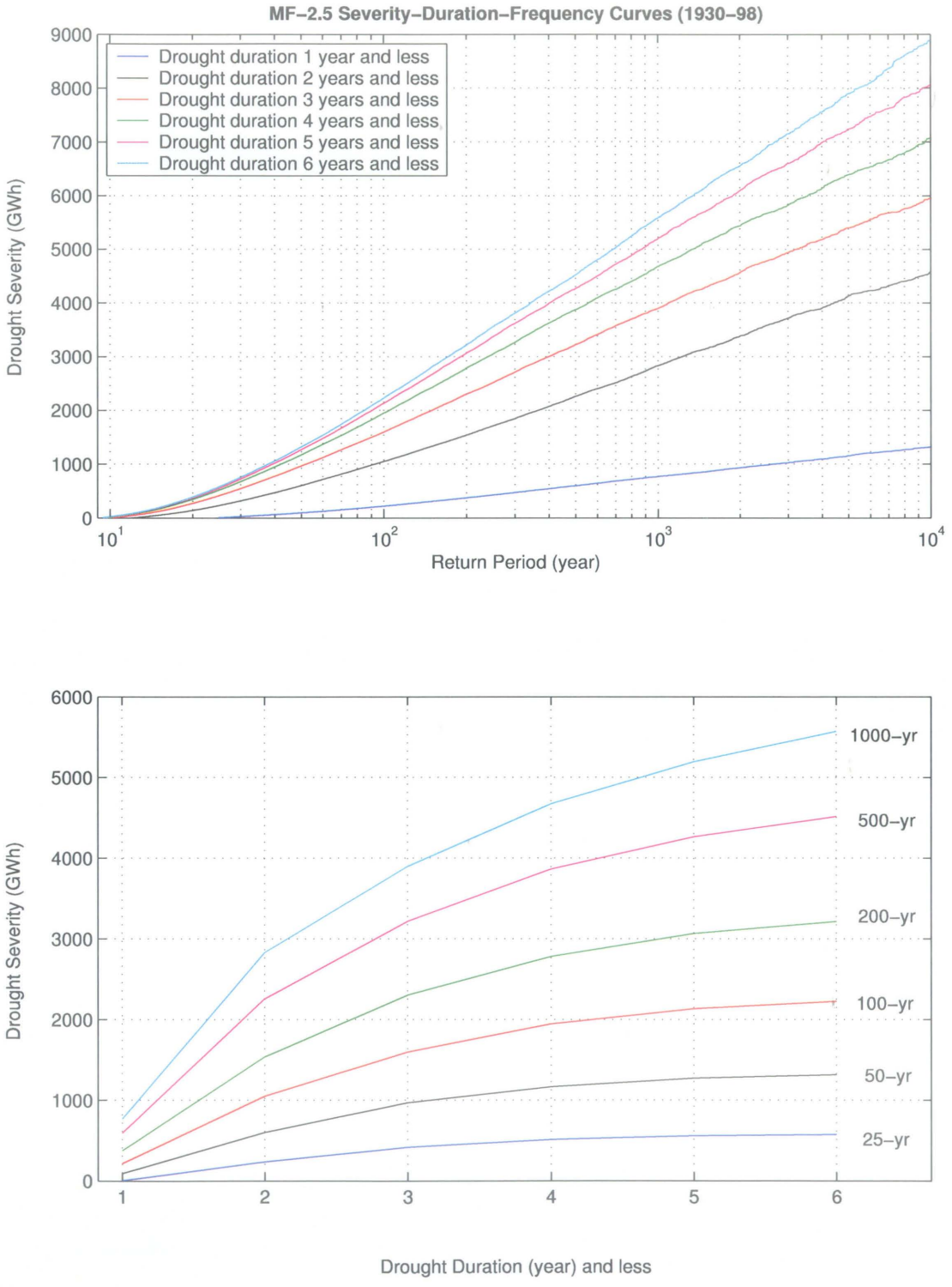


Figure N.18: Conditional Drought Frequency Curves obtained from MF-2.5 (1930-98).