DISSERTATION

CHARACTERIZING THE SEVERITY OF HYDROLOGIC DROUGHTS

Submitted by

Zeyad Salem Tarawneh

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer, 2006

UMI Number: 3233376

COLORADO STATE UNIVERSITY

June 6, 2006

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ZEYAD SALEM TARAWNEH ENTITLED CHARACTERIZING THE SEVERITY OF HYDROLOGIC DROUGHTS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Advisor

Department Head

ABSTRACT OF DISSERTATION

CHARACTERIZING THE SEVERITY OF HYDROLOGIC DROUGHTS

Planning and management studies of water resources systems generally take into account the temporal and spatial variability of hydrological processes such as streamflows, particularly periods of extreme floods and droughts. This dissertation concerns with research on multiyear droughts that may occur in a given river system. One may obtain some relevant information about droughts from the observed streamflows of the river basin under consideration, however it has been shown in literature that the information one may obtain from the observed records is quite limited and uncertain because of the typical short sample sizes that are generally available.

The main objective of this research is to characterize the severity of multiyear drought events at one or more sites in a river basin. In particular, mathematical models are developed to determine the return period and risk of extreme droughts. Also simulation experiments are conducted to verify the results obtained from the models. A deficit occurs where the yearly streamflow is smaller than a certain threshold water demand, otherwise surplus occurs. A drought of a given length l occurs where a consecutive number of deficits persists over the time period l. The sequence of surplus and deficit has been modeled by a discrete autoregressive or discrete autoregressive moving average processes. The single year deficit and the drought amount over the length l have been fitted using the beta distribution function. The drought occurrence probability in a given time period (risk) has been determined assuming that the first

occurrence probability of drought is geometric. Furthermore, approximate expressions for determining the return period of drought have been developed. The results obtained from the referred models have been compared and verified using simulation experiments based on annual streamflow data of the Poudre, Colorado, and Niger rivers. It is concluded that the models developed are quite reasonable and provide fairly good estimates of drought properties. In addition, the referred models have been extended to characterize regional droughts and it is concluded that the proposed models give fairly good estimates of drought properties on a regional context.

Lastly, tree ring indices have been utilized to reconstruct streamflows. A multivariate multiple linear regression model (MREXTN) including an autoregressive term and the random (noise) component was developed in such a way that streamflow variances, autocorrelations, and cross-correlations are properly taken into account. In addition, a procedure was developed that enables one selecting a single reconstructed flow sequence at each site from the multiple sequences that may arise because of using the noise term. The proposed model was tested against competing models using simulation experiments. It is concluded that model MREXTN compares favorably versus alternative models. Furthermore, the model was applied to reconstruct flows at the Colorado River system. It has been shown that the reconstructed flows provide useful information that enables one to analyze drought properties more accurately than using historical records alone.

Zeyad S. Tarawneh
Civil Engineering Department
Colorado State University
Fort Collins, CO80523
Summer 2006

ACKNOWLEDGMENT

At the beginning, all praises and gratitude are due to Allah the most gracious, the most merciful.

Upon completing the graduate work that leads to this dissertation, the author deeply appreciates the guidance, encouragement, and the great support of his advisor Dr. Jose D. Salas, Professor of Civil and Environmental Engineering. The author wishes also to express his sincere gratitude to other members of his graduate committee represented by Dr. Neil Grigg and Dr. Jorge Ramirez, Professors of Civil and Environmental Engineering, Dr. Jay Breidt, Professor of Statistics, and Dr. Connie Woodhouse, NOAA Paleoclimatology Program for their invaluable efforts that enhanced the quality of this work. In addition special thanks to Dr. D. C. Boes for his comments and accessibility.

The financial support offered by the Hashemite University, Jordan is appreciated. The research regarding my doctoral studies has been partially funded by the projects "Predicting the severity of low flows and droughts for agricultural systems in Colorado", Project COL 00701, Agricultural experiment station and "development of stochastic hydrology for the Colorado river system" U.S. Bureau of Reclamation. The support is gratefully acknowledged.

During his study, the author highly appreciates the encouragement and the faithful support of his parents and his wife Nuha. Their patience, encouragement and the full support have greatly motivated the author to succeed and finish his study.

TABLE OF CONTENTS

CHAPTER I 1
INTRODUCTION 1
1.1 Summary on Drought Characterization 1
1.2 Objectives of the Study 6
1.3 Summary of the Study Chapters 8
CHAPTER II 10
LITERATURE REVIEW 10
2.1 Introduction 10
2.2 Definitions of Drought Events: Review and Basic Concepts 12
2.3 Review on DARMA (p, q) Models 14
2.4 Characterization of Single Year Drought Events 19
2.5 Characterization of Multiyear Drought Events 23
2.6 Characterization of Multiyear Regional Droughts 27
2.7 Defining the Areal Aspect of Regional Droughts 28
2.8 Joint Analysis of Multi-Sites Under Drought 29
2.9 Drought Characterization Through Using Tree Ring Reconstructed Flows 31
2.10 Multi-site Record Extension Methods for Reconstructing Streamflows 32
CHAPTER III 46
CHARACTERIZATION OF DROUGHTS: SINGLE SITE ANALYSIS 46
3.1 Introduction 46
3.2 Occurrence Probability of Run of Deficits 47
3.3 Distribution of Drought Deficits 50
3.4 Occurrence Probability of Multiyear Drought Events 54
3.5 Risk of Multiyear Drought Events 55
3.6 Return Period of Multiyear Drought Events 58
3.7 Applications 60
3.8 Results and Discussion 64
3.9 Summary and Conclusions 72
3.10 Appendix 3A 90
3.11 Appendix 3B 92
CHAPTER IV 93
CHARACTERIZING REGIONAL HYDROLOGIC DROUGHTS 93
4.1 Introduction 93
4.2 Basic Definitions and Assumptions 95
4.3 Converting Multi-Sites Deficit Occurrence to Single Occurrence 96
4.4 Occurrence Probability of Run of Regional Deficits 102
4.5 Distribution of the Regional Drought Deficit 103
4.6 Occurrence Probability of Multiyear Regional Drought Events 106

4.7 Risk of Multiyear Regional Droughts 107
4.8 Return Period of the Multiyear Regional Drought 109
4.9 Spatial Patterns of Regional Hydrologic Droughts 110
4.10 Application 113
4.11 Results and Discussion 114
4.12 Summary and Conclusions 126
4.13 Appendix 4A 138
4.14 Appendix 4B 139
CHAPTER V 142
IMPROVING DROUGHT CHARACTERIZATION USING TREE-RING RECONSTRUCTED STREAMFLOWS 142
5.1 Introduction 142
5.2 Multi-Site Record Extension Model with Noise (MREXTN) 145
5.3 The Performance of the Record Extension Model MRXETN 148
5.4 Reconstructed Streamflows of the Colorado River 152
5.5 Characterizing Droughts of the Colorado River 157
5.6 Summary and Conclusions 165
CHAPTER VI 185
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS 185
REFERENCES 189

CHAPTER I

INTRODUCTION

1.1 Summary on Drought Characterization

Over time natural water supply systems may fall short to fulfill the water requirements for urban, agricultural, industrial uses, and environmental needs. This water shortage may extend for several years and cover large areas leading to considerable economic and social consequences. Technically, the word "drought" has been used for a long time to describe water shortage over time and space. The frequent and irregular occurrence of droughts has been the motive behind a considerable number of studies aimed at understanding and better characterizing the severity of droughts, quantifying drought risk and estimating the return period of droughts.

Many operational definitions of drought have been proposed in literature. For example, Yevjevich (1967) defined the drought event as a succession of consecutive periods at which the water supply remains below the threshold where that consecutive periods are preceded and succeeded by at least one period at which the water supply is equal to or greater than the threshold. On the other hand, a drought event may be defined following the definition of runs as in Feller (1968). In that case, a drought is defined as a run of successive periods at which the water supply remains below the threshold as long as these runs are not overlapped. Alternatively, Fernandez and Salas (1999) employed the definition of runs as in Schwager (1983) to define the drought as a run of consecutive periods at which the water supply remains below the threshold regardless where the run starts as long as it satisfies the continuity. The later definition may be useful to capture droughts of extreme magnitudes regardless where the event starts. In this study, the characterization of drought events will be made considering that drought events are defined according to the third definition (to be called the alternative definition of drought events).

The characterization of extreme drought events is made mainly through evaluating drought properties like the occurrence probability, return period, and the risk of drought occurrence during specified time span. Generally, the return period of an event can be defined in different ways for different applications (Fernandez and Salas, 1999; Douglas et al, 2002). It can be defined as the expected value of the waiting time for a specified event to occur for the first time, or it can be defined as the expected value of the interarrival time between any two recurrent events of specific criteria (Fernandez and Salas 1999). The risk is defined as the probability that a specified event has occurred at least once during a given time span (Chung and Salas 2000).

Centering the attention on the drought length and considering that drought events are defined according to the alternative definition, Fernandez and Salas (1999) derived the distribution of first occurrence time, the recurrence time, return period, and the risk of single time step drought events given that these events arise from either independent or simple Markov process. Likewise, Chung and Salas (2000) found the distribution of the first occurrence time and the distribution of the recurrence time, the expected value and the variance of the first arrival time of drought events, and the risk of drought events assuming that these events arise from either simple Markov process or processes that exhibit longer time dependence structure, i.e. DARMA(1,1) process. Summarizing, the studies mentioned above (e.g. Fernandez and Salas,1999; Chung and Salas, 2000) focused on characterizing drought events considering only the length of the drought. Studies that consider both the length and magnitude have been also developed (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003, Salas et al, 2005). One purpose of this study is to extend the mathematical procedures developed so far to include both the length and the magnitude of drought in characterizing extreme drought events.

The magnitude of a drought event is quantified by accumulating the single deficits over the length of the drought (Salas et al, 2005). Since deficits are correlated, deriving the distribution of the drought magnitude, i.e. the sum of the single deficits, is a complex problem (unless single deficits are normally distributed). Also the distribution of drought magnitude can be obtained by fitting the empirical distribution obtained from the historical data using certain probability distribution function (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003). Alternatively the single deficits may be assumed to be independent and identically distributed and in this case the distribution of their sum conditioned to the fixed length of the drought event can be found under certain conditions (e.g. Salas at al, 2005). Generally, distributions such as the gamma and the exponential are employed to represent the drought deficit (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003, Biondi et al, 2005, Salas et al, 2005). However since the single deficits or their conditional sum are bounded by lower and upper values, then the upper tail unbounded distributions such as the gamma or the exponential may not represent well the distribution of the single deficits or their sum. In this research, the single deficits and their conditional sum will be fitted using the Beta distribution.

Droughts are also characterized by their spatial extent. Generally, a regional drought event is defined as the continuous time interval during which the deficit area equals to or greater than a predefined critical area (Tase, 1976). In literature, many attempts have been made to characterize regional droughts by describing this temporal and spacial variability (e.g. the Palmer Drought Severity Index, Palmer Hydrological Drought Index, and Drought Monitor). However, these drought indices do not give information about the frequency or the recurrence time of drought events (Rossi and Cancelliere, 2003; Salas et al, 2005). Statistical techniques have been employed in literature for the purpose of better characterizing regional droughts. In general, two main approaches have been suggested in literature for characterizing regional droughts. The first considers the proportion of the area covered by the regional deficit (e.g. Tase, 1976; Sen, 1980; Santos, 1983; Kingery, 1992; Shin and Salas, 2000; Hisdal and Tallaksen, 2003), while the second considers the joint analysis of the study variable at many sites under drought conditions (e.g. Yevjevich, 1972; Guerrero-Salazar and Yevjevich, 1975; Sen, 1979; Sen, 1998; Bayazit and Onoz, 2005).

In characterizing regional droughts based on areal deficit concept, the study region has to be divided into grid cells and then to be analyzed for the occurrence of deficit at each cell separately (e.g. Tase, 1976). The regional drought event is considered to occur when the sum of the cells (sub-areas) where deficits are observed exceeds certain critical area, i.e. the deficit area A exceeds a certain critical area. Also other useful indices were introduced in the literature for analyzing regional droughts (e.g. Tase, 1976; Santos, 1983). The probability distribution functions of the deficit area A, the areal deficit D, and their joint distribution were developed in other studies (e.g. Sen, 1980).

The concept of joint analyzing many sites under drought condition is an alternative procedure to characterize regional droughts. Assuming bivariate processes of simple Markov each that are serially and mutually dependent, a new process was defined such that a deficit would occur in the new process when deficits occurred simultaneously in the given original processes (e.g. Guerrero-Salazar and Yevjevich, 1975). The attractiveness of such concept appears when dealing with many sites at the same time for the purpose of characterizing extreme regional droughts. This study also provides other useful procedures to characterize the spatial development of regional droughts making use of the possible conversion of the deficit occurrence in many processes to deficit occurrence in single process based on the work of Guerrero-Salazar and Yevjevich (1975).

The characterization of droughts at either single site or regional basis can be made through the analysis of the available streamflow records. However, when the observed records are relatively short, like the case of most observed streamflow records, then the characterization of the extreme hydrologic droughts may not be reliable because only a few drought events may be observed (Salas et al, 2006). The characterization of droughts can be improved using longer records that can be generated using stochastic simulation models (Salas et al, 2005). Likewise, extending the short records back in time using longer tree ring records as proxies enhances the reliability of drought characteristics over what can be achieved when using the observed flows only (e.g. Woodhouse, 2001; Gonzales and Valdes, 2003; Salas et al, 2006).

Many record extension models are used in literature to reconstruct streamflows back in time using tree ring indices. Among these is the traditional linear regression model with least squares (e.g. Cleaveland, 2000; Meko et al, 2001; Woodhouse, 2001; Gedalof et al, 2004). Generally, reconstructed streamflows using the traditional regression models underestimate the natural variability and persistence in the observed streamflows. Other modified regression models (e.g. Meko, 1995; Gonzales and Valdes, 2003; Salas et al, 2006) have tackled problems of the underestimation and persistence by adding a noise term and term to account for previous flows.

The record extension model REXTN that has been developed by Salas et al (2006) to generate and select a single reconstructed streamflow trace will be expanded in this study to consider the reconstruction of streamflow records at many sites simultaneously while maintaining most of the key statistics of the original short streamflow records. The extended reconstructed flows will be used to characterize hydrologic droughts. Drought properties obtained from analyzing the reconstructed flows will be compared with properties obtained from analyzing the observed flow records.

1.2 Objectives of the Study

Two main objectives are considered in this study. The first objective is to develop mathematical models to characterize the severity of hydrologic droughts at the single site and regional scale assuming that the process behind the occurrence of deficits and surpluses is either DAR(1) or DARMA(1,1) process. The second objective is to improve the characterization of severe hydrologic droughts utilizing reconstructed streamflow records that can be obtained using tree ring indices data as the source of information over the extension period. The specific objectives of this study are:

1- To model the distribution of drought deficits using more realistic probability distribution function that has been utilized before. In addition, mathematical expressions to estimate the mean and the variance of deficits sum as a function of the single deficits moments will be developed. Analytical procedures will be developed to estimate the first occurrence and occurrence probabilities, return period, and the risk of drought events considering jointly the magnitude and the length of droughts.

2- To extend the mathematical procedures to be developed in 1 above to characterize regional hydrologic droughts, i.e. obtaining their first occurrence and occurrence probabilities, return period, and risk of regional droughts. In addition, probability of regional drought growth from one given region towards other regions will be provided.

3- To develop a record extension model that can be used to reconstruct streamflow records back in time at many sites simultaneously. The reconstructed streamflows will be used to characterize the severity of hydrologic drought events and the obtained results of drought properties will be compared with the results obtained from the analysis of the relatively short observed streamflow records.

4- To apply the procedures to be developed in 1 – 3 above to drought analysis and streamflow reconstruction using the Colorado River, Poudre River, and the Niger River flows.

1.3 Summary of the Study Chapters

Chapter 2 reviews the definitions of drought events provided in the literature. The basic characteristics of the low order autoregressive moving average processes, i.e. DAR(1) or DARAM(1,1) processes, that can be used to model the series of deficits and surpluses resulted from truncating the original process is also reviewed. Focusing the attention on the alternative definition of drought events (e.g. Schwager, 1983; Fernandez and Salas, 1999), the characterization of the single site drought events as a single deficit event and as a run of many consecutive deficits is utilized through evaluating the occurrence probability, return period, and the risk considering only the length of the drought event. The main approaches and procedures that have been used to characterize regional drought are also reviewed in Chapter 2. Moreover, summary on existing record extension models and procedures used to reconstruct streamflow records back in time using tree ring indices as information source is also provided in the last section of Chapter 2.

Chapter 3 presents the development of mathematical procedures that can be used to characterize droughts for evaluating the occurrence probabilities, return period, and the risk when the drought event is characterized by the length only and assuming the series of deficits and surpluses in which drought events arise is modeled using DAR(1) or DARMA(1,1) process. The developed analytical procedures have been generalized to include the magnitude of the drought event besides its length as well and have been used to characterize the severity of the multiyear hydrologic droughts that may occur in the Poudre, the Colorado, and the Niger River basins.

Introducing the new process Z_t that represents the occurrence of the regional deficit, the characterization of the regional hydrologic droughts has been approached in Chapter 4 through converting the multi-sites deficit occurrence to single case deficit occurrence in the process Z_t. Mathematical procedures to estimate the mean and the variance of the process Z_t are also provided and verified using simulation experiments. The procedures developed earlier in Chapter 3 have been extended to evaluate the occurrence probabilities, return period, and the risk of regional droughts. Procedure to evaluate the probability in which the regional drought may grow from one region to the other is provided in Chapter 4. The developed procedures have been used to characterize regional hydrologic droughts in the Upper Colorado River basin.

In Chapter 5, the generalization of the record extension model REXTN is made to consider the reconstruction of streamflow records at many sites simultaneously back in time using tree ring indices as information source over the extension period. The multivariate version of the record extension model REXTN developed here was used to reconstruct streamflows of the Colorado River back to the year 1490 for the purpose of improving the characterization of droughts in the Colorado River basin. Finally, Chapter 6 outlines the summary of results, conclusions and recommendations.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

The term drought has been used intensively in literature to refer to the natural phenomenon in which the water supply falls short to fulfill the demand for water over the time, i.e. days, weeks, moths, or years. In literature, the definitions of the term drought are categorized into conceptual definitions and operational definitions based on the purpose of the researcher. The conceptual definitions are mainly used to help in understanding the concept of drought, for example, drought is defined as the period of time lasting weeks, months or years, during which the actual moisture supply consistently falls short of the climatically expected moisture supply (Ponce et al, 2000). On the other hand, the operational definitions are aimed at characterizing droughts by referring to the beginning, end, and the degree of drought severity, for example, referring to one of the operational definitions, drought event is defined as a succession of consecutive periods (runs) at which the water supply remains below the demand in which the run is preceded and succeeded by water supply that is equal to or bigger than the demand value (Salas et al, 2005). Ultimately, the operational definitions of droughts are of most interest because the degree of water shortage (severity) and the length of shortage time affect the urban water supply, modern complex industries, economic and social sectors, and the natural habitats (Heim, 2002).

All recent studies consider the drought to be the most complex hydrologic phenomenon. Thus for the purpose of quantifying droughts, drought indices as quantitative measures have been developed. The most common drought indices used in the United States are the Palmer Drought Severity Index, the Standardized Precipitation Index, and the Drought Monitor, however drought indices are not suitable to characterize extreme historical droughts (Rossi and Cancelliere, 2003; Salas et al, 2005). Alternatively, other studies have made attempts toward characterizing droughts employing statistical techniques that are ultimately enable one to obtain frequency statements regarding properties of severe droughts based on single site analysis (e.g. Yevjevich, 1967; Sen, 1980; Sharma, 1997; Fernandez and Salas, 2001; Shiau and Shen, 2001; Gonzalez and Valdes, 2003; Salas et al, 2005) and based on regional analysis of droughts (e.g. Guerrero-Salazar and Yevjevich, 1975; Tase, 1976; Santos, 1983; Kingery, 1992; Sen, 1998; Shin and Salas, 2000; Hisdal and Tallaksen, 2003; Bayazit and Onoz, 2005).

When the observed hydrologic data, i.e. streamflow or precipitation, is used to characterize extreme historical droughts then one may obtain unreliable drought properties in the light that only few drought events may be observed when analyzing the relatively short observed records. Therefore, one may use longer proxy data to reliably extend the length of the short observed records for the purpose of improving the characterization of extreme droughts. One purpose to do so is the utilization of the long record tree ring indices to reconstruct long hydrologic records through reliable record extension models (e.g. Woodhouse, 2001; Gonzalez and Valdes, 2003; Salas et al, 2006).

The general purpose of this chapter is to review the literature related to drought characterization procedures based on single site and regional analysis. In addition, review on the basics of the discrete autoregressive moving average processes used to model the sequence of wet and dry states is also included. Moreover, procedures related to streamflow reconstruction techniques used for the purpose of improving drought characterization are also reviewed.

2.2 Definitions of Drought Events: Review and Basic Concepts

Let Y_t be the time series of the hydrologic variable under consideration (streamflow or precipitation). If the series Y_t is truncated at the level y_o, then series of deficits when $Y_t < y_o$ and surpluses when $Y_t \geq y_o$ is resulted. The continuous time interval when $Y_t < y_o$ is called a run of deficits. Generally in literature, drought events are defined conceptually according to the way the run of successive deficits is considered. Yevjevich (1967) defined the drought of length l as the event that is made of a run of l successive deficits that are preceded and followed by at least on surplus. Figure 2.1 depicts three typical drought events defined following the first definition, i.e. one two years, one three years and one four years drought event. This definition of drought events has been used widely in literature to characterize the severity of extreme droughts (e.g. Sen, 1976; Dracup et al, 1980; Sen, 1980; Sharma, 1998; Shiau and Shen, 2001; Cancelliere et al, 2003; Salas et al, 2005).

The second possible definition of drought events is to the definition of runs as in Feller (1968). The l years drought is defined as the event that is made of any possible run of l successive deficits as long as runs do not overlap each other. Figure 2.2 shows 4 typical two years drought events labeled A, B, C, and D defined according to the second definition. On the other hand, the third definition of drought events is basically derived from the alternative definition of runs (e.g. Schwager, 1983; Fernandez and Salas, 1999). It defines the drought event of length l as the event that is made of any possible run of l successive deficits regardless where the run starts as long as it satisfies the continuity of l successive deficits. In this case it differs from the second definition by allowing adjacent runs to overlap each other. Figure 2.3 shows possible two years drought events defined according to the third definition, i.e. the events labeled A, B, C, D, E, and F.

Regardless of the definition used, the main components of the drought event, i.e. the length L and the magnitude D (Figure 2.1) are defined the same. The first is defined as the period of time during which the variable of interest is continuously below the truncation level, while the later is defined as the amount of the deficit over the drought length. These components are random and show to be significantly correlated to each other (Shiau and Shen, 2001; Salas et al, 2005). If the variable Y_t is truncated at the level y_o, then the single year deficit d_t is (Salas et al, 2005):

$$d_t = y_o - Y_t \tag{2.1}$$

Related to the single deficits d_t the drought magnitude D is given as:

$$D = \sum_{t=1}^{l} d_t \tag{2.2}$$

Besides its length and magnitude, drought events may also be characterized by its initiation, termination, return period and spatial extent (Salas et al, 2005). The return period of an event can be defined in different ways for different applications (Fernandez and Salas, 1999; Douglas et al, 2002). It can be defined as the expected value of the waiting time N to the first occurrence of an event that is greater than or equal particular event (Fernandez and Salas, 1999). It can be defined also as the expected value of the interarrival time W between any two successive occurrences of an event that is greater than or equal particular event (Fernandez and Salas, 1999). Figure 2.4 shows typical definition of the first arrival time N and the recurrence time W between recurrent drought events. Moreover, drought event is associated also with risk of occurrence during the time span S. Chung and Salas (2000) defined the risk as the probability that the particular drought event has exceeded the critical event at least once during the time span S. If the underlying variable Y_t is truncated at the level y_o, then the resulted sequence X_t of deficits and surpluses is defined as follows:

$$X_t = \begin{cases} 0 & \textit{when } Y_t < y_o \\ 1 & \textit{otherwise} \end{cases} \tag{2.3}$$

where the state 0 denotes the occurrence of deficit and the state 1 denotes the occurrence of surplus. The main characteristics of the discrete autoregressive moving average (DARMA) processes used to model the 0,1 process X_t will be discussed next.

2.3 Review on DARMA (*p, q*) Models

The family of the discrete autoregressive moving average DARMA(p,q) models were developed by Jacobs and Lewis (1977, 1978a, 1978b) with p as the order of the autoregressive component and q as the order of the moving average component. The DARMA(p,q) models are the discrete counterparts of the well known ARMA(p,q) models (Chung and Salas, 2000). When q the order of the moving average component is zero, then the DARMA(p,q) process reduces to the DAR(p) process. If the process X_t has finite discrete state space D, then the DAR(1) process is given by:

$$X_t = V_t X_{t-1} + (1 - V_t) Y_t \tag{2.4}$$

where V_t is an independent Bernoulli (0,1) process with parameter λ such that $P[V_t = 1] = \lambda$ with $0 \leq \lambda \leq 1$, and Y_t is an independent identically distributed random variable such that $P[Y_t = i] = \pi_i$ with $0 \leq \pi_i \leq 1$. When the discrete state space $D \in \{0,1\}$ then each of variables V_t and Y_t is 0,1 process and consequently X_t is a 0,1 process too. The process X_t can be written as (Chung, 1999):

$$X_t = \begin{cases} X_{t-1} & \textit{with probability}\, \lambda \\ Y_t & \textit{with probability}\, 1-\lambda \end{cases} \tag{2.5}$$

the expected value of the process X_t is given as (Chung, 1999):

$$E[X_t] = E[Y_t] = \sum_{a \in D} a\pi_a \tag{2.6}$$

and the variance of X_t is (Chung, 1999):

$$V[X_t] = E[X_t^2] - \{E[X_t]\}^2 = \sum_{a \in D} a^2 \pi_a - \left(\sum_{a \in D} a\pi_a \right)^2 \tag{2.7}$$

the autocorrelation function of the of the process X_t is (Chung, 1999):

$$\rho_k(X) = \lambda^k \qquad \text{for } k \geq 1 \tag{2.8}$$

The DAR(1) process X_t given by (2.4) is a first order Markov chain with transition probabilities (Chung, 1999):

$$P[X_{t+1} = j \mid X_t = i] = p_{ij} = \begin{cases} (1-\lambda)\pi_j & i \neq j \\ \lambda + (1-\lambda)\pi_j & i = j \end{cases} \tag{2.9}$$

and for $i, j \in \{0, 1\}$ the transition probability matrix Q is:

$$Q = \begin{bmatrix} \lambda + (1-\lambda)\pi_0 & (1-\lambda)\pi_1 \\ (1-\lambda)\pi_0 & \lambda + (1-\lambda)\pi_1 \end{bmatrix} \tag{2.10}$$

The DARMA(1,1) process is expressed as follows (Chung, 1999):

$$X_t = U_t Y_t + (1-U_t) Z_{t-1} \tag{2.11}$$

where, U_t is an independent Bernoulli (0,1) process with parameter β, such that $P[U_t = 1] = \beta$ with $0 \leq \beta \leq 1$, and Y_t is as defined previously, and Z_{t-1} is a DAR(1) process with parameter λ as in (2.4). The variables X_t, Y_t and Z_t are stationary and have the same probability distribution π_0 and π_1. The DARMA(1,1) process X_t can be written as (Chung, 1999):

$$X_t = \begin{cases} Y_t & \textit{with probability } \beta \\ Z_{t-1} & \textit{with probability } 1-\beta \end{cases} \tag{2.12}$$

The expected value and variance of the DARMA(1,1) process X_t can be obtained using (2.6) and (2.7) respectively. Its autocorrelation function is (Chung, 1999):

$$\rho_k(X) = c\lambda^{k-1} \qquad \text{for } k \geq 1 \tag{2.13}$$

where c is a coefficient that is given by (Chung, 1999):

$$c = (1-\beta)(\lambda + \beta - 2\lambda\beta) \tag{2.14}$$

If the discrete state space D is defined as $D = \{0,1\}$, then each of Y_t, Z_t , and X_t is 0,1 process. The DARMA(1,1) process X_t defined by (2.12) is not Markov chain (Chung, 1999), however the bivariate sequence $\{X_t , Z_t\}$ is Markov chain. For any u, v, k and m the transition probabilities of this bivariate sequence may be expressed as (Chung, 1999):

$$H_k(u,v) = P[X_{t+1} = k, Z_{t+1} = v \mid X_t = m, Z_t = u] \tag{2.15}$$

referring to (2.11), the pair (X_t , Z_t) is independent of X_{t-1}, thus:

$$H_k(u,v) = P[X_{t+1} = k, Z_{t+1} = v \mid Z_t = u] \tag{2.16}$$

The one-step transition probability matrices of the two-state bivariate Markov chain are (Chung, 1999):

$$H_0(u,v) = \begin{bmatrix} \lambda(1-\beta) + [1-\lambda(1-\beta)]\pi_0 & (1-\beta)(1-\lambda)\pi_1 \\ \beta(1-\lambda)\pi_1 & \beta\lambda\pi_0 \end{bmatrix} \tag{2.17}$$

and

$$H_1(u,v)=\begin{bmatrix}\beta\lambda\pi_1 & \beta(1-\lambda)\pi_1\\ (1-\beta)(1-\lambda)\pi_0 & \lambda(1-\beta)+[1-\lambda(1-\beta)]\pi_1\end{bmatrix} \tag{2.18}$$

The two-state bivariate sequence $\{X_t, Z_t\}$ can be converted into four state simple Markov chain W_t by setting $W_t = 2X_t + Z_t$. The values of X_t, Z_t and the corresponding values of W_t are (Chung, 1999):

X_t	Z_t	The corresponding values of W_t
0	0	0
0	1	1
1	0	2
1	1	3

The four state simple Markov chain W_t has transition probabilities given as follows for $a, b \in \{0, 1, 2, 3\}$:

$$p_w(a,b)=P[W_{t+1}=b\,|\,W_t=a] \tag{2.19}$$

for example the probability $p_w(0,1)$ is obtained as follows:

$$p_w(0,1)=P[W_{t+1}=1\,|W_t=0]=P[X_{t+1}=0,Z_{t+1}=1\,|\,X_t=0,Z_t=0]$$

$$=P[X_{t+1}=0,Z_{t+1}=1\,|Z_t=0]=H_0(0,1)$$

following the same concept, the transition probability matrix Q of the univariate Markov chain W_t becomes (Chung, 1999):

$$Q=\begin{array}{c} \\ 0\\1\\2\\3\end{array}\!\!\begin{array}{c}\begin{array}{cccc}\;\;0\;\;\;\;\;\; & \;\;\;\;1\;\;\;\;\;\; & \;\;\;\;2\;\;\;\;\;\; & \;\;\;\;3\;\;\;\;\end{array}\\ \begin{bmatrix}H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1)\\ H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1)\\ H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1)\\ H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1)\end{bmatrix}\end{array} \tag{2.20}$$

and the marginal distribution of the state 0 in the process W_t is obtained as:

$$
\begin{aligned}
P[W_t = 0] &= P[X_t = 0, Z_t = 0] \\
&= P[X_t = 0, Z_t = 0 \mid Z_{t-1} = 0] P[Z_{t-1} = 0] \\
&+ P[X_t = 0, Z_t = 0 \mid Z_{t-1} = 1] P[Z_{t-1} = 1] \\
&= H_0(0,0)\pi_0 + H_0(1,0)\pi_1
\end{aligned}
$$

following the same concept, the complete results are (Chung, 1999):

$$
\begin{aligned}
P[W_t = 0] &= H_0(0,0)\pi_0 + H_0(1,0)\pi_1 \\
P[W_t = 1] &= H_0(0,1)\pi_0 + H_0(1,1)\pi_1 \\
P[W_t = 2] &= H_1(0,0)\pi_0 + H_1(1,0)\pi_1 \\
P[W_t = 3] &= H_1(0,1)\pi_0 + H_1(1,1)\pi_1
\end{aligned}
\tag{2.21}
$$

From the Markov process W_t the states 0 and 1 imply the state 0 in the process X_t, while the states 2 and 3 imply the state 1 in X_t. The marginal discrete probability π_a, where $a \in D$, is estimated as follows (Chang et al, 1984):

$$
\hat{\pi}_a = \frac{\overline{T}_a}{\sum_{j \in D} \overline{T}_j} \tag{2.22}
$$

where $\overline{T}_a$ is the mean run length of state a. If the discrete state space $D \in \{0,1\}$ then the marginal discrete probabilities π_0 and π_1 can be estimated by $\hat{\pi}_0$ and $\hat{\pi}_1$ using (2.22) as $\hat{\pi}_0 = \overline{T}_0 / (\overline{T}_0 + \overline{T}_1)$ and $\hat{\pi}_1 = \overline{T}_1 / (\overline{T}_0 + \overline{T}_1)$ respectively. The parameter λ of the DAR(1) process may be estimated from (2.8) by replacing the sample autocorrelation coefficient $\hat{\rho}_1(X)$ at lag k =1 and solving for λ. For the DARMA(1,1) process, the parameter c may be estimated from (2.13) by replacing the sample autocorrelation coefficient $\hat{\rho}_1(X)$ at lag k =1 as follows (Chung, 1999):

$$
c = \hat{\rho}_1(X) \tag{2.23}
$$

substituting (2.23) in (2.13), the parameter λ for the DARMA(1,1) process may by estimated as (Chung, 1999):

$$\hat{\lambda} = \frac{\hat{\rho}_2(X)}{c} = \frac{\hat{\rho}_2(X)}{\hat{\rho}_1(X)} \tag{2.24}$$

Buishand (1978) used the high order autocorrelations to estimate the parameter λ for the DARMA(1,1) process that minimizes the equation:

$$f(\lambda) = \sum_{k=1}^{m} \left[r(k) - c\lambda^{k-1} \right]^2 \tag{2.25}$$

Chung (1999) used the Newton-Raphson technique with the initial guess $\lambda = r(2)/r(1)$ to solve for the parameter λ. The parameter β can be estimated using (2.14) by substituting the proper values for λ and c (Chung, 1999).

2.4 Characterization of Single Year Drought Events

The single year drought event is defined as the event that is made of a run of one year where the study variable (precipitation or streamflow) is below the truncation level (Chung and Salas, 2000). Assuming stationary process, Fernandez and Salas (1999) provided analytically the distribution of the first arrival time, the distribution of the recurrence time, return period, and the risk of the single time step drought event when the underlying 0,1 process X_t is either independent or simple Markov, i.e. DAR(1) process. For single year events arise from the simple Markov process X_t, the probability $f_{n,e}$ that the event e occurs for the first time at the nth trial is (Fernandez and Salas, 1999):

$$f_{n,e} = \begin{Bmatrix} 0 & for\, n = 0 \\ p_0 & for\, n = 1 \\ (1 - p_0) p_{11}^{n-2} p_{10} & for\, n \geq 2 \end{Bmatrix} \tag{2.26}$$

The return period T as the expected value of the first occurrence time N is (Fernandez and Salas, 1999):

$$T = E(N) = 1 + \frac{1 - p_0}{p_{10}} \tag{2.27}$$

where the probability $p_i = P[X_t = i]$ is the state i unconditional probability of X_t, and the probability $p_{ij} = P[X_t = j \mid X_{t-1} = i]$ is the conditional state probability of X_t, i.e. the probability to get the state j given that the previous state is i. The unconditional probability p_i of the simple Markov process X_t can be determined using (2.22), i.e. $p_i = \pi_i$.

For the two states process, Sen (2003) suggested that p_i can be determined as the ratio $N_i / (N_i + N_j)$ where N_i and N_j are respectively the number of occurrences of states i and j in the process X_t. The conditional probabilities p_{00}, p_{01}, p_{10} and p_{11} can be determined using (2.10) for the simple Markov process X_t. As an alternative to (2.10), the probability p_{00} can be evaluated as (Cramer and Leadbetter, 1967):

$$p_{00} = p_0 + \frac{1}{2\pi p_0} \int_0^{\rho} \frac{\exp\left[-q_c^2 / (1+z)\right]}{\sqrt{1-z^2}} dz \tag{2.28}$$

where $p_0 = P[Y_t \le y_o] = P[X_t = 0]$ and z is the dummy variable of the integration, ρ is lag1 serial correlation coefficient of the process Y_t, and the constant q_c is defined as the quantile of the cumulative normal distribution such that $q_c = \Phi^{-1}(P[Y_t \le y_c])$. The probability p_{01} can be evaluated as $1 - p_{00}$. As an alternative to (2.10), the probability p_{10} can be found as (Sen, 1976):

$$p_{10} = \frac{p_0}{1 - p_0}(1 - p_{00}) \tag{2.29}$$

and the probability $p_{11} = 1 - p_{10}$. For single events arise from the Markov process X_t, the occurrence probability $f_{w,e}$ of the recurrent event e is (Fernandez and Salas, 1999):

$$f_{w,e} = \begin{cases} 0 & for\, w = 0 \\ p_{00} & for\, w = 1 \\ p_{01} p_{11}^{w-2} p_{10} & for\, w \geq 2 \end{cases} \tag{2.30}$$

The return period as the expected value of the recurrence time W is T (Fernandez and Salas, 1999):

$$T = E(W) = p_{00} + \frac{p_{01}}{p_{10}}(1 + p_{10}) \tag{2.31}$$

For single events, if the process X_t is independent, then (2.27) and (2.31) state clearly that the expected value of the first occurrence time is the same as the expected value of the recurrence time, i.e. $T = E(N) = E(W) = 1/p_0$. This result is expected because, for the single event e, the probability of the first occurrence is the same as the probability of the occurrence when X_t is independent. In that case, the probability $f_{w,e}$ does not depend on the history at or before w-1 time steps, i.e. does not depend on the initial conditions as in (2.30) so basically $f_{w,e}$ is the same as $f_{n,e}$. The risk R_e that at least one single drought event e has occurred during the time span S is given as (Fernandez and Salas, 1999):

$$R_e = 1 - (1 - p_0)(1 - p_{10})^{S-1} \tag{2.32}$$

If the process X_t is not Markov, then (2.26) and (2.30) cannot be used to characterize the single drought event arises form X_t. Let X_t be DARMA(1,1) process, then as mentioned previously, the two-state bivariate sequence $\{X_t, Z_t\}$ can be converted into four state simple Markov chain by defining W_t. Chung and Salas (2000) gave the probability distribution of the first occurrence time of the single drought event A as follows:

$$f_{n,A} = \begin{cases} 0 & for\, n = 0 \\ P(A) & for\, n = 1 \\ P(B)G(B,B)^{n-2}G(B,A); & for\, n \geq 2 \end{cases} \tag{2.33}$$

where $P(A) = P[W_t = 0] + P[W_t = 1]$, $P(B) = P[W_t = 2] + P[W_t = 3]$, and the event B is defined as non-drought event. The expected value of the first arrival time N of the event A is (Chung and Salas, 2000):

$$T(A) = E(N) = \sum_{n=0}^{\infty} nf_{n,A} = 1 + \frac{P(B)}{G(B,A)} \tag{2.34}$$

The probability distribution of the recurrence time of the event A is (Chung and Salas, 2000):

$$f_{w,A} = \begin{cases} 0 & for\, w = 0 \\ G(A,A) & for\, w = 1 \\ G(A,B)G(B,B)^{w-2}G(B,A); & for\, w \geq 2 \end{cases} \tag{2.35}$$

and the expected value of the recurrence interval of the event A is (Chung and Salas, 2000):

$$T(A) = E(W) = \sum_{w=0}^{\infty} wf_{w,A} = G(A,A) + \frac{G(A,B)}{G(B,A)}[1 + G(B,A)] \tag{2.36}$$

The probability $G(I,J)$ is defined as the probability to arrive to J from I, which is given as:

$$G(I,J) = P[W_{t+1} \in J \mid W_t \in I] = \frac{\sum_{r \in J} \sum_{k \in I} P[W_t = k] g^{(1)}(k,r)}{P(I)} \tag{2.37}$$

and generally the probability $g^{(n)}(i,j)$ is:

$$g^{(n)}(i,j) = P[W_{t+n} = j \mid W_t = i] \tag{2.38}$$

for $n = 1$, $g^{(1)}(i,j)$ is given by (2.19).

The risk R_A that at least one single drought event A has occurred within the time period 0 - S is given as (Chung and Salas, 2000):

$$R_A = \sum_{n=0}^{S} f_{n,A} = P(A) + P(B)[1 - G(B,B)^{S-1}] \tag{2.39}$$

2.5 Characterization of Multiyear Drought Events

The multiyear drought event is defined as the event that spans over many years where deficits are continuously observed. The drought event of interest is the event { A_l } where exactly there are a run of l years continuous deficit regardless where the run starts. The first occurrence probability f_{n,A_l} of the multiyear drought event { A_l } is defined as the probability that the drought event { A_l } occurs for the first time at the nth trial (Fernandez and Salas, 1999). For drought events arise from simple Markov process X_t, the probability f_{n,A_l} for the multiyear drought event { A_l } is (Fernandez and Salas, 1999):

$$f_{n,A_l} = \begin{bmatrix} 0 & \text{for } n < l \\ p_0 (p_{00})^{l-1} & \text{for } n = l \\ u_{0,n-l} p_{00}^{l} + u_{1,n-l} p_{10} p_{00}^{l-1} - \sum_{i=1}^{l-1} f_{n-i,A_l} p_{00}^{i} & \text{for } n > l \geq 2 \end{bmatrix} \tag{2.40}$$

The term $u_{j,n}$ is defined as the probability that run l does not occur at or before trial n on which state j was observed,

$$u_{0,n} = u_{0,n-1} p_{00} + u_{1,n-1} p_{10} - f_{n,A_l} \tag{2.41}$$

and

$$u_{1,n} = u_{0,n-1} p_{01} + u_{1,n-1} p_{11} \tag{2.42}$$

The initial probabilities are $u_{0,1} = p_0$ and $u_{1,1} = p_1$. For the drought event { A_l }, the risk $R_l(S)$ that event { A_l } has occurred at least once at or before the trial $n = S$ is (Chung and Salas, 2000):

$$R_{A_l}(S) = \sum_{n=0}^{S} f_{n,A_l} \tag{2.43}$$

If the process X_t shows to have longer memory dependence structure, for example X_t is DARMA(1,1) process, then (2.43) can not be used to evaluate the risk of multiyear drought events arising from the DARMA(1,1) process X_t. Assuming that X_t, is DARMA(1,1) process, then it has been shown that the two-state bivariate sequence {X_t, Z_t} given by (2.11) can be converted into four state simple Markov chain $W_t \in \{0,1,2,3\}$, with states 0 and 1 in W_t imply state 0 in the process X_t, and states 2 and 3 in W_t imply state 1 in X_t. Since the drought event is defined as run of dry states, then this run can be a run of all 0s, all 1s or any combination of 0s and 1s, i.e. there are 2^l run patterns of dry states, therefore the l year drought event is possible to occur if any of the 2^l run patterns occur (Chung and Salas, 2000). Moreover, the first occurrence of the multiyear drought event is sensitive to the type of the run pattern (the type of the 0, 1 mixture) where some runs could have shorter first occurrence time and hence higher risk of occurrence while other patterns could show longer first occurrence time and hence lower risk of occurrence (Chung and Salas, 2000). Based on that Chung and Salas (2000) considered the weight $w^{(i)}$ of each run pattern i in deriving the analytical procedure to calculate the risk $R^*_{A_l}(S)$

of the multiyear drought event$\{A_l\}$ when the process 0,1 process X_t is DARMA(1,1). The risk $R^*_{A_l}(S)$ is given as (Chung and Salas, 2000):

$$R^*_{A_l}(S) = 1 - \prod_{i=1}^{2^l} \left[1 - w^{(i)} R^{(i)}_{A_l}(S)\right] \tag{2.44}$$

For the run pattern i the risk $R^{(i)}_{A_l}(S)$ is given by (2.43) while the weight $w^{(i)}$ is given as (Chung and Salas, 2000):

$$w^{(i)} = \frac{T^{(m)}}{T^{(i)}} \tag{2.45}$$

$T^{(i)}$ is the expected first arrival time of run of pattern i, and $T^{(m)}$ is the shortest expected first arrival time among all patterns given by (Chung and Salas, 2000):

$$T^{(m)} = \min[T^{(1)}, T^{(2)}, \ldots\ldots, T^{(2^l)}] \tag{2.46}$$

The expected first arrival time $T^{(i)}$ of the run of pattern i can be evaluated using the algorithm proposed by Schwager (1983) for runs in Markov dependent process which can be summarized as follows: Let the run r of the pattern i be the run of length l that appears in simple Markov process where the outcome probability at certain trial depends only on the outcome of the previous trial. Define $p(h, j)$ as the transition probability to reach state h given that the previous state is j, while $u_{j,n}$ as probability that run r does not occur at or before trial n where state j was observed. The probability that the last j entries of run r occur while the last $(j+1)^{\text{th}}$ trial is h is given by:

$$q(h, j) = p[h, r(l - j + 1)] \prod_{k=l-j+2}^{l} p[r(k-1), r(k)] \qquad j \le l \tag{2.47}$$

with $r(k)$ as the k^{th} entry of the run r that has the length l and the pattern i. For the initial conditions, Schwager (1983) assumed $h = 0$, thus the first occurrence of the run of l entries is given by (Schwager, 1983):

$$f^{(n)} = \begin{bmatrix} 0 & for\, n < l \\ q(0,l) & for\, n = l \\ \sum_{h=1}^{v} u_{h,n-l}\, q(h;l) - \sum_{k \in K} f^{(n-k)} q[r(k),l] / q[r(k),l-k] & for\, n > l \end{bmatrix} \tag{2.48}$$

and the probability $u_{j,n}$ is given by:

$$u_{j;n} = \begin{bmatrix} \sum_{h=1}^{v} u_{h,n-1}\, p(h;j); & for\ j \neq r(l), n \geq 2 \\ \sum_{h=1}^{v} u_{h,n-1}\, p(h;j) - f^{n} & for\ j = r(l) \end{bmatrix} \tag{2.49}$$

where the upper limit v is the total number of possible states, and

$$K = \{k \mid 1 \leq k \leq l-1,\ r(k+1,l) = r(1,l-k)\}$$

The expected first arrival time of run pattern i is:

$$T^{(i)} = E(N) = \sum_{n=0}^{\infty} n f^{(n)} \tag{2.50}$$

Given that the process X_t is DRAMA(1,1) and the risk $R^{*}_{A_l}(S)$ from (2.44), the first occurrence probability f_{n,A_l} for the multiyear drought event $\{A_l\}$ is (Chung and Salas, 2000):

$$f_{n,A_l} = \begin{bmatrix} R^{*}_{A_L}(n) & n = 1 \\ R^{*}_{A_L}(n) - R^{*}_{A_L}(n-1) & n \geq 2 \end{bmatrix} \tag{2.51}$$

2.6 Characterization of Multiyear Regional Droughts

Droughts as extreme events are characterized by having random magnitude that is associated to the random length, however in reality droughts are considered to be regional incidents as result of having the ability to develop randomly in the space, i.e. droughts are characterized also by spatial coverage. Therefore, the definition of drought events should include the areal coverage aspect (Hisdal et al, 2000). The characterization of regional droughts has been a major concern for many researchers over the past few decades that resulted in developing several procedures for characterizing regional droughts. Among procedures that can be used to describe the time and space variability of droughts are the drought indices. However, they fall short to provide frequency statements on drought occurrences (Salas et al, 2005). Statistical techniques have been used to characterize regional droughts considering the areal aspect of droughts. Two main approaches have been suggested in literature in which the characterization of regional droughts incorporates the regional aspect of droughts. The first relies on the proportion of the area covered by the regional deficit, i.e. the study area is divided into cells where the study variable is analyzed per cell, ultimately reasonable analysis on regional droughts may be utilized by referring to the proportion of the area covered by the drought relative to some reference (e.g. Tase, 1976; Sen, 1980; Santos, 1983; Kingery, 1992; Shin and Salas, 2000; Hisdal and Tallaksen, 2003). On the other hand, the second relies on the joint analysis of the study variable at sites where regional drought conditions are observed (e.g. Yevjevich, 1972; Guerrero-Salazar and Yevjevich, 1975; Sen, 1979; Sen, 1998; Bayazit and Onoz, 2005).

2.7 Defining the Areal Aspect of Regional Droughts

The first consideration of the areal aspect in characterizing regional droughts was introduced through the work of Tase (1976). Dividing the study region into grid of sub-areas (cells) of the same size and considering precipitation as the study variable, Tase (1976) introduced the areal aspect of regional droughts through the analysis of deficits in each cell. In that case, a regional drought occurs if the sum of the single cells where deficits are observed exceeds certain critical area. The deficit area A is defined as the number of grid cells at which deficit occurs, i.e. with precipitation below certain threshold ξ_o (Tase, 1976):

$$A = \sum_{i=1}^{M} I_{(\xi \le \xi_o)}(\xi_i) \tag{2.52}$$

where $I_{(\xi \le \xi_o)}(\xi_i)$ is an indicator function describing the presence of deficit in the site i:

$$I_{(\xi \le \xi_o)}(\xi_i) = \begin{cases} 1 & if\ \xi_o \le \xi_i \\ 0 & if\ \xi_o > \xi_i \end{cases} \tag{2.53}$$

Other useful indices were introduced also by Tase (1976), for example the areal deficit D, and the maximum deficit intensity I. On the basis of random fields concept and probability theory, Sen (1980) gave theoretically the probability distribution functions of the deficit area A, the areal deficit D, and their joint distribution under the assumption that the random field is consisting of spatially independent hydrological variable at the m sites forming the whole region under study. Considering the work of Tase (1976) as a base line, Santos (1983) elaborated more in characterizing regional droughts by redefining and creating new set of indices that are called the auxiliary variables of the regional droughts. These indices were used to define the regional drought properties, i.e. the drought duration, the regional areal deficit, the proportion of temporal drought area, and the temporal total regional areal deficit. Complete analytical framework that shows the first two moments of the auxiliary variables and regional drought properties besides their approximate distribution functions can be found in literature (e.g. Santos,1983).

Regardless of the magnitude of the regional deficit, generally the regional drought event can be defined as the continuous time interval during which the deficit area equals to or greater than a predefined critical area (Santos, 1983). This critical area can be defined as a significant proportion of the total region area, where values as large as 50% or 75% of the total region area may be taken as the critical area (Santos, 1983).

2.8 Joint Analysis of Multi-Sites Under Drought

In characterizing regional hydrological droughts, i.e. the study variable is streamflow, dividing the study area into cells of the same size may not be utilized easily due to the nature of the flow contributing drainage areas, i.e. usually flow drainage areas are irregular in shape and have different size. Therefore, the concept of joint analysis of many sites under drought conditions seems to be more realistic than the concept of the areal aspect, i.e. grid of cells of equal size, to characterize regional hydrologic droughts.

Considering the concept the joint analysis, Guerrero-Salazar and Yevjevich (1975) introduced the bivariate case of joint analysis. Assuming that the 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$ at sites (1) and (2) respectively are simple Markov each resulting from truncating the normally distributed variables $Y_t^{(1)}$ and $Y_t^{(2)}$ at the levels $y_o^{(1)}$ and $y_o^{(2)}$ respectively, and that $X_t^{(1)}$ and $X_t^{(2)}$ show to be serially and mutually dependent, Guerrero-Salazar and Yevjevich (1975) aimed at obtaining the probability distribution of the longest negative run for such case by defining the variable Z_t such that:

$$Z_t = 1 - [(1 - X_t^{(1)})(1 - X_t^{(2)})] \qquad (2.54)$$

The occurrence of deficit in $X_t^{(i)}$,$\{i = 1, 2\}$ is assigned the value 0 therefore the process Z_t will take the value 0 whenever deficits occur at $X_t^{(1)}$ and $X_t^{(2)}$ at the same time and takes 1 otherwise. Later, considering the joint analysis of truncated hydrological variable at n sites, Sen (1979) derived analytically the probability of occurrence that n sites are under drought conditions considering different cases of the serial and mutual correlations. However for the case that n series are serially and mutually dependent, the analytical procedure provides an approximate solution.

Based on an independent dry and wet spell occurrences, Sen (1998) investigated theoretically through two different probabilistic models the probability distribution function of the regional drought spatial and temporal extents. The first model assumes that once the drought hits part of the region then that part will remain under drought conditions and then as the time passes the drought will hit a new sub area until the whole region is covered. The second model assumes that drought stricken area might recover and have a wet period during the coarse of time, i.e. as the time passes some dry spells may stay dry or may become wet.

Bayazit and Onoz (2005) assumed that the probability distribution of the regional drought length could be modeled as geometric with parameter θ. The parameter θ was derived theoretically assuming that truncated flows in many sites are cross-correlated and first order Markov process each.

2.9 Drought Characterization Through Using Tree Ring Reconstructed Flows

The characterization of severe hydrologic droughts may be utilized through analyzing the historical streamflow records if they are long enough. However, the historical records are limited in their length and they may not provide a reliable picture about the total natural variability that affects the characterization of extreme hydrologic droughts (Woodhouse, 2001). Therefore it is extremely important to reliably lengthen the short historical streamflow records to a length much longer than the length of the current historical records (Salas et al, 2006). One useful technique to lengthen short records is to build using the current historical flows a stochastic model that synthesizes long series of streamflows that may yield and exhibit additional features and scenarios of possible extreme events that may occur in the future (Salas et al, 2005). Another useful technique that has been suggested in literature is to extend short historical flow records back in time based on reliable record extension model using tree ring indices as a source of information over the extension period (e.g. Woodhouse, 2001; Gonzales and Valdes, 2003; Salas et al, 2006).

Tree ring data have proven to be useful proxies for extending short records of hydrologic variables back in time (Loaiciga et al, 1993; Woodhouse, 2001; Salas et al, 2006). The usefulness of trees as proxies in reconstructing streamflows comes from being sensitive recorders to the natural climate variability. Usually trees assimilate into their growing rings the same set of climatic factors like precipitation and evapotranspiration that affect the amount of streamflow (Meko et al, 1995). Besides of being good recorders of the climatic variability, tree ring data has other advantages over other types of proxy data by being more accurate dating to the year and have the ability to preserve the low frequency and high frequency variations (Meko et al, 1995). The extension process of short records can be utilized through using different record extension models for different purposes. One practical purpose behind extending the short streamflow records at one or many sites utilizing the information that is contained in nearby sites with longer records is to improve the characterization of extreme hydrologic droughts. This can be achieved through the improvement in the reliability of parameters estimates. The reliability of parameter estimator is measured by the variance of that estimator. Generally the lower the variance of the estimator the better the corresponding parameter estimate represents the population parameter.

In literature it has been found that there are two main objectives behind extending short records. One objective is to improve the estimates of the population parameters of the underlying variable (e.g. the mean and the variance of the short record variable) while the other has been to maintain (preserve) the estimates of the parameters of the short record. Thus, several record extension models were proposed focusing on one objective or the other. The following section reviews record extension models that can be used to extend short streamflow records at many sites using information contained in longer tree ring records for the purpose of improving the characterization of droughts.

2.10 Multi-site Record Extension Methods for Reconstructing Streamflows

For extending one short variable given one long variable, a comprehensive and collective review on the purposes and procedures of the extension can be found in the literature (e.g. Salas et al, 2006). In summary, consider a short record denoted by the variable y_t, t=1, ..., N_1 and a longer record denoted by the variable x_t, t=1, ..., N_1, N_1+1,

..., N_1+N_2 where N_1 and N_2 are the lengths of the short and long records, respectively. Assuming that the variables y_t and x_t are bivariate normally distributed Matalas and Jacobs (1964) suggested a linear regression model with a noise term to extend the short records. They gave the unbiased estimators of the population mean and variance and the criteria so as to improve the estimators of he mean and the variance of y_t. However, Matalas and Jacobs' model neglected the serial correlation of y_t and the use of the noise term leads to multiple records extensions. To overcome the latter limitation, Hirsch (1982) introduced the MOVE1 and MOVE2 extension models intended to maintain the basic key statistics of the short record while Vogel and Stedinger (1985) introduced the MOVE3 and MOVE4 extension models that are considered an improvement over the MOVE1 and MOVE2 models, respectively. Also Grygier and Stedinger (1989) suggested the generalized MOVE (GMOVE) for extending short records intended to preserve the serial correlation of the short record and the cross correlation between y_t and x_t.

Salas et al (2006) developed recently a record extension model (REXTN), which is a linear regression model that includes an autoregressive term to account for the effect of the previous flows (serial correlation) and the noise term. The suggested model REXTN yields an extended record that maintains the basic statistics of the original series y_t, such as the mean, variance, and serial correlation, and the cross correlation between x_t and y_t. In addition, the method includes a procedure for selecting a single record extension from the many traces that may be generated using the random term.

Consider the case of p short records (p streamflow sites) defined by the vector $\boldsymbol{y}_t$ = $[y_t^{(1)} \ y_t^{(2)} \cdots y_t^{(p)}]^T$ having with the length N_1 and m longer records (m tree ring indices) defined by the vector $\boldsymbol{x}_t$ = $[x_t^{(1)} \ x_t^{(2)} \cdots x_t^{(m)}]^T$ having the length $N = N_1 + N_2$. The multivariate version of the extension procedures described above should be used to extend all short variables $\boldsymbol{y}_t$ simultaneously considering all the $\boldsymbol{x}_t$ variables. Such a multivariate multiple linear regression model may be expressed as:

$$\hat{\boldsymbol{y}}_t = \overline{\boldsymbol{y}}_1 + b_1(x_t^{(1)} - \overline{x}_1^{(1)}) + \cdots + b_m(x_t^{(m)} - \overline{x}_1^{(m)}) = \overline{\boldsymbol{y}}_1 + \boldsymbol{b}^T(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_1) \tag{2.55}$$

where $\overline{\boldsymbol{y}}_1$ is the vector of means of $\boldsymbol{y}_t$ over the period N_1, $\boldsymbol{b} = [b_1 \ b_1 \cdots b_m]^T$ is a vector of the model parameters with the superscript T stands for the transpose, and $\overline{\boldsymbol{x}}_1$ is the vector of means of the longer records $\boldsymbol{x}_t$ over the period N_1. The least squares estimate of the vector $\boldsymbol{b}$ is given by (Johnson and Wichern, 2005):

$$\hat{\boldsymbol{b}} = [(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_1)^T(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_1)]^{-1}(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_1)^T(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_1) \tag{2.56}$$

where the superscript -1 stands for the inverse of the matrix. Model (2.55) with least squares estimators will be referred to as MLRLS. Fiering (1962) studied the case where p=1 and m=2 and derived the maximum likelihood estimators of the mean and the variance of the short record.

Gilroy (1970) studied the record extension problem where p = 1 and $m \geq 1$. In this case the mean of the combined series $\tilde{y}_t = \{y_1, y_2, \cdots, y_{N_1}, \hat{y}_{N_1+1}, \cdots, \hat{y}_{N_1+N_2}\}$ with $\hat{y}_t$ obtained from (2.55) is the maximum likelihood unbiased estimator of the population mean, while the variance of $\tilde{y}_t$ is a biased estimator of the population variance. Gilroy (1970) extended the work of Matalas and Jacobs (1964) and derived the expressions for the unbiased estimators of the population mean and variance. It required that model (2.55) includes the error term ε_t. Thus the record extension model is given by:

$$\hat{y}_t = \bar{y}_1 + \boldsymbol{b}^T(\boldsymbol{x}_t - \bar{\boldsymbol{x}}_1) + \alpha(1-\hat{R}^2)^{1/2} s_{y1}\varepsilon_t \tag{2.57}$$

where s_{y1}^2 is variance of the short record y_t over the period N_1, $\hat{R}^2$ is the multiple correlation coefficient of y_t and $\boldsymbol{x}_t$, and ε_t is a normally distributed noise with mean 0 and variance 1.

The maximum likelihood unbiased estimators of the population mean μ_y and variance σ_y^2 are given by (Gilroy, 1970):

$$\bar{y} = \bar{y}_1 + \frac{N_2}{N}\boldsymbol{b}^T(\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}}_1) \tag{2.58}$$

and

$$s_y^{\,2} = \frac{1}{N-1}\left\{\begin{aligned}&(N_1-1)s_{y1}^{\,2} + (N_2-1)\boldsymbol{b}^T\boldsymbol{S}_{x2x2}\boldsymbol{b} + (N_2-1)\alpha^2(1-\hat{R}^2)s_{y1}^{\,2}\\ &+\frac{N_1N_2}{N}\boldsymbol{b}^T(\bar{\boldsymbol{x}}_2-\bar{\boldsymbol{x}}_1)(\bar{\boldsymbol{x}}_2-\bar{\boldsymbol{x}}_1)^T\boldsymbol{b}\end{aligned}\right\} \tag{2.59}$$

where $\bar{\boldsymbol{x}}_2$ is the sample mean vector of $\boldsymbol{x}_t$ over the period N_2, $\boldsymbol{S}_{UV}$ represents the lag-0 covariance matrix between variable vectors U and V^T, and the coefficient α is given by:

$$\alpha = \left[\frac{N_2(N_1-2m-2)(N_1-1)}{(N_2-1)(N_1-m-2)(N_1-m-1)}\right]^{1/2} \tag{2.60}$$

in which the coefficient α has been added to the regression equation (2.55) in order to remove the bias in the estimation of the population variance. Moran (1974) noticed that the unbiased estimator of the population variance given by Gilroy (1970) could be expressed in a less complicated form as

$$s_y^2 = \boldsymbol{b}^T \boldsymbol{S}_{xx} \boldsymbol{b} + \left[1 - \frac{m(N-m-2)}{(N_1-m)(N-1)}\right] s_{y.x}^2 \tag{2.61}$$

where $\boldsymbol{S}_{xx}$ is variance-covariance matrix of $\boldsymbol{x}_t$ defined over the whole period N, and $s_{y.x}^2$ is the residual mean square for the regression of y_t on $\boldsymbol{x}_t$ that is given by $s_{y.x}^2 = [s_{y1}^2 - \boldsymbol{b}^T \boldsymbol{P}_{xy}]/(N_1 - m - 1)$, in which $\boldsymbol{P}_{xy}$ is the cross product of $\boldsymbol{x}_t$ and y_t over the concurrent period N_1, i.e. $\boldsymbol{P}_{xy} = (\boldsymbol{x}_t - \bar{\boldsymbol{x}}_1)(y_t - \bar{y}_1)^T$.

Hirsch (1982) noticed that although the extension model of Matalas and Jacobs (1964), i.e. the bivariate case of (2.57), can be used to extend the short record over the extension period (N_2) which leads to the unbiased estimates of the population mean and variance, however, that procedure is based on adding a noise term which is not practical because it will lead to multiple record extensions. Moreover, that referred extension model does not preserve the serial correlation (Hirsch, 1982). Similarly, one would conclude the same for the multivariate extension model (2.57). Salas et al (2006) developed a record extension model that includes the noise term and an additional term to account for the serial correlation. The model denoted as REXTN is defined as (Salas et al, 2006):

$$\hat{y}_t = \bar{y}_1 + a(\hat{y}_{t-1} - \bar{y}_1) + \boldsymbol{b}^T(\boldsymbol{x}_t - \bar{\boldsymbol{x}}_2) + c\,\varepsilon_t \tag{2.62}$$

where a, $\boldsymbol{b}$, and c are the model parameters, and ε_t is the noise term that is normally distributed with mean 0 and variance 1. The parameters a, $\boldsymbol{b}$, and c are given by (Salas et al, 2006):

$$\hat{a} = [\boldsymbol{S}_{y1y1}(1) - \boldsymbol{S}_{y1x1}\boldsymbol{S}_{x1x1}^{-1}\boldsymbol{S}_{x1y1}(1)][\boldsymbol{S}_{y1y1} - \boldsymbol{S}_{x1y1}^T(1)\boldsymbol{S}_{x1x1}^{-1}\boldsymbol{S}_{x1y1}(1)]^{-1} \tag{2.63}$$

$$\hat{\boldsymbol{b}} = [\boldsymbol{S}_{y1x1} - \hat{a}\boldsymbol{S}_{x1y1}^T(1)]\boldsymbol{S}_{x1x1}^{-1} \tag{2.64}$$

and

$$\hat{c}^2 = \boldsymbol{S}_{y1y1}(1) - \hat{a}\,\boldsymbol{S}^T_{y1y1}(1) - \hat{\boldsymbol{b}}\,\boldsymbol{S}_{x1y1} \tag{2.65}$$

where the notation $\boldsymbol{S}_{UV}$ was defined previously and $\boldsymbol{S}_{UV}(1)$ represents the covariance matrix between variable vectors U_t and V^T_{t-1} (i.e. lag-1 covariance). Note that $\boldsymbol{S}_{x1y1} = \boldsymbol{S}^T_{y1x1}$ and when $p = 1$ then $\boldsymbol{S}_{y1y1}(1)$ is the lag-1 covariance of y_t over the period N_1.

Since (2.62) includes the random term ε_t multiple record extensions are possible. Salas et al (2006) detailed a procedure for selecting a single realization among the many possible realizations based on the statistical information that can be extracted from the longer record x_t. Considering the ratios $\frac{\bar{\hat{y}}_2}{\bar{y}_1}$, $\frac{\bar{x}_2}{\bar{x}_1}$, $\frac{s_{\hat{y}2}}{s_{y1}}$, and $\frac{s_{x2}}{s_{x1}}$, where $\bar{x}_1$, s_{x1} and $\bar{x}_2$, s_{x2} are the means and the standard deviations of x_t over the periods N_1 and N_2 respectively, and $\bar{y}_1$, s_{y1} and $\bar{\hat{y}}_2$, $s_{\hat{y}2}$ are respectively, the means and standard deviations of y_t based on the extended record and the short record. Salas et al (2006) suggested that, among many possible extensions of $\hat{y}_t$ over the extension period N_2, one could select the extended series $\hat{y}_t$ such that the quantity

$$z = \left\{0.5 \times \left|\frac{\bar{\hat{y}}_2}{\bar{y}_1} - \frac{\bar{x}_2}{\bar{x}_1}\right| + 0.5 \times \left|\frac{s_{\hat{y}2}}{s_{y1}} - \frac{s_{x2}}{s_{x1}}\right|\right\} \tag{2.66}$$

is minimum, where $|A|$ stands for the absolute value of A, and 0.5 is the weight considered. Salas et al (2006) showed that the REXTN model performed quite well and compared favorably against competing record extension models.

The record extension methods discussed above have considered the extension of one short variable y_t, i.e. p = 1, using one or more longer records $\boldsymbol{x}_t$, i.e. $m \geq 1$. However, there are many practical cases where one may need to extend several (short) variables jointly. Grygier and Stedinger (1989) considered this case and introduced the multivariate record extension model GMOVE as:

$$\hat{\boldsymbol{y}}_t = \overline{\boldsymbol{y}}_1 + \boldsymbol{A}(\boldsymbol{y}_{t-1} - \overline{\boldsymbol{y}}_1) + \boldsymbol{B}(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_2) \tag{2.67}$$

where the parameters matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be estimated by:

$$\boldsymbol{S}_{y1y1} - \boldsymbol{S}_{y1x1}\boldsymbol{S}_{x1x1}^{-1}\boldsymbol{S}_{y1x1}^{T} = \hat{\boldsymbol{A}}[\boldsymbol{S}_{y1y1} - \boldsymbol{S}_{x1y1}^{T}(1)\boldsymbol{S}_{x1x1}^{-1}\boldsymbol{S}_{x1y1}(1)]\hat{\boldsymbol{A}}^{T} \tag{2.68}$$

$$\hat{\boldsymbol{B}} = [\boldsymbol{S}_{y1x1} - \hat{\boldsymbol{A}}\boldsymbol{S}_{x1y1}^{T}(1)]\boldsymbol{S}_{x1x1}^{-1} \tag{2.69}$$

As noted, model (2.67) considers the term $\boldsymbol{y}_{t-1}$ but does not consider the random term ε_t. Salas et al (2006) compared the performance of model (2.67) for the case where p = 1 and m = 1 relative to those of other record extension models using simulation experiments. It has been found that GMOVE estimates efficiently the mean, variance, and the cross-correlation between the short and longer records but its performance is less efficient for reproducing the lag-1 serial correlation and predicting the extended record.

In many cases the length of the observed streamflow data is limited relative to the total number of parameters to be estimated using multivariate record extension models for extending streamflows at a large number of sites, i.e. the model may have a large number of parameters to be estimated relative to the available data and consequently the principle of parameters parsimony may not be preserved. For example, this may be the case of the Colorado River system where record extensions may be needed at the majority of the 29 sites. In these cases, a more reasonable approach may be to use a multivariate record extension model to extend the streamflows at a few sites (e.g. at certain key stations) then the streamflows at the remaining sites (substations) can be obtained using spatial disaggregation models. Let z_t be the vector of flows at some key stations obtained using a parsimonious multivariate record extension model and $\hat{y}_t$ the vector of flows to be estimated at the substations. The spatial disaggregation model to estimate the flows at the substations is (Salas et al, 1980):

$$\hat{\boldsymbol{y}}_t = \boldsymbol{A z}_t + \boldsymbol{B \varepsilon}_t + \boldsymbol{C}\, \hat{\boldsymbol{y}}_{t-1} \tag{2.70}$$

where it is assumed that the means have been subtracted from z_t and y_t, and $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are the model parameters, which are given by:

$$\hat{\boldsymbol{A}} = [\boldsymbol{S}_{y1z1} - \boldsymbol{S}_{y1y1}(1)\boldsymbol{S}_{y1y1}^{-1}\boldsymbol{S}_{z1y1}^{T}(1)][\boldsymbol{S}_{z1z1} - \boldsymbol{S}_{z1y1}(1)\,\boldsymbol{S}_{y1y1}^{-1}\,\boldsymbol{S}_{z1y1}^{T}(1)]^{-1} \tag{2.71}$$

$$\hat{\boldsymbol{C}} = [\boldsymbol{S}_{y1y1}(1) - \hat{\boldsymbol{A}}\,\boldsymbol{S}_{z1y1}(1)]\,\boldsymbol{S}_{y1y1}^{-1} \tag{2.72}$$

$$\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}^{T} = \boldsymbol{S}_{y1y1} - \hat{\boldsymbol{A}}\,\boldsymbol{S}_{z1y1} - \hat{\boldsymbol{C}}\,\boldsymbol{S}_{y1y1}^{T}(1) \tag{2.73}$$

a procedure for estimating $\hat{\boldsymbol{B}}$ from $\hat{\boldsymbol{B}}\hat{\boldsymbol{B}}^{T}$ can be found in Bras and Rodriguez (1993).

The parameters $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ may be estimated by (2.71)- (2.73) using the sample estimates of $\boldsymbol{S}_{UV}$ and $\boldsymbol{S}_{UV}(1)$ for any U and V. However, Lane (1982) reported that the use of the sample estimates $\boldsymbol{S}_{z1y1}(1)$ and $\boldsymbol{S}_{y1y1}(1)$ instead of the population $\boldsymbol{S}_{ZY}(1)$ and $\boldsymbol{S}_{YY}(1)$, respectively would result in model parameters $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ that may not preserve the linear relationship between z_t and y_t. Thus, Lane (1982) suggested that the covariance matrices $\boldsymbol{S}_{ZY}(1)$ and $\boldsymbol{S}_{YY}(1)$ in (2.71)- (2.73) must be estimated as:

$$\boldsymbol{S}_{ZY}(1) = \boldsymbol{S}_{z1z1}(1)\,\boldsymbol{S}_{z1z1}^{-1}\,\boldsymbol{S}_{z1y1} \tag{2.74}$$

and

$$\boldsymbol{S}_{YY}(1) = \boldsymbol{S}_{y1y1}(1) + \boldsymbol{S}_{y1z1}\,\boldsymbol{S}_{z1z1}^{-1}[\boldsymbol{S}_{ZY}(1) - \boldsymbol{S}_{z1y1}(1)] \tag{2.75}$$

The flows $\hat{y}_t$ obtained from the disaggregation model (2.70) do not necessarily preserve the spatial proportion between the flows at the upstream stations (substations) and the flows at the downstream stations (key stations). Therefore, the flows obtained from model (2.70) may need to be adjusted to achieve the spatial proportionality. Denoting by $\hat{y}_t^{(i)}$ the t-th flow value at substation i, the adjusted flow $\hat{y}_t^{*(i)}$ may be obtained as:

$$\hat{y}_t^{*(i)} = \hat{y}_t^{(i)} + (\hat{r}_t \hat{z}_t - \sum_{j=1}^{n} \hat{y}_t^{(j)}) \frac{\hat{\sigma}^2(i)}{\sum_{j=1}^{n} \hat{\sigma}^2(j)} \tag{2.76}$$

where $\hat{z}_t$ is the flow at the key downstream station, $\sum_{j=1}^{n} \hat{y}_t^{(j)}$ is the sum of the flows for the n upstream stations, $\hat{\sigma}^2(i)$ is the estimated variance of $\hat{y}_t^{(i)}$. The term $\hat{r}_t$ over the period of the extension can be obtained by simulating the ratios r_t using a model, e.g. the AR(1) model, that can be fitted to the observed historical flows recorded at the key and substations. The sequence ratio r_t is determined as $r_t = (\sum_{j=1}^{n} y_t^{(j)}) / z_t$, where z_t and $y_t^{(j)}$ are the t-th observed flow values at the key and the j-th substation, respectively.

Several record extension models have been employed in literature to extend streamflow records using tree-ring indices data. For example, Woodhouse (2001) used multiple linear regression models and the stepwise regression technique to select the tree-ring indices to be used for reconstructing the streamflows of the Middle Boulder Creek – Colorado. Likewise, Meko et al (2001) applied Principal Component Analysis (PCA) to reconstruct streamflows and selected a model that gave the lowest mean square error for validation. Hidalgo et al (2000) used the t-test and sign test to select the most effective principal components to be included in the reconstruction model, while Gedalof et al (2004) selected principal components of tree-ring variables with the criterion that only principal components with eigenvectors that exhibited physically meaningful loadings and had eigen-values that were statistically greater than 1.0 were selected in the final streamflow reconstruction model.

The reconstruction models discussed previously assumes a linear relationship between the independent variable (e.g. streamflow) and the explanatory variables (e.g. tree-ring indices). For several variables it may be necessary that the underlying variable be transformed to normal to improve the relationship between the independent and the explanatory variables. For example, Meko et al (2001) used log-10 transformation for flows to adjust for the heteroscedasticity (a feature that is visible as increased scatter at higher flows in plots of flow against tree-ring indices) in the relationship between streamflows and tree-ring data. Cleaveland (2000) used a multiple linear regression model with quadratic transformation of the tree-ring data and the model accounted for 68% of the total flow variance compared to a model using untransformed data, which accounted for 62% of the total variance. Also Salas et al (2006) used a model based on the log-transformed flows of the Poudre River, which explained around 60% of the variance compared to 53% obtained without transforming the data.

The reconstructed streamflow records enable one observing a wider range of flow scenarios that may be obtainable from the historical records alone. For example, Woodhouse (2001) observed that the reconstructed streamflows of the Middle Boulder Creek showed that the low flow events that occurred in the past were more persistent than those found from the analysis of the historical records. Similar other studies of tree ring reconstructed flows indicate that droughts of more severe magnitude and longer durations had occurred in the past compared to droughts occurred during the historical period (e.g. Meko et al, 1995; Meko et al, 2001; Gedalof et al, 2004; Salas et al, 2006).

In literature, usually the process of building flow reconstruction model goes through several statistical procedures (e.g. Woodhouse, 2001; Gonzalez and Valdes, 2003; Salas et al, 2006). Chapter 5 shows a detailed procedure that can be followed in building and testing the skills of the reconstruction model. Summarizing, firstly the relationship between the variable to be reconstructed (independent variable) and the tree ring variables (predictors) is assessed, for example through cross correlation. Secondly, predictors that statistically explain significant part of the variance of the independent variable are selected using proper statistical procedure, for example the stepwise regression analysis. Finally, the reconstruction model is calibrated and then validated using selected statistics. The validation statistic that is usually used to assess the performance of reconstruction models is the reduction error RE. It measures the skill of the model to produce reasonable predicted streamflows. The RE statistic is defined as:

$$RE = 1 - \frac{\sum_{t=1}^{N_v} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{N_v} (y_t - \bar{y}_c)^2} \qquad (2.77)$$

where, N_v is the size of the validation sample, $\bar{y}_c$ is the mean of the flow data used in the calibration period and $\hat{y}_t$ is the predicted value within the validation period based on the model.

The examination of the previous studies dealing with streamflow reconstructions based on tree ring indices indicate that the record extension models utilized have been applicable for one dependent variable, e.g. the streamflow data at a given stream gaging station. However, in practice it may be necessary to reconstruct the streamflows at many sites simultaneously. In this study the record extension model REXTN developed by Salas et al (2006) will be generalized to consider the reconstruction of streamflows at many sites simultaneously using tree ring indices as predictors.

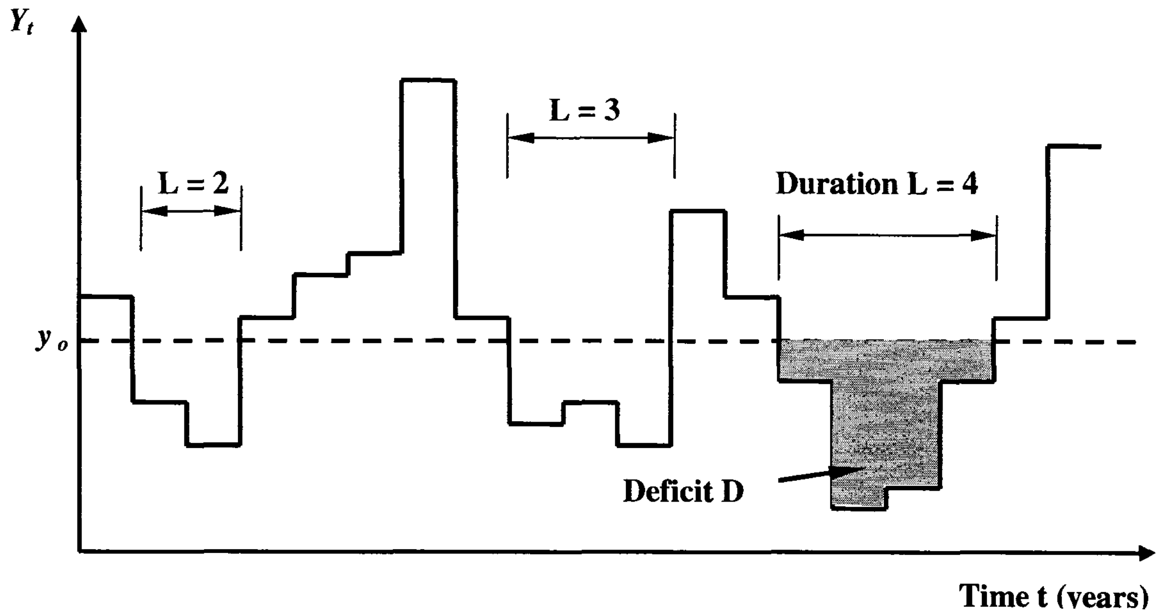

Fig 2.1 Typical drought events defined according to the first definition (Yevjevich 1967). The figure shows one 2 years, one 3 years, and one 4 years drought with the basic components of drought event shown on the far right event.

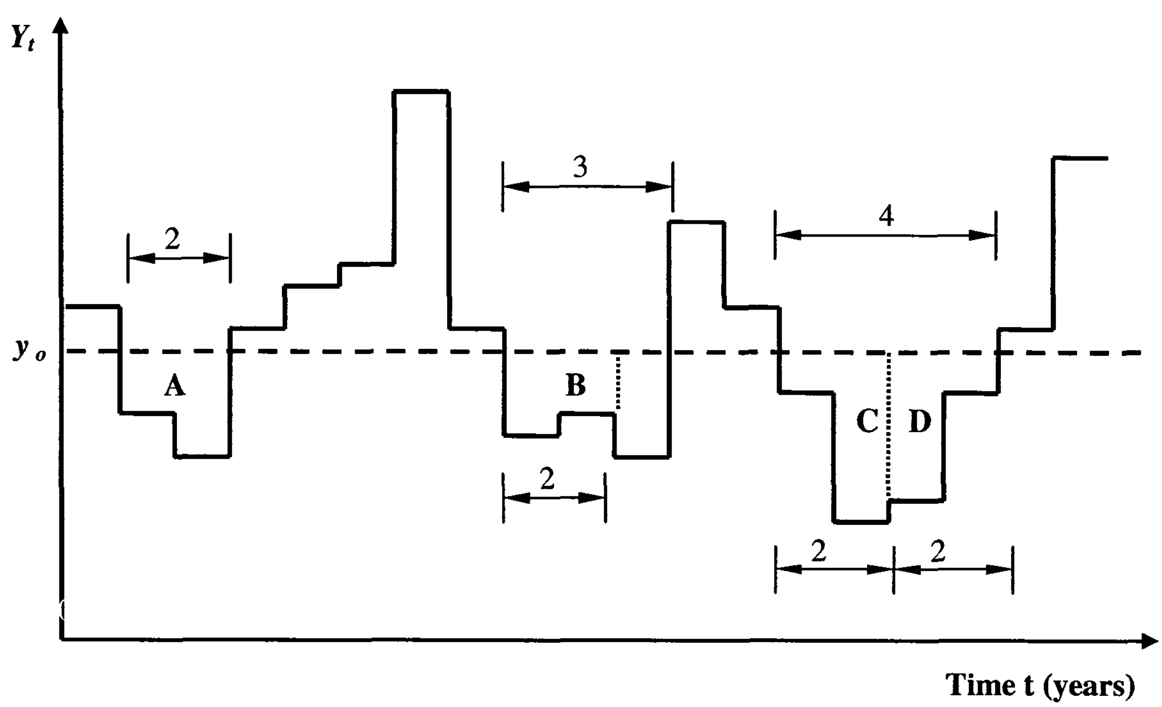

Fig 2.2 Typical two years drought events defined according to the second definition (Feller 1986). The figure depicts the 2 years events labeled A, B, C, and D.

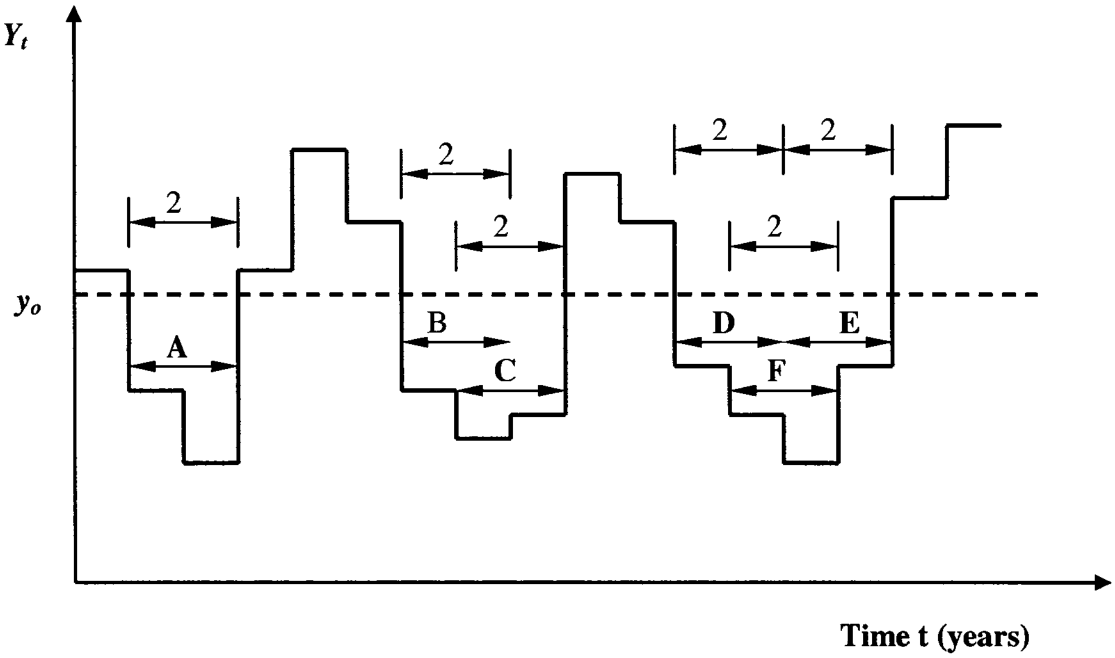

Fig 2.3 Typical two years drought events defined according to the third definition (Schwager 1983, Fernandez and Salas 1999). The figure depicts the 2 years events labeled A, B, C, D, E and F.

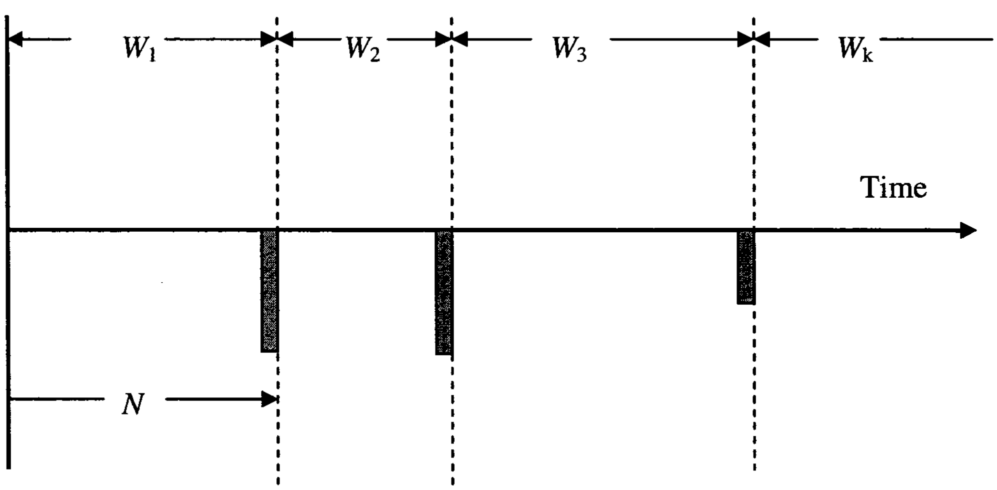

Fig 2.4 The definition of first arrival time N and the interarrival time W between repeated drought events. Possible drought events are shown by the rectangles.

CHAPTER III

CHARACTERIZATION OF DROUGHTS: SINGLE SITE ANALYSIS

3.1 Introduction

Drought is a complex natural phenomenon that originates from deficiency in the water supply over certain time period to meet the demand. Usually droughts of long duration and severe magnitude could lead to significant impacts and consequences such as failure to meet the demand for agriculture and municipal uses, deterioration in the water quality of rivers and lakes, the ensuing failure to meet the standards for aquatic life, recreation, etc. Consequently, characterizing severe drought events that may occur during the project life of any water resources system provides useful information for designing and operating such system (Sen, 1980). The characterization of the severe multiyear droughts is considered a complex physical and mathematical problem in which the complexity may be associated to several reasons such as the random evolution of the drought phenomenon, the strong correlation between drought components, i.e. the duration and the magnitude, the influence of the large-scale atmospheric-oceanic variability, and the relative short historical record lengths that make the estimation of extreme drought properties unreliable and less informative at higher drought lengths.

In literature many attempts were dedicated towards characterizing the severity of multiyear droughts (e.g. Yevjevich, 1967; Sen, 1980; Loaiciga et al, 1992; Loaiciga and

Leipnik, 1996; Sharma, 1997; Shiau and Shen, 2001; Gonzalez and Valdes, 2003; Cancelliere et al, 2003; Salas et al, 2005). The forgoing studies defined drought events according to the definition of runs as in Yevjevich (1967).

Schwager (1983) developed an algorithm to evaluate the first occurrence probability of a run of specific outcomes that satisfies certain length when the underlying process is either independent or has simple Markov structure. Based on that and focusing on drought duration, Fernandez and Salas (1999) derived the distribution of the first occurrence time and retrun period of single year and multiyear droughts arising from Markov dependent trials. Chung and Salas (2000) extended the work of Fernandez and Salas (1999) to include drought events arising from process that exhibits longer time dependence structure. In this chapter, drought events are defined according to the definition of runs as in Schwager (1983). Mathematical procedures to evalaute the first occurrence probability, occurrence probability, return period, and risk of multiyear drought events will be derived considering the length and the magnitude of the drought. The analytical procedures developed will be used to characterize the severity of multiyear droughts in the Poudre, the Colorado, and the Niger river basins.

3.2 Occurrence Probability of Run of Deficits

At a certain year t, if the underlying process Y_t (precipitation or streamflow) is below specified truncation level y_o, then the year t is a deficit year. Let X_t be the sequence resulted from truncating the underlying process Y_t at the level y_o such that X_t takes 0 whenever the year t is a deficit year and takes 1 otherwise,

$$X_t = \begin{cases} 0 & for\, Y_t < y_o \\ 1 & otherwise \end{cases} \tag{3.1}$$

Following the alternative definition of runs (e.g. Schwager, 1983), the drought event { $A_{l,t}$ } is said to occur whenever a run of l successive deficits occurs starting at the position t. Based on that, the probability $P[A_{l,t}]$ that a run of l successive deficits (l successive 0 states) occurs starting at the time t is:

$$P[A_{l,t}] = P[X_t = 0, X_{t+1} = 0, ..., X_{t+l-1} = 0] \tag{3.2}$$

The occurrence probability $P[A_{l,t}]$ can be obtained mathematically knowing the characteristics of the 0,1 process X_t. If X_t is assumed to be simple Markov stationary process, then using the law of conditional probability the probability $P[A_{l,t}]$ is written as:

$$P[A_{l,t}] = P[X_t = 0]\left(P[X_{t+1} = 0 \mid X_t = 0\right)^{l-1} \tag{3.3}$$

Since X_t is assumed stationary, then time index t can be dropped from the expression $P[A_{l,t}]$ and in general the occurrence probability $P[A_l]$ of the drought event { A_l } is:

$$P[A_l] = p_0 \left(p_{00}\right)^{l-1} \tag{3.4}$$

where $p_0 = P\left[X_t = 0\right]$ and $p_{00} = P[X_{t+1} = 0 \mid X_t = 0]$. If the drought event { A_l } arises from the process X_t that is independent then (3.4) can be used to evaluate the probability $P[A_l]$ noting that $p_{00} = p_0$. In general (3.4) is similar to the expression of the occurrence probability of drought defined according to the definition of runs as in Yevjevich (1967) as shown by other studies (e.g. Loaiciga and Leipnik, 1996; Shiau and Shen, 2001; Salas et al, 2005).

If the stationary process X_t shows to have longer time dependence structure, for example X_t is DARMA(1,1), then (3.4) can not be used to evaluate the occurrence probability $P[A_l]$. Assuming that X_t is DARMA(1,1) process, then it has been shown that the two state bivariate sequence $\{X_t, Z_t\}$ given by (2.11) can be converted into a four states simple Markov chain W_t, such that $W_t \in \{0,1,2,3\}$. The states 0 and 1 in W_t imply the state 0 in X_t while the states 2 and 3 in W_t imply the state 1 in X_t. Since the multiyear drought event $\{ A_l \}$ is defined as a run of l continuous dry states, then this run appears in the simple Markov process W_t with pattern of either 0s at all l steps, 1 at all l steps, or any possible combination of 0s and 1s.

Assuming X_t as DARMA(1,1) stationary process, then the occurrence probability $P[A_{l,t}]$ of the multiyear drought event $\{ A_l \}$ given by (3.2) is written as:

$$\begin{aligned} P[A_{l,t}] = P[X_{t+l-1} = 0 \mid X_{t+l-2} = 0,\ldots,X_t = 0]P[X_{t+l-2} = 0 \mid X_{t+l-3} = 0,\ldots,X_t = 0] \\ \ldots.P[X_{t+1} = 0 \mid X_t = 0] \times P[X_t = 0] \end{aligned} \tag{3.5}$$

The probability $P[X_t = 0]$ can be obtained easily from the characteristics of W_t, i.e. since the states 0 and 1 in W_t imply the state 0 in X_t, then $P[X_t = 0]$ is calculated as:

$$P[X_t = 0] = P[W_t = 0] + P[W_t = 1] \tag{3.6}$$

substituting (3.6) in (3.5) and noting that the sequence X_t defined by (2.11) can form the bivariate Markov chain $\{X_t, Z_t\}$ and that the pair (X_t, Z_t) is independent of X_{t-1}, then (3.5) can be written as:

$$\begin{aligned} P[A_{l,t}] = P[W_t = 0]\left[P(X_{t+1} = 0, Z_{t+1} = 0 \mid Z_t = 0) + P(X_{t+1} = 0, Z_{t+1} = 1 \mid Z_t = 0)\right]^{l-1} \\ + P[W_t = 1]\left[P(X_{t+1} = 0, Z_{t+1} = 0 \mid Z_t = 1) + P(X_{t+1} = 0, Z_{t+1} = 1 \mid Z_t = 1)\right]^{l-1} \end{aligned} \tag{3.7}$$

replacing the equivalent terms from (2.21) and (2.16) in (3.7), then the occurrence probability $P[A_l]$ of the multiyear drought event $\{A_l\}$ when X_t is DARMA(1,1) stationary process is:

$$
\begin{aligned}
P[A_l] = &[H_0(0,0)\pi_0 + H_0(1,0)\pi_1] \times [H_0(0,0) + H_0(0,1)]^{l-1} \\
&+ [H_0(0,1)\pi_0 + H_0(1,1)\pi_1] \times [H_0(1,0) + H_0(1,1)]^{l-1}
\end{aligned} \tag{3.8}
$$

where the transition probability $H_k(i, j)$, for i, j, and $k \in \{0, 1\}$, are given in (2.17) and (2.18).

3.3 Distribution of Drought Deficits

When the 0,1 process X_t (3.1) takes 0, i.e. Y_t is below the level y_o, then the amount of the single year deficit d_t is given by:

$$d_t = y_o - Y_t \tag{3.9}$$

In this case, the single deficit random variable d_t takes minimum value of 0 when $Y_t = y_o$ and maximum value of y_o when $Y_t = 0$. If continuous single deficits are observed over the length l of the drought event $\{A_l\}$, then the magnitude D of the drought is the accumulated deficits over the length l,

$$D = \sum_{t=1}^{l} d_t \tag{3.10}$$

It is convenient to define a specific value of the deficit D_o as:

$$D_o = \gamma y_o \tag{3.11}$$

where γ is constant. Since the drought event $\{A_l\}$ is defined as the run over which the deficits are observed, then the magnitude D that is given by (3.10) is the sum of the autocorrelated single deficits d_t over the length l because d_t is originated from Y_t that is autocorrelated (Salas et al, 2005). Since deficits are autocorrelated, deriving the distribution of the drought magnitude D, i.e. the sum of the autocorrelated single deficits, is a complex problem unless the single deficits are normally distributed (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003; Salas et al, 2005). The distribution of the drought magnitude can be obtained by fitting the empirical distribution obtained from the historical data using certain probability distribution function (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003). Alternatively, the single deficits may be assumed to be independent and identically distributed and in this case the distribution for their sum conditioned to the fixed length of the drought event $\{ A_l \}$ can be found under certian conditions (e.g. Salas at al, 2005).

Generally, the gamma and the exponential distributions are employed to fit the distribution of the single deficits and their sum (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003; Biondi et al, 2005; Salas et al, 2005). However since the single deficit random variable and its sum is bounded by lower and upper values as can be seen from (3.9), then the unbounded upper tailed distributions such as the gamma or the exponential may not represent well the single deficits and consequently their sum D. Referring to (3.9), the single year deficit random variable d_t is bounded by lower value of a=0 and upper value of b=y_o (Figure 3.1), therefore one may assume that the single deficit random variable d_t is Beta distributed with density function given by:

$$f_d(s) = \frac{1}{B(\alpha,\beta)} \frac{(s-a)^{\alpha-1}(b-s)^{\beta-1}}{(b-a)^{\alpha+\beta-1}} \quad , \qquad a \le d \le b \qquad (3.12)$$

where a and b are the lower and the upper bounds respectively and α and β are two additional parameters. It should be noted that deriving an exact expression to represent the distribution of the deficits sum conditioned to the occurrence of the event A_l, i.e. the distribution of $D \mid A_l$, considering all the features of the underlying process is not an easy task. To simplify the derivation, the single deficits d_t may be assumed to be independent and identically distributed (e.g. Salas et al, 2005) and therefore knowing the moment generating function of the single deficit random variable d_t, the distribution of the random variable D (the sum of the *iid* single deficits) can be derived exactly for some distributions, for example the Gamma distribution. The moment generating function of the Beta distributed variable d_t is rather complex function that can not be handled easily to derive the moment generating function of the variable $D \mid A_l$, consequently the derivation of the exact distribution of the variable $D \mid A_l$ assuming that the single deficit is Beta distributed may not be utilized here. However the random variable $D \mid A_l$ is also bounded by lower value of zero and an upper value of lb, where b equals the amount of the truncation level y_o (Figure 3.1).

Apparently, the assumption that the random variable $D \mid A_l$ is Beta distributed with parameters (α_l, β_l, a, lb) appears to be plausible. The probability density function of the Beta distributed variable $D \mid A_l$ is:

$$f_{D|A_l}(s) = \frac{1}{B(\alpha_l, \beta_l)} \frac{(s-a)^{\alpha_l - 1}(lb - s)^{\beta_l - 1}}{(lb - a)^{\alpha_l + \beta_l - 1}} \quad , \quad a \leq D \leq lb \tag{3.13}$$

Eventually, the probability distribution of the conditional deficits $P[D > D_o \mid A_l]$ can be evaluated as:

$$P[D > D_o \mid A_l] = \int_{D_o}^{lb} \frac{1}{B(\alpha_l, \beta_l)} \frac{(s-a)^{\alpha_l - 1}(lb - s)^{\beta_l - 1}}{(lb - a)^{\alpha_l + \beta_l - 1}} ds \tag{3.14}$$

The parameters α_l and β_l of the Beta distribution can be estimated using the method of moments given the estimates of $\mu_{D|A_l}$ and $\sigma^2_{D|A_l}$, i.e. the mean and the variance of the deficit D corresponding to the drought event $\{A_l\}$ of the fixed length l. The moments $\mu_{D|A_l}$ and $\sigma^2_{D|A_l}$ can be determined either empirically using the observed data assuming the number of drought events is sufficiently large or using the long synthetic data (e.g. Shiau and Shen, 2001) or as a function of the mean and the variance of the single deficits, i.e. μ_d and σ^2_d that can be determined from the historical data. Assuming that the series of the soingle deficits d_t is AR(1) process, then the moments $\mu_{D|A_l}$ and $\sigma^2_{D|A_l}$ of the deficit D corresponding to the drought event of the fixed length l are given as (Sen, 1980):

$$\mu_{D|A_l} = l\,\mu_d \tag{3.15}$$

and

$$\sigma^2_{D|A_l} = \sigma^2_d \left\{ l + 2\rho \frac{l(1-\rho)-(1-\rho^l)}{(1-\rho)^2} \right\} \tag{3.16}$$

where ρ is the lag1 serial correlation of the AR(1) single deficits. If the series of deficits is ARMA(1,1) process, then it can be shown that the mean $\mu_{D|A_l}$ of D is given by (3.15) and the variance $\sigma^2_{D|A_l}$ of D is (Appendix 3A):

$$\sigma^2_{D|A_l} = \sigma^2_d \left\{ l + 2\rho_1 \frac{l(1-\phi)-1+\phi^l}{(1-\phi)^2} \right\} \tag{3.17}$$

where ρ_1 is the lag1 serial correlation between the deficits and ϕ is the autoregressive parameter of the deficits ARMA(1,1) process. The relation between ϕ and the lag-k autocorrelation coefficient is $\rho_k = \phi\,\rho_{k-1}$, for k $\geq$ 2. If the deficit D, $D \in [0,\ lb]$, is adjusted to $\tilde{D}$, such that $\tilde{D} \in [0,1]$, with mean and variance of $m_{\tilde{D}|A_l}$ and $s^2_{\tilde{D}|A_l}$ respectively, then the moment estimators of the Beta parameters α_l and β_l are given as:

$$\hat{\alpha}_l = \frac{m^2_{\tilde{D}|A_l}(1 - m_{\tilde{D}|A_l})}{s^2_{\tilde{D}|A_l}} - m_{\tilde{D}|A_l} \tag{3.18}$$

$$\hat{\beta}_l = \frac{(1 - m_{\tilde{D}|A_l})\hat{\alpha}_l}{m_{\tilde{D}|A_l}} \tag{3.19}$$

For any l, the mean and variance of $\tilde{D}$ are related to mean and variance of D as $m_{\tilde{D}|A_l} = \dfrac{\mu_{D|A_l}}{lb}$ and $s^2_{\tilde{D}|A_l} = \dfrac{\sigma^2_{D|A_l}}{(lb)^2}$.

3.4 Occurrence Probability of Multiyear Drought Events

In general, drought events are characterized by the run length l over which deficits are observed that is associated with the magnitude D of the drought. The characterization of droughts based on the separate analysis of the drought length or the drought magnitude alone cannot reveal the significant correlation between them (Shiau and Shen, 2001). Similar to Salas et al (2005), to characterize droughts considering the joint analysis of both drought components, i.e. the length l and the magnitude D, one may define the drought event $\{ D > D_o \cap A_l \}$ as the event that is made of a run of the length l that has total deficit D that is greater than the critical deficit D_o. Given the drought event $\{ D > D_o \cap A_l \}$ the drought occurrence probability $P[D > D_o \cap A_l]$ can be evaluated using the law of the conditional probability. Assuming that the probability distribution $P[D > D_o \mid A_l]$ of the deficits sum conditioned to the occurrence of the event $\{ A_l \}$ that has the fixed length l is known, then given the occurrence probability of the drought event { A_l }, i.e. $P[A_l]$, the occurrence probability $P[D > D_o \cap A_l]$ is be written as:

$$P[D > D_o \cap A_l] = P[A_l] P[D > D_o \mid A_l] \tag{3.20}$$

The conditional probability $P[D > D_o \mid A_l]$ is the probability distribution of D conditioned to the occurrence of { A_l } that can be evaluated using (3.14). Assuming that X_t is stationary process, then the occurrence probability of the event { $D > D_o \cap A_l$ } is:

$$P[D > D_o \cap A_l] = P[A_l] \int_{D_o}^{lb} \frac{1}{B(\alpha_l, \beta_l)} \frac{(s-a)^{\alpha_l - 1} (lb - s)^{\beta_l - 1}}{(lb - a)^{\alpha_l + \beta_l - 1}} ds \tag{3.21}$$

where $P[A_l]$ is given by (3.4) and (3.8) for X_t as simple Markov and DRAMA(1,1) process respectively.

3.5 Risk of Multiyear Drought Events

The risk that particular multiyear drought event occurs within a specified time period 0 - S is defined as the probability that the drought event has occurred at least once within the time period 0 – S (Chung and Salas, 2000). The time period S can be the project design life or any period of time in which the decision maker is interested. If the 0,1 process X_t is simple Markov and considering drought events are characterized by the run length l only, i.e. the event { A_l }, the risk $R_{A_l}(S)$ that the event { A_l } has occurred at or before the trial $n = S$ is given as (Chung and Salas, 2000):

$$R_{A_l}(S) = \sum_{n=0}^{S} f_{n, A_l} \tag{3.22}$$

where f_{n, A_l} is the probability that the event { A_l } occurs for the first time at the trial n which is given by (2.40) assuming that the process X_t is either simple Markov or an independent process (Fernandez and Salas, 1999). In fact f_{n,A_l} given by (2.40) gives the distribution of the first arrival time N for the event $\{ A_l \}$ considering just the length l assuming that the time domain is discrete, however obtaining the distribution of N considering both the length l and the magnitude D of the drought and following the same algorithm at which (2.40) was obtained seems to be rather complicated. An approximate distribution for the time N will be used as the approach to obtain an expression to evaluate the risk of the multiyear drought event $\{ D > D_o \cap A_l \}$ considering both the length and the magnitude of the drought.

From the basics of the stochastic theory, if the event $\{ B \}$ occurs in continuous time domain, then the number of occurrences of the event $\{ B \}$ can be assumed as Poisson process and in that case the distribution of the time M required for the event $\{ B \}$ first occurrence is exponential (Taylor and Karlin, 1998). If drought events occur on discrete time domain, then the time N required for the drought event $\{ A_l \}$ to occur for the first time at the trial $N = n$ is a discrete random variable that takes values between l to ∞, i.e. $N \in [l, \infty)$. Moreover, referring to the definition of N it can be assumed that the drought event $\{ A_l \}$ that occurs for the first time at $N = n$ is independent from drought events occurred in the past (many years before the beginning of the time N). Therefore, considering the drought event $\{ A_l \}$, the assumption that N is geometrically distributed with parameter θ_{A_l} might be justified, i.e. the counter part of the exponential distribution of the first arrival time M in the continuous time domain (Taylor and Karlin, 1998). The parameter θ_{A_l} may be defined as the probability associated to the occurrence of the event

$\{A_l\}$ within the time period 0 - S. Knowing that the variable $N \in [l, \infty)$, the probability that $N = n$ is given as:

$$P[N = n] = \theta_{A_l}[1 - \theta_{A_l}]^{n-l} \tag{3.23}$$

the probability $P[N \leq S]$ is:

$$P[N \leq S] = \sum_{n=l}^{S} \theta_{A_l}[1 - \theta_{A_l}]^{n-l} \tag{3.24}$$

and using the algebra of the power series (3.24) is written as (Appendix 3B):

$$P[N \leq S] = 1 - [1 - \theta_{A_l}]^{S+1-l} \tag{3.25}$$

Referring to the definition of the risk, the probability $R_{A_l}(S)$ is equivalent to the probability $P[N \leq S]$, i.e. $R_{A_l}(S) = P[N \leq S]$, therefore (3.25) is written as:

$$R_{A_l}(S) = 1 - [1 - \theta_{A_l}]^{S+1-l} \tag{3.26}$$

Rearranging (3.26), the unknown parameter θ_{A_l} can be evaluated as:

$$\theta_{A_l} = 1 - [1 - R_{A_l}(S)]^{1/(S+1-l)} \tag{3.27}$$

the term $R_{A_l}(S)$ can be calculated using (3.22). When the drought event $\{D > D_o \cap A_l\}$ is under consideration, then following the same approach one may still assume that N, i.e. the first arrival time for the event $\{D > D_o \cap A_l\}$, is geometrically distributed with parameter $\theta_{D>D_o \cap A_l}$. For drought event $\{D > D_o \cap A_l\}$, the parameter $\theta_{D>D_o \cap A_l}$ can be assumed for simplicity as:

$$\theta_{D>D_o, A_l} = P_D(l)\theta_{A_l} \tag{3.28}$$

where the parameter θ_{A_l} is calculated using (3.27) and the probability $P_D(l)$ is the conditional probability $P[D > D_o \mid A_l]$ that is given by (3.14). Similar to (3.26), the risk

$R_{D>D_o\cap A_l}(S)$ that the drought event $\{ D > D_o \cap A_l \}$ occurs at least once within the time period $0 - S$ is:

$$R_{D>D_o\cap A_l}(S) = 1 - \left[1 - P_D(l)\theta_{A_l}\right]^{S+1-l} \tag{3.29}$$

Replacing (3.27) in (3.29) and rearranging, the risk $R_{D>D_o\cap A_l}(S)$ that the event $\{ D > D_o \cap A_l \}$ occurs at least once within the interval $0 - S$ is:

$$R_{D>D_o,A_l}(S) = 1 - \left[1 - P_D(l) + P_D(l)[1 - R_{A_l}(S)]^{1/(S+1-l)}\right]^{S+1-l} \tag{3.30}$$

Similar to Chung and Salas (2000), knowing the probability $R_{D>D_o\cap A_l}(n)$ at any time step n, one may be able to evaluate the first occurrence probability $f_{n,\{D>D_o\cap A_l\}}$ for the drought event $\{ D > D_o \cap A_l \}$ as:

$$f_{n,\{D>D_o\cap A_l\}} = \begin{bmatrix} R_{D>D_o\cap A_l}(n) & n = 1 \\ R_{D>D_o\cap A_l}(n) - R_{D>D_o\cap A_l}(n-1), & n \geq 2 \end{bmatrix} \tag{3.31}$$

If the process X_t is DARMA(1,1), then the risk $R_{D>D_o\cap A_l}(S)$ of the drought event $\{ D > D_o \cap A_l \}$ is given by (3.30) replacing $R_{A_l}(S)$ by the total risk $R^*_{A_l}(S)$ that is given by (2.44), and then the first occurrence probability $f_{n,\{D>D_o\cap A_l\}}$ is computed using (3.31).

3.6 Return Period of Multiyear Drought Events

The return period of the drought event is defined in different ways for different applications (Fernandez and Salas, 1999; Douglas et al, 2002). It is defined as the expected value of the time interval N required by the drought event to occur for first time, or it is defined as the expected value of the interarrival time W between recurrent drought events of the same or different kinds. The return period of the drought event

$\{ D > D_o \cap A_l \}$ based on the first definition is given mathematically similar to Fernandez and Salas (1999) as:

$$T = E(N) = \sum_{n=0}^{\infty} n \, f_{n,\{D>D_o \cap A_l\}} \tag{3.32}$$

where the probability $f_{n,\{D>D_o \cap A_l\}}$ for the drought event $\{ D > D_o \cap A_l \}$ is given in (3.31).

To analytically evaluate the return period of the drought event $\{ D > D_o \cap A_l \}$ as the expected value of the interarrival time W between recurrent drought events, the concepts of the renewal theory is applied on drought as renewal phenomena (Loaiciga et, al 1992). The fundamental result of the renewal theory states that the expected number $E[N_{\{D>D_o \cap A_l\},n}]$ of drought events of the kind $\{ D > D_o \cap A_l \}$ in a given time period n divided by the period n converges to $1/T$ assuming that the time period n is sufficiently large, where T is the expected value of the renewal time W (Loaiciga et al, 1992). Following the alternative definition of drought events (e.g. Schwager, 1983), the number $N_{\{D>D_o \cap A_l\},n}$ of drought events $\{ D > D_o \cap A_l \}$ that may arise from the sequence of the length n is:

$$N_{\{D>D_o \cap A_l\},n} = \sum_{t=1}^{n-l+1} \mathrm{I}_t \tag{3.33}$$

where I_t is an indicator function that is defined as:

$$\mathrm{I}_t = \begin{cases} 1 & \textit{when } \{D > D_o \cap A_l\} \textit{ occurs starting at the time } t \\ 0 & \textit{otherwise} \end{cases} \tag{3.34}$$

The upper limit of the sum term in (3.33), i.e. n-l+1, is the maximum number of drought events of the type $\{ D > D_o \cap A_l \}$ that may occur in a sequence of length n when these events are defined following the alternative definition (e.g. Schwager 1983). The expected number of drought events $E[N_{\{D>D_o \cap A_l\},n}]$ in the sequence of length n can be obtained by taking the expectation on both sides of (3.33):

$$E[N_{\{D>D_o \cap A_l\},n}] = (n - l + 1)\, E[\mathrm{I}_t] \tag{3.35}$$

and the expected value $E[\mathrm{I}_t]$ is:

$$E[\mathrm{I}_t] = \sum_{i \in \{0,1\}} i \; P[\mathrm{I}_t = i] = P[D > D_o \cap A_l] \tag{3.36}$$

If n is sufficiently large, i.e. $n \approx n\text{-}l\text{+}1$, then based on the fundamental result of the renewal theory the return period of the recurrent drought events of the type $\{ D > D_o \cap A_l \}$ as the expected value of the interarrival time W is:

$$T = E(W) = \frac{n}{E[N_{\{D>D_o \cap A_l\},n}]} = \frac{1}{P[D > D_o \cap A_l]} \tag{3.37}$$

The probability $P[D > D_o \cap A_l]$ of the recurrent drought events is given by (3.21). Generally, equation (3.37) is valid regardless of the structure of the process at which the drought event evolves.

3.7 Applications

The mathematical procedures developed here will be used to evaluate the first occurrence probability, occurrence probability, return period and risk of multiyear drought events that are characterized by the length l and the magnitude D assuming drought events are defined following Schwager (1983). The derived procedures will be used to characterize droughts in the Poudre, the Colorado, and the Niger river basins. The annual streamflow series of the Poudre, the Colorado, and the Niger rivers are shown in Figures 3.2a, 3.2b, and 3.2c respectively. The long term mean of each streamflow series is also shown.

The annual streamflow series of the Poudre river at the mouth of Canyon gaging station consists of 119 annual streamflow observations covering the period 1884 – 2002. The basic streamflow statistics of the mean, standard deviation, skewness, lag-1 autocorrelation coefficient and longest drought are 299011 acre-ft, 106512 acre-ft, 0.98, 0.153 and 8 years respectively. The annual streamflow series of the Colorado river at the Lees Ferry gaging station consists of 98 annual streamflow observations covering the period 1906 - 2003 with mean, standard deviation, skewness, lag-1 autocorrelation coefficient and longest drought as 15076151 acre-ft, 4444186 acre-ft, 0.17, 0.259 and 5 years respectively. The annual streamflow series of the Niger river at Koulikoro covers the period 1907 – 1999 shows mean, standard deviation, skewness, lag-1 autocorrelation coefficient, lag-2 autocorrelation coefficient and longest drought as 1373.66 CMS (Cubic Meter/Second), 395.9 CMS, 0.17, 0.66, 0.63, and 14 years respectively.

For the purpose of drought characterization, each streamflow series was truncated at the level of the long-term mean of the observed data. The resulted series of deficits and surpluses was converted to series of 0s and 1s by assigning the state 0 when deficit is observed and the state 1 otherwise. The unconditional and conditional state probabilities of the truncated annual streamflow series for each of the Poudre and the Colorado rivers were determined according to the procedure described in section 2.3. The parameter λ was estimated using (2.8) while the marginal discrete state probabilities π_0 and π_1 were estimated using (2.22). The conditional state probabilities p_{ij} for $i,j \in \{0,1\}$ were estimated using (2.9) given $\hat{\lambda}$, $\hat{\pi}_0$, and $\hat{\pi}_1$. For example, given the Poudre streamflow data, the 0,1 process X_t is obtained at first by truncating the annual streamflow of the Poudre river at the long term mean, then the parameter λ is estimated from (2.8) by replacing the lag1 sample autocorrelation coefficient $\hat{\rho}_1(X)$ of the DAR(1) process X_t. From Poudre data, $\hat{\lambda}$ = 0.0961 while $\hat{\pi}_0 = \overline{T}_0 /(\overline{T}_0 + \overline{T}_1) = 0.563$ and $\hat{\pi}_1 = \overline{T}_1 /(\overline{T}_0 + \overline{T}_1) =$ 0.437. The unconditional and conditional state probabilities are $p_0 = 0.563$, $p_1 = 0.437$ and the states transition matrix is given by $\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.605 & 0.395 \\ 0.509 & 0.491 \end{bmatrix}$. Similarly, for the Colorado river $p_0 = 0.50$, $p_1 = 0.50$ and $\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.583 & 0.417 \\ 0.417 & 0.583 \end{bmatrix}$.

The annual streamflow series of the Niger river was converted to a sequence of 0s and 1s by truncating the series at the level of the long-term mean. The state 0 denotes the dry state while the state 1 denotes the wet state. The marginal discrete probabilities of the dry and wet states, π_0 and π_1 respectively were estimated using (2.22). The marginal discrete probability of the dry state is $\hat{\pi}_0$ = 0.462 while for the wet state is $\hat{\pi}_1$ = 0.538. For the same river, with the annual streamflow series that covers the period 1907 – 1990, Chung and Salas (2000) showed that the DARMA(1,1) is the best model to fit the 0,1 sequence resulted from truncating the annual streamflow series of the Niger river at the long-term mean. Therefore, DARMA(1,1) model was fitted to the 0,1 sequence resulted from truncating the annual streamflow of the Niger river over the period 1907 – 1999 at the long term mean of the observed flows. The parameters λ and β of the DARAM(1,1) process were evaluated using (2.25) and (2.14) respectively. The estimates of λ and β are 0.848 and 0.242 respectively. The one-step transition probability matrices of the two-state bivariate Markov chain are given by (2.17) and (2.18), thus:

$$H_0(u,v)=\begin{bmatrix}0.8077 & 0.0621\\ 0.0170 & 0.0948\end{bmatrix},\ \text{and}\ H_1(u,v)=\begin{bmatrix}0.1103 & 0.0198\\ 0.0534 & 0.8348\end{bmatrix}$$

The discrete marginal probabilities of the simple Markov process W_i were estimated using (2.21) and the estimated values are $P[W=0]$ = 0.3825, $P[W=1]$= 0.0797, $P[W=2]$= 0.0797, and $P[W=3]$= 0.4581.

The short observed streamflow records may not provide a reliable picture about the total natural variability that affect the characterization of extreme hydrologic droughts that may last for many years, thus it is extremely important to reliably lengthen the short observed streamflow records to a length much longer than the length of the current observed records (Salas et al, 2005). A useful technique to lengthen short records is to build using the current observed records a stochastic simulation model that generates long records of synthetic streamflows that may yield and exhibit additional features and scenarios of possible extreme events that may occur in the future (Salas et al 2005). In order to generate long synthetic flows, a suitable ARMA(p,q) model was first fit to each of the annual streamflow series under consideration based on its characteristics and consequently 50000 annual streamflow points were generated for each of the rivers under consideration. For example, annual streamflows of the Poudre river can be sufficiently generated using AR(1) model (e.g. Salas et al, 2005), annual streamflows of the Colorado river can be generated using AR(1) model (e.g. Tarboton, 1995), while annual streamflows of the Niger river can be sufficiently generated using ARMA(1,1) model (e.g. Chung and Salas, 2000). Empirical results of drought properties obtained from analyzing the generated data will be used to verify the results obtained from the derived mathematical procedures.

3.8 Results and Discussion

Considering the multiyear drought event $\{ D > D_o \cap A_l \}$ that is characterized by the length l and the magnitude D, analytical procedures to evaluate the first occurrence probability, occurrence probability, return period and the risk have been developed here assuming that the sequence of deficits and surpluses X_t is either DAR(1) or DARMA (1,1) stationary process. To include deficits in characterizing the severity of hydrologic droughts jointly with the length of the drought, assumptions on the distribution of the single year deficits and drought amount over the length l is required.

Because it is bounded, the single year deficit variable is assumed to be Beta distributed with parameters (α, β, a, b) with a = 0 and b = y_o (Figure 3.1). The distribution of the deficit D, i.e. the sum of single deficits conditioned to the occurrence of the drought event $\{ A_l \}$, is also assumed to be Beta with parameters (α_l, β_l, a, lb) where a and b are as defined previously. The parameters α_l and β_l were determined using (3.18) and (3.19) respectively given the mean $m_{\tilde{D}|A_l}$ and the variance $s^2_{\tilde{D}|A_l}$ of $\tilde{D}$, $\tilde{D} \in [0,1]$. The mean $m_{\tilde{D}|A_l}$ and the variance $s^2_{\tilde{D}|A_l}$ of $\tilde{D}$ are related to the mean $\mu_{D|A_l}$ and the variance $\sigma^2_{D|A_l}$ of the deficit D, $D \in [0, lb]$, that can be obtained using (3.15) and (3.16 or 3.17) respectively as a function of the mean μ_d and the variance σ^2_d of the single deficits obtained from truncating the observed flows at its long-term mean. Given the deficits series resulted from truncating the Poudre river flows, the estimated values of μ_d, σ_d, and ρ are 74257 AF, 47865 AF, and 0.082, while using the Colorado river flows the estimated values of μ_d, σ_d, and ρ are 3.64 MAF, 2.29 MAF, and 0.18. Given the deficits series resulted from truncating the flows of the Niger river the estimated values of μ_d, σ_d^2, ρ_1, ρ_2, and ϕ are 328.5 m^3/sec, 47405 m^3/sec, 0.44, 0.39, and 0.88 respectively. The estimated values of α_l and β_l are not shown here.

Given the length of the drought, once the event $\{ D > D_o \cap A_l \}$ has a magnitude D that exceeds D_o then it is considered severe event and will be accounted for characterization. At γ values of 0, 0.5 and 1.0 the occurrence probability of the multiyear drought event $\{ D > D_o \cap A_l \}$ was obtained using (3.21) given the truncated annual streamflow series of the Poudre, the Colorado, and the Niger rivers. Figures 3.3a – 3.3c show the occurrence probability for the drought event $\{ D > D_o \cap A_l \}$ at γ values of 0, 0.5 and 1.0. Generally it can be seen from the figures that the analytical procedure (3.21) to evaluate the occurrence probability gives reasonable results when compared with the empirical occurrence probability results obtained from analyzing droughts using the 50000 simulated and observed streamflows to some extent. Generally Figures 3.3a – 3.3c indicate the uncertainty in estimating the occurrence probability of droughts using only the short historical flows. For example, it can be noticed that using the historical flows, the estimated values of the occurrence probability depart from expectations, i.e. theoretical results and results from simulation, which can be explained as a result of the limited number of droughts obtained from analyzing the historical flows. Moreover, estimates of the occurrence probability from analyzing the historical flows were not available when the drought length is relatively large, i.e. more than 5 years (Figure 3.3b).

For drought events considering only the length l, i.e. $\gamma = 0$, the occurrence probability $P[A_l]$ decreases as the drought length l increases. This results is expected since droughts become more extreme and hence their occurrence probability decreases as the their length increases. On the other hand, considering the magnitude D of the drought besides its length l, i.e. $\gamma > 0$, then as the magnitude D increases to exceed the increasing D_o (the increasing γ), generally droughts become less frequent to occur which is expected due to the high unusual drought magnitude. Moreover, given specific critical drought level D_o, then at short length (say 1 year) droughts are considered more extreme (less frequent) and after that as the drought length l increases droughts become more frequent to occur, i.e. at moderate drought lengths (say 2 – 4 years) droughts occur more often, however at longer drought lengths (say 6 years or more), droughts become less frequent as a result of the extreme drought length. In that case, the magnitude of the drought event is not critical in making droughts as extreme events.

Besides enabling one to estimate the chance that the drought event $\{ D > D_o \cap A_l \}$ occurs, the occurrence probability shown in Figures 3.3a, 3.3b and 3.3c show an other useful application by enabling one to extract information about the length of the most frequent drought. For example, referring to Figure 3.3a, it can be noticed that at $\gamma = 0.5$, the most frequent drought event has a length of 2 – 3 years. Moreover, it can be noticed that varying the amount of the critical drought, i.e. varying γ, would result in different lengths of the most frequent drought. Table 3.1 shows the results of such analysis. Generally, streamflows that show long persistence in their dependence structure, for example the streamflows of the Niger river, the length of the most frequent drought is higher than that in streamflows that show short persistence in their dependence structure like the streamflows of the Poudre river for example. Similar results regarding the length of the most frequent drought were obtained from analyzing both the Poudre river and the Colorado river, which is expected since both rivers mainly belong to the same climatic region.

The approach used to derive the risk expression (3.30) is based on the assumption that the first arrival time N of the drought event is geometrically distributed with parameter $\theta_{D>D_o \cap A_l}$. To assess the robustness of that assumption, the empirical distribution of the first arrival time N required for the drought event $\{ D > D_o \cap A_l \}$ to occur for the first time obtained from simulation was compared with the theoretical fit of N as geometric with parameter $\theta_{D>D_o \cap A_l}$ estimated using (3.28). For simulation purposes, annual flows of the Poudre river were simulated using AR(1) model until the first event $\{ D > D_o \cap A_l \}$ has occurred and then the time N associated to the event first occurrence was recorded. At each of the desired lengths of 1 and 8 years and magnitude D that exceeds D_o with $\gamma = 0.5$, this simulation procedure was repeated 10000 times leading to a 10000 simulated values of N. Figures 3.4a and 3.4b show the empirical probability $P[N = n]$ obtained from simulation versus the theoretical $P[N = n]$ obtained by fitting N as geometric with parameter $\theta_{D>D_o \cap A_l}$ estimated using (3.28) for l = 1 and 8 years respectively. Generally, it can be concluded from both figures that the assumption that N is geometrically distributed appears to be quite reasonable as compared with results from simulation that may justify the approximation made in deriving the risk equation (3.30).

The risk that the multiyear drought event $\{ D > D_o \cap A_l \}$ occurs during the project life S is obtained using (3.30). Assuming that S = 25 years, the results obtained using (3.30) are compared with the results obtained from the empirical analysis using the simulated and observed streamflows (Figures 3.5a – 3.5c). The results from figures show generally when γ = 0, the risk of drought occurrence decreases as the drought length increases. Moreover, increasing the magnitude of the drought (increasing γ) the risk that the event $\{D > D_o \cap A_l\}$ occurs decreases, while within the same level of the critical drought (same γ > 0), the probability that an extreme drought occurs is low compared to the most frequent droughts, i.e. droughts of moderate lengths (nearly 2 – 3 years at γ = 0.5 and nearly 3 – 4 years at γ = 1) as shown by Figure 3.5b for example. Extreme droughts here are referred to drought events of short length and considerable magnitude (γ > 0) or drought events of long duration (6 years or more) regardless of the magnitude. Comparing Figures 3.5a and 3.5c it can be concluded that droughts arise from series of deficits and surpluses of longer time persistent show to have higher risk especially for droughts of longer duration (6 years or more). Figures 3.5a – 3.5c generally indicate the uncertainty in characterizing droughts using the short historical flows when compared with the more reliable results obtained from the analytical procedures or simulation.

The return period as the expected value of the first arrival time of the drought event $\{D > D_o \cap A_l\}$ to occur, i.e. $T = E(N)$, was obtained using (3.32) considering the critical drought of magnitude D_o at γ = 0, 0.5 and 1. Figures 3.6a – 3.6c show the return period T of the multiyear drought event $\{D > D_o \cap A_l\}$ arising from truncating the annual streamflows of the Poudre, the Colorado, and the Niger rivers respectively. Generally, the results obtained from the analytical procedure (3.32) developed to evaluate the return period as the expected value of the first arrival time for the drought event of the kind $\{D > D_o \cap A_l\}$ to occur agree with the results obtained from the empirical analysis of the simulated streamflows. The results shown in Figures indicate generally that the return period of drought events increases by increasing the drought magnitude resulted from increasing the level D_o, i.e. increasing γ, which is expected because droughts become more extreme and consequently their return period will be high. Moreover, the return period increases as the drought length l increases where at long drought lengths (say 6 years or more) the magnitude of the drought D under the given values of D_o, i.e. γ, is not contributing to make the drought as an extreme event, while the case may be the opposite with droughts of short lengths (1 – 3 years) where the effect of the magnitude D is considerably huge.

For drought events of the kind $\{D > D_o \cap A_l\}$, the return period T as the expected value of the interarrival time $E(W)$ between recurrent droughts was determined using the mathematical procedure (3.37) at γ values of 0, 0.5 and 1. Figures 3.7a – 3.7c show the return period T as $E(W)$ for the referred drought event arising from truncating the annual streamflows of the Poudre, the Colorado, and the Niger rivers respectively. Generally, the results obtained from the procedure (3.37) agree with the results obtained from the empirical analysis of the simulated and the observed streamflows to some extent. The results shown in Figures 3.7a – 3.7c indicate generally that the return period of droughts increases by increasing the magnitude D due to the increase in the level D_o, i.e. increasing γ, which is expected because droughts become more extreme and consequently their return period will be high. Generally, the return period increases as the drought length l increases where at long drought lengths (say 6 years or more) the magnitude D of the drought is not contributing in making the drought event an extreme event, while the case may be the opposite with droughts of short length (1 – 2 years) where the effect of the magnitude D is considerably huge. Comparing Figures 3.7a or

3.7b with 3.7c, the effect of streamflow persistence on the return period is clear, for example, comparing Figures 3.7a and 3.7c one observes that the average interarrival time of drought events arise from truncating streamflows of longer time persistence is much shorter. This is expected due to the longer persistence in the flows, i.e. a year of low flow would have more chance to be followed by another year or years of low flow as flows get more dependence in their structure (more persistent). The General behavior of the return period as shown in Figures 3.7a – 3.7c is similar to the findings of other studies (e.g. Salas et al 2005) although using different definition of drought events.

Comparing results obtained for the return period using the two definitions stated above, generally it can be noticed that the return period calculated as $E(N)$ is always greater than or equal to the return period calculated as $E(W)$. This conclusion is expected, because the waiting time N to the event first occurrence is a random variable that takes minimum value of l (the length of the of the desired drought event) but not less than, whereas the interarrival time W between the recurrent drought events may take minimum value of 1 year regardless of drought length l, i.e. recall that recurrent drought events are defined according to Schwager (1983) where events may overlap. Thus it is expected that $T = E(N)$ is always greater than or equal to $T = E(W)$.

Generally results shown by Figures 3.3, 3.5, 3.6, and 3.7 indicate the uncertainty in characterizing the severity of hydrologic droughts when observed streamflows are used alone in the analysis of droughts. The uncertainty is associated to the short length of the observed flows where usually only few historical drought events may be observed when using the observed streamflows for the analysis of the extreme hydrologic droughts. In fact, in some cases droughts of higher lengths may not be observed at all within flow records of around 100 years. Referring to Figure 3.7a it can be seen from the results that the return periods obtained using the observed flow records depart from their expectation, i.e. from the analytical results or simulation results, specially when $\gamma = 1$ and become less reliable and less informative as the drought length increases. Similar conclusions are also found from the results of the occurrence probability and the risk for the referred rivers flows.

The analysis of hydrologic droughts obtained using the analytical procedures to estimate drought properties based on the assumption that the drought deficits are Beta distributed. To assess the robustness of fitting the deficits as Beta distributed compared to other distributions used in the literature, for example the Gamma (e.g. Shiau and Shen 2001, Gonzalez and Valdes 2003, Salas et al 2005), the return period as $E(W)$ has been estimated using (3.37) for the event $\{ D > D_o \cap A_l \}$ assuming that deficits are Beta distributed in one case and Gamma distributed on the other. Figure 3.8 shows the return period as $E(W)$ of the event $\{ D > D_o \cap A_l \}$ arises by truncating the annual streamflow of the Poudre river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical solution (3.37) assuming deficits are Beta distributed in one case while Gamma distributed on the other versus the empirical results obtained from analyzing the observed and the 50000 simulated flows.

Generally it can be noticed from Figure 3.8 that at $\gamma = 0.5$ both distributions quite well represent the deficits, however when $\gamma = 1$, the return period obtained from (3.37) with deficits as Beta distributed are more closer to the empirical results obtained from simulation than those obtained from (3.37) with deficits as Gamma distributed. In fact when $\gamma = 1$, the analytical solution (3.37) with the deficits as Gamma distributed results in return period for the drought event $\{ D > y_o \cap A_1 \}$, i.e. the drought event that has 1 year length and magnitude D greater than y_o, of around 1300 years which is practically unrealistic since any single year drought event can not have a deficit that is greater than y_o (Figure 3.1). This result is expected because theoretically the Gamma distribution furnishes a probability value greater than 0 to the single deficit to have a magnitude greater than y_o which is practically unrealistic, i.e. that probability should be zero because the magnitude of the single deficit is bounded by a maximum value of y_o.

3.9 Summary and Conclusions

The multiyear drought event of length l is defined as a run of l consecutive deficits regardless where the run starts as long as the run satisfies the continuity. One reason to follow the referred definition of drought events is the ability to capture extreme hydrologic droughts regardless where the event starts, i.e. there could be a particular drought event that is extreme in terms of the magnitude and not necessarily is headed and followed by at least one surplus that is a requirement of other definitions. The multiyear drought events under consideration are characterized by the fixed length l and the magnitude D that exceeds given critical deficit. Mathematical models to evaluate the occurrence probability, return period, and risk of multiyear droughts are presented here. The series of deficits and surpluses resulted from truncating the annual flows is modeled using low order discrete autoregressive moving average processes. The bounded drought deficit is assumed to be Beta distributed with parameters α_l and β_l.

The developed mathematical procedures were used to characterize the severity of extreme hydrologic droughts in the Poudre, the Colorado, and the Niger river systems.

Drought properties in terms of the occurrence probability, return period, and risk estimated empirically from analyzing the historical streamflows of the referred rivers have shown to be quite uncertain, i.e. they deviate from the expectation, especially for droughts of short duration and considerable magnitude ($\gamma > 0$). Moreover, estimates of the referred drought properties were not available from analyzing the historical flows when the drought length is relatively large, i.e. 5 years or more for the Colorado River case.

The analytical procedures developed here provide more reliable estimates of drought properties when compared with the uncertain and the less informative results obtained from analyzing the historical streamflow, therefore these procedures can be used to characterize the severity of multiyear extreme droughts. Drought properties obtained from the analytical procedures were verified through comparison with empirical results obtained from simulation. In addition, results of the return period analysis show in general that fitting deficits as Beta improves the characterization of short extreme droughts, i.e. droughts of short length and considerable magnitude ($\gamma = 1$), as compared to other distributions used in literature like the Gamma. However when droughts get longer (6 years or more) the estimated drought properties were not sensitive to the kind of deficits distribution and the extremeness in the drought properties is associated only to the extreme drought length.

Table 3.1 The duration of the most frequent drought event $\{ D > D_o \cap A_l \}$ at γ values of 0.5 and 1 obtained from the procedure (3.21) using annual streamflows of the Poudre, the Colorado, and the Niger rivers.

Annual Streamflow of	Duration the most frequent drought (years)	
	$\gamma = 0.5$	$\gamma = 1$
The Poudre River	2 – 3	4
The Colorado River	2 – 3	4
The Niger River	3 – 4	5 – 6

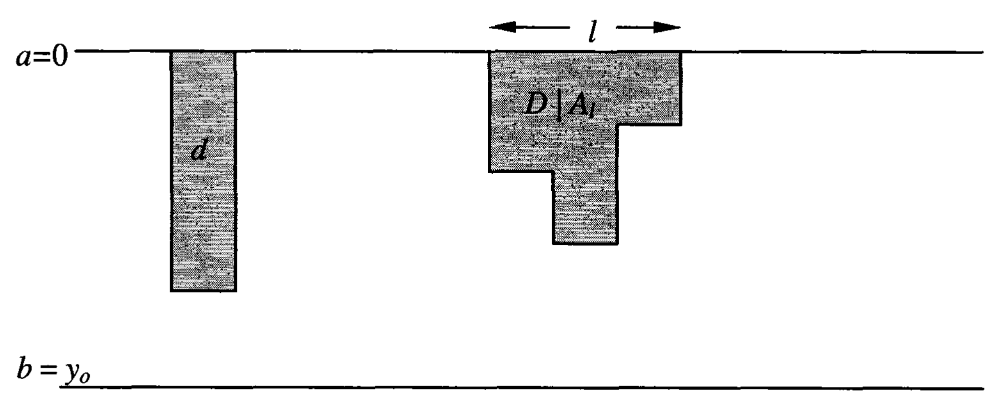

Fig 3.1 The bounded single deficit random variable d and the variable $D \mid A_l$ (the conditional sum of deficits). The lower bound of d and $D \mid A_l$ is $a = 0$, the upper bound of d is $b = y_o$ while the upper bound of $D \mid A_l$ is $lb = ly_o$.

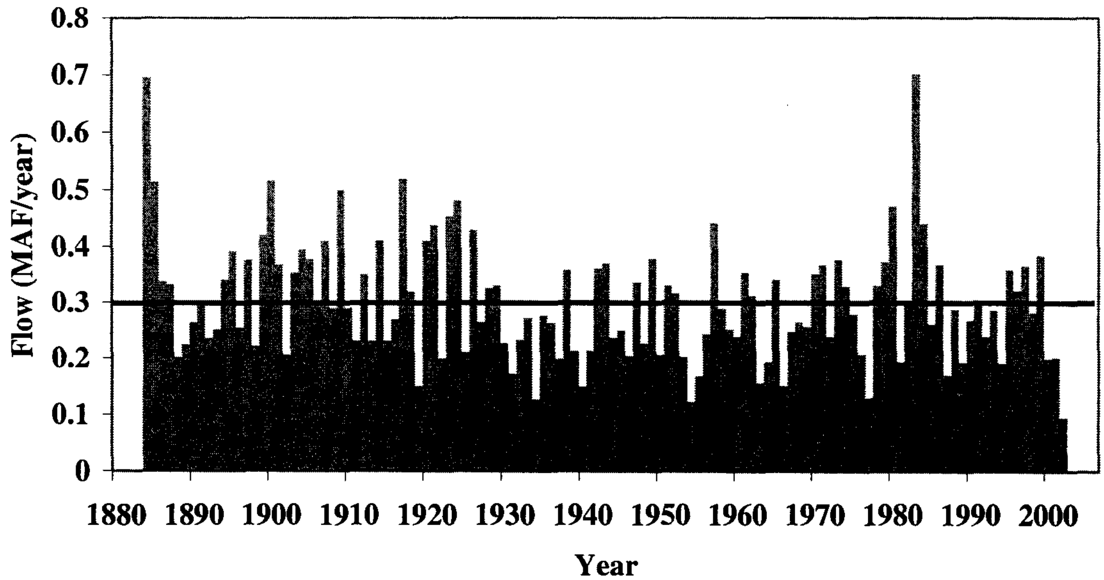

Fig 3.2a The annual streamflow series of the Poudre river at the mouth of canyon gaging station covering the period 1884 - 2002. The line represents the long-term mean of the observed streamflows.

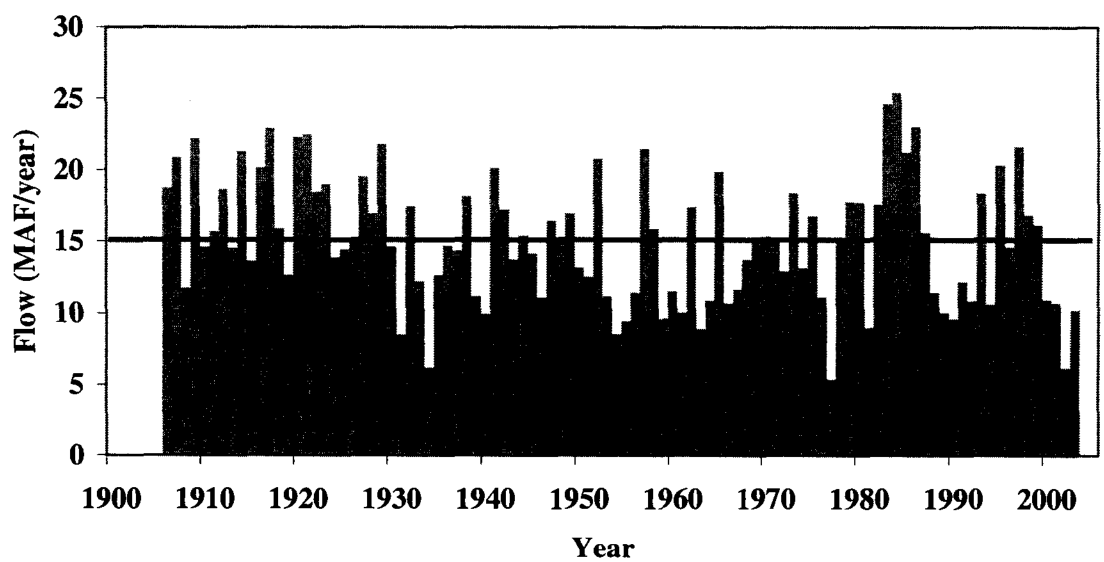

Fig 3.2b The annual streamflow series of the Colorado river at Lees Ferry gaging station covering the period 1906 - 2003. The line represents the long-term mean of the observed streamflows.

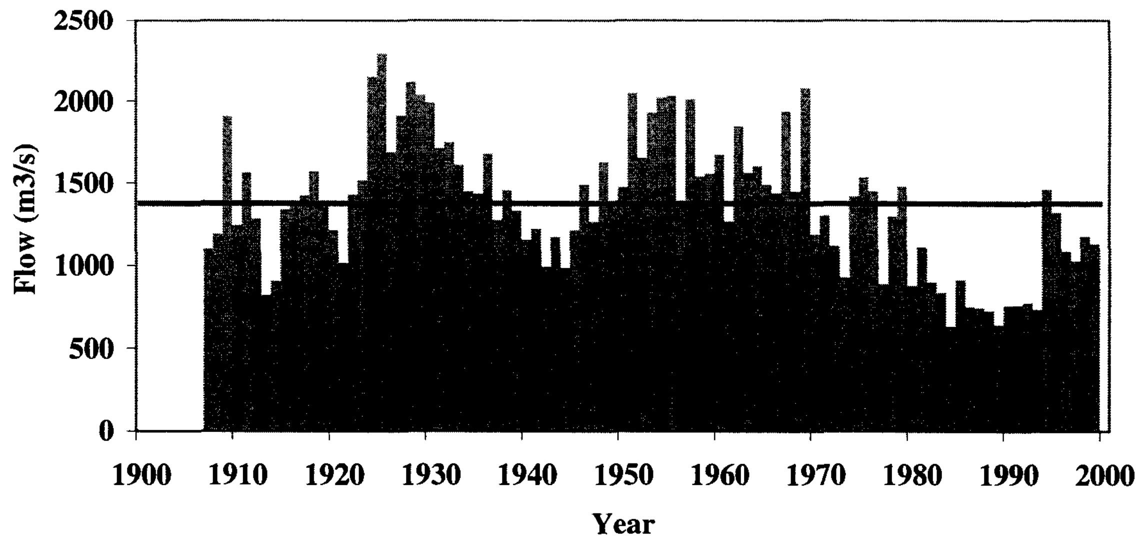

Fig 3.2c The annual streamflow series of the Niger river at Koulikoro gaging station covering the period 1907 – 1999. The line represents the long-term mean of the observed streamflows.

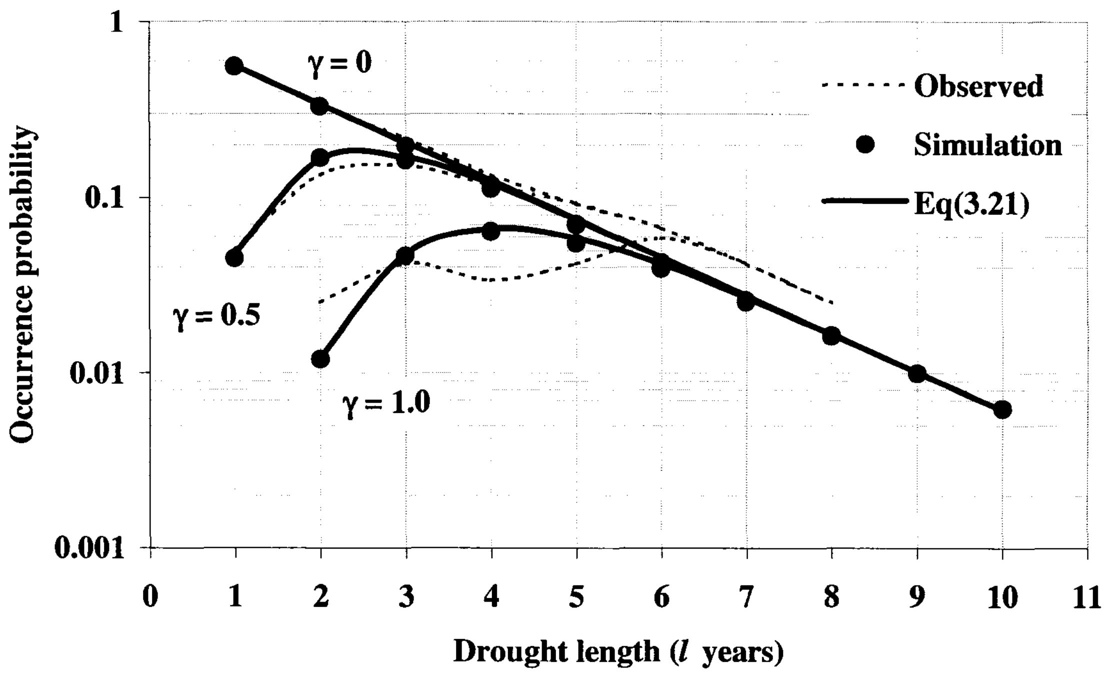

Fig. 3.3a Occurrence probability of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Poudre river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical solution (3.21) versus empirical results obtained from analyzing the observed and the 50000 simulated flows.

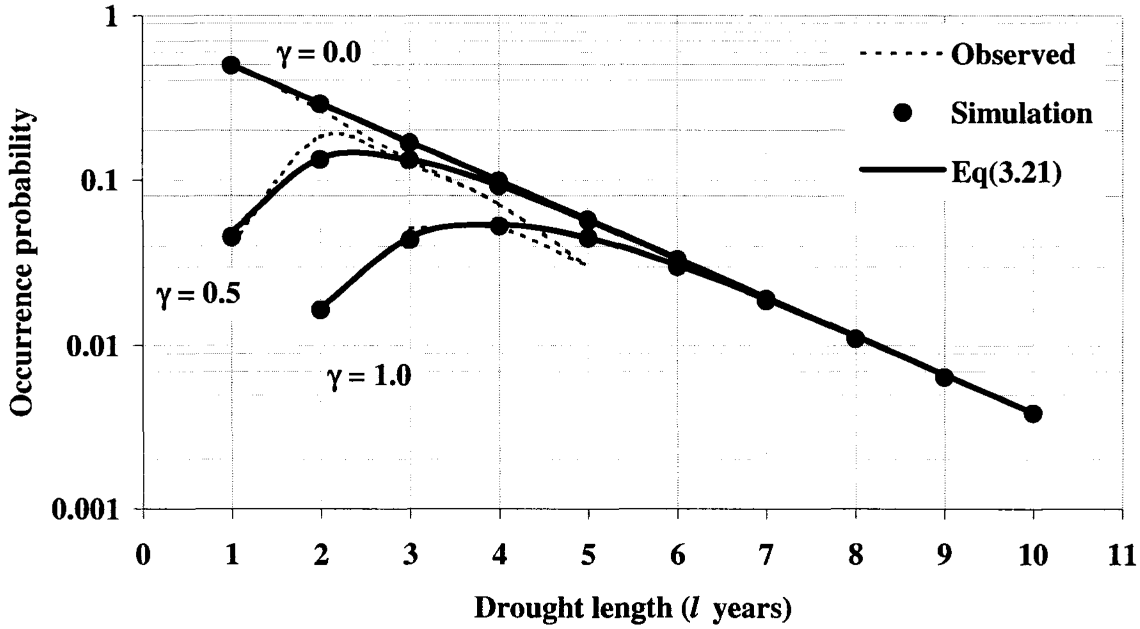

Fig. 3.3b Occurrence probability of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Colorado river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical solution (3.21) versus empirical results obtained from analyzing the observed and the 50000 simulated flows.

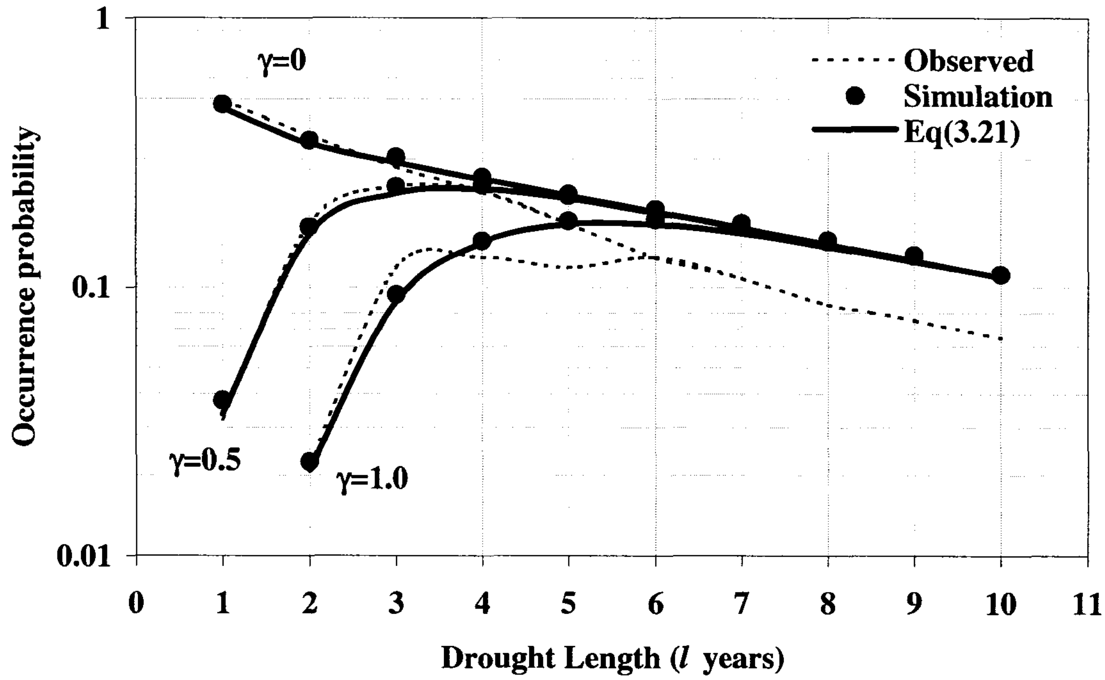

Fig. 3.3c Occurrence probability of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Niger river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical solution (3.21) versus empirical results obtained from analyzing the 50000 simulated flows.

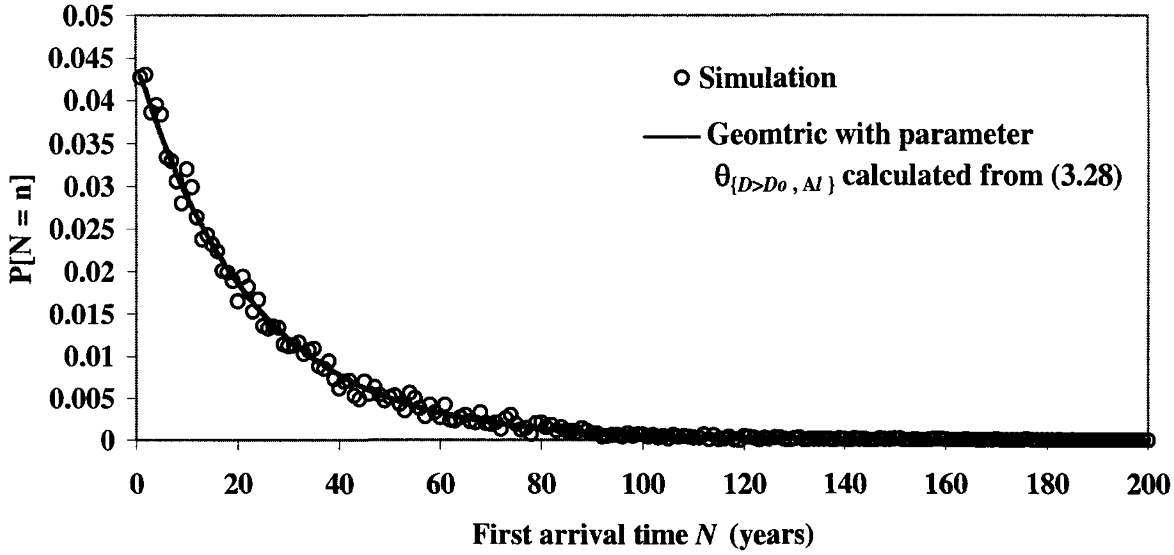

Fig. 3.4a Empirical distribution of the first arrival time N obtained from simulating the Poudre River flows to the event $\{ D > 0.5y_o \cap A_l \}$ first occurrence versus the geometric fit of N with parameter $\theta_{D>D_o \cap A_l}$ estimated using (3.28). Drought length $l = 1$ year.

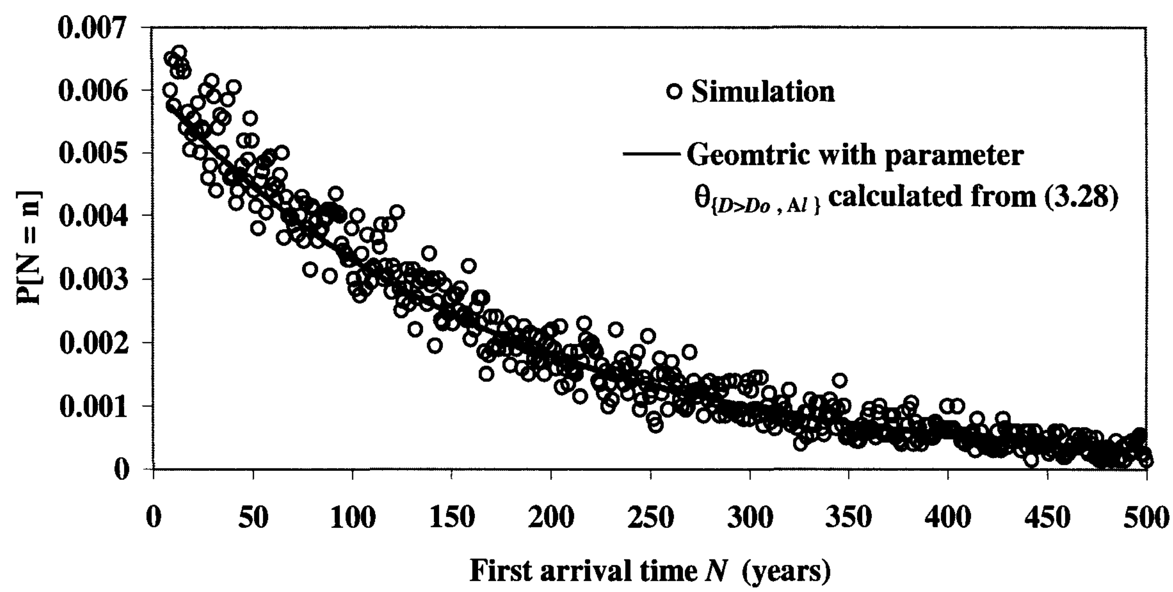

Fig. 3.4b Empirical distribution of the first arrival time N obtained from simulating the Poudre River flows to the event $\{ D > 0.5y_o \cap A_l \}$ first occurrence versus the geometric fit of N with parameter $\theta_{D>D_o \cap A_l}$ estimated using (3.28). Drought length $l = 8$ year.

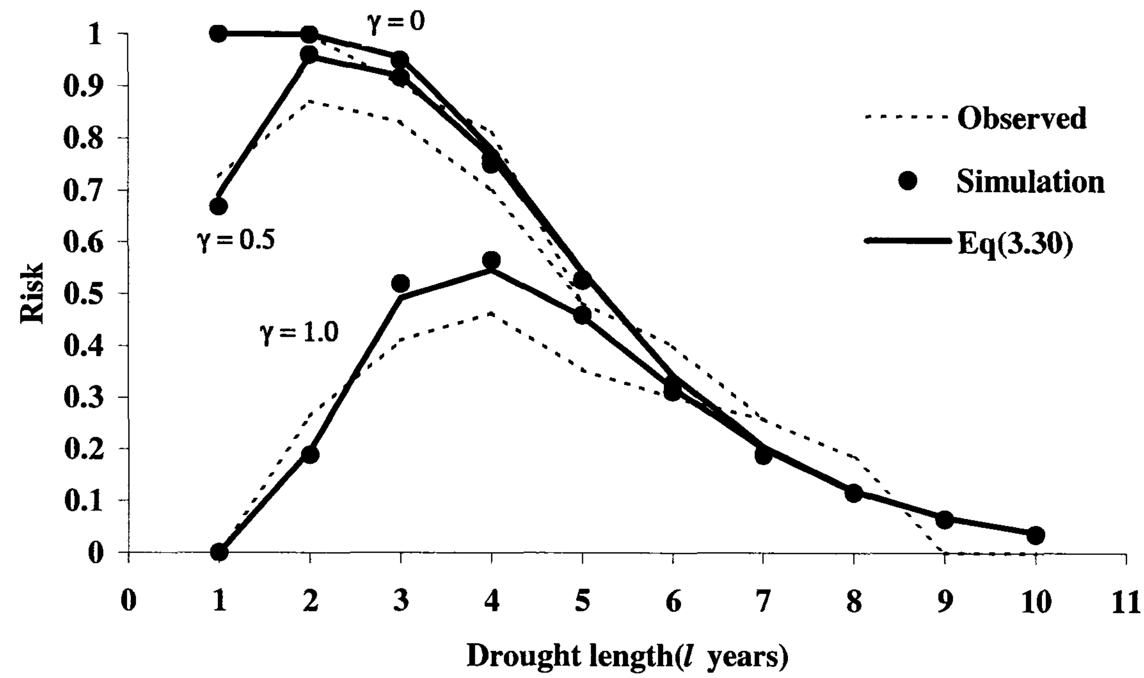

Fig. 3.5a Risk of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Poudre river at the long term mean. At γ values of 0, 0.5 and 1.0 the figure shows the analytical result using (3.30) versus the empirical results obtained from analyzing the observed and the 50000 simulated flows. The time span S =25 years.

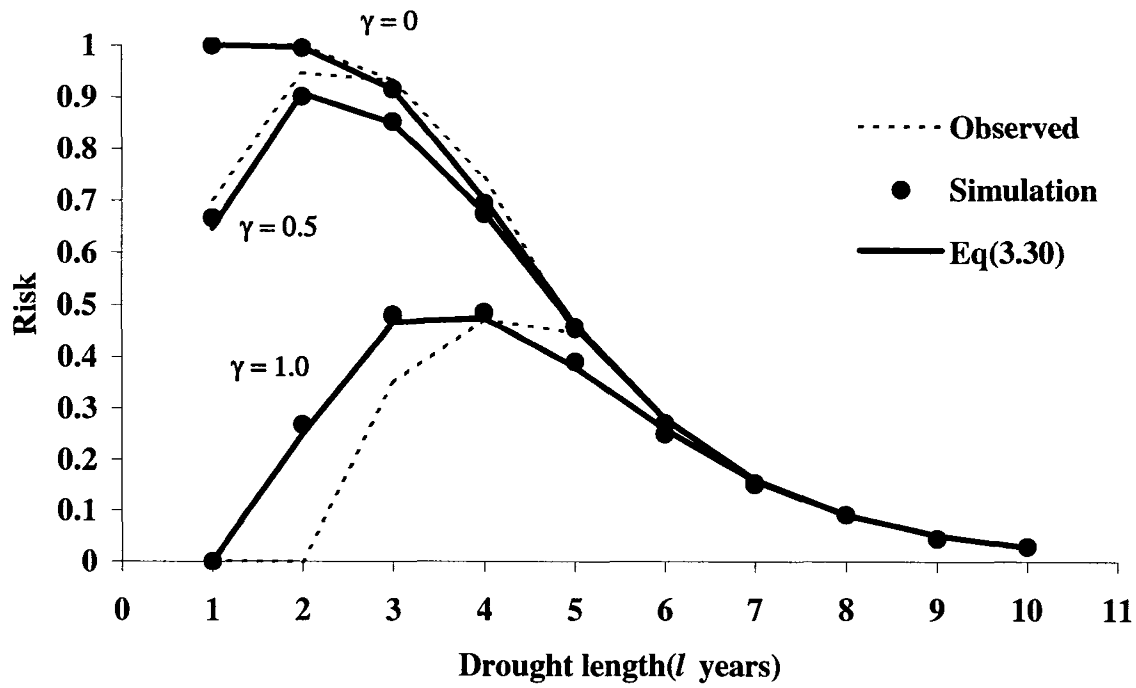

Fig. 3.5b Risk of the multiyear drought event { $D > D_o \cap A_l$ } arises from truncating the annual streamflows of the Colorado river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical result using (3.30) versus the empirical results obtained from analyzing the observed and the 50000 simulated flows. The time span S =25 years.

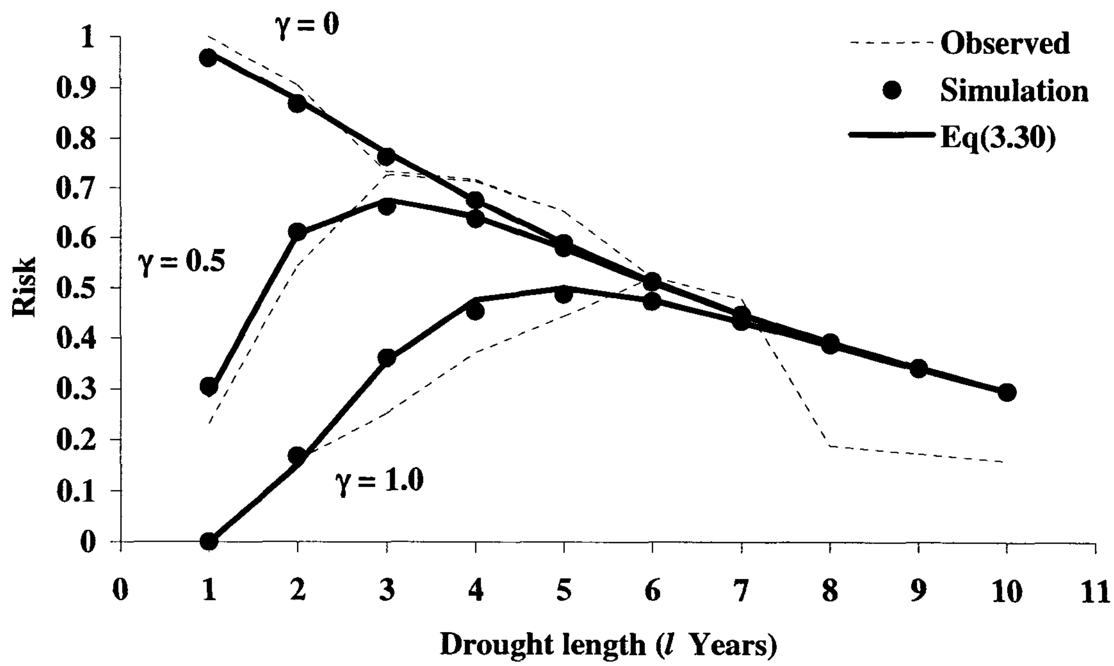

Fig. 3.5c Risk of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Niger river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical results using (3.30) versus the empirical results obtained from analyzing the 50000 simulated flows. The time span S =25 years.

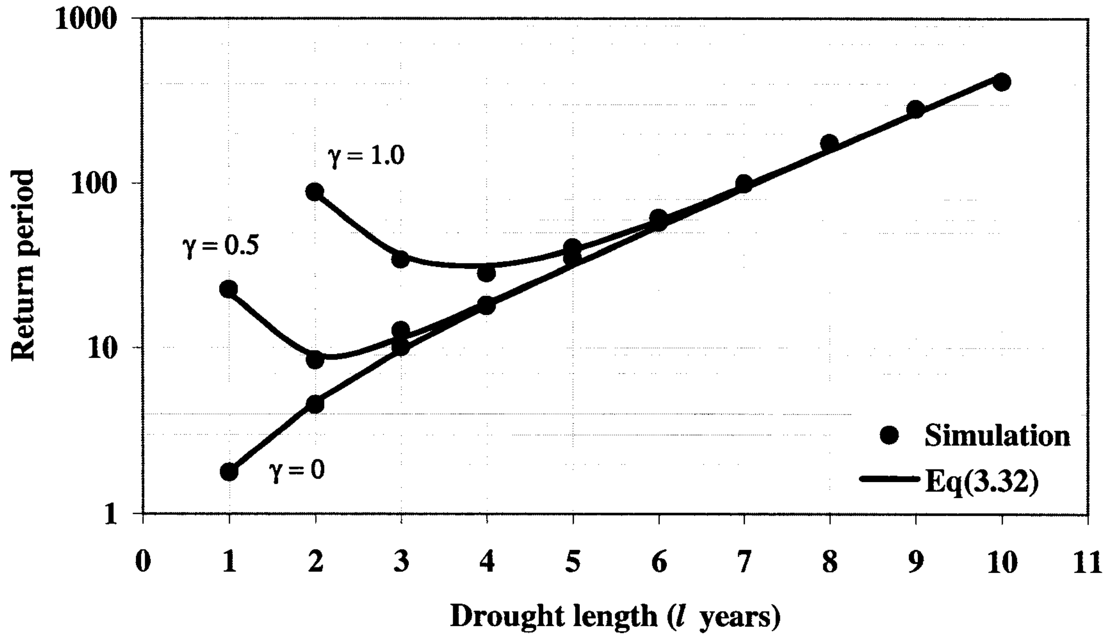

Fig. 3.6a Return period of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Poudre river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(N)$ obtained using (3.32) versus empirical results obtained from analyzing the 50000 simulated flows.

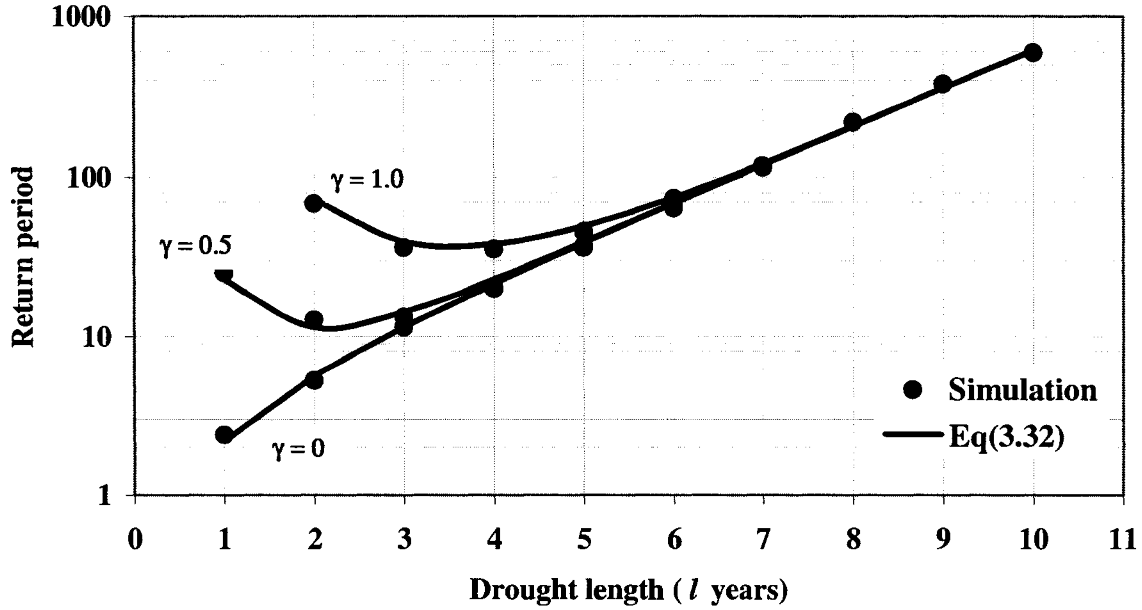

Fig. 3.6b Return period of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Colorado river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(N)$ obtained using (3.32) versus empirical results obtained from analyzing the 50000 simulated flows.

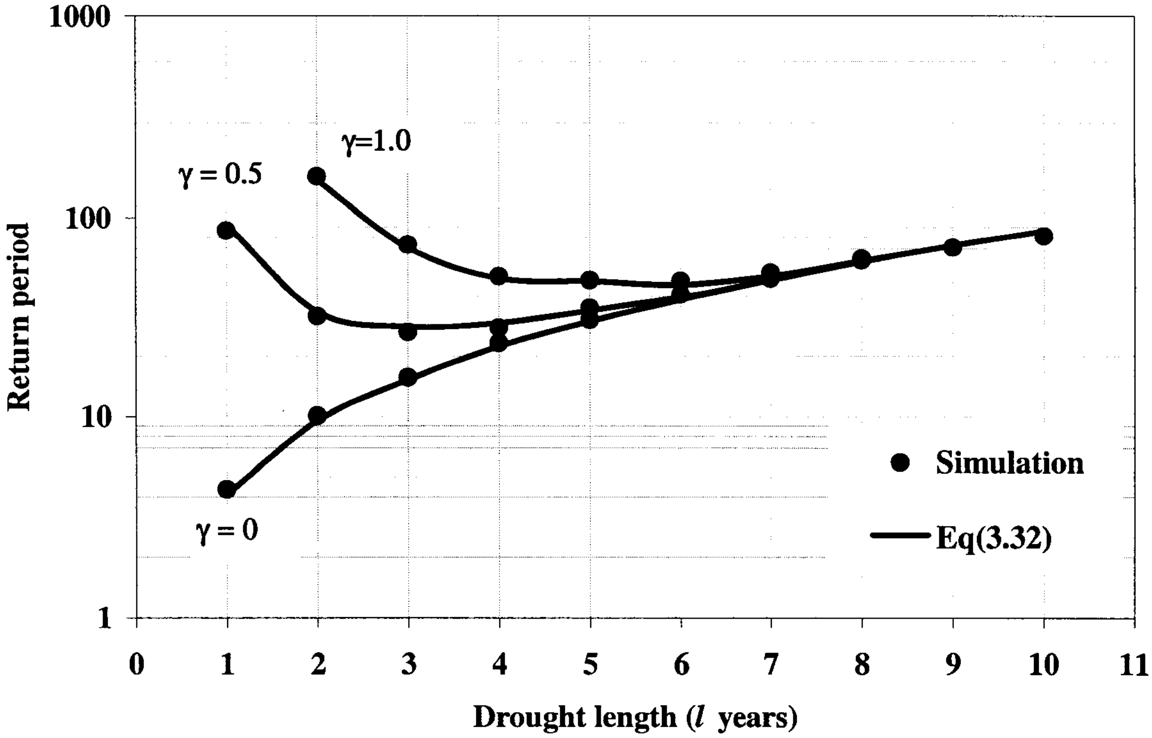

Fig. 3.6c Return period of the multiyear drought event $\{D > D_o \cap A_l\}$ arises from truncating the annual streamflows of the Niger river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(N)$ obtained using (3.32) versus empirical results obtained from analyzing the 50000 simulated flows.

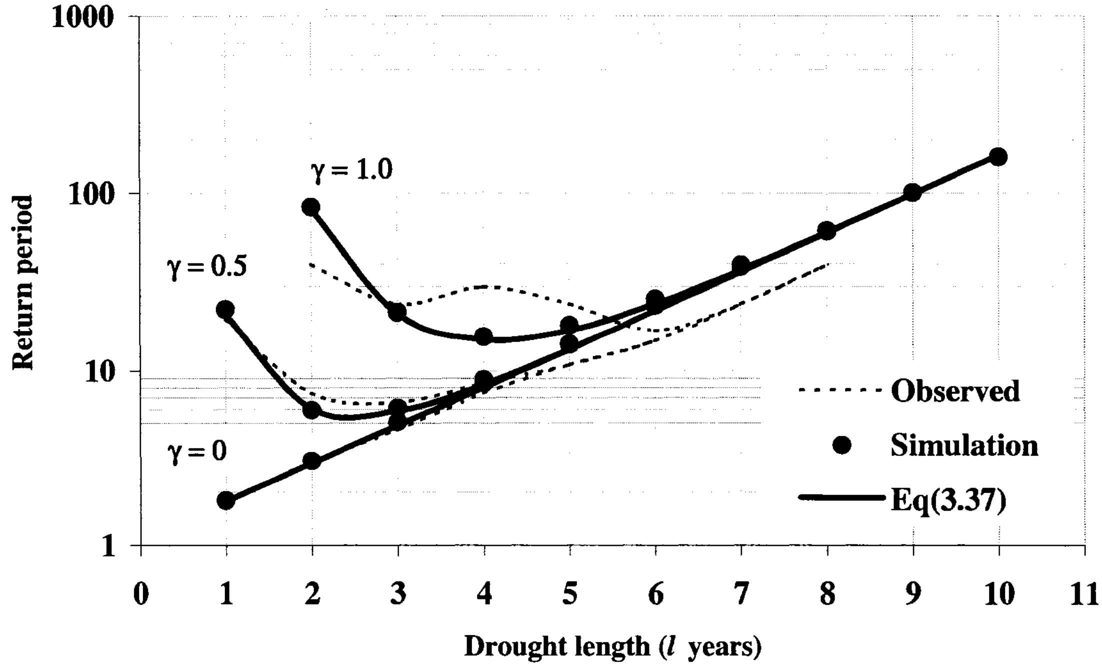

Fig. 3.7a Return period of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Poudre river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(W)$ obtained using (3.37) versus empirical results obtained from analyzing the observed and the 50000 simulated flows.

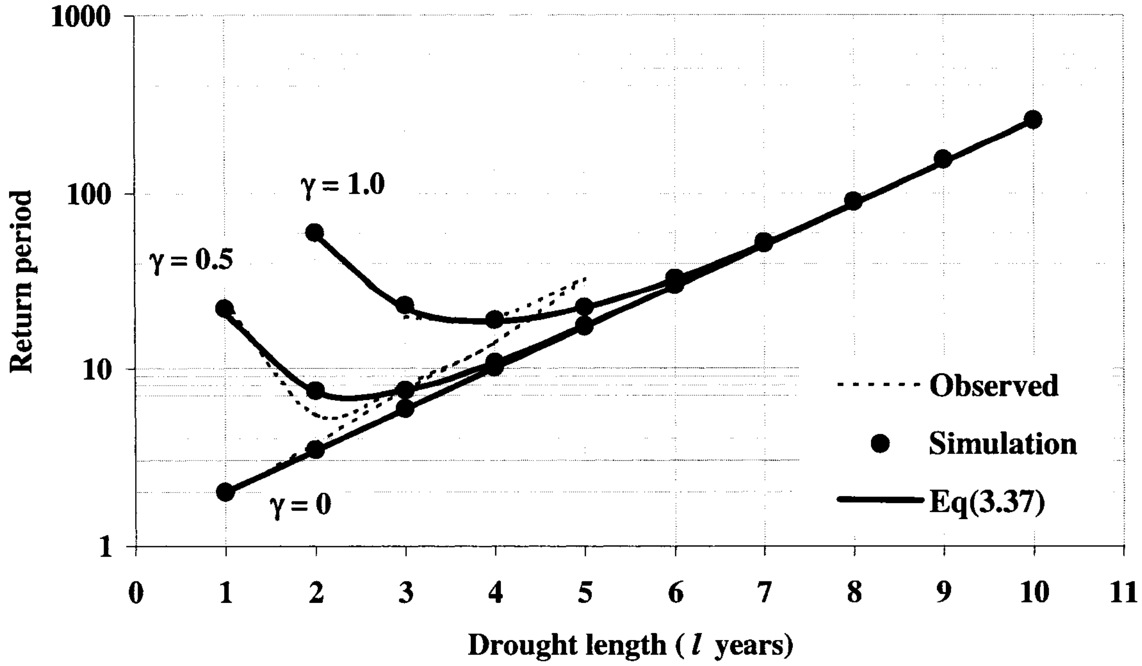

Fig. 3.7b Return period of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Colorado river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(W)$ obtained using (3.37) versus empirical results obtained from analyzing the observed and the 50000 simulated flows.

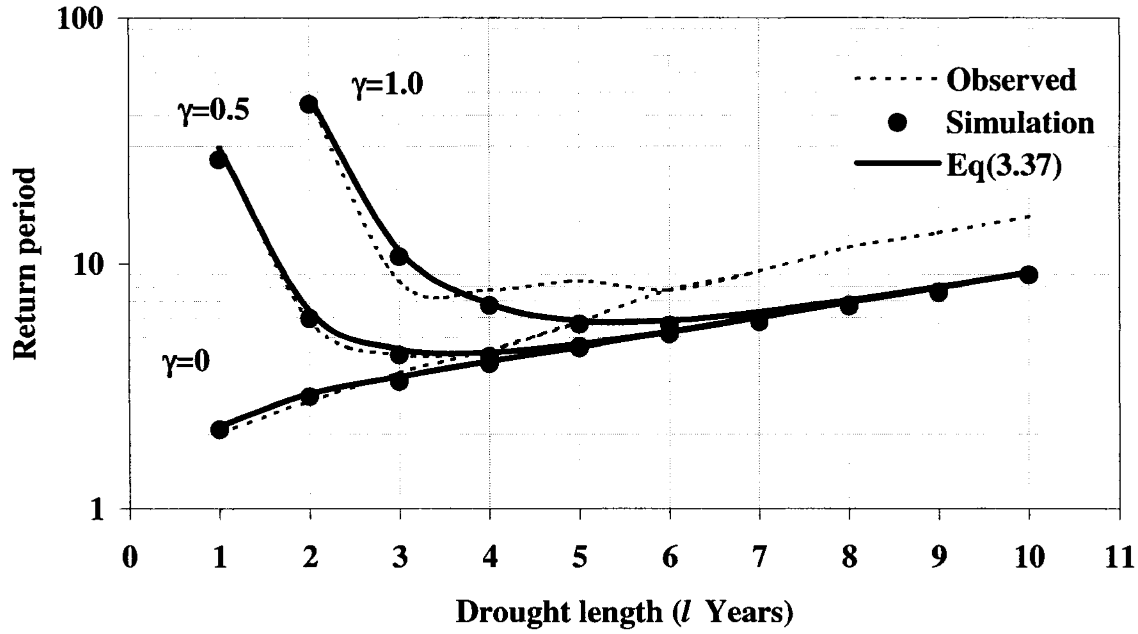

Fig. 3.7c Return period of the multiyear drought event $\{ D > D_o \cap A_l \}$ arises from truncating the annual streamflows of the Niger river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(W)$ obtained using (3.37) versus empirical results obtained from analyzing the 50000 simulated flows.

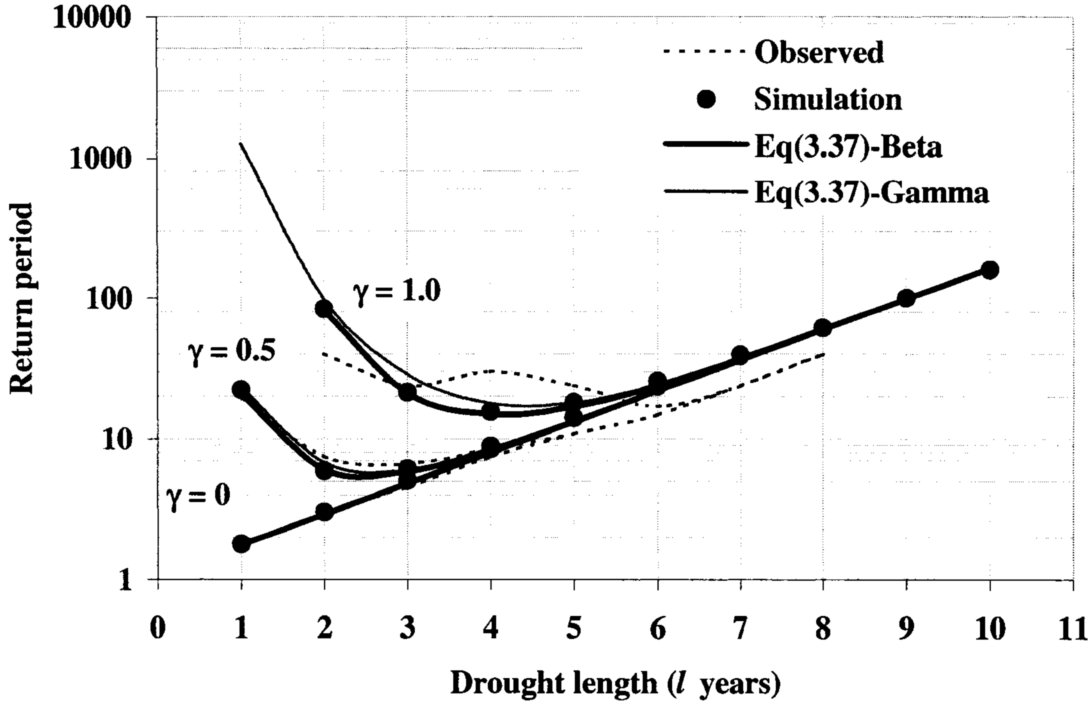

Fig. 3.8 Return period of the multiyear drought event $\{D > D_o \cap A_l\}$ arises from truncating the annual streamflows of the Poudre river at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the return period as $E(W)$ obtained using (3.37) with deicits as Beta and as Gamma versus empirical results obtained from analyzing the observed and the 50000 simulated flows.

3.10 Appendix 3A

Assume that the series of deficits, d_1, d_2, d_3, …, d_l is ARMA(1,1) process. The deficit D as the sum of single deficits conditioned to the drought length l is given as:

$$D = \sum_{t=1}^{l} d_t \tag{3.38}$$

Taking the expectation on both sides of (3.38) then:

$$E[D] = E[d_1] + E[d_2] + ... + E[d_l] \tag{3.39}$$

then assuming stationarity in the series of deficits then $E[d_1]$ = $E[d_2] = E[d_l] = \mu_d$ and

$$\mu_{D|A_L} = l\mu_d \tag{3.40}$$

The variance of D is:

$$V[D] = \sum_{i=1}^{l}\sum_{j=1}^{l} COV[d_i, d_j]$$

$$= lV[d_i] + 2(l-1)COV[d_i, d_{i-1}] + 2(l-2)COV[d_i, d_{i-2}] + ... + 2COV[d_i, d_{i-l+1}]$$

then assuming stationarity in series of deficits, then

$$V[D] = V[d_i]\{l + 2(l-1)\rho_1 + 2(l-2)\rho_2 + ... + 2\rho_{l-1}] \tag{3.41}$$

where ρ_k is the lag-k autocorrelation coefficient. (3.41) can be written as:

$$\sigma^2_{D|A_l} = \sigma^2_d\{l + 2\sum_{k=1}^{l}(l-k)\rho_k] \tag{3.42}$$

For ARMA(1,1) process, the lag-k autocorrelation is $\rho_k = \phi\rho_{k-1}$ for k ≥ 2 or simply can be written as $\rho_k = \rho_1\phi^{k-1}$. Then (3.42) is written as:

$$\sigma^2_{D|A_l} = \sigma^2_d\{l + 2\rho_1\sum_{k=1}^{l}(l-k)\phi^{k-1}\} \tag{3.43}$$

for simplicity, the sum term $\sum_{k=1}^{l}(l-k)\phi^{k-1}$ in (3.43) is written as:

$$\sum_{k=1}^{l}(l-k)\phi^{k-1}=\sum_{k=1}^{l}l\,\phi^{k-1}-\sum_{k=1}^{l}k\,\phi^{k-1} \tag{3.44}$$

the term $\sum_{k=1}^{l}l\,\phi^{k-1}=l\sum_{j=0}^{l-1}\phi^{j}$, and by changing variables, i.e. j = k-1, then from the

analytics of the power series the term $l\sum_{j=0}^{l-1}\phi^{j}$ can be written as:

$$l\sum_{j=0}^{l-1}\phi^{j}=l\left[\frac{1}{1-\phi}-\sum_{j=l}^{\infty}\phi^{j}\right] \tag{3.45}$$

then changing variables one more time, i.e. $m = j - l$, then

$$l\sum_{j=0}^{l-1}\phi^{j}=l\left[\frac{1}{1-\phi}-\sum_{m=0}^{\infty}\phi^{m+l}\right]=l\left[\frac{1}{1-\phi}-\frac{\phi^{l}}{1-\phi}\right] \tag{3.46}$$

Similarly, using the analytics of the power series then it can be shown that the term

$$\sum_{k=1}^{l}k\,\phi^{k-1}=\frac{\phi-\phi^{l+1}}{(1-\phi)^{2}}+\frac{(1-\phi^{l})-l\phi^{l}}{1-\phi} \tag{3.47}$$

Combining (3.46) and (3.47) in (3.44) and later substituting (3.44) in (3.43) and simplifying, the variance is:

$$\sigma_{D|A_l}^{2}=\sigma_{d}^{2}\left\{l+2\rho_{1}\frac{l(1-\phi)-1+\phi^{l}}{(1-\phi)^{2}}\right\} \tag{3.48}$$

3.11 Appendix 3B

Assuming that the first arrival time N for the event A_l to occur is geometrically distributed with parameter θ_{A_l} and assuming that events arrived independently, the probability $P[N \leq S]$ is written as:

$$P[N \leq S] = \sum_{k=l}^{S} \theta_{A_l} \left[1-\theta_{A_l}\right]^{k-l} = 1 - \sum_{k=S+1}^{\infty} \theta_{A_l} \left[1-\theta_{A_l}\right]^{k-l} \tag{3.49}$$

let $j = k - (S+1)$, then $k = j + S + 1$ and (3.49) becomes:

$$P[N \leq S] = 1 - \sum_{j=0}^{\infty} \theta_{A_l} \left(1-\theta_{A_l}\right)^{(S+j+1)-l} = 1 - \sum_{j=0}^{\infty} \theta_{A_l} \left(1-\theta_{A_l}\right)^{j} \left(1-\theta_{A_l}\right)^{S+1-l}$$

$$= 1 - \left(1-\theta_{A_l}\right)^{S+1-l} \theta_{A_l} \sum_{j=0}^{\infty} \left(1-\theta_{A_l}\right)^{j}$$

The term $\sum_{j=0}^{\infty} (1-\theta_{A_l})^{j} = \dfrac{1}{1-(1-\theta_{A_l})} = \dfrac{1}{\theta_{A_l}}$, thus

$$P[N \leq S] = 1 - (1-\theta_{A_l})^{S+1-l} \tag{3.50}$$

CHAPTER IV

CHARACTERIZING REGIONAL HYDROLOGIC DROUGHTS

4.1 Introduction

Regional hydrologic droughts are extreme events that are characterized by regional deficits that cover large area and may extend over long time. The degree of water deficiency and shortage time dramatically affect natural habitats, economic and social sectors, urban water supply, and the modern complex industries (Heim, 2002). Generally droughts as extreme hydrologic events are characterized by having random magnitude that is associated to its random duration. However, in reality droughts are considered to be regional phenomenon as result of having the ability to develop randomly in space, i.e. droughts have spatial coverage, therefore the definition of droughts should include the areal coverage aspect (Hisdal et al, 2000).

The characterization of regional droughts can be made using drought indices (e.g. the Palmer Drought Severity Index, the Palmer Hydrological Drought Index, the Drought Monitor). In general, drought indices are useful tools to monitor and describe the spatial variability of droughts, however common weaknesses of current drought indices have been reported in literature (e.g. Byun et al, 1999). Moreover, drought indices are not suitable tools to characterize historical droughts in terms of the disability to provide statements on how droughts frequently occur (Rossi and Cancelliere, 2003; Salas et al, 2005). Alternatively, the characterization of regional droughts can be made effectively using statistical approaches. In that case, regional droughts can be described by the occurrence probability, average recurrence time, and risk of occurrence given fixed time period. Two main statistical approaches that incorporate the regional aspect of droughts have been suggested in the literature. The first approach focuses on the proportion of the area covered by deficit as the key term in characterizing regional droughts (e.g. Tase, 1976; Sen, 1980; Santos, 1983; Kingery, 1992; Shin and Salas, 2000; Hisdal and Tallaksen, 2003). The second approach relies on joint analyzing the study variable at many sites where simultaneous deficits are observed as the key term in characterizing regional droughts (e.g. Yevjevich, 1972; Guerrero-Salazar and Yevjevich, 1975; Sen, 1979; Sen, 1998; Bayazit and Onoz, 2005).

For regional hydrologic droughts, i.e. when the study variable is streamflow, the characterization of regional droughts with areal aspect following the first approach (e.g. Tase, 1976; Santos, 1983) may not appear an attractive approach in the essence that dividing the study area into grid of uniform cells may not be utilized easily, i.e. usually streamflows are measured or calculated at gaging stations at which the flows contributing areas are not uniform in shape nor equal in the size. Therefore, in this Chapter the proposed analysis to characterize regional hydrologic droughts will be based on the second approach, i.e. the joint analysis of simultaneous deficits that appear by truncating the streamflow at many sites, following Guerrero-Salazar and Yevjevich (1975).

The characterization of multiyear regional hydrologic droughts will include the evaluation of the occurrence probability, return period of the regional events, and the risk that specific regional drought may occur at least once within given time horizon. The probability that the regional drought evolves spatially from one sub-region towards the other will be also utilized for the purpose of characterizing the spatial patterns of regional droughts. The procedures developed here will be used to characterize regional hydrologic droughts in the Upper Colorado River basin.

4.2 Basic Definitions and Assumptions

Let the region under study consists of m streamflow gaging sites, $i = 1, 2, \ldots, m$, where $Y_t^{(i)}$ is the study variable (streamflow) at site(i). If $Y_t^{(i)}$ is truncated at the level $y_o^{(i)}$, then the resulted 0,1 process $X_t^{(i)}$ for site(i) is defined as:

$$X_t^{(i)} = \begin{cases} 0 & when\ Y_t^{(i)} < y_o^{(i)} \\ 1 & otherwise \end{cases} \tag{4.1}$$

where the state 0 indicates the occurrence of single year deficit. The multiyear regional drought event is defined following the alternative definition of runs of deficits (e.g. Schwager, 1983). In that case, the multiyear regional drought event { $A_{l,t}^m$ } that covers m sites and has fixed length of l years occurs whenever m continuous runs each of l successive deficits occur jointly in the m sites starting at the time t. Figure 4.1 depicts two typical 3 years regional drought events that occur simultaneously in sites a and b assuming that the regional drought event is defined according to the alternative definition.

The regional deficit of the regional drought is defined as the sum of the m sites deficits over their common period of occurrence. Similar to the univariate case, the return period of the multiyear regional drought event is defined as the expected value of the time N required for the specified regional drought event to occur for first time. It is defined also as the expected value of the time interval W between recurrent regional drought events. Return period of regional droughts will be considered using both definitions stated above. Figure 4.1 shows a graphical definition of the first arrival time N and the interarrival time W between recurrent regional droughts. The risk that regional drought occurs within the time period S is defined as the probability that the multiyear regional drought event covers many sites occurs at least once within the time period S.

In any single site in the study region, the variable $Y_t^{(i)}$ is assumed to have simple temporal dependence structure, i.e. can be sufficiently modeled using AR(1) model, and after truncating $Y_t^{(i)}$ at the level $y_o^{(i)}$ the resulted 0,1 process $X_t^{(i)}$ is approximately simple Markov process. The variables $Y_t^{(1)}$, $Y_t^{(2)}$, ..., $Y_t^{(m)}$ are assumed to be in space correlated and it is assumed that MAR(1) model is sufficient to generate flows to the desired length.

4.3 Converting Multi-Sites Deficit Occurrence to Single Occurrence

Referring to (4.1), at a certain time t if one observes deficits over m sites simultaneously, then regardless of the magnitude of the regional deficit, the case is considered as regional drought. If the variables $Y_t^{(i)}$, $i = 1, 2, \ldots, m$, are mutually independent (uncorrelated in space), then the characterization of the regional drought can be obtained easily knowing the marginal probability of occurrence of the drought event $\{ A_l \}$ that has the fixed length l at each site. The single site probability of occurrence of the drought event $\{ A_l \}$ is discussed in Chapter 3. However in most cases, hydrologic variables $Y_t^{(i)}$, $i = 1, 2, \ldots, m$, show correlation in space. In that case, the characterization of multiyear regional droughts may be considered a simple task if one visualizes the multi-site deficits occurrence as a single deficit occurs in a single process which can be achieved considering the work of Guerrero-Salazar and Yevjevich (1975) as a basis.

Let the bivariate 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$, i.e. $m = 2$, as defined by (4.1) be resulted from truncating the bivariate normally distributed variables $Y_t^{(1)}$ and $Y_t^{(2)}$ at the levels $y_o^{(1)}$ and $y_o^{(2)}$ respectively. If $X_t^{(1)}$ and $X_t^{(2)}$ show to be serially and mutually dependent and assuming that each of $X_t^{(1)}$ and $X_t^{(2)}$ is simple Markov process, Guerrero-Salazar and Yevjevich (1975) derived the probability distribution of the longest negative run considering the bivariate case by defining the variable Z_t such that:

$$Z_t = 1 - [(1 - X_t^{(1)})(1 - X_t^{(2)})] \tag{4.2}$$

Since the occurrence of deficit in each of $X_t^{(1)}$ and $X_t^{(2)}$ is assigned the value 0, then referring to (4.2) the new process Z_t takes 0 (deficit occurrence) at the time t whenever deficits occur simultaneously in $X_t^{(1)}$ and $X_t^{(2)}$ at that time t, otherwise Z_t takes 1. The characteristics of the 0,1 process Z_t can be obtained empirically given the 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$, however some characteristics of the process Z_t can be obtained also theoretically as a function of the characteristics of $X_t^{(1)}$ and $X_t^{(2)}$. For example, considering the bivariate case $X_t^{(1)}$ and $X_t^{(2)}$, and knowing that the Z_t, $X_t^{(1)}$, and $X_t^{(2)}$ are all 0,1 processes then it can be shown (Appendix 4A) that the expected value of the process Z_t is:

$$E[Z_t] = 1 - P[X_t^{(1)} = 0]P[X_t^{(2)} = 0] - r_{X^1X^2}\sqrt{P[X_t^{(1)} = 1]P[X_t^{(1)} = 0]P[X_t^{(2)} = 1]P[X_t^{(2)} = 0]} \tag{4.3}$$

where $r_{X^1X^2}$ is the lag-0 cross correlation coefficient between $X_t^{(1)}$ and $X_t^{(2)}$.

If $X_t^{(1)}$ and $X_t^{(2)}$ are obtained by truncating the bivariate normal $Y_t^{(1)}$ and $Y_t^{(2)}$ at the level of the mean, then similar to Salas and Chung (2001), it can be shown that $r_{X^1X^2}$

is related to $r_{Y^1Y^2}$ as $r_{X^1X^2} = \frac{2}{\pi}\sin^{-1}(r_{Y^1Y^2})$ and $r_{Y^1Y^2}$ is the lag-0 cross correlation between the pair $Y_t^{(1)}$ and $Y_t^{(2)}$. Generally it can be shown that the kth moment of the process Z_t is given by:

$$E[Z_t]^k = P[Z_t = 1] \qquad \text{for} \quad k = 1, 2, 3, \ldots\ldots \tag{4.4}$$

and making use of (4.4), the variance of the process Z_t is:

$$V[Z_t] = P[Z_t = 1] - (P[Z_t = 1])^2 \tag{4.5}$$

Moreover, the probabilities $P[Z_t = 0]$ and $P[Z_t = 1]$ can be found easily from (4.3) and (4.4) as:

$$P[Z_t = 1] = 1 - P[X_t^{(1)} = 0]P[X_t^{(2)} = 0] - r_{X^1X^2}\sqrt{P[X_t^{(1)} = 1]P[X_t^{(1)} = 0]P[X_t^{(2)} = 1]P[X_t^{(2)} = 0]} \tag{4.6}$$

and

$$P[Z_t = 0] = P[X_t^{(1)} = 0]P[X_t^{(2)} = 0] + r_{X^1X^2}\sqrt{P[X_t^{(1)} = 1]P[X_t^{(1)} = 0]P[X_t^{(2)} = 1]P[X_t^{(2)} = 0]} \tag{4.7}$$

The lag-k serial autocorrelation coefficient $\rho_Z(k)$ of the process Z_t can be found as:

$$\rho_Z(k) = \frac{E[Z_t Z_{t-k}] - E[Z_t]E[Z_{t-k}]}{V[Z_t]} \tag{4.8}$$

The expectation term $E[Z_t Z_{t-k}]$ in (4.8) can be obtained as follows:

$$E[Z_t Z_{t-k}] = \sum_{i=0}^{1}\sum_{j=0}^{1} i\,j\; P[Z_{t-k} = i, Z_t = j] = P[Z_{t-k} = 1, Z_t = 1] \tag{4.9}$$

then substituting (4.4), (4.5) and (4.9) in (4.8), the autocorrelation coefficient $\rho_Z(k)$ is:

$$\rho_Z(k) = \frac{P[Z_{t-k}=1, Z_t=1] - \left(P[Z_t=1]\right)^2}{P[Z_t=1] - \left(P[Z_t=1]\right)^2} \tag{4.10}$$

If the variables Z_t and Z_{t-k} are independent of each other for all k and they are identically distributed, then their joint distribution $P[Z_{t-k}=1, Z_t=1]$ simply equals the product of their marginal distributions $P[Z_{t-k}=1] \times P[Z_t=1] = \left(P[Z_t=1]\right)^2$ and in that case the lag-k correlation coefficient $\rho_Z(k)$ given by (4.10) is zero for all k. If the variables Z_t and Z_{t-k} are correlated, then referring to (4.2) it can be seen that the probability $P[Z_{t-k}=1, Z_t=1]$ is:

$$P[Z_{t-k}=1, Z_t=1] = \sum_{i,j,n,o} P[X_{t-k}^{(1)}=i, X_{t-k}^{(2)}=j, X_t^{(1)}=n, X_t^{(2)}=o] \tag{4.11}$$

where i, j, n and o are all possible combinations of 0s and 1s the variables $X_{t-k}^{(1)}$, $X_{t-k}^{(2)}$, $X_t^{(1)}$, and $X_t^{(2)}$ may take so that each of Z_{t-k} and Z_t takes the value 1. Table 4.1 shows such possible combinations.

The estimation of $\rho_Z(k)$ in (4.10) requires the evaluation of the probability term $P[Z_{t-k}=1, Z_t=1]$ in (4.11) considering all the possible arrangements of i, j, n, and o (Table 4.1) the variables $X_{t-k}^{(1)}$, $X_{t-k}^{(2)}$, $X_t^{(1)}$, and $X_t^{(2)}$ may take. Since the variables $X_{t-k}^{(1)}$, $X_{t-k}^{(2)}$, $X_t^{(1)}$ and $X_t^{(2)}$ are serially and cross correlated to each other, then the normal distribution may be employed to evaluate such joint probability. For example the probability term $P[X_{t-k}^{(1)}=0, X_{t-k}^{(2)}=1, X_t^{(1)}=1, X_t^{(2)}=1]$ with $i=0, j=1, n=1$, and $o=1$, i.e. the i, j, n, and o combinations given by the first row in Table 4.1, can be evaluated as:

$$P[X_{t-k}^{(1)}=0, X_{t-k}^{(2)}=1, X_t^{(1)}=1, X_t^{(2)}=1] = P[Y_{t-k}^{(1)} < y_o^{(1)}, Y_{t-k}^{(2)} \geq y_o^{(2)}, Y_t^{(1)} \geq y_o^{(1)}, Y_t^{(2)} \geq y_o^{(2)}] \tag{4.12}$$

If the variables $Y_{t-k}^{(1)}$, $Y_t^{(1)}$, $Y_{t-k}^{(2)}$ and $Y_t^{(2)}$ are assumed to be joint normally distributed with vector of means $\boldsymbol{\mu}$ and variance-covariance matrix □ as:

$$\boldsymbol{\mu} = \begin{bmatrix} E[Y_t^{(1)}] \\ E[Y_{t-k}^{(1)}] \\ E[Y_{t-k}^{(2)}] \\ E[Y_t^{(2)}] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \end{bmatrix} \tag{4.13}$$

and

$$\Sigma = \begin{bmatrix} V[Y_t^{(1)}] & C[Y_{t-k}^{(1)}, Y_t^{(1)}] & C[Y_{t-k}^{(1)}, Y_{t-k}^{(2)}] & C[Y_{t-k}^{(1)}, Y_t^{(2)}] \\ C[Y_t^{(1)}, Y_{t-k}^{(1)}] & V[Y_t^{(1)}] & C[Y_t^{(1)}, Y_{t-k}^{(2)}] & C[Y_t^{(1)}, Y_t^{(2)}] \\ C[Y_{t-k}^{(2)}, Y_{t-k}^{(1)}] & C[Y_{t-k}^{(2)}, Y_t^{(1)}] & V[Y_t^{(2)}] & C[Y_{t-k}^{(2)}, Y_t^{(2)}] \\ C[Y_t^{(2)}, Y_{t-k}^{(1)}] & C[Y_t^{(2)}, Y_t^{(1)}] & C[Y_t^{(2)}, Y_{t-k}^{(2)}] & V[Y_t^{(2)}] \end{bmatrix} \tag{4.14}$$

and given the joint normal density function (Johnson and Wichern 2002):

$$f(y) = \frac{1}{(2\pi)^{m/2} \,|\, \Sigma \,|^{1/2}} e^{-(y-\mu)^T \Sigma^{-1} (y-\mu)/2} \tag{4.15}$$

then probability $P[X_{t-k}^{(1)} = 0, X_{t-k}^{(2)} = 1, X_t^{(1)} = 1, X_t^{(2)} = 1]$ of (4.12) can be evaluated as:

$$P[X_{t-k}^{(1)} = 0, X_{t-k}^{(2)} = 1, X_t^{(1)} = 1, X_t^{(2)} = 1] = \int_{-\infty}^{y_o^{(1)}} \int_{y_o^{(1)}}^{\infty} \int_{y_o^{(2)}}^{\infty} \int_{y_o^{(2)}}^{\infty} f(y) dy \tag{4.16}$$

Similarly, the rest of probabilities that may result from other possible combinations of i, j, n and o (Table 4.1) can be evaluated. In that case, the lag-k serial correlation coefficient $\rho_Z(k)$ of the process Z_t can be estimated theoretically under the assumptions stated above. The estimation of $\rho_Z(k)$ either empirically given the 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$ or theoretically assuming $Y_{t-k}^{(1)}$, $Y_t^{(1)}$, $Y_{t-k}^{(2)}$ and $Y_t^{(2)}$ are normal as shown above may give one a clue about the dependence structure of the process Z_t. In general, if the process Z_t is assumed to behave as low order DARMA process, i.e. DAR(1) or DARMA(1,1), then one may use the analytical solution of the single site drought characterization developed in Chapter 3 to characterize regional hydrologic droughts considering two sites.

The bivariate case in (4.2) can be extended to handle the case of simultaneously observing deficits at m sites at the time t. The variable Z_t is then defined as:

$$Z_t = 1 - [(1 - X_t^{(1)})(1 - X_t^{(2)}) \cdots (1 - X_t^{(j-1)})(1 - X_t^{(m)})] \tag{4.17}$$

Upon browsing (4.17), it can be seen that the multi-sites deficits joint occurrence in $X_t^{(1)}$, $X_t^{(2)} \ldots X_t^{(m)}$ has been converted to a single deficit occurs in Z_t. The estimation procedures of Z_t characteristics developed for the bivariate case can be extended to consider m sites, however one should expect rather complicated expressions for $E[Z_t]$, $V[Z_t]$, and $\rho_Z(k)$ as the number of sites m increases over 2. The definition of the process Z_t over m sites and therefore applying the single site drought characterization techniques developed in chapter 3 would simplify the characterization of regional droughts over m sites. Even though it is rather complicated to obtain theoretically the properties of the process Z_t, these properties can be estimated empirically given the 0,1 processes $X_t^{(1)}$, $X_t^{(2)} \ldots X_t^{(m)}$. For the purpose of this study it will be assumed that the process Z_t resulted from $X_t^{(1)}$, $X_t^{(2)} \ldots X_t^{(m)}$ is simple Markov, i.e. can be modeled as DAR(1) process. The marginal probabilities $P[Z_t = i]$ and the transition probabilities $P[Z_t = j \mid Z_{t-1} = i]$ for $i, j \in \{0,1\}$ can be estimated using (2.22) and (2.10) respectively.

4.4 Occurrence Probability of Run of Regional Deficits

At a certain time *t*, if the underlying processes $Y_t^{(1)}$, $Y_t^{(2)} \ldots Y_t^{(m)}$ are below their truncation levels $y_o^{(1)}$, $y_o^{(2)}$, $\ldots y_o^{(m)}$ then that year *t* is a single year regional deficit. Given the corresponding 0,1 processes $X_t^{(1)}$, $X_t^{(2)} \ldots X_t^{(m)}$ from (4.1), the occurrence probability $P[A_{l,t}^m]$ of the multiyear regional drought event $\{ A_{l,t}^m \}$ that covers jointly *m* sites and has length of *l* years starting at the time *t* can be expressed as:

$$P[A_{l,t}^m] = P\begin{bmatrix} X_t^{(1)} = 0, X_{t+1}^{(1)} = 0, \cdots, X_{t+l-1}^{(1)} = 0 \\ X_t^{(2)} = 0, X_{t+1}^{(2)} = 0, \cdots, X_{t+l-1}^{(2)} = 0 \\ X_t^{(3)} = 0, X_{t+1}^{(3)} = 0, \cdots, X_{t+l-1}^{(3)} = 0 \\ \vdots \qquad \vdots \qquad \vdots \\ X_t^{(m)} = 0, X_{t+1}^{(m)} = 0, \cdots, X_{t+l-1}^{(m)} = 0 \end{bmatrix} \tag{4.18}$$

When streamflows are used to characterize regional droughts, then in the same region streamflows are usually spatially and temporally correlated. In that case close form solution to (4.18) under the assumption of serial and mutual dependence may not utilize easily. Alternatively, one may use the process Z_t (4.17) to convert this complicated multi-deficits occurrence to single deficit occurs in Z_t. Assuming that the resulted 0,1 process Z_t is approximatly simple Markov stationary process, then (4.18) may be expressed as:

$$P[A_{l,t}^m] = P[Z_t = 0, Z_{t+1} = 0, \cdots, Z_{t+l-1} = 0] \tag{4.19}$$

Employing the Markovian property, then:

$$P[A_l^m] = p_0^{(Z)} [p_{00}^{(Z)}]^{\,l-1} \tag{4.20}$$

where, $p_0^{(Z)}$ and $p_{00}^{(Z)}$ are the unconditional and conditional probabilities of the 0 state in the process Z_t. This approximate solution will be used to obtain the occurrence probability of the regional run of l successive deficits that occur jointly in m sites.

4.5 Distribution of the Regional Drought Deficit

Drought components, i.e. the length and the magnitude, are interrelated to each other and therefore it is required to consider the joint analysis of both components when characterizing droughts (Salas et al, 2005). The multiyear regional drought event $\{\boldsymbol{D} > \boldsymbol{D}_o \cap A_l^m\}$ is defined as the event that jointly covers m sites and have length of l years with regional deficit $\boldsymbol{D}$ that exceeds the critical deficit $\boldsymbol{D}_o$. In this case, the critical deficit $\boldsymbol{D}_o$ is defined as the vector containing $D_o^{(i)}$ components for i = 1, 2, ...m where $D_o^{(i)}$ is defined by (3.11) for site i. The single year deficit $d_t^{(i)}$ is:

$$d_t^{(i)} = y_o^{(i)} - Y_t^{(i)} \tag{4.21}$$

When continuous single deficits are observed in site i over the length l of the drought event, then the magnitude of the drought deficit $D^{(i)}$ is measured by accumulating the single deficits $d_t^{(i)}$ over the length l,

$$D^{(i)} = \sum_{t=1}^{l} d_t^{(i)} \tag{4.22}$$

In any site i, it should be noted that the magnitude $D^{(i)}$ (4.22) is the sum of autocorrelated deficits $d_t^{(i)}$ over the length l (Salas et al, 2005). Unless the deficits are normally distributed, it is very hard to obtain the distribution of $D^{(i)}$ given that the single deficits are autocorrelated (Shiau and Shen, 2001; Salas et al, 2005). Usually, the observed or simulated data at site i is used to estimate empirical parameters of the fitted distribution (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003). Alternatively, single deficits $d_t^{(i)}$ may be assumed independent and identically distributed to derive the distribution of $D^{(i)}$ (e.g. Salas at al, 2005). Generally for the single site case, the gamma and exponential distributions are employed to fit the distribution of drought deficits (e.g. Shiau and Shen, 2001; Gonzalez and Valdes, 2003; Biondi et al, 2005; Salas et al, 2005).

In case that $D^{(i)}$, i = 1, 2, ...m, are observed in m sites over the regional drought length l, then the regional drought deficit $\boldsymbol{D}$ is the vector containing $D^{(i)}$, i.e. $\boldsymbol{D} = [D^{(1)}\ D^{(2)} \cdots D^{(m)}]^T$ where T stands for the transpose. Similar to the single site case, since the distribution of $D^{(i)}$ can be approximated conditioned to the occurrence of the single site drought event { A_l } of the fixed length l, then the distribution of $\boldsymbol{D}$ conditioned to the occurrence of the regional event { A_l^m } that has the length l may be given as:

$$P[\boldsymbol{D} > \boldsymbol{D}_o \mid A_l^m] = P[D^{(1)} > D_o^{(1)}, D^{(2)} > D_o^{(2)}, ... D^{(m)} > D_o^{(m)} \mid A_l^m] \qquad (4.23)$$

If the variables $Y_t^{(1)}$, $Y_t^{(2)} \ldots Y_t^{(m)}$ are mutually independent then it can be assumed that the deficits in sites are independent as well. Therefore the conditional distribution of $\boldsymbol{D}$ in (4.23) is obtained as the product of the marginal distributions of $D^{(i)}$ conditioned to the occurrence of the drought event { A_l^m } of the fixed length l at the different sites as:

$$P[\boldsymbol{D} > \boldsymbol{D}_o \mid A_l^m] = P[D^{(1)} > D_o^{(1)} \mid A_l^m] P[D^{(2)} > D_o^{(2)} \mid A_l^m] ... P[D^{(m)} > D_o^{(m)} \mid A_l^m] \qquad (4.24)$$

Usually in the same region, the streamflows $Y_t^{(1)}$, $Y_t^{(2)} \ldots Y_t^{(m)}$ are spatially correlated, therefore it is expected that the deficits $D^{(1)}, D^{(2)},..., D^{(j)}$ will be spatially correlated as well which makes the evaluation of $P[\boldsymbol{D} > \boldsymbol{D}_o \mid A_l^m]$ in (4.23) a hard task. Assuming that $D^{(1)}, D^{(2)},..., D^{(m)}$ are joint normally distributed then (4.23) can be evaluated theoretically, however the normal approximation of the joint distribution of $D^{(1)}, D^{(2)},..., D^{(m)}$ conditioned to the length of the regional drought may not represent well the deficits since in reality deficits $D^{(1)}, D^{(2)},..., D^{(m)}$ show to be highly skew.

Alternatively, modeling the single year regional deficit d_t' and consequently their sum D' appears an attractive option. The single year regional deficit d_t' is defined as:

$$d_t' = \sum_{i=1}^{m} d_t^{(i)} \tag{4.25}$$

Over the regional drought length l, the regional drought deficit D' is:

$$D' = \sum_{t=1}^{l} d_t' \tag{4.26}$$

For the single site case, the single deficit $d_t^{(i)}$ and their sum $D^{(i)}$ conditioned to the drought length can be fitted as Beta distributed in the light that deficit random variables is bounded by lower and upper values (Fig. 3.1). Similarly, it can be said that the single year regional deficit variable d_t' is bounded by lower value of 0 and upper value of $b' = \sum_{i=1}^{m} y_o^{(i)}$, i.e. the sum of the truncation levels over the m sites. Moreover, the variable D' conditioned to the occurrence of the regional event of the length l is also bounded variable that takes minimum value of 0 and maximum value of lb'. Similar to the univariate case, assuming that $D' \mid A_l^m$ is Beta Distributed, then

$$P[D' > D'_o \mid A_l^m] = \int_{D'_o}^{lb'} \frac{1}{B(\alpha'_l, \beta'_l)} \frac{(s - a')^{\alpha'_l - 1}(lb' - s)^{\beta'_l - 1}}{(lb' - a')^{\alpha'_l + \beta'_l - 1}} ds \tag{4.27}$$

where s is the dummy variable of the integration, $a' = 0$, b' is as defined above, α'_l and β'_l are the parameters of the Beta distribution, and $D'_o = \gamma \sum_{i=1}^{m} y_o^{(i)}$ with γ as a constant.

The parameters α'_l and β'_l of the Beta distribution can be estimated using the method of moments given $E[D' \mid A_l^m]$ and $V[D' \mid A_l^m]$. Similar to the univariate case, the moments $E[D' \mid A_l^m]$ and $V[D' \mid A_l^m]$ can be determined either empirically using the observed data or analytically employing (3.15) and (3.16) respectively as a function of the moments of the single year regional deficits, i.e. $E[d']$ and $V[d']$ that can be determined easily from the historical data.

4.6 Occurrence Probability of Multiyear Regional Drought Events

Considering the regional drought of length l and magnitude expressed by the regional deficit D' observed over l, and defining the regional drought event $\{D' > D'_o \cap A_l^m\}$ as the event that jointly covers m sites and has length of l years and regional deficit D' that exceeds the critical deficit D'_o, the occurrence probability $P[D' > D'_o \cap A_l^m]$ can be evaluated employing the law of the conditional probability. Assuming that the probability distribution of the regional deficits sum D' conditioned to the occurrence of the regional event { A_l^m } that has the fixed length l is available, i.e. $P[D' > D'_o \mid A_l^m]$, and given the occurrence probability of the regional drought event { A_l^m }, i.e. $P[A_{l,t}^m]$, the occurrence probability $P[D' > D'_o \cap A_l^m]$ of the regional drought event $\{D' > D'_o \cap A_l^m\}$ is:

$$P[D' > D'_o \cap A_l^m] = P[D' > D'_o \mid A_l^m]P[A_l^m] \tag{4.28}$$

Substituting $P[D' > D'_o \mid A_l^m]$ from (4.27) in (4.28) then,

$$P[D' > D'_o \cap A_l^m] = P[A_l^m]\int_{D'_o}^{lb'} \frac{1}{B(\alpha'_l,\beta'_l)} \frac{(s-a')^{\alpha'_l-1}(lb'-s)^{\beta'_l-1}}{(lb'-a')^{\alpha'_l+\beta'_l-1}} ds \tag{4.29}$$

where the probability $P[A_{l,t}^m]$ is given by (4.20) assuming that Z_t is simple Markov.

4.7 Risk of Multiyear Regional Droughts

The risk of the multiyear regional drought event is defined as the probability that the drought event covers m sites in a region has occurred at least once within the time period 0 - S. The analysis of the regional drought risk is similar to the analysis developed for the single site case. The risk $R_{A_l^m}(S)$ that the multiyear regional drought event { A_l^m } of the fixed length l occurs at least once within time period $0 - S$ is:

$$R_{A_l^m}(S) = \sum_{n=0}^{S} f_{n,A_l^m} \tag{4.30}$$

where f_{n,A_l^m} is probability that the regional drought event { A_l^m } of the fixed length l occurs for the first time at the step n. Similar to the case of the occurrence probability (4.18), obtaining direct and close form solution to the regional drought first occurrence probability f_{n,A_l^m} considering the spatial and serial correlation seems to be rather complicated case. However, assuming that the occurrence of the regional drought in many sites simultaneously is well represented by the occurrence of run of deficits in the process Z_t and assuming that Z_t is approximately Markov process, then the probability f_{n,A_l^m} can be computed approximately as,

$$f_{n,A_l^m} = \begin{bmatrix} 0 & for\ n < l \\ p_0^{(Z)}\left(p_{00}^{(Z)}\right)^{l-1} & for\ n = l \\ u_{0,n-l}\left(p_{00}^{(Z)}\right)^{l} + u_{1,n-l}\,p_{10}^{(Z)}\left(p_{00}^{(Z)}\right)^{l-1} - \sum_{k=1}^{l-1} f_{n-k,A_l^m}\left(p_{00}^{(Z)}\right)^{k} & for\ n > l \geq 2 \end{bmatrix} \tag{4.31}$$

where $u_{0,n} = u_{0,n-1}p_{00}^{(Z)} + u_{1,n-1}p_{10}^{(Z)} - f_{n,A_l^m}$ and $u_{1,n} = u_{0,n-1}p_{01}^{(Z)} + u_{1,n-1}p_{11}^{(Z)}$.

Considering the magnitude of the regional drought expressed by the regional deficit D', then similar to the univariate case developed in Chapter 3, the risk $R_{D'>D'_o \cap A_l^m}(S)$ that the multiyear regional drought event $\{D' > D'_o \cap A_l^m\}$ has occurred at least once within the period 0 - S is:

$$R_{D'>D'_o \cap A_l^m}(S) = 1 - \left[1 - P_{D'}(l) + P_{D'}(l)[1 - R_{A_l^m}(S)]^{1/(S+1-l)}\right]^{S+1-l} \tag{4.32}$$

where $P_{D'}(l) = P[D' > D'_o \mid A_l^m]$ and $R_{A_l^m}(S)$ are given by (4.27) and (4.30) respectively.

Given the risk of the regional drought event $\{D' > D'_o \cap A_l^m\}$, then its first occurrence probability $f_{n,\{D'>D'_o, A_l^m\}}$ is given as:

$$f_{n,\{D'>D'_o, A_l^m\}} = \begin{bmatrix} R_{D'>D'_o, A_l^m}(n) & n = 1 \\ R_{D'>D'_o, A_l^m}(n) - R_{D'>D'_o, A_l^m}(n-1), & n \geq 2 \end{bmatrix} \tag{4.33}$$

In fact, solutions given by (4.32) and (4.33) are an approximate solutions for the risk and first occurrence probability of the regional drought event $\{D' > D'_o \cap A_l^m\}$ respectively, due to the assumption that Z_t is approximately Markov process.

4.8 Return Period of the Multiyear Regional Drought

The return period of the regional drought event can be defined as the expected value of the time N (Figure 4.1) required by the regional drought event $\{D' > D'_o \cap A_l^m\}$ to occur for first time. In addition, it can be defined as the expected value of the interarrival time W (Figure 4.1) between recurrent regional drought events of the type $\{D' > D'_o \cap A_l^m\}$. Following the first definition, the return period T is:

$$T = E(N) = \sum_{n=0}^{\infty} n.f_{n,\{D'>D'_o \cap A_l^m\}} \tag{4.34}$$

where the probability $f_{n,\{D'>D'_o,A_l^m\}}$ is given in (4.33).

Similar to the single site case, the return period as the expected value of the interarrival time W between the recurrent regional drought events of the type $\{D' > D'_o \cap A_l^m\}$ can be evaluated applying the concepts of the renewal theory on droughts as renewal phenomena. In that case, the expected number $E[N_{\{D'>D'_o,A_l^m\},n}]$ of the regional drought event $\{D' > D'_o \cap A_l^m\}$ that has the fixed regional run length of l years and regional magnitude D' that exceeds D'_o in a given time period n divided by the period n converges to $1/T$ when the time period n is sufficiently large (Loaiciga et al, 1992). The number $N_{\{D'>D'_o,A_l^m\},n}$ of regional droughts that may arise from a sequence of length n is:

$$N_{\{D'>D'_o, A_l^m\},n} = \sum_{t=1}^{n-l+1} \mathrm{I}'_t \quad (4.35)$$

where I'_t is an indicator function that takes 1 when l successive regional deficits occur in m sites starting at the time t with regional deficits D' that exceeds D'_o and 0 otherwise. The expected number $E[N_{\{D'>D'_o, A_l^m\},n}]$ of regional droughts of the kind $\{D' > D'_o \cap A_l^m\}$ is:

$$E[N_{\{D'>D'_o \cap A_l^m\},n}] = (n - l + 1)\, E[\mathrm{I}'_t] \quad (4.36)$$

and the $E[\mathrm{I}'_t]$ is:

$$E[\mathrm{I}'_t] = P[D' > D'_o \cap A_l^m] \quad (4.37)$$

Assuming n is sufficiently large, i.e. $n \approx n\text{-}l\text{+}1$, then based on the fundamental result of the renewal theory the return period T as the expected value of the interarrival time W between recurrent regional drought events of the type $\{D' > D'_o \cap A_l^m\}$ is:

$$T = E(W) = \frac{n}{E[N_{\{D'>D'_o \cap A_l^m\},n}]} = \frac{1}{P[D' > D'_o \cap A_l^m]} \quad (4.38)$$

The probability $P[D' > D'_o \cap A_l^m]$ of the recurrent drought event is given by (4.29).

4.9 Spatial Patterns of Regional Hydrologic Droughts

Besides the ability to evolve over the time, droughts in reality grow in space randomly. In this section, the proposed procedure to analyze the spatial development of regional droughts assumes that the regional drought starts at first at a specific sub-region and later with the time it grows toward other sub-regions in the main region. The objective behind this procedure is to evaluate in terms of probability the chance, based on yearly time scale, that drought develops from one specific sub-region towards the other, for example from the specific sub-region denoted by B towards the specific sub-region denoted by C. To evaluate such probability of drought growth it will be necessary to assume that the year $t-2$ was the end of the wet period over sub-regions B. In addition, it is needed to assume that at the year $t-1$ the regional drought has occurred in the specific sub-region B only while region C is still under wet condition. At the year t, the interest is to evaluate the probability that the regional drought would evolves towards the specific sub-region C while the sub-region B is still under drought condition as well.

For the bivariate case, for example the sub-region B contains site 1 with 0,1 process $X_t^{(1)}$ while sub-region C contains site 2 with 0,1 process $X_t^{(2)}$. Assuming that each of $X_t^{(1)}$ and $X_t^{(2)}$ is simple Markov process and that $X_t^{(1)}$ is spatially correlated with $X_t^{(2)}$, then under the assumption that drought has occurred firstly in site1 only in the year $t-1$, the probability $P_{1\to 2}$ that this drought would evolve from site 1 towards site 2 in the year t while site 1 is still under drought condition is given as:

$$P_{1\to 2} = P[(X_t^{(1)}=0, X_t^{(2)}=0) \mid (X_{t-1}^{(1)}=0, X_{t-1}^{(2)}=1)] \tag{4.39}$$

If the joint process $Z_t^{(1,2)}$ is defined from $X_t^{(1)}$ and $X_t^{(2)}$ using (4.2), then it can be said that $P[Z_{t-1}^{(1,2)}=0] = P[X_{t-1}^{(1)}=0, X_{t-1}^{(2)}=0]$ and that:

$$P[Z_{t-1}^{(1,2)}=1] = P[X_{t-1}^{(1)}=0, X_{t-1}^{(2)}=1] + P[X_{t-1}^{(1)}=1, X_{t-1}^{(2)}=0] + P[X_{t-1}^{(1)}=1, X_{t-1}^{(2)}=1] \tag{4.40}$$

Equation (4.40) shows all the possible combinations of 0s and 1s the processes $X_t^{(1)}$ and $X_t^{(2)}$ may take such that the joint process $Z_t^{(1,2)}$ takes 1 (non-drought).

However, based on the fact that the conditioning term in (4.39), i.e. ($X_{t-1}^{(1)} = 0, X_{t-1}^{(2)} = 1$), is just one part of (4.40), then the probability $P_{1\to 2}$ in (4.39) may be expressed as the conditional probability $P[Z_t^{(1,2)} = 0 \mid Z_{t-1}^{(1,2)} = 1]$ corrected for the contribution of the case ($X_{t-1}^{(1)} = 0, X_{t-1}^{(2)} = 1$) only. In that case:

$$P_{1\to 2} = P[Z_t^{(1,2)} = 0 \mid Z_{t-1}^{(1,2)} = 1] \times \frac{P[X_{t-1}^{(1)} = 0, X_{t-1}^{(2)} = 1]}{P[Z_{t-1}^{(1,2)} = 1]} \tag{4.41}$$

where the probabilities $P[Z_t^{(1,2)} = 1]$ and $P[Z_t^{(1,2)} = 0 \mid Z_{t-1}^{(1,2)} = 1]$ can be found from the characteristics of the process $Z_t^{(1,2)}$ assuming it can be fitted as DAR(1) process, while the probability $P[X_t^{(1)} = 0, X_t^{(2)} = 1]$ can be found empirically by jointly analyzing the 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$. However, when $X_t^{(1)}$ and $X_t^{(2)}$ are originated from the bivariate normal pair $[Y_t^{(1)}, Y_t^{(2)}]$, then the probability $P[X_t^{(1)} = 0, X_t^{(2)} = 1]$ can be computed using the bivariate normal distribution as $P[Y_t^{(1)} < y_o^{(1)}, Y_t^{(2)} \geq y_o^{(2)}]$.

The procedure (4.41) can be extended generally to evaluate the probability that the regional drought would grow spatially from one sub-region contains many sites to another different sub-region of many sites. For example, if drought has occurred at first in sub-region *B* that contains sites *b*1, *b*2, ... *bm*, and one would like to calculate the probability that this drought would evolve towards the sub-region *C* that contains sites *c*1, *c*2, ...*cn*, then similar to (4.41) this probability is:

$$P_{B\to C} = P[Z_t^{(B,C)} = 0 \mid Z_{t-1}^{(B,C)} = 1] \times \frac{P[U_{t-1}^{(B)} = 0, U_{t-1}^{(C)} = 1]}{P[Z_{t-1}^{(B,C)} = 1]} \tag{4.42}$$

where the processes $Z_t^{B,C}$, U_t^B and U_t^C are defined using (4.17) as follows:

$$U_t^B = 1 - [(1 - X_t^{(b1)})(1 - X_t^{(b2)}) \cdots (1 - X_t^{(bm)})] \tag{4.43}$$

$$U_t^C = 1 - [(1 - X_t^{(c1)})(1 - X_t^{(c2)}) \cdots (1 - X_t^{(cn)})] \tag{4.44}$$

$$Z_t^{B,C} = 1 - [(1 - X_t^{(b1)})(1 - X_t^{(b2)}) \cdots (1 - X_t^{(bm)})(1 - X_t^{(c1)})(1 - X_t^{(c2)}) \cdots (1 - X_t^{(cn)})] \tag{4.45}$$

If the processes, $X_t^{(a1)}$, ..., $X_t^{(am)}$, $X_t^{(b1)}$, ..., $X_t^{(bn)}$ are simple Markov each and $Z_t^{B,C}$ is assumed to be simple Markov, then the probability $P[Z_t^{B,C} = 0 \mid Z_t^{B,C} = 1]$ and $P[Z^{B,C} = 1]$ can be found by fitting the 0,1 process $Z_t^{B,C}$ as DAR(1) process. The term $P[U_t^{(B)} = 0, U_t^{(C)} = 1]$ can be estimated similar to the bivariate case $P[X_t^{(1)} = 0, X_t^{(2)} = 1]$.

The probability $P_{B \to C}$ gives the chance that a drought would grow from sub-region *B* towards sub-region *C*. In addition, if the probability $P_{B \to C}$ is compared with the probability $P_{C \to B}$ then one may be able to describe the regional drought growth direction that could be of interest in regional planning of water resources systems.

4.10 Application

Multiyear regional hydrological droughts that occur in the Upper Colorado River basin (UCRB) will be characterized by evaluating the first occurrence probability, occurrence probability, return period, and the risk using the proposed procedures developed in this Chapter. The UCRB consists of 20 streamflow gaging sites with annual streamflow records for the entire 20 sites cover the period 1906 – 2003 as estimated by the U. S. Bureau of Reclamation. Figure 4.2 shows the spatial distribution of the 20 sites over of the study area. The UCRB was selected as a study region for its importance as a source of water supply for agriculture, hydroelectric power generation, municipal, and industrial uses in the western United States. Many water resources projects like the Front Range Diversions in Colorado, Lake Powell and Lake Mead are located along the river.

The historical streamflow records of length 98 years are relatively short to reliably characterize droughts in the UCRB, therefore simulated streamflows of 50000 years long have been obtained for the 20 sites in the UCRB. Practically, the length of the observed streamflow data is limited relative to the total number of parameters to be estimated when using multivariate models to simulate at once the flows at all the 20 sites in UCRB, i.e. the principle of parameters parsimony may not be preserved. In that case, the reasonable approach is to use the multivariate model to simulate streamflows at few sites (e.g. key sites) and then streamflows at the remaining sites (subsites) can be obtained using the spatial disaggregation model (2.70). The key sites are selected spatially such that the spatial disaggregation schemes can be applied easily to get flows at the rest of the sites. In the case study, the selected key sites are 20, 19, 18, 17, 16, and 8.

Fitting MAR(1) model to the historical flows at sites 20, 19, 18, 17, 16, and 8, 50000 years of simulated flows were firstly obtained. The generated flows at the key sites 8 and 16 were spatially disaggregated to obtain the 50000 years of generated flows at up stream subsites 15, 14, 13, 12, 11, 7, 6, and 2 (Figure 4.2). Similarly, the generated flows at rest of sites (the most up stream) were obtained by disaggregating the upper level flows.

4.11 Results and Discussion

This study targets the characterization of the multiyear regional drought event $\{D' > D'_o \cap A_l^m\}$ in the UCRB that has 20 streamflow gaging sites. For the multiyear regional drought event $\{A_l^m\}$, the occurrence probability $P[A_l^m]$ that this event covers the UCRB, m = 20 sites, is given generally by (4.18). Usually in the same region, i.e. the UCRB, streamflows over space and through time show to be significantly correlated and in that case the close form solution of (4.18) if existed seems to be rather complicated. However, (4.20) is considered an approximate solution for (4.18) which can be obtained utilizing the predefined 0,1 process Z_t given by (4.17). Specifically, one should be able to convert the multisite simultaneous deficit occurrence in the processes $X_t^{(1)}$, $X_t^{(2)}$, ..., $X_t^{(m)}$ to an equivalent single deficit occurrence in Z_t and consequently one can use the analytical procedures developed for the single site drought analysis to characterize multiyear regional droughts over the Colorado River.

Considering the bivariate case of 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$ that arise from truncating $Y_t^{(1)}$ and $Y_t^{(2)}$ at $y_o^{(1)}$ and $y_o^{(2)}$, then useful properties of the process Z_t can be obtained theoretically given the properties of $X_t^{(1)}$ and $X_t^{(2)}$. For example, the expected value and the variance of Z_t can be obtained using (4.3) and (4.5) respectively given the characteristics of $X_t^{(1)}$ and $X_t^{(2)}$ and their cross correlation $r_{X^1X^2}$, while the lagk correlation coefficient of Z_t can be obtained using (4.10). To show the applicability of (4.3), (4.5), and (4.10) simulation experiments were used. For the purpose of simulation, the bivariate normal pair $Y_t^{(1)}$ and $Y_t^{(2)}$ was generated using MAR(1) model to a length of 50000 points. Both of $Y_t^{(1)}$ and $Y_t^{(2)}$ has mean = 0, variance = 1, and the lag1 serial correlation $\phi = 0.3$, i.e. $\phi(Y_t^{(1)}) = \phi(Y_t^{(2)}) = 0.3$. The lag0 cross correlation $r_{Y^{(1)}Y^{(2)}}$ between the pair $Y_t^{(1)}$ and $Y_t^{(2)}$ was varying between 0 to 1. The 0,1 processes $X_t^{(1)}$ and $X_t^{(2)}$ were obtained by truncating $Y_t^{(1)}$ and $Y_t^{(2)}$ at the mean level $y_o^{(1)} = y_o^{(2)} = 0$. The cross correlation between the processes $X_t^{(1)}$ and $X_t^{(2)}$ was obtained given the lag-0 cross correlation $r_{Y^{(1)}Y^{(2)}}$ between the pair $Y_t^{(1)}$ and $Y_t^{(2)}$ using $r_{X^1X^2} = \frac{2}{\pi}\sin^{-1}[r_{Y^1Y^2}]$ as shown by Salas and Chung (2001). Figures 4.3, 4.4 and 4.5 show the expected value, the variance, and the lag1 serial correlation coefficient of the process Z_t obtained using (4.3), (4.5), and (4.10) respectively compared to the results obtained from simulation.

Generally, Figures 4.3, 4.4 and 4.5 indicate that when $X_t^{(1)}$ and $X_t^{(2)}$ are obtained by truncating the normally distributed pair $Y_t^{(1)}$ and $Y_t^{(2)}$ at the mean level $y_o^{(1)} = y_o^{(2)} = 0$, then (4.3), (4.5), and (4.10) well represent the expected value, the variance, and the lag1 serial correlation coefficient of the process Z_t given by (4.2). Moreover, Figures 4.3, 4.4 and 4.5 reflect the variation in the $E(Z_t)$, $V(Z_t)$, and $\phi(Z_t)$ as the cross correlation $r_{X^1X^2}$ varies. For example, the $E(Z_t)$ varies linearly with $r_{X^1X^2}$ from 0.75 when $r_{X^1X^2} = 0$ to 0.5 when $r_{X^1X^2}$ is nearly 1, while the $V(Z_t)$ varies non-linearly with $r_{X^1X^2}$, i.e. negative polynomial of second order, from 0.1875 when $r_{X^1X^2} = 0$ to 0.25 when $r_{X^1X^2}$ is nearly 1. The bottom line is, as long as the properties of $X_t^{(1)}$ and $X_t^{(2)}$ are known from the historical data, then (4.3), (4.5) and (4.10) can still be used to obtain the $E(Z_t)$, $V(Z_t)$, and $\phi(Z_t)$ of the process Z_t regardless of the distribution of $Y_t^{(1)}$ and $Y_t^{(2)}$. Generally as the cross correlation $r_{X^1X^2}$ increases ($r_{Y^{(1)}Y^{(2)}}$ increases), the characteristics of the process Z_t get closer to the characteristics of $X_t^{(1)}$ and $X_t^{(2)}$. This result may justify the general assumption that if $X_t^{(1)}$ and $X_t^{(2)}$ are simple Markov each and highly cross correlated, then the resulted process Z_t may be well approximated by simple Markov process. However, to further investigate the structure of the process Z_t, simulation beyond the bivariate case has been employed. In that case, using MAR(1) model, the normally distributed variables $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ have been generated assuming cross correlation between any two variables $r_{Y^{(i)}Y^{(j)}}$, for $i \neq j$, is the same and varying at 0.25, 0.5, and 0.75. The serial correlation in each variable $\phi(Y_t^{(i)})$, $i = 1, 2, 3$, and 4, has been considered to be the same and varying to take 0.1, 0.25, and 0.5. At each combination of the cross and serial correlation, the variables $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ have been generated to the length of 5000 points and then truncated at the mean to obtain the 0,1 processes $X_t^{(1)}$, $X_t^{(2)}$, $X_t^{(3)}$, and $X_t^{(4)}$ and the concurrent 0,1 process Z_t from (4.17).

The empirical autocorrelation function (ACF) of Z_t has been fitted against the theoretical ACF assuming Z_t as DAR(1) and DARMA(1,1) at the different values of cross and serial correlation as shown by the Figures 4.1B – 4.3B (Appendix 4B). Generally, it can be concluded from Figures 4.1B – 4.3B that when the serial correlation in the variables $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is low (0.25 or less) then the resulted 0,1 process Z_t obtained by truncating the variables $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ may be approximated as DAR(1) and DARMA(1,1) process as well. However when the serial correlation in the variables $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is high (more than 0.25) then the resulted process Z_t may be approximated by DARMA(1,1) more accurately than the DAR(1). Moreover as the cross correlation between the variables $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ increases the resulted process Z_t behaves as DAR(1) although it is well represented by DARMA(1,1) (Figure 4.3B).

Beyond the bivariate case, i.e. the case of having m sites in the study region, the characteristics of the 0,1 process Z_t can be obtained empirically given the input processes $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(m)}$. In the case study, since the UCRB contains 20 sites, i.e. $m = 20$, the main characteristics of the process Z_t can be estimated empirically given the processes $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(20)}$ assuming that Z_t is approximately simple Markov. The processes

$X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(20)}$ were obtained by truncating the historical flows $Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(20)}$ of the entire 20 sites in the UCRB at the level of the long term mean for each.

To characterize regional hydrological droughts in the UCRB, the process Z_t has been defined over the entire 20 gaging sites forming the river basin. Further check on the dependence structure at which the process Z_t may take assuming that $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(20)}$ are simple Markov each has been carried out. Figure 4.6 shows the autocorrelation function (ACF) of the 0,1 process Z_t obtained from the historical processes $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(20)}$ of the UCRB versus the theoretical autocorrelation given by (2.8) and (2.13) assuming Z_t is DAR(1) and DRAMA(1,1) process respectively. Generally, it can be seen from Figure 4.6 that approximating Z_t as DAR(1) process is fair enough for the purpose of this study. Assuming that each of the 0,1 processes $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(20)}$ of the UCRB is simple Markov, then the result shown in Figure 4.6 may suggest that Z_t is approximately simple Markov process. Furthermore, as shown by Chung and Salas (2000), if the process Z_t is adequately fitted using DAR(1), then the run length distribution of the dry state (state 0) obtained theoretically from the characteristics of the process Z_t as DAR(1) has to represent the run length distribution of the historical or the simulated dry states at the region. In Chung and Salas (2000) runs of dry state are defined according to the definition of Yevjevich (1967).

Figure 4.7 shows the run length distribution of the dry state obtained theoretically (e.g. Chung and Salas 2000) from the characteristics of Z_t as DAR(1) compared to the empirical distribution obtained from simulation. Generally, it can be seen from Figure 4.7 that, assuming Z_t as DAR(1) process well represented the run length distribution of the dry state obtained by truncating the streamflows of sites 1 – 20 in the UCRB which also supports the assumption that Z_t is approximately simple Markov process. The unconditional and conditional state probabilities of Z_t as simple Markov process have been estimated using procedures shown in Chapter 2. For example, considering the entire region of the UCRB, the parameter λ of the DAR(1) process has been estimated using (2.8) with estimated value of 0.23. The unconditional probability of the dry state $p_0^{(Z)}$ has been estimated using (2.22) to have 0.2653 while the conditional probability of the dry state $p_{00}^{(Z)}$ has been estimated using (2.10) to have 0.434.

The occurrence probability of the multiyear regional drought event $P[A_l^m]$, m = 20, that covers the entire UCRB can be estimated using (4.20) given the characteristics of Z_t. However, in reality it is desired to characterize multiyear regional hydrologic droughts in terms of the magnitude and the length. The magnitude D' (4.26) of the multiyear regional drought event is expressed by the total deficit over the area that is under drought. Given the expected value and the variance of the historical single year regional deficit $E[d']$ and $V[d']$, the expected value and the variance of the sum D' conditioned to the occurrence of the regional drought of the length l, i.e. $E[D' \mid A_l^m]$ and $V[D' \mid A_l^m]$, were estimated using (3.15) and (3.16) respectively. For the purpose of this study, the at site truncation level $y_o^{(i)}$ has been taken as the long term mean of the historical streamflows, while the regional critical drought D'_o has been calculated as $D'_o = \gamma \sum_{i=1}^{m} y_o^{(i)}$ at γ values of 0, 0.5, and 1. Similar to the case of the single site drought characterization, since the single year regional deficit random variable d' and hence its conditional sum D' is bounded by lower and Upper values, the variables d' and D'

were assumed to be Beta distributed. The parameters of the Beta distribution have been estimated using the method of moments given $E[D' \mid A_l^m]$ and $V[D' \mid A_l^m]$.

The occurrence probability of the multiyear regional drought event $\{D' > D'_o \cap A_l^m\}$ that has the length l was computed using (4.29). Figure 4.8 shows the occurrence probability of the multiyear regional drought event $\{D' > D'_o \cap A_l^m\}$, m = 20, that covers the entire UCRB evaluated using (4.29) versus the empirical occurrence probability obtained from analyzing the simulated and the observed flows. From the results shown in Figure 4.8, generally it can be concluded that regional drought properties obtained from analyzing the observed flows alone are quite uncertain. The analytical procedure to evaluate the occurrence probability of the referred regional drought event gives quite reasonable results as compared with the empirical results obtained from analyzing the simulated flows. The results show more specifically that, for regional drought events considering just the duration, i.e. γ = 0, the occurrence probability of the regional drought event decreases as the drought length increases. In that case, the regional drought becomes extreme as a result of the increasing length. On the other hand, when considering the magnitude of the regional drought with the length, i.e. γ > 0, as the magnitude D' increases (due to the increasing γ) and the length increases then generally regional droughts become less frequent to occur. Moreover, given specific critical regional deficit D'_o (γ = 0.5), then at short drought length (say 2 years) droughts are considered less extreme (more frequent) and later to some extent as the drought length increases, droughts become less frequent to occur, i.e., at longer duration (say 4 years or more), regional droughts start becoming less frequent to occur due to their extreme length. In that case, the magnitude of the regional drought event is not contributing significantly like its extreme length to make droughts as extreme events.

In addition to estimate the chance that the regional drought event occurs, the occurrence probability shown in Figure 4.8 provides more information about the duration of the most frequent regional drought event. For example, referring to Figure 4.8 it can be noticed that at $\gamma = 0.5$, the most frequent regional drought event has length of 2 years, while at $\gamma = 1$ the most frequent regional drought event has length of 3 – 4 years. Moreover, it can be noticed that varying the amount of the critical deficit, i.e. varying γ, would result in different lengths of the most frequent regional drought, which is expected due to the contribution of the drought magnitude.

The return period calculated as the expected value of the interarrival time W between recurrent multiyear regional drought events of the type $\{D' > D'_o \cap A_l^m\}$ that covers the entire UCRB is given by (4.38). Figure 4.9 shows the return period as $E(W)$ for the drought event $\{D' > D'_o \cap A_l^m\}$, m = 20, that covers the entire UCRB obtained from (4.38) versus the empirical return period obtained from the analysis the 50000 years simulated and the observed streamflows. Generally the results obtained through the analytical procedure (4.38) developed to evaluate the return period as $E(W)$ agree quite well with the results obtained from the empirical analysis of the simulated streamflows and to some extent the observed. Generally the figure also indicates the uncertainty in the retrun period as estimated from the analysis of the observed flows alone.

Similar to the univariate case, the results from Figure 4.9 indicate that for regional drought events considering the length only, i.e. $\gamma = 0$, the return period increases as the drought length increases, which is expected because regional droughts become more extreme as the duration increases, i.e. droughts of longer duration has less chance to occur and therefore its return period will be eventually high and vise versa. Moreover the return period of regional droughts increases by increasing the magnitude of the drought, i.e. due to the increasing γ, which is expected because droughts turn to be more extreme due to the severe magnitude and consequently their return period will be high, however the magnitude of droughts with length 4 years or more is not significant as their extreme lengths in making the regional drought as an extreme event. Practically, the results in Figure 4.9 show that droughts of length more than 4 years are less important for planning purposes since these droughts are less frequent to occur, i.e. their return period exceeds the practical planning horizons (most of the cases 50 years or less). For example, regional droughts of length 5 years or more have relatively long average recurrence time (more than 100 years) compared with droughts of 4 years or less that have average recurrence time nearly 50 years or less.

The return period as the expected value of the first arrival time N needed for the multiyear regional drought event $\{D' > D'_o \cap A_l^m\}$ to occur for the first time at the trial n is given by (4.34) where the probability $f_{n,\{D'>D'_o,A_l^m\}}$ is given by (4.33). Figure 4.10 shows the return period of the multiyear regional drought event $\{D' > D'_o \cap A_l^m\}$, m = 20, that covers the entire UCRB obtained as $E(N)$ using (4.34) versus the empirical return period obtained from the analysis of the simulated streamflows. Generally, the results shown in Figure 4.10 obtained using (4.34) agree with the results obtained from the empirical analysis of the simulated streamflows. The results from Figure 4.10 indicate generally that the return period of regional drought events increases by increasing the magnitude of the drought, i.e. increasing γ, because droughts become extreme events due to their extreme magnitude and consequently their return period will be high (less frequent), while at drought length of 4 years or more the magnitude of the regional drought is not contributing in making the drought event extreme, i.e. for regional droughts of length greater than 4 years, then regardless of the drought magnitude as the drought length increases the return period increases.

Comparing the results obtained for the return period shown in Figures 4.9 and 4.10 using the two definitions of the return period then in general it can be noticed that the return period evaluated as $E(N)$ is always greater than or equal to the return period calculated as $E(W)$. This can be explained as result of the fact that the first arrival time for the drought to occur for the first time regardless of the past takes minimum value that equals the length of the desired drought but not less than. On the other hand, the interarrival time W between recurrent regional events may take minimum value of 1 year regardless of drought length l simply because drought events are defined according to the alternative definition of runs (e.g. Schwager 1983) where runs of deficits may overlap. Therefore it is expected that $T = E(N)$ is always greater than or equal to $T = E(W)$.

The risk that the multiyear regional drought occurs during the project life is an important issue in the operation and the planning of water resources systems. For regional drought events that are characterized by the length only, i.e. the regional drought event $\{A_l^m\}$, then (4.30) can be used to evaluate the risk that such drought event has occurred at least once during the specified time span S. Assuming that the time span S is 25 years, then for the regional drought event $\{D' > D'_o \cap A_l^m\}$ the risk that such drought has occurred at least once within 25 years was obtained using the procedure (4.32). The risk results obtained from (4.32) for droughts of the kind $\{D' > D'_o \cap A_l^m\}$ are compared with the results obtained from the empirical analysis using the simulated and the observed flows (Figure 4.11). Mainly, the results show that generally as the regional drought length increases, the risk that the drought occurs with time horizon of 25 years decreases. Moreover, as the level of the critical deficit increases, i.e. γ increases, the risk that the referred regional drought to occur within the 25 years decreases. Practically, the results show that regional drought events of length 5 years or more are less likely to occur, i.e. their chance to occur is very low (nearly 10%) and decreasing as l increases. When $\gamma = 0$, the regional droughts of lengths 1, 2, and 3 years are the most risky events that may cover the entire UCRB at least once within the 25 years period with probabilities of around 99%, 85%, and 50% respectively. While γ increases, the risk of occurrence during the time horizon of 25 years generally decreases, however when γ is less than 1, for example $\gamma = 0.5$, then regional droughts of 1 or 2 years get the chance in the range of 65 - 80% to occur in the UCRB at least once within 25 years.

In general, drought properties obtained from the analysis of the observed flows show to be quite uncertain when compared with results obtained from the proposed analytical procedures or from simulation. The uncertainty can be associated to the shortness in the observed flows to give reliable drought properties compared to their theoretical expectations specially when the case is extreme drought event. In fact, in some cases droughts of higher lengths may not be observed at all within the analysis of flow records of around 100 years, i.e. the case of the Colorado river. For example, referring to Figure 4.9 it can be seen that return periods obtained using the observed flows depart from their expectation, i.e. the analytical or simulation results, specially when $\gamma > 0$. Moreover, the analysis of the observed flows is in fact less informative as the drought length increases (no information about droughts of length more than 4 years). Similar conclusions are also found from the results of the occurrence probability and the risk for the referred river flows.

To get a clue of how regional droughts may evolve spatially from one sub-region to the other, i.e. from sub-region B to sub-region C, the probability $P_{B \to C}$ that the regional drought may extend from the sub-region B towards sub-region C can be obtained using (4.42). Considering the UCRB, if sub-regions B and C are defined such that the sub-region B contains sites 2, 4, 5 and 6 while the sub-region C contains sites 7, 8, 14, 15, and 16, then the probability $P_{B \to C}$ that the regional drought grows this year from the sub-region B towards sub-region C has the value of 0.041 given that the past year only sub-region B was under drought condition, while the probability $P_{C \to B}$ has the value of 0.011. In general, the estimated probabilities $P_{B \to C}$ and $P_{C \to B}$ seems to be almost the same and low (0.04 or less) which indicates that most of times the regional drought occurs in both regions B and C at the same time after the wet season. Moreover, since sub-regions B and C are east west in direction relative to each other, then from the results of $P_{B \to C}$ and $P_{C \to B}$ one may conclude that there is no clear pattern about drought growth along the east west direction in the UCRB.

The same analysis has been carried out to evaluate the probability that the regional drought may extend from sub-region D that contains sites 9, and 10 towards sub-region E that contain sites 11, 12, and 13, then the probability $P_{D \to E}$ that the drought grows from the sub-region D towards the southern sub-region E has the value of 0.1, while $P_{E \to D}$ has the value 0. In this case, may be it can be concluded that in general regional droughts in the UCRB may have clear pattern to grow along the north south direction. In particular, the probability $P_{D \to E}$ has a considerable value (10 %) compared to $P_{E \to D}$ which indicates that the direction of the drought growth is north to south more than south to north.

4.12 Summary and Conclusions

In this study, the regional hydrologic drought event is defined as the event that is made of a run of successive deficits occur simultaneously over many sites in a region. One main reason to follow this definition of drought events is due to its ability to capture extreme regional hydrologic drought events regardless where the events start, i.e. there could be a particular drought event that is extreme in terms of the magnitude and not necessarily is headed and followed by at least one surplus. The general objective is to characterize the multiyear regional hydrologic droughts. A single year regional deficit occurs, i.e. single deficit occurs in the process Z_t, whenever concurrent single year deficits in many sites in the region occur. If the structure of the process Z_t is known, then one may use analytical procedures developed to characterize multiyear droughts at single site to characterize multiyear regional droughts.

Assuming that each of the spatially correlated 0,1 processes $X_t^{(1)}$, $X_t^{(2)}$,.... $X_t^{(j)}$ is simple Markov, the resulted 0,1 process Z_t may be fairly approximated by simple Markov. The magnitude of the regional drought deficit D' is defined as the sum of the single year regional deficits d' over the length of the regional drought. The single year regional deficit d' and the regional drought magnitude D' are bounded by lower and upper values and therefore they were assumed to be Beta distributed with parameters α_l' and β_l'. The parameters α_l' and β_l' were determined using the method of moments given the first two moments of D'.

Analytical procedures developed to evaluate the occurrence probability, return period, and the risk were used to characterize the multiyear regional hydrologic droughts occurring in the Upper Colorado River basin. The proposed procedures show to give reasonable drought properties when compared with uncertain results obtained empirically from the analysis of the observed streamflows. The uncertainty in the properties of drought as obtained empirically from analyzing the observed streamflows is associated to the shortness of the observed flow records. The results obtained from the mathematical procedures were verified by results from simulation. In addition, procedure to analyze the spatial patterns of regional droughts has showed generally that regional droughts evolve spatially in the Upper Colorado basin most likely from north to south in direction.

Table 4.1 Sets of possible 0,1 combinations the variables $X_{t-k}^{(1)}$, $X_{t-k}^{(2)}$, $X_t^{(1)}$ and $X_t^{(2)}$ may take such that Z_{t-k} and Z_t take the value 1.

i	j	o	n
0	1	1	1
0	1	0	1
0	1	1	0
1	0	1	1
1	0	1	0
1	0	0	1
1	1	0	1
1	1	1	0
1	1	1	1

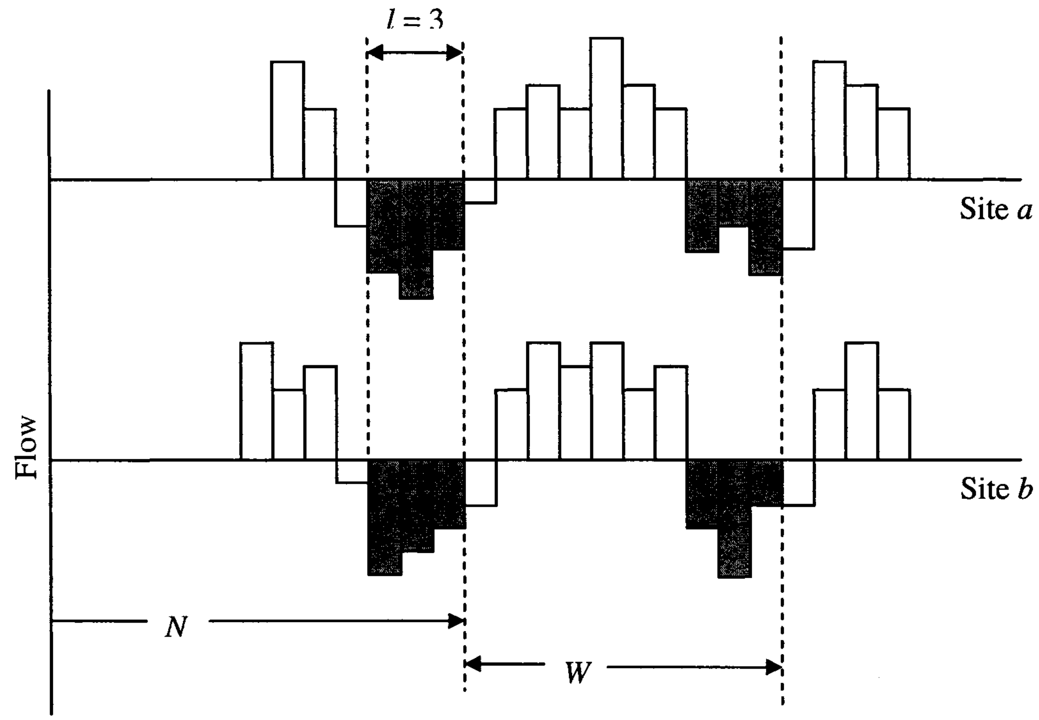

Fig 4.1 Definition of the regional drought event. Two possible regional drought events of 3 years each defined according to the alternative definition (e.g. Schwager 1983) that occur simultaneously in sites a and b are shown by dark rectangles. The first arrival time N and the interarrival time W between recurrent regional drought events are also shown.

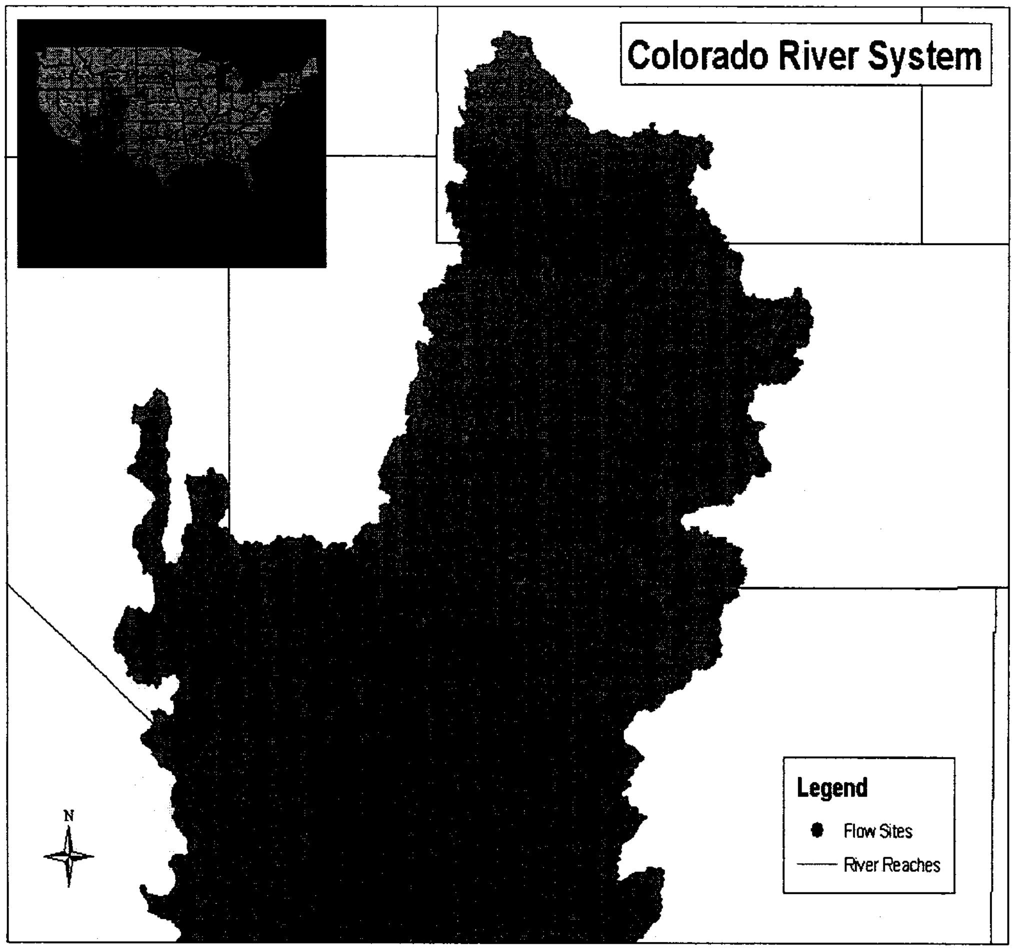

Fig 4.2 The Colorado River basin. The upper Colorado basin contains streamflow gaging sites 1 – 20 while the lower basin contains streamflow gaging sites 21 – 29.

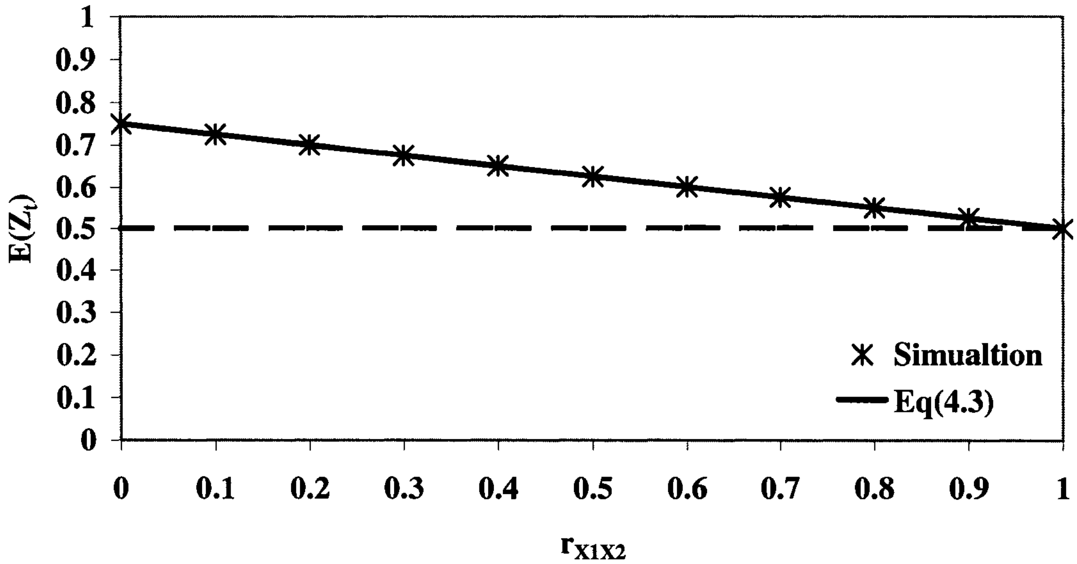

Fig 4.3 The expected value of the process Z_t defined from the pair $X_t^{(1)}$ and $X_t^{(2)}$ that is resulted from truncating the normally distributed pair $Y_t^{(1)}$ and $Y_t^{(2)}$ at $y_o^{(1)} = y_o^{(2)} = 0$. The dashed line represents $E(X_t^{(1)}) = E(X_t^{(2)}) = 0.5$.

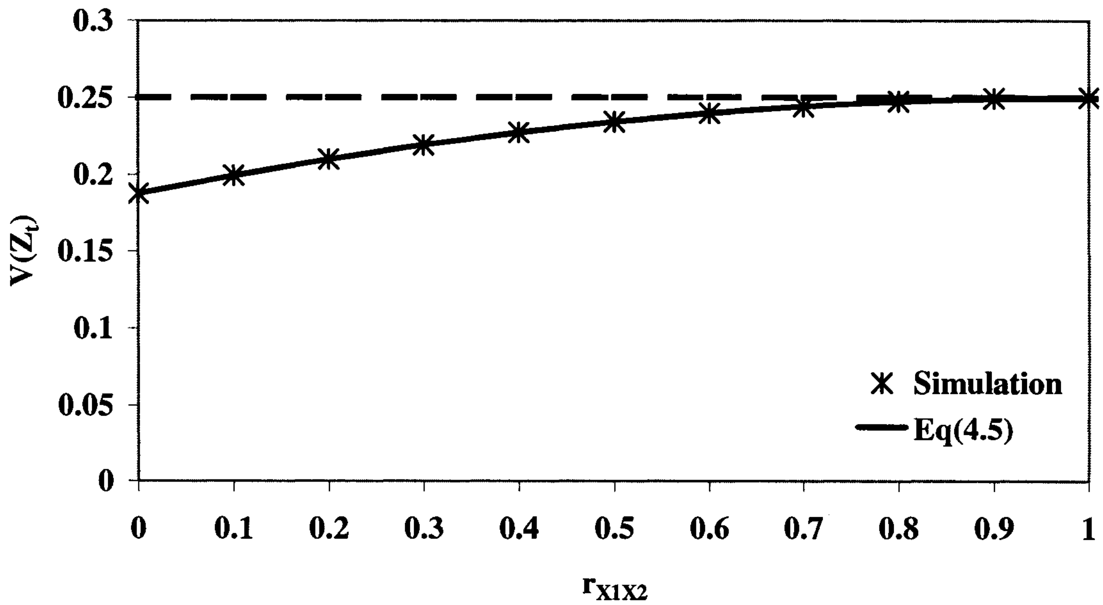

Fig 4.4 The variance of the process Z_t defined from the pair $X_t^{(1)}$ and $X_t^{(2)}$ that is resulted from truncating the normally distributed pair $Y_t^{(1)}$ and $Y_t^{(2)}$ at $y_o^{(1)} = y_o^{(2)} = 0$. The dashed line represents $V(X_t^{(1)}) = V(X_t^{(2)}) = 0.25$.

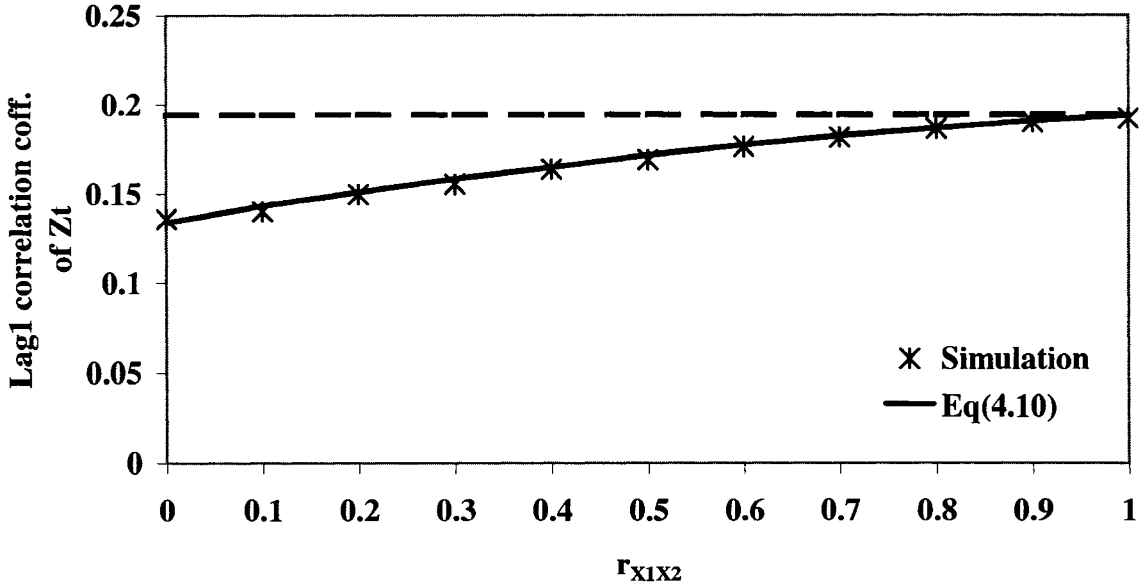

Fig 4.5 The lag1 correlation coefficient of the process Z_t defined from the pair $X_t^{(1)}$ and $X_t^{(2)}$ that is resulted from truncating the normally distributed pair $Y_t^{(1)}$ and $Y_t^{(2)}$ at $y_o^{(1)} = y_o^{(2)} = 0$. The dashed line represents $\phi(X_t^{(1)}) = \phi\,(X_t^{(2)}) = 0.194$.

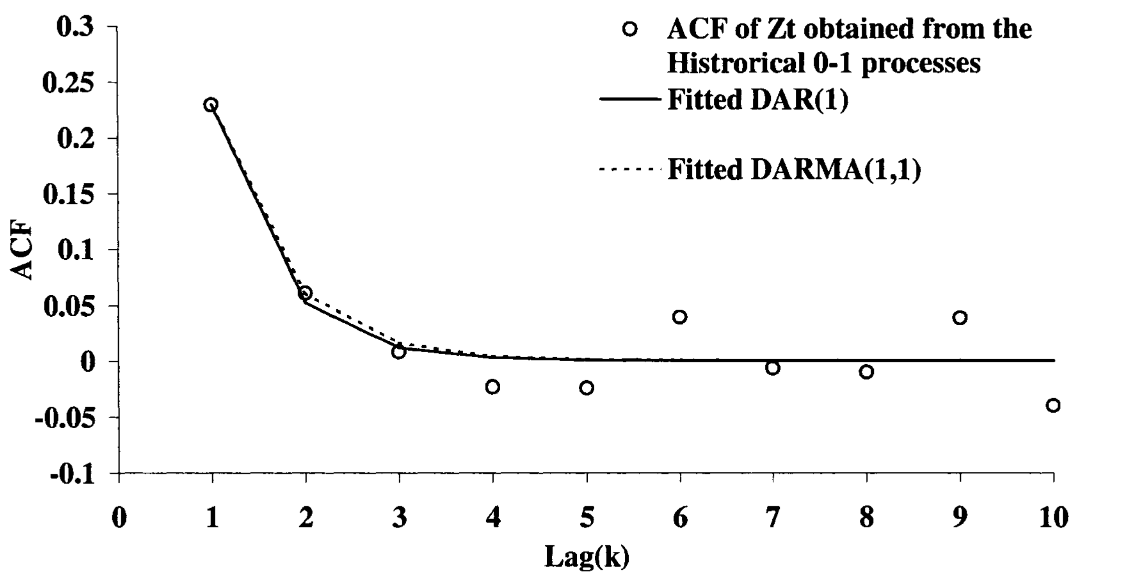

Fig 4.6 The autocorrelation function of the process Z_t defined from the 0,1 processes resulted from truncating the flows of the entire 20 sites in the Upper Colorado River basin versus the theoretical autocorrelation function assuming Z_t as DAR(1) and DARMA(1,1) process.

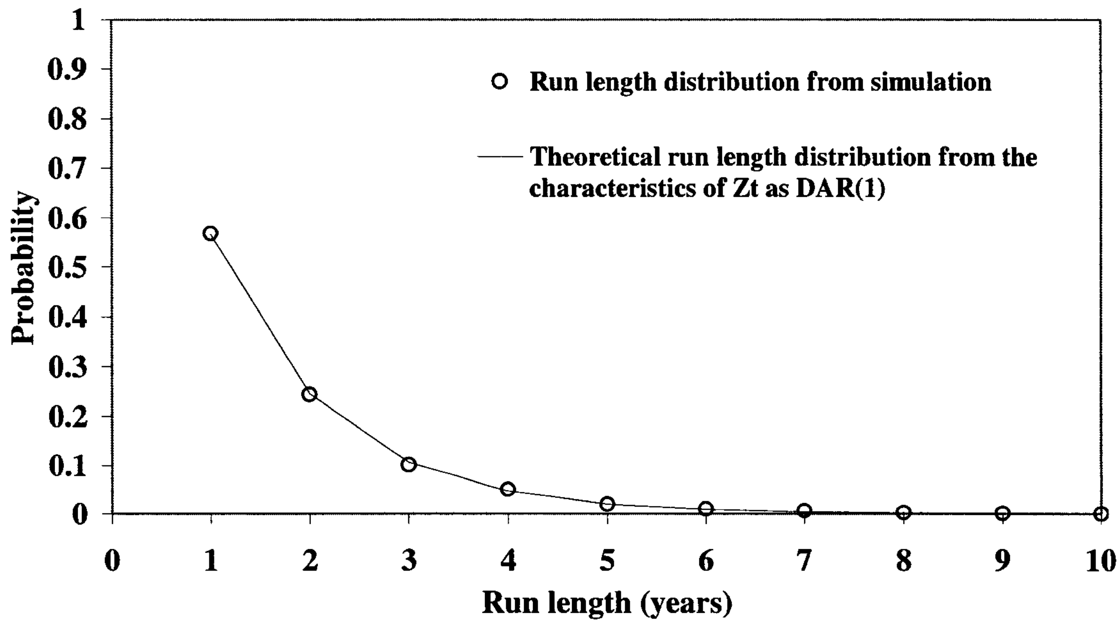

Fig 4.7 The run length distribution of the dry state obtained theoretically from the characteristics of Z_t as DAR(1) process compared to the empirical distribution obtained from simulated flows of the 20 sites in the upper Colorado River Basin. Runs are defined according to the definition of Yevjevich (1967).

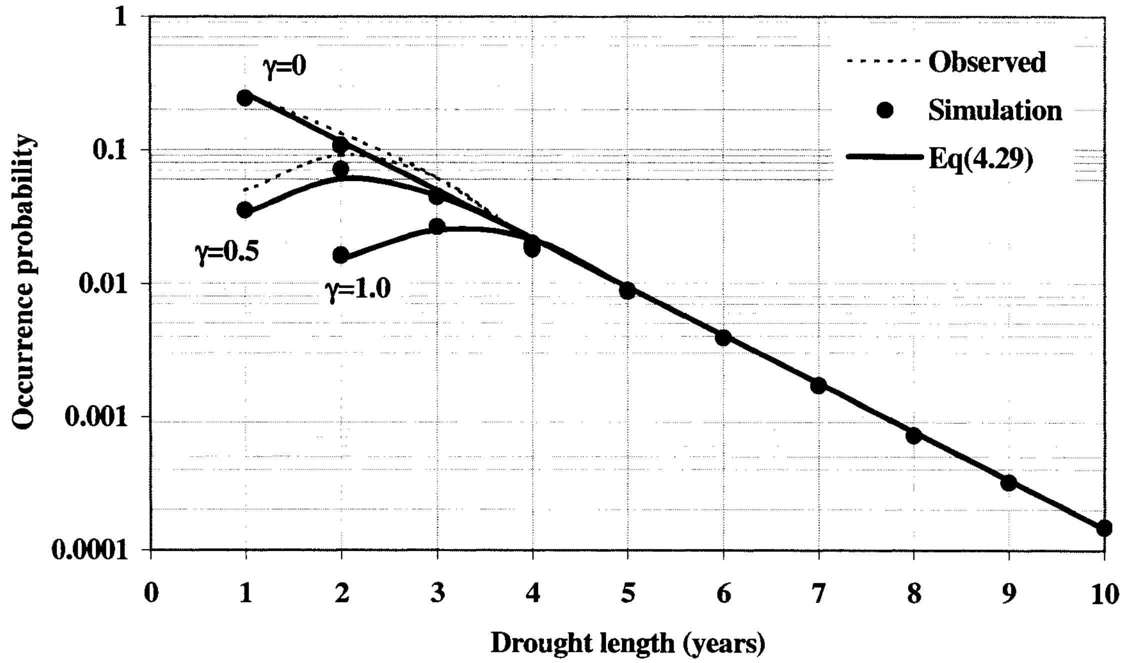

Fig 4.8 Occurrence probability of the multiyear regional drought $\{D' > D'_o \cap A_l^m\}$ arises from truncating the annual streamflow series of the 20 sites in the Upper Colorado River at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical results (4.29) versus the empirical results obtained from analyzing the 50000 years simulated flows and the observed flows.

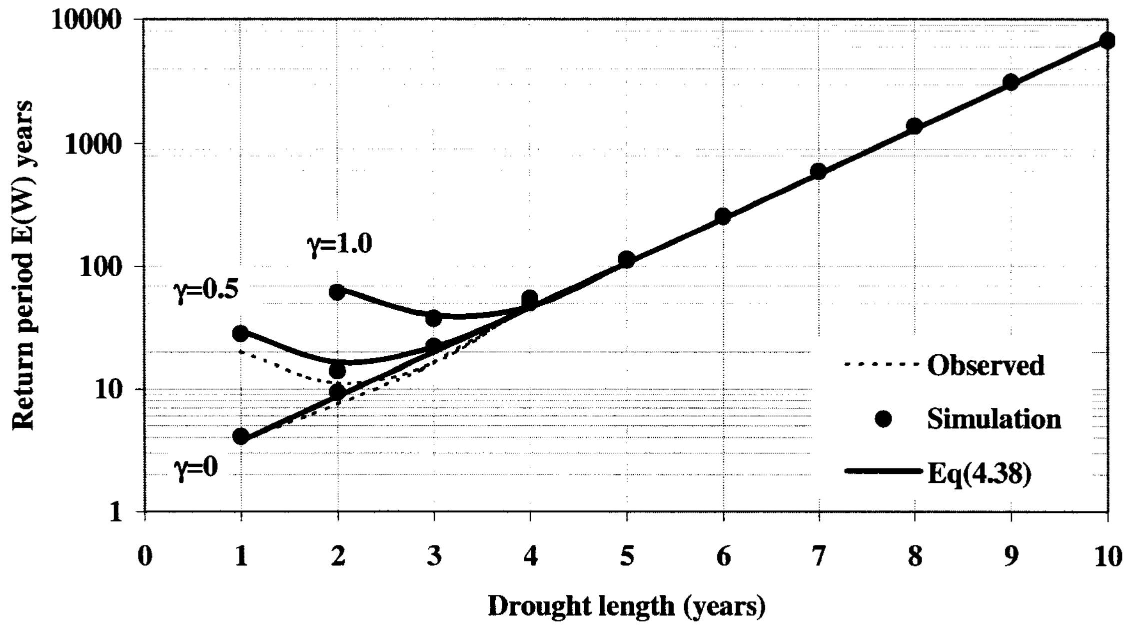

Fig 4.9 Return period of the multiyear regional drought $\{D' > D'_o \cap A_l^m\}$ arises from truncating the annual streamflow series of the 20 sites in the Upper Colorado River at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical results (4.38) of T as $E(W)$ versus the empirical results obtained from analyzing the 50000 years simulated flows and the observed flows.

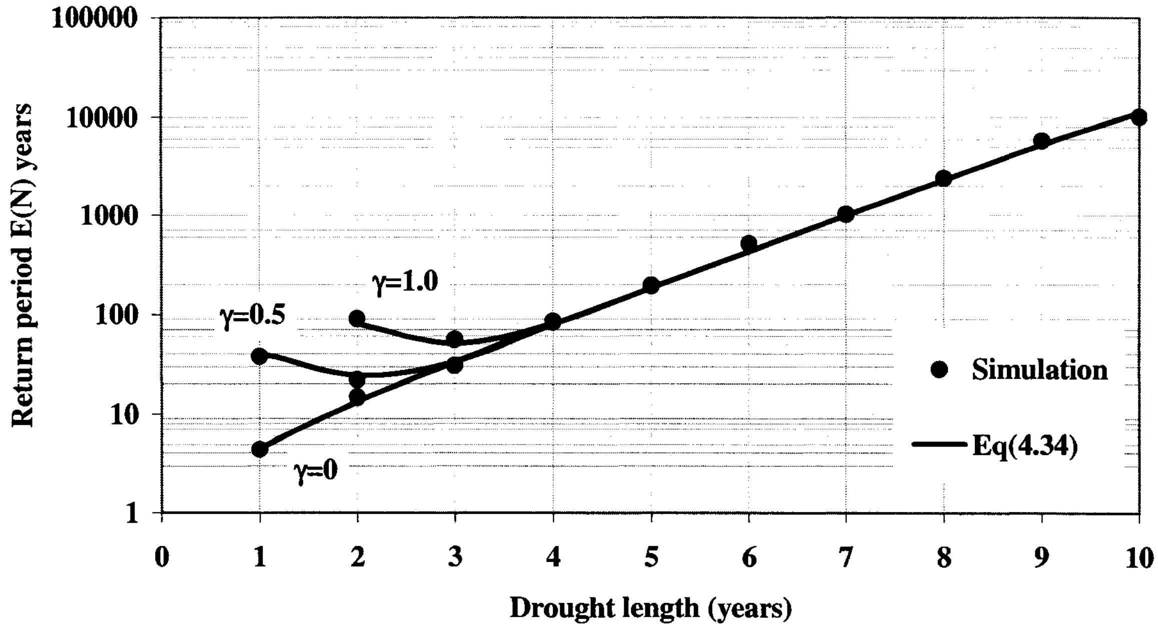

Fig 4.10 Return period of the multiyear regional drought $\{D' > D'_o \cap A_l^m\}$ arises from truncating the annual streamflow series of the 20 sites in the Colorado River at the long term mean. At γ values of 0, 0.5 and 1.0, the figure shows the analytical results (4.34) of T as $E(N)$ versus the empirical results obtained from analyzing the 50000 years of simulated flows.

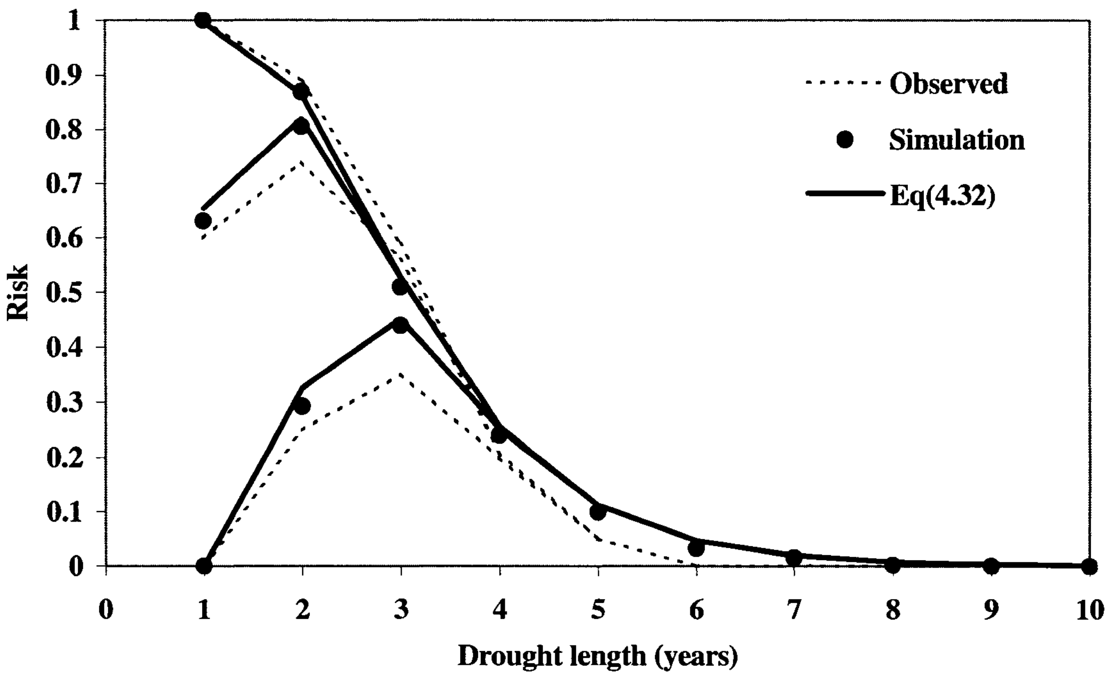

Fig 4.11 Risk of the multiyear regional drought $\{D' > D'_o \cap A_l^m\}$ arises from truncating the annual streamflow series of the 20 sites in the Colorado River at the long term mean occurs at least once within planning time of 25 years. At γ values of 0, 0.5 and 1.0, the figure shows the analytical result (4.32) versus the empirical results obtained from analyzing the 50000 years of simulated flows.

4.13 Appendix 4A

Considering the bivariate case $X_t^{(1)}$ and $X_t^{(2)}$, the process Z_t is defined as:

$$Z_t = 1 - [(1 - X_t^{(1)})(1 - X_t^{(2)})] \tag{4.46}$$

and equation (4.46) can be written also as:

$$Z_t = X_t^{(1)} + X_t^{(2)} - X_t^{(1)} X_t^{(2)} \tag{4.47}$$

Taking the expectation on both sides of (4.47), then:

$$E[Z_t] = E[X_t^{(1)}] + E[X_t^{(2)}] - E[X_t^{(1)} X_t^{(2)}] \tag{4.48}$$

Since, $X_t^{(1)}$ is 0,1 process, then the m moment is given by:

$$E[X_t^{(1)}] = \sum_{a \in \{0,1\}} a^m P[X_t^{(1)} = a] \tag{4.49}$$

from (4.49), $\quad E[X_t^{(1)}] = P[X_t^{(1)} = 1] \quad$ (4.50)

and $\quad V[X_t^{(1)}] = E[X_t^{(1)}]^2 - \left(E[X_t^{(1)}]\right)^2$

$$= P[X_t^{(1)} = 1] - \left(P[X_t^{(1)} = 1]\right)^2 = P[X_t^{(1)} = 1] P[X_t^{(1)} = 0] \tag{4.51}$$

Similarly, $\quad E[X_t^{(2)}] = P[X_t^{(2)} = 1] \quad$ (4.52)

and $\quad V[X_t^{(2)}] = P[X_t^{(2)} = 1] P[X_t^{(2)} = 0] \quad$ (4.53)

The expectation term $E[X_t^{(1)} X_t^{(2)}]$ of (3) can be written as:

$$E[X_t^{(1)} X_t^{(2)}] = E[X_t^{(1)}] E[X_t^{(2)}] + r_{X^1 X^2} \sqrt{V[X_t^{(1)}] V[X_t^{(2)}]} \tag{4.54}$$

replacing results of (5) - (9) in (3) and simplifying, then

$$E[Z_t] = 1 - P[X_t^{(1)} = 0] P[X_t^{(2)} = 0] - r_{X^1 X^2} \sqrt{P[X_t^{(1)} = 1] P[X_t^{(1)} = 0] P[X_t^{(2)} = 1] P[X_t^{(2)} = 0]}$$

(4.55)

4.14 Appendix 4B

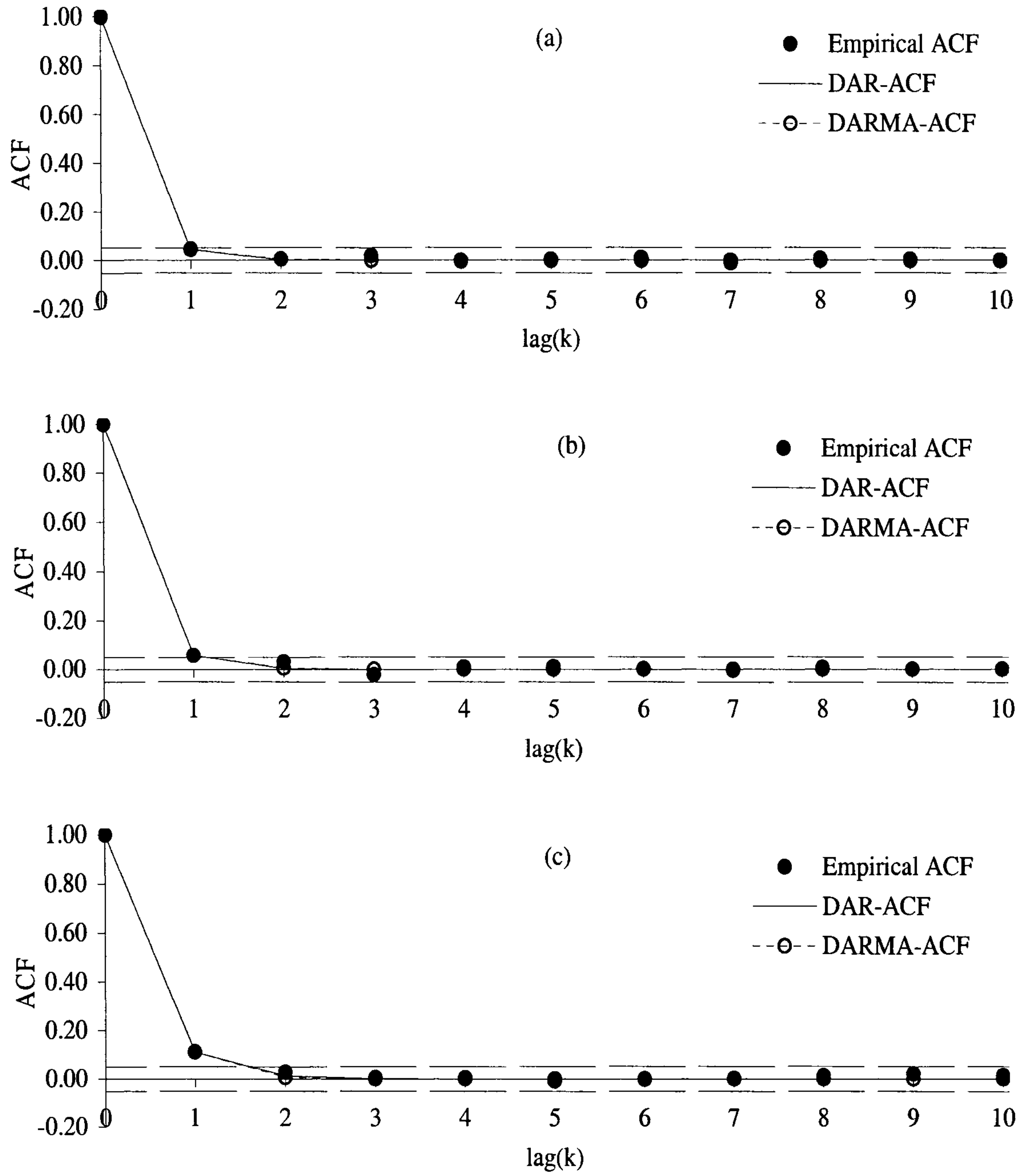

Fig. 4.1B The autocorrelation function (ACF) of the 0,1 process Z_t. Empirical results from simulation versus the theoretical as DAR(1) and DARMA(1,1) when the serial correlation in each of $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.1, (a) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.25, (b) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.5, and (c) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.75.

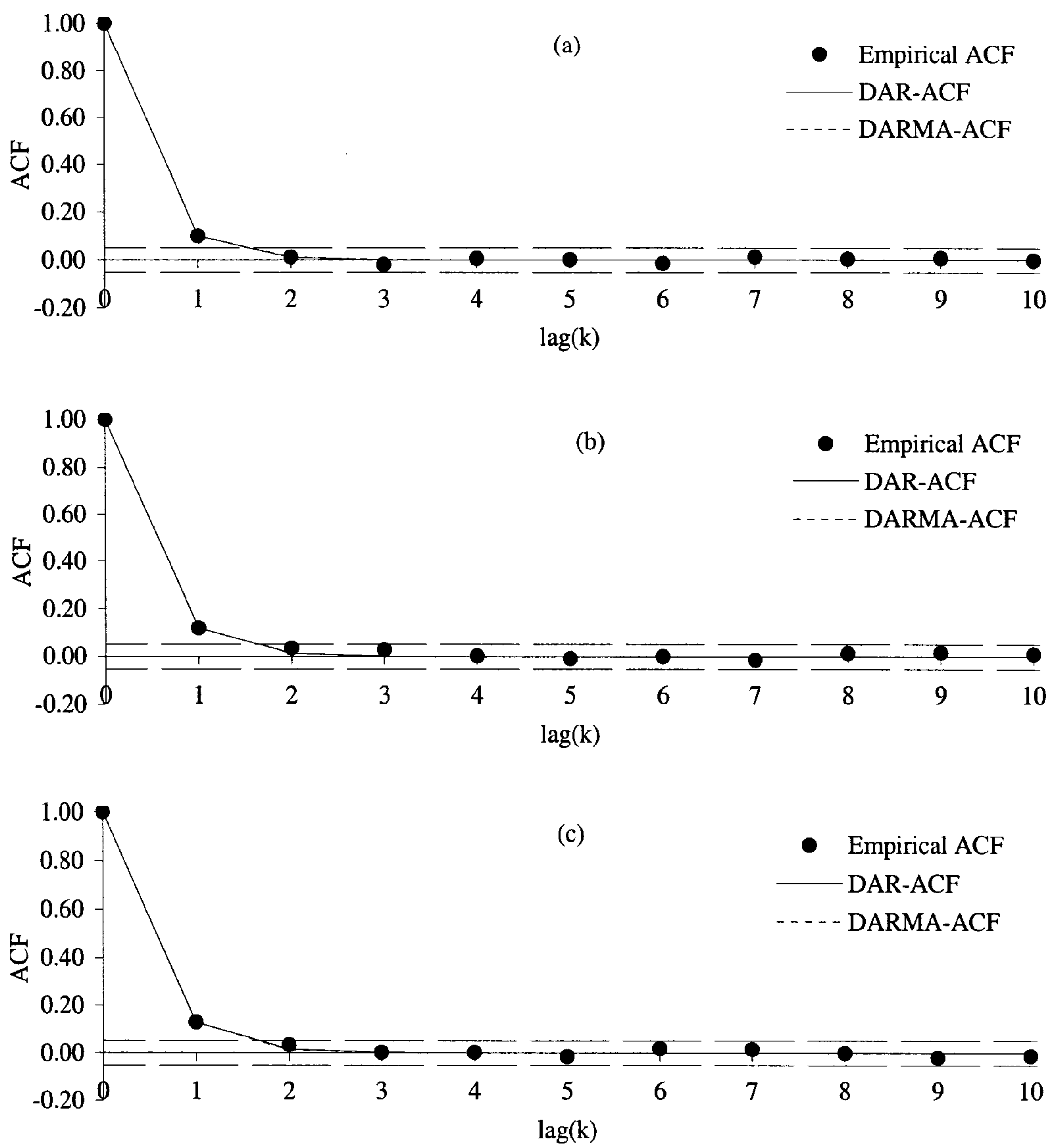

Fig. 4.2B The autocorrelation function (ACF) of the 0,1 process Z_t. Empirical results from simulation versus the theoretical as DAR(1) and DARMA(1,1) when the serial correlation in each of $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.25, (a) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.25, (b) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.5, and (c) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.75.

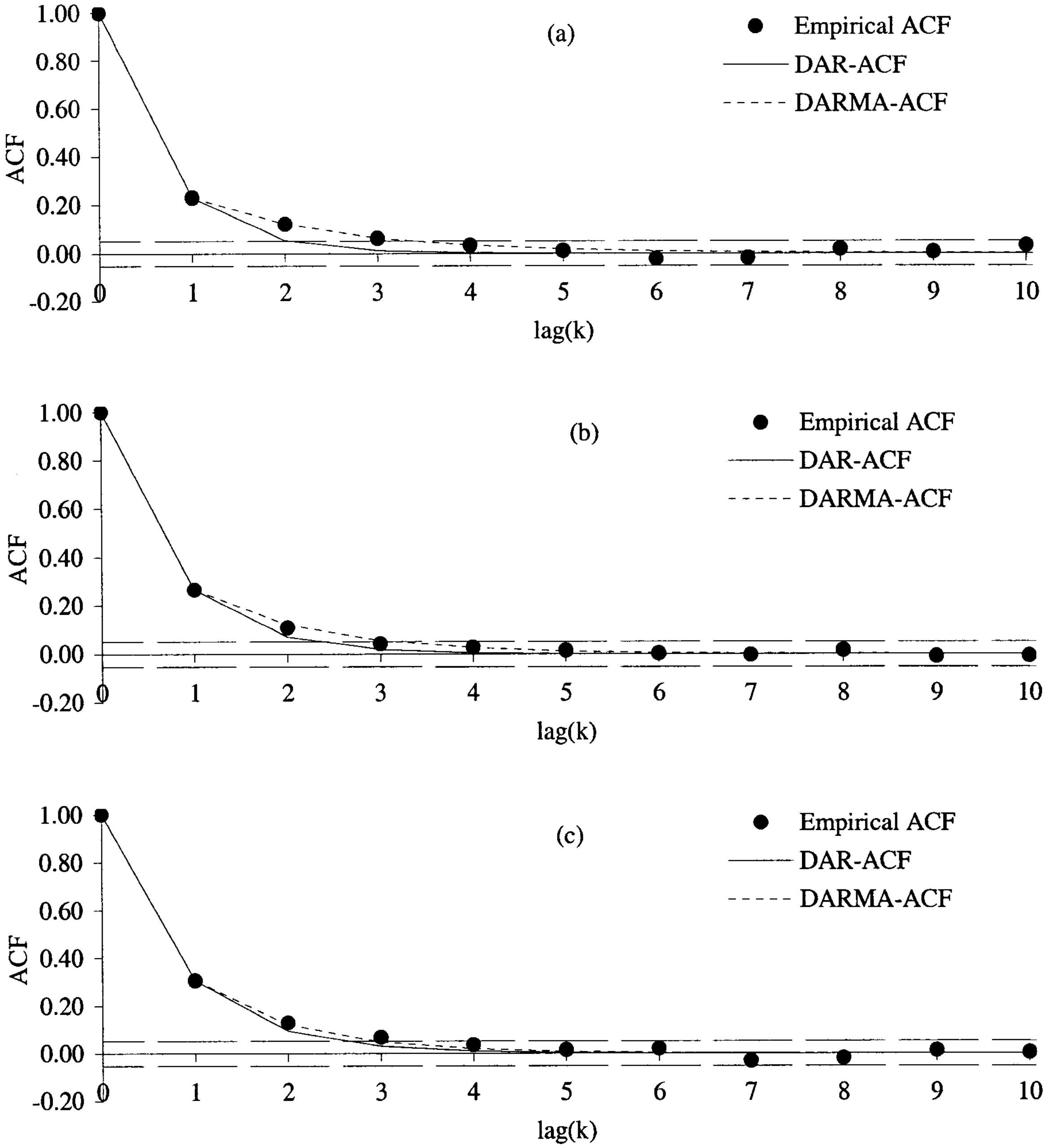

Fig. 4.3B The autocorrelation function (ACF) of the 0,1 process Z_t. Empirical results from simulation versus the theoretical as DAR(1) and DARMA(1,1) when the serial correlation in each of $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.5, (a) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.25, (b) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.5, and (c) the cross correlation between $Y_t^{(1)}$, $Y_t^{(2)}$, $Y_t^{(3)}$, and $Y_t^{(4)}$ is 0.75.

CHAPTER V

IMPROVING DROUGHT CHARACTERIZATION USING TREE-RING RECONSTRUCTED STREAMFLOWS

5.1 Introduction

Water resources operation policies and management plans of rivers rely most of the time on the current observed streamflows records. However, these streamflow records are limited in their length and can reflect only portion of the total of natural variability (Woodhouse, 2000). For example, the current length of the observed streamflow records of the Colorado River that is considered an important source of water supply for many users in the western United States extends to around 98 years coving the period 1906 – 2003. Ultimately, it can be urged that the short observed streamflow records may not provide a reliable picture of the total natural variability which affect the characterization of extreme hydrologic events like droughts that may last for many years, therefore it is extremely important to reliably lengthen the short observed streamflow records to a length much longer than the length of the current observed records (Salas et, al 2006). Using the observed records, stochastic simulation models can be built to provide reliable long records of synthetic streamflows that can be used to characterize extreme events like droughts (Salas et al, 2005).

Alternatively, one may use reliable record extension models with tree ring indices as the source of information over the extension period to extend short observed records back in time for long enough to capture more extreme drought events than what can be observed from the historical records (e.g. Woodhouse, 2001; Gonzales and Valdes, 2003; Salas et al, 2006). Trees are useful for reconstructing streamflows because they are sensitive recorders to the natural climate variability. In addition to its accuracy to date to the year, usually trees assimilate into their growth rings the same set of climatic factors like precipitation and evapotranspiration that affect the amount of the streamflows (Meko et al, 1995).

Usually procedures for extending short variable records rely on the information that is contained in longer records, e.g. streamflows at other sites, precipitation, or proxy data such as tree ring indices. The quality of the information gain is related proportionally to the cross correlation between the variable to be extended and the variable(s) that will be used in the extension (Matalas and Jacobs, 1964). Generally, when streamflow records are short, the estimated parameters from such records may be poor estimates of the population parameters (Fiering, 1963). Therefore one motivation behind extending short streamflow records at one site or many sites simultaneously is to improve the reliability of parameters estimates. The reliability of the parameter estimator is measured by the variance of that estimator where the lower the variance of the estimator the better the corresponding parameter estimate represents the population parameter. In general, the objectives behind extending short records are either to improve estimates of the population parameters or maintain parameters estimates of the short record.

Several extension models have been employed in literature to reconstruct streamflows back in time using tree-ring data as source of information over the extension period. For example, the traditional multiple linear regression models is widely used to reconstruct streamflows at one site for several decades in the past using tree-ring indices as predictors (e.g. Cleaveland, 2000; Hidalgo et al, 2000; Meko et al, 2001; Woodhouse, 2001; Gedalof et al, 2004). Except for the studies by Gonzalez and Valdes (2003) and Salas et al (2006) generally the traditional multiple regression models that are used in literature for reconstructing streamflows from tree ring indices do not incorporate the previous values of the streamflows, i.e. the serial correlation of the streamflow variable is not considered. Moreover, except for the studies by Meko et al (2001), Gonzalez and Valdes (2003), and Salas et al (2006), the traditional multiple regression models that are used in literature do not include the random term that is intended to capture the variability of the observed flows. However, record extension models that include the random term give multiple record extensions of the streamflow variable due to the random generation of the noise term. This is the procedure utilized by Meko et al (2001) and Gonzalez and Valdes (2003) for examining possible flow and drought scenarios that may have occurred in the past.

Practically among the multiple record extensions that are produced due to the noise addition, one may simply select one of the multiple record extensions. Yet another alternative may be to select a single record extension that meets certain statistical criterion as advocated by Salas et al (2006). In the latter study a criterion was developed to select a single record extension so that key statistics, such as the mean and the standard deviation estimated for the short record to be extended are maintained. This procedure is further developed in this study so as to include multiple sites and multiple variables.

Generally, the foregoing streamflow extension models consider the case of extending streamflow at one site using tree-ring. In this chapter, a multivariate version of the one site record extension model called REXTN (Salas et al, 2006) is developed to extend the streamflows at many sites. In order to test the applicability of the proposed model its performance is evaluated and compared with alternative models based on simulation experiments. The evaluation is based on a number of key statistics that include the mean, standard deviation, serial correlation, and correlations between sites. The extension model with a spatial disaggregation schemes will be used to simultaneously extend the streamflows at some sites within the Colorado River using tree ring indices. In addition, the reconstructed streamflows are utilized to extract drought statistics and drought scenarios that may have occurred in the past centuries in the Colorado River basin. These drought statistics and scenarios are compared with those obtained from the short historical records.

5.2 Multi-Site Record Extension Model with Noise (MREXTN)

The record extension model with the noise term (REXTN) given by (2.62) can be used to extend the streamflows at one site (p = 1) using one or several predictor variables $\boldsymbol{x}_t$ ($m \geq 1$). The record extension model (REXTN) can be generalized to consider the extension of many short variables $\boldsymbol{y}_t$ ($p \geq 1$) using one or many predictor variables $\boldsymbol{x}_t$ ($m \geq 1$). The multivariate record extension model MREXTN is given by:

$$\hat{\boldsymbol{y}}_t = \overline{\boldsymbol{y}}_1 + \boldsymbol{A}(\boldsymbol{y}_{t-1} - \overline{\boldsymbol{y}}_1) + \boldsymbol{B}(\boldsymbol{x}_t - \overline{\boldsymbol{x}}_2) + \boldsymbol{C}\boldsymbol{\varepsilon}_t \qquad (5.1)$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are the model parameters, and the noise $\boldsymbol{\varepsilon}_t$ is multivariate normally distributed vector of p terms that are serially independent with mean vector of elements = 0 and variance-covariance matrix with diagonal elements = 1 and 0 for off diagonals, i.e. spatially independent. The parameters $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are considered the general case for the parameters of the model REXTN given by (2.63) - (2.65). The estimates of the parameters $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ can be obtained by generalizing the results in (2.63) - (2.65) as follows:

$$\hat{\boldsymbol{A}} = [\boldsymbol{S}_{y1y1}(1) - \boldsymbol{S}_{y1x1}\boldsymbol{S}_{x1x1}^{-1}\boldsymbol{S}_{x1y1}(1)][\boldsymbol{S}_{y1y1} - \boldsymbol{S}_{x1y1}^{T}(1)\boldsymbol{S}_{x1x1}^{-1}\boldsymbol{S}_{x1y1}(1)]^{-1} \tag{5.2}$$

$$\hat{\boldsymbol{B}} = [\boldsymbol{S}_{y1x1} - \hat{\boldsymbol{A}}\,\boldsymbol{S}_{x1y1}^{T}(1)]\,\boldsymbol{S}_{x1x1}^{-1} \tag{5.3}$$

and

$$\hat{\boldsymbol{C}}\hat{\boldsymbol{C}}^{T} = \boldsymbol{S}_{y1y1}(1) - \hat{\boldsymbol{A}}\,\boldsymbol{S}_{y1y1}^{T}(1) - \hat{\boldsymbol{B}}\,\boldsymbol{S}_{x1y1} \tag{5.4}$$

The proposed model MREXTN may result in many simultaneous possible extensions of the short records $\boldsymbol{y}_t$ due to the random generation of the noise term $\boldsymbol{\varepsilon}_t$. However a modified procedure that is based on the procedure described by Salas et al (2006) for selecting a single realization among many possible realizations of $\hat{\boldsymbol{y}}_t$ is presented here by considering an additional weight a_{ij}. The weight a_{ij} is defined as the proportion of the explained variance of the variable $y_t^{(i)}$ due to the contribution of the variable $x_t^{(j)}$ alone relative to the total explained variance of $y_t^{(i)}$ due to the contribution of all $\boldsymbol{x}_t$ variables. If the proportion of the explained variance of the variable $y_t^{(i)}$ due to the contribution of the variable $x_t^{(j)}$ alone is defined by the $R^2_{y^{(i)}x^{(j)}}$ which is the coefficient of the determination resulted by regressing $y_t^{(i)}$ on $x_t^{(j)}$ alone using (2.55), and if the total explained variance of $y_t^{(i)}$ due to the contribution of all $\boldsymbol{x}_t$ variables is defined by the $R^2_{y^{(i)}\boldsymbol{x}}$ which is the coefficient of the determination resulted by regressing $y_t^{(i)}$ on $\boldsymbol{x}_t$ using (2.55), then the weight a_{ij}, for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, m$, is given as:

$$a_{ij} = \frac{R^2_{y^{(i)}x^{(j)}}}{R^2_{y^{(i)}x}} \tag{5.5}$$

In Salas et al (2006), the criteria was to select a single noise sequence such that the equalities $\frac{\bar{\hat{y}}_2}{\bar{y}_1} = \frac{\bar{x}_2}{\bar{x}_1}$ and $\frac{s_{\hat{y}2}}{s_{y1}} = \frac{s_{x2}}{s_{x1}}$ hold or practically the quantity z defined in (2.66) is minimum given that the extension model is used to extend one short variable y_t. Similarly, one may redefine the quantity z to incorporate the weight a_{ij} that consider the part of $y_t^{(i)}$ variance explained by the contribution of each $x_t^{(j)}$ for $j = 1, 2, \ldots, m$, noting that there are several short variables $\boldsymbol{y}_t = [y_t^{(1)} \;\; y_t^{(2)} \cdots y_t^{(p)}]^T$. The quantity z is given as:

$$z = 0.5\left[\sum_{i=1}^{p}\left|\left(\frac{\bar{\hat{y}}_2^{(i)}}{\bar{y}_1^{(i)}} - \sum_{j=1}^{m} a_{ij}\frac{\bar{x}_2^{(j)}}{\bar{x}_1^{(j)}}\right)\right|\right] + 0.5\left[\sum_{i=1}^{p}\left|\left(\frac{s_{\hat{y}_2}^{(i)}}{s_{y_1}^{(i)}} - \sum_{j=1}^{m} a_{ij}\frac{s_{x_2}^{(j)}}{s_{x_1}^{(j)}}\right)\right|\right] \tag{5.6}$$

where |A| stands for the absolute value of A. In this case, the criterion is to select among many possible extensions of $\hat{\boldsymbol{y}}_t$ the unique extension that gives streamflows statistics $\bar{\hat{y}}_2^{(i)}$ and $s_{\hat{y}_2}^{(i)}$ over the period N_2 for $i = 1, 2, \ldots, p$ such that the quantity z is minimum.

Currently for the purpose of drought characterization, the streamflows at many sites, for example the Colorado River, can be extended using one of the following multi-site record extension models: the multivariate GMOVE, the multivariate linear regression model with least squares (MLRLS), or the proposed multivariate record extension model with the noise (MREXTN). However, selecting one extension model that may result in a plausible streamflows is a crucial issue, i.e. it is required to extend the streamflows at many sites within the Colorado River for example such that the majority of the key statistics of the observed flows are maintained. In this case, the selection criterion of the extension model that will be used to extend the streamflows of the Colorado River will be based on the performance of each of the competing models mentioned previously.

5.3 The Performance of the Record Extension Model MRXETN

The performance of the model MREXTN (5.1) relative to other existing record extension models like the MLRLS (2.55) and the multivariate GMOVE (2.67) can be assessed in terms of reproducing some specified key (target) statistics using simulation experiments. It should be noted that in our case the key statistics to be reproduced are the statistics of the short record. For the purpose of the simulation, multivariate normal variables $y_t^{(1)}$, $y_t^{(2)}$ and x_t were generated considering that $y_t^{(1)}$ and $y_t^{(2)}$ are the short records while x_t is the longer record. The variable $y_t^{(1)}$ was assumed to have population mean, standard deviation, and lag-1 serial correlation coefficient as: $\mu_{y^{(1)}}$ = 100, $\sigma_{y^{(1)}}$ = 35, and $\rho_1(y^{(1)})$= 0.15, respectively, and the variable $y_t^{(2)}$ was assumed to have population mean, standard deviation, and lag-1 serial correlation coefficient as: $\mu_{y^{(2)}}$ = 200, $\sigma_{y^{(2)}}$ = 70, and $\rho_1(y^{(2)})$ =0.2, respectively, while the variable x_t was assumed to have population mean, standard deviation, and lag-1 serial correlation coefficient as: μ_x = 50, σ_x = 20, and $\rho_1(x)$ = 0.3, respectively.

In addition, the population cross-correlation coefficient $\rho_{y_t^{(1)}y_t^{(2)}}$ between $y_t^{(1)}$ and $y_t^{(2)}$ was taken fixed at 0.8. For simplification, it was assumed that the population cross-correlation coefficient $\rho_{x_t y_t^{(1)}} = \rho_{x_t y_t^{(2)}}$ and that both takes the values 0.5, 0.7, and 0.9. Furthermore, the length N_1 of the short record y_t was fixed at 50, while N_2 varied as 50, 100, and 250. The Multivariate data generation was made for variables $y_t^{(1)}$, $y_t^{(2)}$ and x_t for the total record length $N_1 + N_2$. The short and the long records were utilized for estimating the parameters of a given record extension model. The generated records for $y_t^{(1)}$ and $y_t^{(2)}$ over the period N_2 were subsequently utilized to calculate the root mean square error (RMSE) of the extended records. For the referred experiment a total of 1000 multivariate samples were generated for each combination of the parameters $\rho_{x_t y_t^{(1)}}$ and N_2.

Using the different record extension models, the short records $y_t^{(1)}$ and $y_t^{(2)}$ have been extended simultaneously over the length N_2. The performance of the different record extension models has been assessed in terms of reproducing some specific target statistics such as the mean and the standard deviation, the lag-1 serial correlation coefficient of the short record, and the cross correlation between the concurrent short-short and short-long records. Let the whole combined record of the variable $\tilde{y}_t^{(i)}$ be defined as:

$$\tilde{y}_t^{(i)} = \begin{cases} y_t^{(i)} & \text{for } t = 1, 2, \ldots, N_1 \\ \hat{y}_t^{(i)} & \text{for } t = N_1 + 1, \ldots\ldots, N_1 + N_2 \end{cases} \tag{5.7}$$

where $y_1^{(i)}, \ldots, y_{N_1}^{(i)}$ is the generated short record and $\hat{y}_{N_1+1}^{(i)}, \ldots, \hat{y}_{N_1+N_2}^{(i)}$ is the extended record for site(i). The performances of the different record extension models were made comparing the mean of the ratios $\bar{\tilde{y}}^{(i)} / \bar{y}_1^{(i)}$, $S^2_{\tilde{y}^{(i)}} / S^2_{y1^{(i)}}$, $r_1(\tilde{y}^{(i)}) / r_1(y1^{(i)})$, $r_{\tilde{y}^{(1)}\tilde{y}^{(2)}} / r_{y1^{(1)}y1^{(2)}}$ and $r_{\tilde{y}^{(i)}x} / r_{x1\,y1^{(i)}}$ which were determined from the 1000 simulated extensions as obtained from the different record extension models. The statistics $\bar{\tilde{y}}^{(i)}$,

$S^2_{\tilde{y}^{(i)}}$, and $r_1(\tilde{y}^{(i)})$ are the mean, variance, and lag-1 serial correlation coefficient of $\tilde{y}_t^{(i)}$, respectively, $r_1(y1^{(i)})$ is the lag-1 serial correlation coefficient of the short record $y_t^{(i)}$ of length N_1, $r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}$ is the cross correlation coefficient between $\tilde{y}_t^{(1)}$ and $\tilde{y}_t^{(2)}$ over the whole length $N_1 + N_2$, while $r_{y1^{(1)}\,y1^{(2)}}$ is the cross correlation coefficient of the short records $y_t^{(1)}$ and $y_t^{(2)}$ over the length N_1, and $r_{x\tilde{y}^{(i)}}$ is the cross correlation coefficient between x_t and $\tilde{y}_t^{(i)}$ over the whole length $N_1 + N_2$, while $r_{x1\,y1^{(i)}}$ is the cross correlation coefficient of the short record $y_t^{(i)}$ and x_t of length N_1. The statistics $\bar{y}_1^{(i)}$ and $S^2_{y1^{(i)}}$ are the target (specified) mean and variance of $y_t^{(i)}$ over the length N_1.

Since the motivation behind the extension of the short record $y_t^{(i)}$ is that the specific target statistics of $y_t^{(i)}$ are either maintained or improved and to preserve the lag-1 serial correlation of $y_t^{(i)}$ and the cross correlation between x_t and each of $y_t^{(i)}$, and between variables $y_t^{(i)}$ themselves for i = 1, 2, 3, ..., then the mean ratios as described above should have the value of 1. Therefore, the best model among the different models is the one that results in mean values close to 1. The results of the comparison are shown in Tables 5.1 – 5.3, while Table 5.4 shows the results of the same experiment shown in Table 5.1 except that $\rho_1(x) = 0$. Furthermore, an additional performance measure of the record extension models has been added by calculating the RMSE of the extended (predicted) series $\hat{y}_t^{(i)}, t = N_1 + 1, \ldots, N_1 + N_2$ relative to the corresponding generated series. Thus for each extended record and corresponding generated sample the RMSE is computed as

$$RMSE = \sqrt{\frac{1}{N_2}\sum_{t=1}^{N_2}(y_t - \hat{y}_t)^2} \tag{5.8}$$

where y_t denotes the generated series and $\hat{y}_t$ the extended series for the period N_2. Thus, based on 1000 generated samples and extended series the mean value of RMSE was determined, i.e. $m(RMSE)$.

The results of the comparison of the performance of the various models based on the referred simulation experiments are shown in Tables 5.1 – 5.4. First of all, in relation to the target mean, it can be noticed that all models performed quite well across the range of values tested for $\rho_{x_t y_t^{(i)}}$ and N_2. Regarding the target variance, all the models perform adequately for $\rho_{x_t y_t^{(i)}} = 0.9$. However, for lower values of $\rho_{x_t y_t^{(i)}}$ the MLRLS model underestimate the variance significantly noting that the underestimation in the variance increases as the N_2 increases. Also as N_2 increases the performance of the multivariate GMOVE appears to deteriorate regardless of the value of $\rho_{x_t y_t^{(i)}}$, while the MREXTN keeps performing well regardless of $\rho_{x_t y_t^{(i)}}$ and N_2. Moreover, the performance of model REXTN compared to all other models stands out regarding the lag-1 serial correlation.

In all cases, and regardless of the values of $\rho_{x_t y_t^{(i)}}$ and N_2, the values of $m[r_1(\tilde{y}^{(i)})/r_1(y1^{(i)})]$ for the MREXTN are significantly closer to 1 than those of the other models. Considering the cross-correlation, the performance of the multivariate GMOVE and MREXTN are quite good and comparable and they are significantly better than those for the MLRLS. Nevertheless the performances of all models improve significantly as $\rho_{x_t y_t^{(i)}}$ increases, in fact the values of $m[r_{\tilde{y}^{(1)}\tilde{y}^{(2)}} / r_{y1^{(1)}y1^{(2)}}]$ and $m[r_{\tilde{y}^{(i)}x} / r_{x1\,y1^{(i)}}]$ are close to

1 for the multivariate GMOVE and MREXTN when $\rho_{xy} = 0.9$. Finally, the comparison of the values of $m(RMSE)$ reveals that the model MREXTN performs somewhat better than the multivariate GMOVE and very comparable to the MLRLS model. In all cases and for all models $m(RMSE)$ becomes smaller as $\rho_{x_t y_t^{(i)}}$ increases which is expected in terms of the amount of the information that can be transformed from the long record to the short one. Summarizing, considering all the performance measures utilized herein, the referred simulation experiments have shown clearly the superiority of model MREXTN for extending short records which is similar to the findings of other studies that considered the bivariate case (e.g. Salas et al, 2006).

5.4 Reconstructed Streamflows of the Colorado River

The observed streamflow records for the Colorado River (Figure 4.2) span over the period 1906 – 2003 which are considered relatively short for the purpose of characterizing extreme hydrologic events like droughts. Based on the results obtained from the analysis of model performance using simulation experiments, the model MREXTN (5.1) will be used to reconstruct the streamflows at some sites within the Colorado River for the sake of improving drought characterization. Specifically, the annual cumulative streamflows at sites 8, 16, 20, and 29 within the Colorado River Basin (Figure 4.2) will be reconstructed using the extension model MREXTN (5.1) and the extension model MLRLS (2.55) with the tree ring indices as information source. Initially, the tree-ring data set that may be used to extend the annual cumulative streamflows at sites 8, 16, 20, and 29 of the Colorado River consists of a pool of annual tree ring data series (chronology residuals) at 55 different tree sites within the region of the Colorado Front Range (Figure 5.2). However, due to the different starting and ending times of the different tree-ring series, only the tree-ring data at 17 different tree sites that covers the period 1490-2002 were used in this study as a pool of possible predictors to extend the streamflows at sites 8, 16, 20, and 29 of the Colorado River. The tree-ring indices data that has been used in the extension was obtained from the Paleoclimatology Branch - NOAA - National Climatic Data Center, Boulder, Colorado.

Firstly, cross-correlation analysis was conducted between the annual streamflows at the selected sites within the Colorado River and the tree-ring indices at the 17 tree sites for the purpose of assessing the strength of the association between the independent variables (streamflows) and the predictors (Tree-ring indices). Moreover, because there may be a significant association between climate in a given year and tree growth in the preceding year and up to the 3 following years due to the biological lags in the tree response (e.g. Earle, 1993), cross-correlation analysis between the annual streamflows at the selected sites within the Colorado River and tree-ring indices at the 17 tree sites including lags of –2, -1, 0, +1, +2, and +3 years of tree-ring indices relative to the given streamflow year was conducted. The result of the cross correlation analysis shows that the annual streamflow at the selected sites within the Colorado River and tree-ring indices are significantly correlated at zero lag only, whereas the correlations at lags different than zero are not significant.

In addition, to avoid over fitting problems in the record extension models due to including many predictors (tree-ring variables), the minimum number of predictors that can explain significant part of the total variability of the streamflows at each site were selected based on stepwise regression analysis. With the proper power transformation, the observed annual streamflows (1906 – 2002) at sites 8, 16, 20, and 29 were first transformed to the normal domain to be consistent with the tree-ring data that is approximately normally distributed. The stepwise regression analysis was applied considering the model (2.55) by regressing the annual transformed streamflows at each site separately on the 17 explanatory variables (tree ring indices). Only tree-ring variables with slope coefficient b that is significantly different from 0 at 5% significance level were selected as possible predictors for the streamflows at that specific site. The analysis shows that only 4 tree-ring indices (predictors) at tree sites indicated by the names PUM, ENC, SLK, and LAN (Figure 5.2) are needed to construct an efficient record extension model to extend the streamflows at the selected four sites within the Colorado River. The total variance explained of the transformed streamflows by the 4 predictors was ranging between 66% for flows at site16 to 76% for flows at site 29 (Table 5.5a). Generally, useful variance explained values typically range from about 40% - 70% (Woodhouse 2000), therefore the 4 tree-ring indices were subsequently used as predictors to simultaneously reconstruct streamflows at the referred sites within the Colorado River using the record extension model MLRLS and the model MREXTN.

To further assess the performance and the skills of the models MLRLS and MREXTN to reconstruct flows at the different sites over the period of the reconstruction, each model was first calibrated and then validated by calculating the appropriate validation statistics in the actual flow domain (i.e. modeling using power transformed flows, extending records in the transformed domain, and then inverting the flows back to the original domain). The calibration and validation was performed using the split sample procedure, i.e. both models was first calibrated over the period 1954 – 2002 and then validated over 1906 – 1953, and another time both models were calibrated over the period 1906 – 1953 and then validated over the period 1954 – 2002 so that estimates of flows during the validation is available for the period 1906 - 2002. The validation statistic that was used to assess the performance of both models is the reduction error *RE* as defined in (2.77) which measures the skill of the model. Table (5.5b) shows the validation statistics *RE* obtained from the foregoing analysis using both models.

Fritts et al (1990) stated that positive values of the *RE* statistic that are close to the value of R^2 are an indication for extension model with good skills to reconstruct streamflows using proxy data. Generally as shown by the results in Table (5.5b) the validation statistics *RE* obtained are comparable for both models, MLRLS and MREXTN, which indicate a reasonable good skill for both models to reconstruct streamflow over the period where observed flows are absent. As an additional validation, one may consider the performance of both models to produce streamflow generates over the reconstruction period so that the key statistics of the observed flows like, the mean, standard deviation, skewness coefficient, and the lag-1 serial correlation coefficient are reproduced. For example, Tables 5.6a and 5.6b show generally that both models produces streamflow reconstructions so that the mean of the short observed record is reproduced, however, in terms of other statistics the model MREXTN performs much well in terms of reproducing the specific key statistics of the Colorado River observed flows.

With the proper power transformation, the original streamflows 1906 - 2002 were transformed to the normal domain and the parameters of the extension model MREXTN and MLRLS have been estimated. Over the period 1490 – 1905 when the observed flows are not available, the extension models MREXTN and MLRLS were used to simultaneously reconstruct the transformed flows (normal domain) at sites 8, 16, 20 and 29 with the 4 tree-ring indices used as information source over the reconstruction period. The reconstructed transformed flows were inverted to its original domain by the back transformation process and after that whenever needed streamflows were adjusted for the spatial proportionality (2.76).

The combination of the observed flows 1906 – 2002 and the reconstructed flows 1490 – 1905 forms the combined flows 1490 – 2002 at each flow site. Figure 5.2 shows the combined flow series for the period 1490 – 2002 at each site obtained using the observed flows and the reconstructed flows resulted from the extension models MREXTN and MLRLS. The basic statistics, i.e. the mean, standard deviation, skewness coefficient, and the lag-1 serial correlation of the observed flows (1906-2002), the combined annual flows based on MREXTN (1490-2002) and the combined annual flows based on MLRLS (1490-2002) for sites 8, 16, 20, and 29 within the Colorado River are shown in Table 5.6a while the cross correlation between flows at sites 8, 16, 20, and 29 using the observed flows (1906-2002), the combined annual flows based on MREXTN (1490-2002), and the combined annual flows based on MLRLS (1490-2002) are shown in Table 5.6b. Generally it can be concluded from the results in Tables 5.6a and 5.6b that the MREXTN model resulted in reconstructed streamflows and later combined flows that reproduce the basic key statistics of the short records more closely than those obtained from the combined flows of the model MLRLS. Specifically, the annual cumulative streamflows produced by the model MLRLS underestimate the mean and the standard deviation, which is expected due to the absence of the stochastic term (noise).

Furthermore, the lag-1 serial correlation is not reproduced in the annual flows obtained using MLRLS compared to the annual flows obtained using MREXTN, which is expected because the model MREXTN includes a term that accounts for the previous flows while the model MLRLS does not.

5.5 Characterizing Droughts of the Colorado River

Although the observed annual streamflow time series may indicate a number of drought events, however the number of these events is still insufficient to estimate drought properties accurately. Therefore one may use stochastic models or record extension models with tree-ring data as information source to provide longer database from which one can obtain many more drought events and drought scenarios (Salas et al, 2006). In this study, the reconstructed annual streamflows of the Colorado River will be used to improve the characterization of droughts within the Colorado River Basin. Two types of analysis of severe droughts of the Colorado River are described in this section. The first one was made to characterize severe droughts based on single site analysis considering the observed annual flows, the combined annual flows based on the model MREXTN, and the combined annual flows based on the model MLRLS for sites 8, 16, 20 and 29 (Table 5.7) where site20 (Lees Ferry gaging station) was selected for detailed analysis. In this study, drought events that are obtained by truncating the single site streamflow series are defined based on the theory of runs as in Yevjevich (1967), i.e. the first definition of drought events as shown in Chapter2. Generally, considering the single site analysis of droughts at each of the sites 8, 16, 20 and 29, then Table 5.7 shows some statistics of the detected drought events regarding the duration ($L \geq 1$ year) and the magnitude as obtained from the analysis of the observed annual cumulative flows, combined annual cumulative flows using MLRLS model, and the combined annual cumulative flows using MREXTN model for sites 8, 16, 20 and 29 of the Colorado River.

Comparing the basic statistics such as the mean, standard deviation of drought duration and magnitude it can be noticed generally that the statistics obtained from the analysis of the combined flows using the model MREXTN reflect the occurrence of more severe droughts in terms of duration and magnitude compared to what can be obtained by analyzing the observed streamflows only. For example, considering the drought statistics obtained by analyzing the flows at site 20, then one observes an average drought duration of 2.41 years and 2 years for droughts obtained from the analysis of the combined flows using the model MREXTN and the observed flows respectively, and an average drought magnitude of 8.49 MAF and 7.32 MAF for combined flows using the model MREXTN and the observed flows respectively.

Similar conclusion can be obtained by comparing the standard deviations of drought duration and magnitude when considering droughts resulted from truncating the combined flows using the model MREXTN and the observed flows. Furthermore, the longest observed drought has duration of 5 years while the longest drought detected by analyzing the combined flows obtained from the model MREXTN has duration of 9 years. Moreover, the magnitude of the most severe observed drought (21.56 MAF) is much less than the magnitude of the most severe drought (30.44 MAF) as been detected by analyzing the combined flows using the model MREXTN. Similar results can be obtained by comparing the drought statistics at other sites (e.g. site 8, 16, and 29) as shown in Table 5.7. Generally, one would expect such results simply because the short observed records are insufficient to estimate drought properties accurately (Salas et al 2006), i.e. the short observed flow records are incapable to capture the whole natural flow variability and therefore characterizing droughts from these short records would give underestimated values of the actual drought properties for that region. On the other hand, referring to Table 5.7, except for the case of the longest drought and the most severe drought magnitude, the analysis of the combined flows using the model MLRLS does not show much improvement in terms of the average drought duration and the average drought magnitude over what is available from the analysis of the observed flows only (1906-2002). For example, considering site20 again, then one observes an average drought duration of 2.08 years obtained from the analysis of the combined flows using the model MLRLS which is just a little higher than the average drought duration of 2 years obtained from the analysis of the observed flows only (in some cases like site 16 and 29 the case is even worse), and an average drought magnitude of 7.06 MAF and 7.32 MAF for the combined flows based on the model MLRLS and the observed flows respectively.

Similar conclusion can be obtained by comparing the standard deviations of drought duration and magnitude when considering droughts resulted from truncating the combined flows using MLRLS model and the observed flows. Generally it can be noticed that there is an underestimation in the mean and standard deviation of both drought duration and magnitude obtained from analysis of the combined flows using MLRLS model compared to what is available from the analysis of the observed flows only noting that the combined flows using MLRLS is 5 times longer than the observed flows record. Specifically, that underestimation is clear for the case of drought magnitude, for example considering site 20, the average drought magnitudes are 7.062 MAF and 7.323 MAF for the combined flows using MLRLS and the observed flows respectively observing the same result at other sites. The underestimation problem of drought statistics obtained by analyzing flows that are generated using traditional regression models has been reported by other similar studies (e.g. Salas et al 2006).

Considering the maximum drought duration and the corresponding maximum drought magnitude for the observed flows, the combined flows based on MREXTN, and the combined flows based on MLRLS (Table 5.7) it can be concluded that augmenting the short observed flows improves the characterization of severe droughts, however, preference may be given to the MREXTN model which generates reconstructed flows that gives more reasonable estimates of critical drought properties that may be expected in a period much longer than the historical period, i.e. the period 1490-2002 (513 years) compared to the observed 1906-2002 (97 years). Two main reasons why the model MREXTN may give reconstructed flows that provide more realistic drought properties and drought scenarios than the model MLRLS does are because the former model includes a noise term which serves in capturing the total natural variability of the flows and an additional term to account for the serial correlation at which both are not considered in the MLRLS model.

For the detailed single site drought analysis the observed annual cumulative flows, the combined annual cumulative flows using MREXTN model, and the combined annual cumulative flows using MLRLS model for site 20 were used. The annual cumulative streamflow data at site20 covers the period 1906–2002 (Figure 5.3), and the basic statistics are: long term mean = 15.13 MAF, standard deviation = 4.44 MAF, skewness coefficient = 0.15, lag-1 serial correlation coefficient = 0.24, longest drought duration = 5 years (1988-1992) that has the maximum total deficit of 21.56 MAF when the annual observed streamflow series was truncated at the level of the long-term mean. The drought analysis for site20 of the Colorado River has been carried out using the observed annual cumulative flows (Figure 5.4a) that covers the period 1906-2002, the combined annual cumulative flows using the model MLRLS that covers the period 1490-1905 (Figure 5.4c), and the combined annual cumulative flows using the model MREXTN that covers the period 1490-1905 (Figure 5.4d). The observed annual cumulative flows, the combined annual cumulative flows using MLRLS, and the combined annual cumulative flows using MREXTN were truncated at the level of the long-term mean for each.

Regarding the single year drought events, the analysis based on the observed flow records of site20 shows that the worst single year drought event has occurred in the year 1977 having a magnitude of 9.72 MAF. On the other hand, the analysis of the combined flows based on the model MREXTN indicates that the 1977 observed drought event has been exceeded 6 times in the past 513 years where the worst single year drought had occurred in the year 1496 having a magnitude of 11.52 MAF. The analysis of the combined annual cumulative flows using MLRLS model shows similar result with the single year worst drought occurred in 1685 with magnitude of 11.5 MAF. Generally, in terms of the analysis of the single year drought, the extension of the observed flows with the tree-ring reconstructed flows improves the characterization of the single year drought where many single year drought events of magnitude greater than the critical observed 1977 drought have been detected. Additional information regarding the severe drought events that had happened in site20 the Colorado River can be obtained from the frequency distribution of drought duration and magnitude as shown in Figures 5.5 and 5.6. For instance, Figure 5.5 shows the frequency distribution of drought duration obtained from the analysis of the observed flows and from the combined flows derived from the MREXTN and MLRLS models for site20. The general pattern is an exponential decay of the frequency distributions that show drought durations obtained from the reconstructed flows extending beyond the longest drought of 5 years obtained from the observed flows (1906-2002). Figure 5.6 shows an exponential decay of the frequency distribution of drought magnitude extending beyond the observed maximum especially for the droughts derived from the combined flows using the model MREXTN.

The analysis of the observed annual cumulative flows shows that the extreme drought in terms of magnitude is the 5 years drought that has happened in 1988 – 1992 and has the maximum total deficit of 21.56 MAF. That particular drought of such magnitude or similar has occurred only once in the 97 years of observed flow record. A practical question that may arise is how frequently a drought of such magnitude or greater occurs? Answers to this question will be difficult or impossible to give based solely on the observed flow record. One may be able to improve the characterization of droughts and perhaps be able to answer the foregoing question by extending the observed records using reconstructed streamflows (Salas et al 2006). Therefore to have a better picture about the distribution of the magnitude of the critical drought in a 97 years time span (the length of the observed record), 417 overlapped sub-samples each 97 years long have been obtained from the combined flow records using the models MREXTN and MLRLS. Then the critical drought magnitude was obtained from each sub-sample and frequency plots were obtained as shown in Figures 5.7 and 5.8. Figure 5.7 shows the distribution of the critical drought magnitude derived from the combined flows using MREXTN model and it also shows the critical drought derived from the observed flow record for comparison purposes. Figure 5.7 shows clearly that the magnitude of the observed 1988-1992 critical drought has been exceeded about 75% of the time. The same analysis has been performed based on the combined flows using MLRLS model and the result is shown in Figure 5.8. The Figure shows that the magnitude of the 1988-1992 critical drought has been exceeded about 38% of the time. Also, the comparison of the frequency plots in Figures 5.7 and 5.8 show a wider variation of the critical droughts obtained from the MREXTN reconstructed flows compared to those obtained from the MLRLS reconstructed flows.

Besides characterizing the duration and magnitude of droughts, the return period of drought events obtained by truncating the observed flows and the combined flows at site 20 has been analyzed. The return period has been defined as the expected value of the interarrival time between drought events (Salas et al, 2005). In this study the drought event E is defined as the event that has magnitude D that exceeds the critical magnitude D_o and duration equals l years (l = 1, 2, 3,..), that is $E = \{D > D_o \cap L = l\}$. The critical magnitude D_o is defined as in (3.11). The return periods of the drought event E at $\gamma = 0$, 0.5, and 1 are shown in Figures 5.9 and 5.10 where y_o has been taken as the long-term mean of the series being analyzed. In general as can be seen from Figures 5.9 and 5.10, as γ or l increases the estimates obtained from the observed data becomes quite uncertain or cannot be obtained. For example for $\gamma = 1$ only estimates of return period for l = 3, 4, and 5 can be obtained from the observed record. Also estimates of the return period for values of $l \geq 4$ years or $\gamma = 1$ based on the combined flows using MLRLS model (Figure 5.10) appear to depart from the expected curves which is expected since the MLRLS model results in flows with low variance and low serial correlation values compared to the observed or flows obtained from MREXTN model.

The second analysis relates to characterizing regional droughts over the Colorado River basin considering both the observed annual cumulative flows and the combined annual cumulative flows (observed plus extended) for sites 2, 6, 8, 9, 13, 16, 20, and 29. The observed annual cumulative flows data are available for sites 2, 6, 8, 9, 13, 16, 20, and 29 for the period 1906 – 2002. However as stated previously for the case of the single site drought analysis, the number of drought events obtained by analyzing the observed flows is still insufficient to estimate drought properties accurately and therefore longer records are required. The annual cumulative flows at sites 8, 16, 20 and 29 were extended using the MREXTN extension model with tree-ring as information source back to the year 1490. The annual cumulative flows at sites 2, 6, 9, and 13 were obtained by disaggregating the extended annual flows at sites 8 and 16 using the disaggregation model (2.70) and later the annual cumulative flows of sites 2, 6, 9, and 13 were adjusted for the spatial proportionality using the spatial adjustment procedure (2.76). In this case the combined annual cumulative flows that cover the period 1490 - 2002, i.e. observed 1906-2002 plus extended 1490-1905, for sites 2, 6, 8, 9, 13, 16, 20, and 29 were utilized for the regional drought analysis.

The regional drought event of duration ($L = l$) is defined as the event that is resulted by the occurrence of simultaneous deficits in all sites within the study region. Figure 5.11 shows the frequency distribution of the regional drought duration obtained by analyzing the observed annual cumulative flows and the combined annual cumulative flows at sites 2, 6, 8, 9, 13, 16, 20, and 29 within the Colorado River basin. The general pattern is an exponential decay of the frequency distributions that show drought durations obtained from the combined flows extending beyond the longest regional drought of 5 years obtained from the observed flows (1906-2002) at sites 2, 6, 8, 9, 13, 16, 20, and 29. Furthermore, Figure 5.12 shows the analysis of the return period of the regional drought event E $(L = l)$ that occurs in sites 2, 6, 8, 9, 13, 16, 20, and 29 simultaneously. It is clear from Figure 5.12 that characterizing return periods of regional drought solely based on the observed flows is quite uncertain compared to what can be achieved from analyzing the reconstructed flows. Generally, it is clear that the characterization of the regional drought properties like the distribution of the regional drought duration and the return period of regional droughts is quite uncertain using the observed flows only compared to the results obtained by analyzing the longer extended records obtained by the record extension model MREXTN.

5.6 Summary and Conclusions

The record extension model with noise REXTN developed by Salas et al (2006) was devoted to reasonably extend one short variable given many predictors (tree ring indices). This chapter generalizes the univariate REXTN to the multivariate REXTN (MREXTN) to extend many short variables simultaneously using tree ring indices. The model MREXTN includes noise term that serves to capture the variability that exists in the natural records, and since the noise term may lead to possible reconstructed traces, a criterion to select a single extension among many is also presented. Simulation experiments are conducted in order to evaluate the performance of the proposed model and compared it with alternative models. The evaluation is based on a number of key statistics that include the mean, standard deviation, serial correlation, and correlations between sites. Tree ring residual chronologies were available for the use as information source to extend streamflows at sites 8, 16, 20, and 29 within the Colorado River back in time. Using stepwise regression, the tree-ring variables that may be used in the final reconstruction model were obtained so that most of the total variance at the selected individual flow sites is explained.

The skills of the reconstruction models MREXTN and MLRLS to reconstruct flows at the selected sites were measured through the validation process. Each model was first calibrated and then validated using the split sample validation procedure. Generally the validation statistics *RE* obtained are comparable for both models, which indicate a reasonable good skill for both models to reconstruct streamflow over the period where observed flows are absent. With the four tree-ring data sets as possible predictors, the extension models MREXTN and MLRLS have been used to reconstruct the streamflows at sites 8, 16, 20 and 29 within the Colorado River back in time to the year 1490. The combined flows (observed 1906-2002 plus reconstructed 1490-1905) for sites 8, 16, 20 and 29 obtained using MREXTN and MLRLS models were available for characterizing droughts. Comparing basic drought statistics like the mean, standard deviation of drought duration and magnitude it can be noticed that statistics obtained from analyzing the reconstructed flows from MREXTN model reflect the occurrence of more extreme droughts than what is observed from analyzing the observed streamflows alone. On the other hand, drought statistics obtained from analyzing the reconstructed flows from MLRLS model do not show much improvement in terms of drought duration and magnitude over what is available from analyzing the observed flows. Based on that it can be concluded that the MREXTN model generates reconstructed flows that reasonable estimates of drought properties.

The analysis of the observed flows at site 20 shows an extreme drought in terms of magnitude is the 5 years drought (1988-1992) that has total deficit of 21.56 MAF. The distribution of the magnitude of 1988 - 1992 drought obtained by analyzing the reconstructed flows from MREXTN shows that the magnitude of the 1988-1992 drought has been exceeded about 75% of the time compared to 38% of the time when analyzing the reconstructed flows from MLRLS model. Analysis of the drought return period shows that generally as λ or l increases, estimates obtained from the observed data becomes quite uncertain or unavailable when compared with the tree ring reconstructed flows.

Table 5.1 The performance statistics $m[\bar{\bar{y}}^{(i)}/\bar{y}_1^{(i)}]$, $m[S^2_{\tilde{y}^{(i)}}/S^2_{y1^{(i)}}]$, $m[r_1(\tilde{y}^{(i)})/r_1(y1^{(i)})]$, $m[r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}/r_{y1^{(1)}y1^{(2)}}]$, $m[r_{\tilde{y}^{(i)}x}/r_{x1\,y1^{(i)}}]$, and $m[RMSE]$ obtained from 1000 generated samples and record extensions for $y_t^{(1)}$ and $y_t^{(2)}$ for the case $\rho_{x,y_t^{(1)}} = \rho_{x,y_t^{(2)}} = 0.5$. Other parameters used in the simulation include: $N_1 = 50$, $\mu_{y^{(1)}} = 100$, $\sigma_{y^{(1)}} = 35$, $\rho_1(y^{(1)}) = 0.15$, $\mu_{y^{(2)}} = 200$, $\sigma_{y^{(2)}} = 70$, $\rho_1(y^{(2)}) = 0.2$, $\mu_x = 50$, $\sigma_x = 20$, $\rho_1(x) = 0.3$, $\rho_{y_t^{(1)}y_t^{(2)}} = 0.8$, and $N_2 = 50, 100, 250$.

Statistic	N_2	MLRLS		Multivariate GMOVE		MREXTN	
		$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
$m\left[\dfrac{\bar{\bar{y}}^{(i)}}{\bar{y}_1^{(i)}}\right]$	50	0.999	0.999	1.007	0.997	0.999	1.000
	100	1.001	1.001	1.001	1.002	1.001	1.001
	250	1.001	1.001	1.001	0.999	0.999	1.000
$m\left[\dfrac{S^2_{\tilde{y}^{(i)}}}{S^2_{y1^{(i)}}}\right]$	50	0.630	0.632	1.083	1.103	0.992	0.988
	100	0.510	0.513	1.102	1.111	0.994	0.994
	250	0.476	0.489	1.188	1.197	0.999	1.000
$m\left[\dfrac{r_1(\tilde{y}^{(i)})}{r_1(y1^{(i)})}\right]$	50	1.329	1.357	1.960	1.931	1.053	1.047
	100	1.510	1.499	2.070	2.085	1.098	1.120
	250	1.883	1.876	2.456	2.401	1.122	1.133
$m\left[\dfrac{r_{\tilde{y}^{(i)}x}}{r_{x1\,y1^{(i)}}}\right]$	50	1.286	1.283	0.981	0.978	0.982	0.987
	100	1.445	1.440	0.989	0.986	0.988	0.994
	250	1.685	1.681	0.981	0.986	0.986	0.988
$m\left[\dfrac{r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}}{r_{y1^{(1)}y1^{(2)}}}\right]$	50	1.000	1.045	1.000	0.987	1.000	1.001
	100	1.000	1.076	1.000	0.979	1.000	1.001
	250	1.000	1.125	1.000	0.970	1.000	1.001
$m[RMSE]$	50	15.02	32.47	19.56	41.29	15.32	30.01
	100	15.12	33.74	19.55	41.97	15.72	32.78
	250	16.36	33.96	20.44	41.96	16.06	32.91

Table 5.2 The performance statistics $m[\bar{\bar{y}}^{(i)}/\bar{y}_1^{(i)}]$, $m[S^2_{\tilde{y}^{(i)}}/S^2_{y1^{(i)}}]$, $m[r_1(\tilde{y}^{(i)})/r_1(y1^{(i)})]$, $m[r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}/r_{y1^{(1)}y1^{(2)}}]$, $m[r_{\tilde{y}^{(i)}x}/r_{x1y1^{(i)}}]$, and $m[RMSE]$ obtained from 1000 generated samples and record extensions for $y_t^{(1)}$ and $y_t^{(2)}$ for the case $\rho_{x_t y_t^{(1)}} = \rho_{x_t y_t^{(2)}} = 0.7$. Other parameters used in the simulation include: $N_1 = 50$, $\mu_{y^{(1)}} = 100$, $\sigma_{y^{(1)}} = 35$, $\rho_1(y^{(1)}) = 0.15$, $\mu_{y^{(2)}} = 200$, $\sigma_{y^{(2)}} = 70$, $\rho_1(y^{(2)}) = 0.2$, $\mu_x = 50$, $\sigma_x = 20$, $\rho_1(x) = 0.3$, $\rho_{y_t^{(1)}y_t^{(2)}} = 0.8$, and $N_2 = 50, 100, 250$.

Statistic	N_2	MLRLS		Multivariate GMOVE		MREXTN	
		$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
$m\left[\frac{\bar{\bar{y}}^{(i)}}{\bar{y}_1^{(i)}}\right]$	50	1.000	1.001	1.004	0.999	1.001	1.000
	100	1.001	1.002	1.000	1.000	1.000	0.999
	250	1.002	1.002	0.999	1.000	1.001	1.000
$m\left[\frac{S^2_{\tilde{y}^{(i)}}}{S^2_{y1^{(i)}}}\right]$	50	0.752	0.757	1.057	1.064	0.998	0.998
	100	0.670	0.677	1.091	1.100	0.995	0.995
	250	0.589	0.596	1.168	1.177	0.998	1.001
$m\left[\frac{r_1(\tilde{y}^{(i)})}{r_1(y1^{(i)})}\right]$	50	1.450	1.426	2.042	2.066	1.100	1.124
	100	1.640	1.662	2.250	2.313	1.145	1.147
	250	1.830	1.792	2.455	2.487	1.166	1.158
$m\left[\frac{r_{\tilde{y}^{(i)}x}}{r_{x1y1^{(i)}}}\right]$	50	1.172	1.169	0.979	0.986	0.986	0.989
	100	1.247	1.240	0.972	0.977	0.989	0.990
	250	1.341	1.333	0.988	0.989	0.984	0.986
$m\left[\frac{r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}}{r_{y1^{(1)}y1^{(2)}}}\right]$	50	1.000	1.080	1.000	1.002	1.000	0.985
	100	1.000	1.116	1.000	0.924	1.000	0.969
	250	1.000	1.170	1.000	0.960	1.000	0.963
$m[RMSE]$	50	12.41	23.44	15.16	29.20	11.75	21.36
	100	12.39	23.80	15.23	29.54	12.88	22.69
	250	12.59	23.95	16.78	30.05	12.65	22.80

Table 5.3 The performance statistics $m[\bar{\bar{y}}^{(i)}/\bar{y}_1^{(i)}]$, $m[S^2_{\tilde{y}^{(i)}}/S^2_{y1^{(i)}}]$, $m[r_1(\tilde{y}^{(i)})/r_1(y1^{(i)})]$, $m[r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}/r_{y1^{(1)}y1^{(2)}}]$, $m[r_{\tilde{y}^{(i)}x}/r_{x1y1^{(i)}}]$, and $m[RMSE]$ obtained from 1000 generated samples and record extensions for $y_t^{(1)}$ and $y_t^{(2)}$ for the case $\rho_{x_t y_t^{(1)}} = \rho_{x_t y_t^{(2)}} = 0.9$. Other parameters used in the simulation include: $N_1 = 50$, $\mu_{y^{(1)}} = 100$, $\sigma_{y^{(1)}} = 35$, $\rho_1(y^{(1)}) = 0.15$, $\mu_{y^{(2)}} = 200$, $\sigma_{y^{(2)}} = 70$, $\rho_1(y^{(2)}) = 0.2$, $\mu_x = 50$, $\sigma_x = 20$, $\rho_1(x) = 0.3$, $\rho_{y_t^{(1)}y_t^{(2)}} = 0.8$, and $N_2 = 50, 100, 250$.

Statistic	N_2	MLRLS		Multivariate GMOVE		MREXTN	
		$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
$m\left[\frac{\bar{\bar{y}}^{(i)}}{\bar{y}_1^{(i)}}\right]$	50	1.000	1.001	1.002	0.999	0.999	1.000
	100	1.001	1.001	0.999	0.999	1.000	0.999
	250	1.003	1.003	1.000	0.999	1.000	0.999
$m\left[\frac{S^2_{\tilde{y}^{(i)}}}{S^2_{y1^{(i)}}}\right]$	50	0.987	0.981	1.033	1.038	1.003	1.005
	100	0.954	0.958	1.056	1.044	1.008	1.012
	250	0.935	0.929	1.147	1.150	1.009	1.011
$m\left[\frac{r_1(\tilde{y}^{(i)})}{r_1(y1^{(i)})}\right]$	50	1.473	1.471	2.222	2.194	1.152	1.487
	100	1.793	1.787	2.355	2.421	1.167	1.168
	250	2.047	2.046	2.564	2.581	1.172	1.178
$m\left[\frac{r_{\tilde{y}^{(i)}x}}{r_{x1y1^{(i)}}}\right]$	50	1.089	1.082	0.988	0.990	0.991	0.992
	100	1.122	1.111	0.978	0.981	0.975	0.973
	250	1.158	1.145	0.975	0.986	0.963	0.964
$m\left[\frac{r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}}{r_{y1^{(1)}y1^{(2)}}}\right]$	50	1.000	1.099	1.000	1.002	1.000	0.987
	100	1.000	1.135	1.000	0.979	1.000	0.974
	250	1.000	1.187	1.000	0.959	1.000	0.956
$m[RMSE]$	50	8.33	17.41	10.20	19.30	8.21	17.53
	100	8.38	17.83	10.05	20.35	8.41	17.35
	250	8.43	17.21	10.95	19.02	8.55	17.95

Table 5.4 The performance statistics $m[\bar{\tilde{y}}^{(i)}/\bar{y}_1^{(i)}]$, $m[S^2_{\tilde{y}^{(i)}}/S^2_{y1^{(i)}}]$, $m[r_1(\tilde{y}^{(i)})/r_1(y1^{(i)})]$, $m[r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}/r_{y1^{(1)}y1^{(2)}}]$, $m[r_{\tilde{y}^{(i)}x}/r_{x1\,y1^{(i)}}]$, and $m[RMSE]$ obtained from 1000 generated samples and record extensions for $y_t^{(1)}$ and $y_t^{(2)}$ for the case $\rho_{x_t y_t^{(1)}} = \rho_{x_t y_t^{(2)}} = 0.5$. Other parameters used in the simulation include: $N_1 = 50$, $\mu_{y^{(1)}} = 100$, $\sigma_{y^{(1)}} = 35$, $\rho_1(y^{(1)}) = 0.15$, $\mu_{y^{(2)}} = 200$, $\sigma_{y^{(2)}} = 70$, $\rho_1(y^{(2)}) = 0.2$, $\mu_x = 50$, $\sigma_x = 20$, $\rho_1(x) = 0$, $\rho_{y_t^{(1)}y_t^{(2)}} = 0.8$, and $N_2 = 50, 100, 250$.

Statistic	N_2	MLRLS		Multivariate GMOVE		MREXTN	
		$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
$m\left[\frac{\bar{\tilde{y}}^{(i)}}{\bar{y}_1^{(i)}}\right]$	50	1.001	1.001	1.000	0.999	1.001	1.000
	100	1.003	1.003	1.002	1.000	1.000	1.001
	250	1.003	1.003	1.001	1.000	1.000	1.000
$m\left[\frac{S^2_{\tilde{y}^{(i)}}}{S^2_{y1^{(i)}}}\right]$	50	0.613	0.618	1.110	1.097	1.014	1.013
	100	0.497	0.496	1.116	1.102	1.021	1.016
	250	0.464	0.475	1.194	1.208	1.035	1.029
$m\left[\frac{r_1(\tilde{y}^{(i)})}{r_1(y1^{(i)})}\right]$	50	0.924	0.939	1.288	1.164	1.044	1.056
	100	0.901	0.917	1.378	1.297	1.089	1.123
	250	0.827	0.862	1.434	1.393	1.133	1.175
$m\left[\frac{r_{\tilde{y}^{(i)}x}}{r_{x1\,y1^{(i)}}}\right]$	50	1.317	1.312	0.951	0.969	1.037	1.022
	100	1.501	1.485	0.952	0.961	1.032	1.022
	250	1.783	1.749	0.963	0.957	1.039	1.035
$m\left[\frac{r_{\tilde{y}^{(1)}\tilde{y}^{(2)}}}{r_{y1^{(1)}y1^{(2)}}}\right]$	50	1.000	1.045	1.000	0.970	1.000	0.973
	100	1.000	1.078	1.000	0.971	1.000	0.974
	250	1.000	1.136	1.000	0.969	1.000	0.974
$m[RMSE]$	50	15.12	31.55	18.47	41.06	15.27	31.03
	100	15.11	31.86	18.51	42.12	15.75	31.31
	250	15.89	32.45	19.22	43.04	16.49	32.37

Table 5.5 The statistics R^2 and the adjusted R^2 obtained by regressing annual streamflows for sites 8, 16, 20, and 29 of the Colorado River on four tree-ring indices selected by stepwise multiple regression analysis. In addition the validation statistic *RE* obtained by the split sample validation procedure are shown.

Site	Statistics obtained from transformed flows		Statistics obtained from original flows		*RE* statistic obtained from flows using Model	
	R^2	Adjusted R^2	R^2	Adjusted R^2	MREXTN	MLRLS
8	0.738	0.724	0.721	0.706	0.606	0.629
16	0.656	0.637	0.638	0.618	0.624	0.634
20	0.753	0.740	0.741	0.727	0.537	0.519
29	0.756	0.742	0.743	0.729	0.665	0.684

Table 5.6 Basic statistic obtained from the observed annual cumulative flows (1906-2002), the combined annual cumulative flows using MREXTN (1490-2002), and the combined annual cumulative flows using MLRLS (1490-2002) for sites 8, 16, 20 and 29 of the Colorado River. Units of the mean and standard deviation are in MAF.

Statistics	Flow record	Site			
		8	16	20	29
Mean	Observed	6.844	5.434	15.126	16.440
	Combined-MREXTN	6.785	5.417	15.001	16.306
	Combined-MLRLS	6.595	5.256	14.567	15.880
Standard deviation	Observed	2.003	1.671	4.440	4.632
	Combined-MREXTN	2.087	1.711	4.518	4.784
	Combined-MLRLS	1.886	1.502	4.122	4.453
Skewness coefficient	Observed	0.23	0.31	0.15	0.17
	Combined-MREXTN	0.09	0.36	0.17	0.16
	Combined-MLRLS	-0.06	0.08	-0.04	-0.05
Lag-1 serial correlation coefficient	Observed	0.26	0.26	0.24	0.23
	Combined-MREXTN	0.23	0.28	0.24	0.24
	Combined-MLRLS	0.08	0.09	0.08	0.07

Table 5.7 Cross correlation coefficient between the observed annual cumulative flows (1906-2002), the combined annual cumulative flows using MREXTN (1490-2002), and the combined annual cumulative flows using MLRLS (1490-2002) for sites 8, 16, 20 and 29.

Site	Flow record	Site			
		8	16	20	29
8	Observed	1	0.853	0.974	0.968
	Combined-MREXTN	1	0.867	0.974	0.973
	Combined-MLRLS	1	0.942	0.99	0.988
16	Observed	0.853	1	0.923	0.908
	Combined-MREXTN	0.867	1	0.953	0.95
	Combined-MLRLS	0.942	1	0.975	0.971
20	Observed	0.974	0.923	1	0.995
	Combined-MREXTN	0.974	0.953	1	0.999
	Combined-MLRLS	0.99	0.975	1	0.999
29	Observed	0.968	0.908	0.995	1
	Combined-MREXTN	0.973	0.95	0.999	1
	Combined-MLRLS	0.988	0.971	0.999	1

Table 5.8 Comparison of various statistics regarding the duration (run length) and magnitude of droughts derived from the observed and from the combined streamflows for sites 8, 16, 20 and 29 – Colorado River. The combined flows are based on observed flows plus reconstructed flows from MLRLS and MREXTN record extension models.

Statistics		Drought statistics obtained from											
		Observed flows at site				Combined flows at site							
						MLRLS				MREXTN			
		8	16	20	29	8	16	20	29	8	16	20	29
No. of Events		24	21	24	23	120	120	120	119	110	104	108	107
Duration	μ^a	2.00	2.43	2.00	2.22	2.08	2.08	2.08	2.10	2.36	2.58	2.41	2.45
	σ^a	1.19	1.47	1.22	1.28	1.40	1.37	1.42	1.44	1.64	1.73	1.61	1.62
	γ	1.18	0.90	1.39	0.94	2.38	1.62	2.31	2.26	1.96	1.40	1.62	1.54
	Cv	0.60	0.60	0.61	0.58	0.67	0.66	0.68	0.69	0.69	0.67	0.67	0.66
	Longest [a]	5	6	5	5	10	7	10	10	10	9	9	9
	Shortest [a]	1	1	1	1	1	1	1	1	1	1	1	1
Magnitude	μ^b	3.33	3.08	7.32	8.01	3.27	2.52	7.06	7.70	3.89	3.26	8.49	9.27
	σ^b	2.79	2.53	6.58	6.82	2.56	1.90	5.57	6.00	3.27	2.66	6.95	7.51
	γ	0.93	1.06	0.87	0.71	1.16	0.95	1.07	1.06	1.25	0.86	0.92	0.91
	Cv	0.84	0.82	0.90	0.85	0.78	0.75	0.79	0.78	0.84	0.81	0.82	0.81
	Max [b]	9.42	9.49	21.56	20.95	13.30	8.54	27.88	30.30	15.42	12.04	30.44	33.16
	Min [b]	0.30	0.17	0.53	0.55	0.05	0.05	0.03	0.43	0.24	0.06	0.38	0.44

μ = mean, σ = standard deviation, γ = skewness coefficient, and Cv = coefficient of variation. [a] units are in years. [b] units are in MAF.

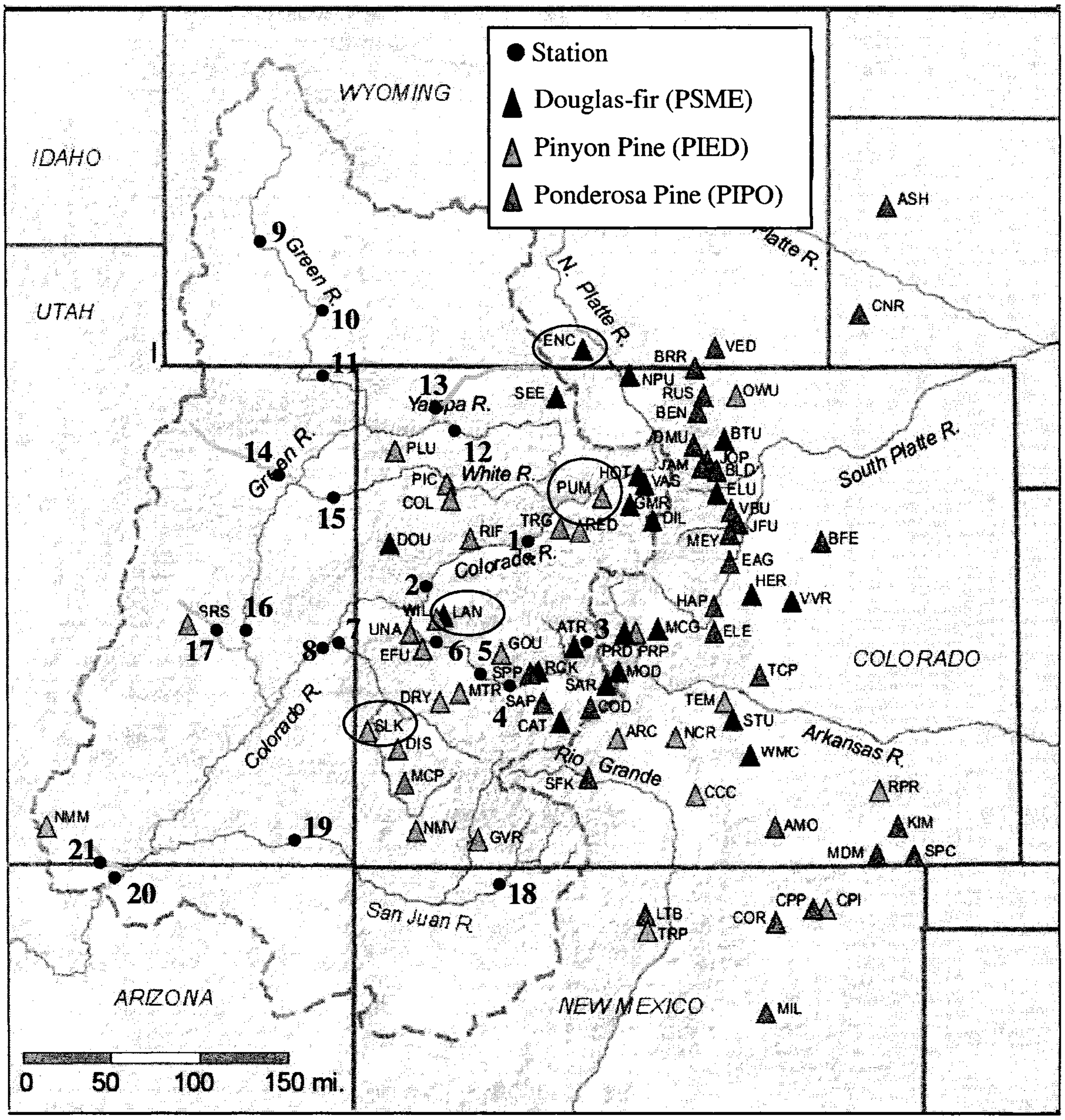

Figure 5.1 The distribution of tree sites within the Colorado River basin. The circle indicates the 4 tree sites PUM, ENC, SLK, and LAN with tree-ring indices used to reconstruct the streamflows of the Colorado River.
(Source: http://www.ncdc.noaa.gov/paleo/streamflow/chronologies.html)

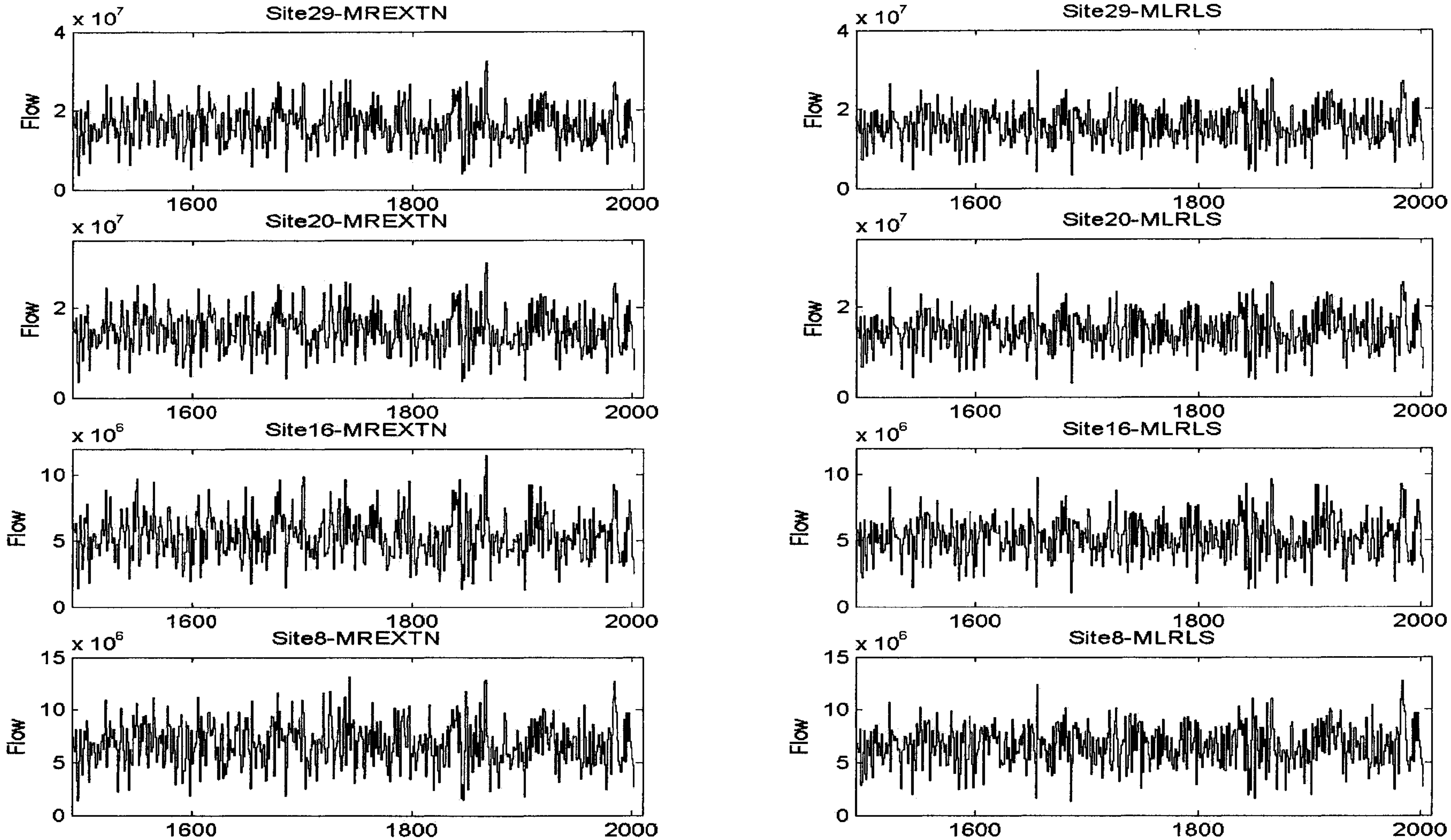

Fig 5.2 Combined annual cumulative flows (observed plus reconstructed using MREXTN or MLRLS models) for sites 8, 16, 20 and 29 of the Colorado River for the period 1490 - 2002 (flow units in acre-ft).

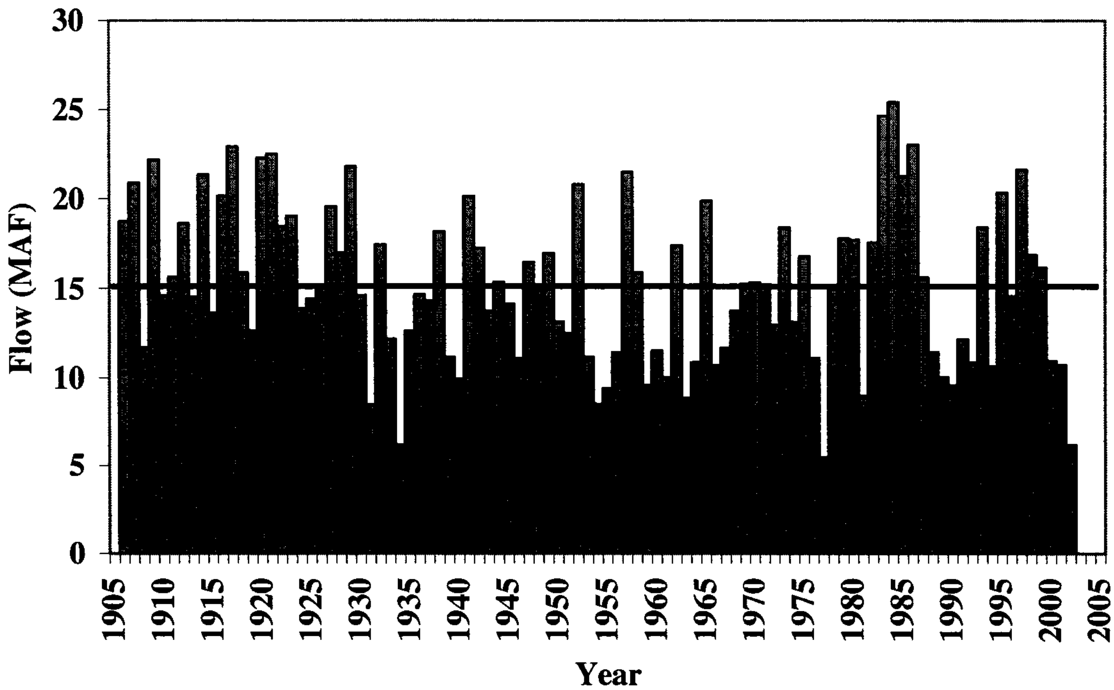

Fig 5.3 Annual streamflow record for site 20 of the Colorado River for the period 1906–2002. The dashed line represents the long-term average (15.126 MAF).

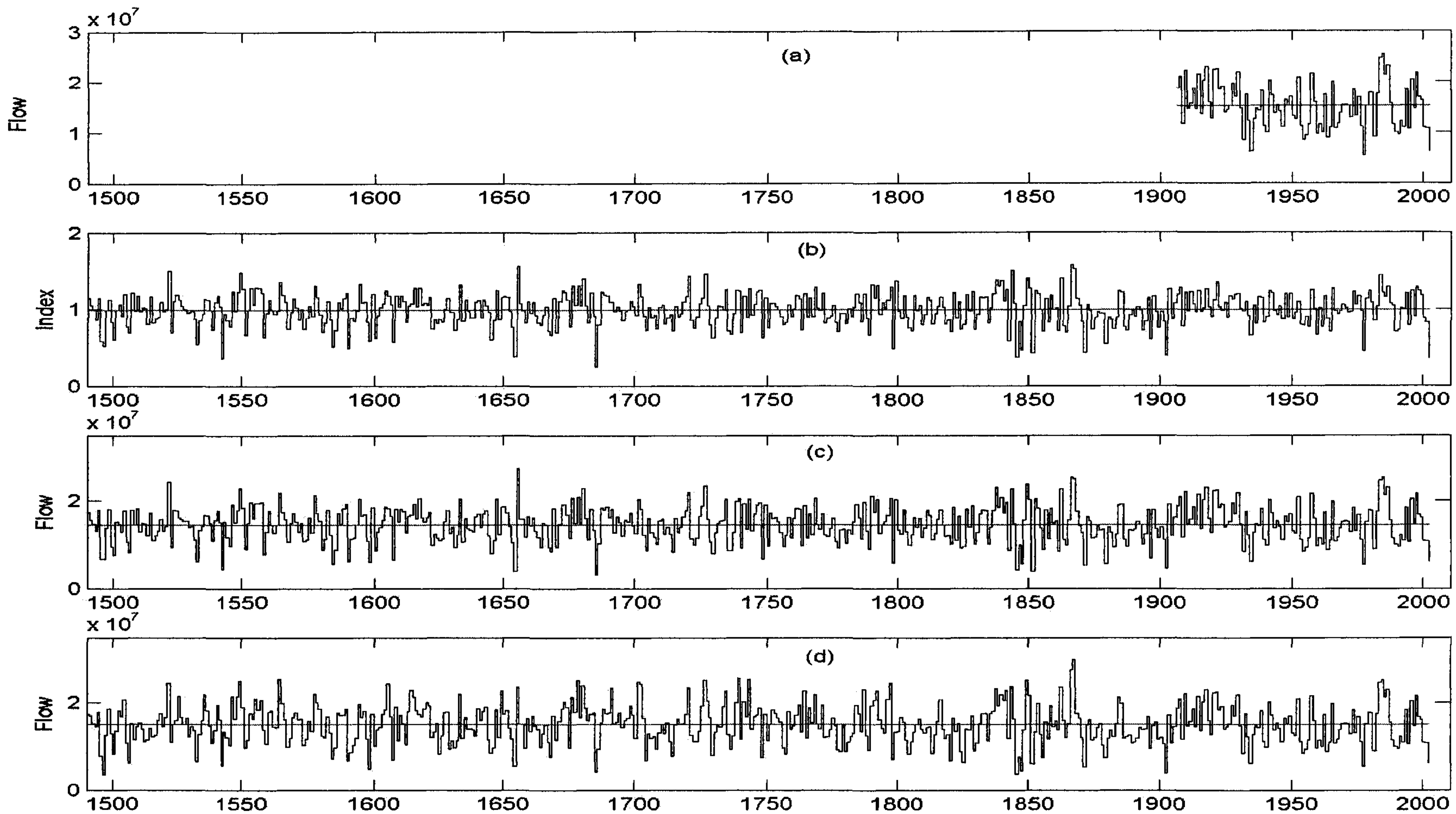

Fig 5.4 Time series of: (a) observed annual cumulative flows for site 20 during the period1906–2002, (b) average tree-ring index over the period 1490–2002, (c) the combined annual cumulative flows for site 20 using MLRLS model, and (d) the combined annual cumulative flows for site 20 using MREXTN model. The solid line represents the long-term mean and unit of flow is acre-ft.

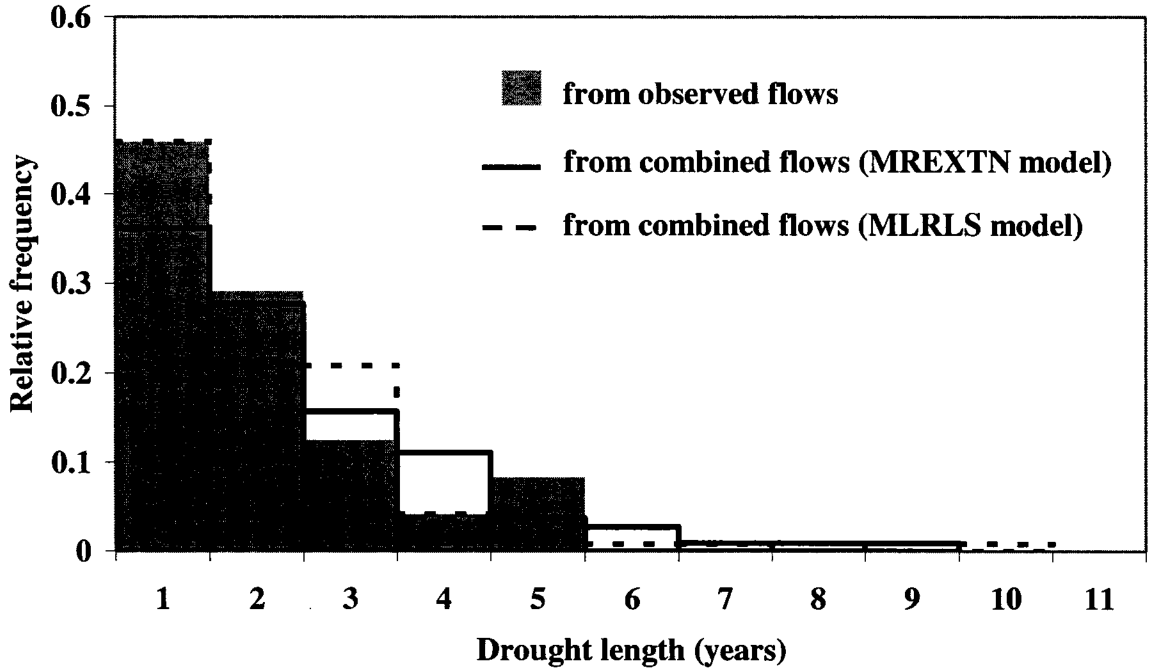

Fig 5.5 Frequency of drought duration obtained from the observed annual cumulative flows and the combined annual cumulative flows based on the MREXTN and MLRLS models for site20 of the Colorado River.

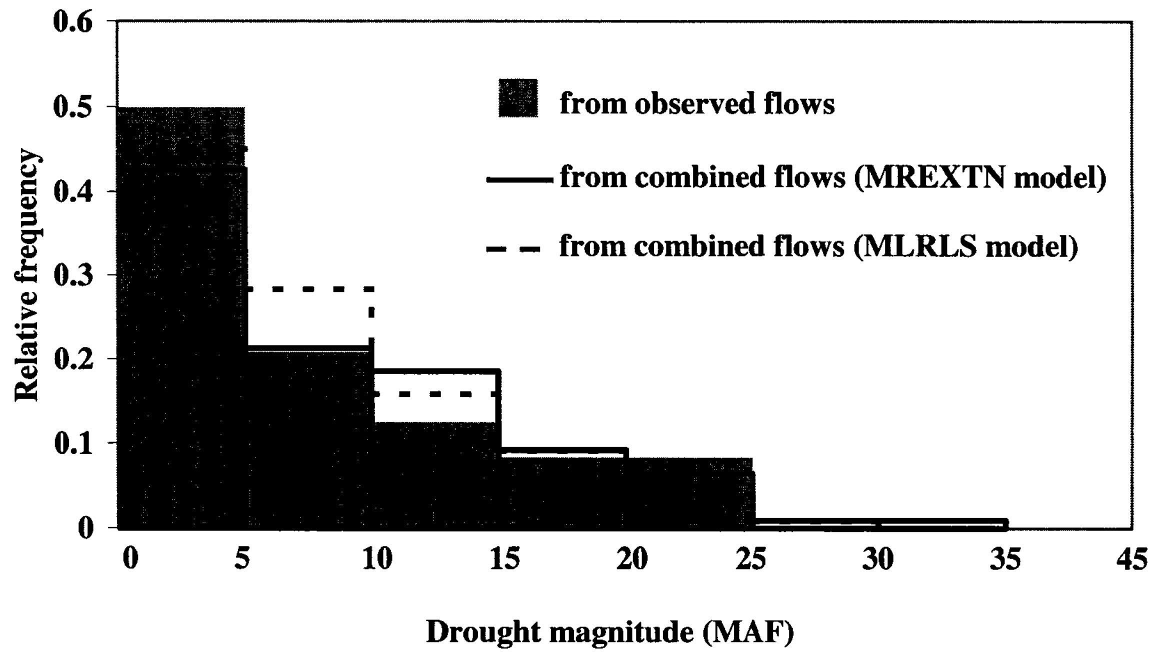

Fig 5.5 Frequency of drought magnitude obtained from the observed annual cumulative flows and the combined annual cumulative flows based on the MREXTN and MLRLS models for site20 of the Colorado River.

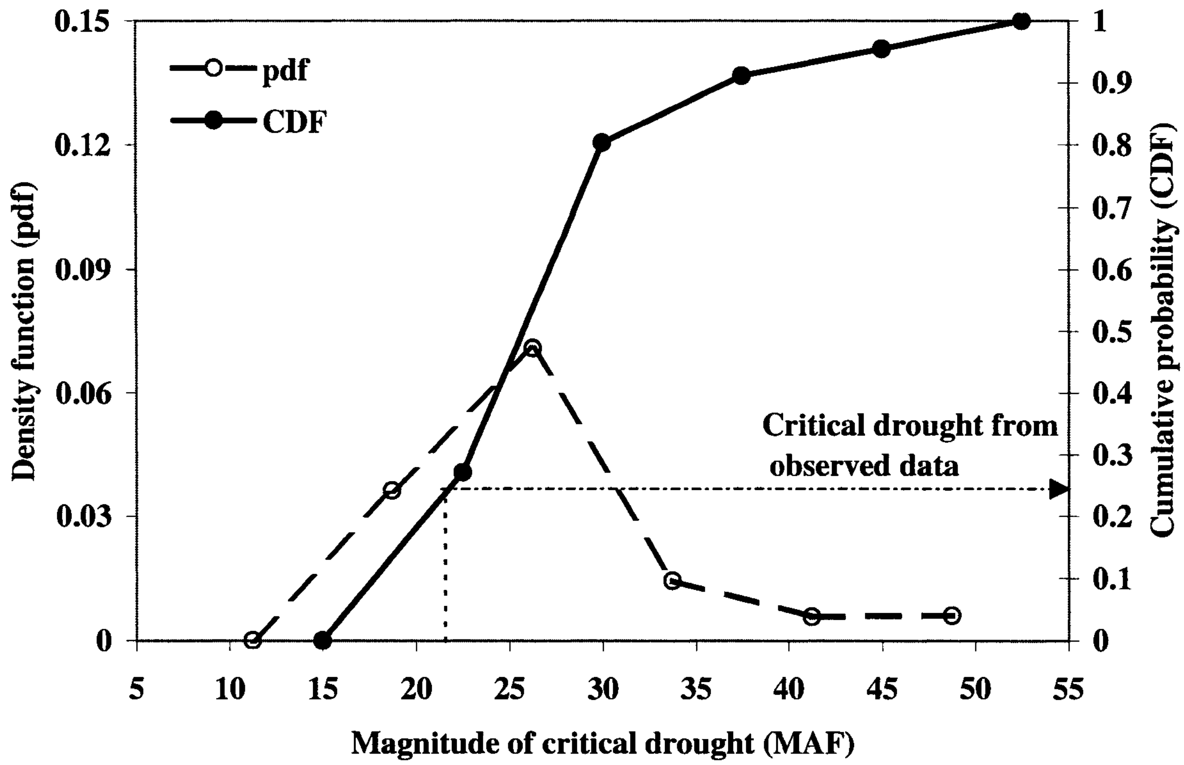

Fig 5.7 Probability distribution of the magnitude of the critical drought obtained from 417 sub-samples of length 97 years assembled from the combined annual cumulative flows based on the MREXTN model for site20 of the Colorado River. For comparison the critical drought obtained from the observed data is also shown.

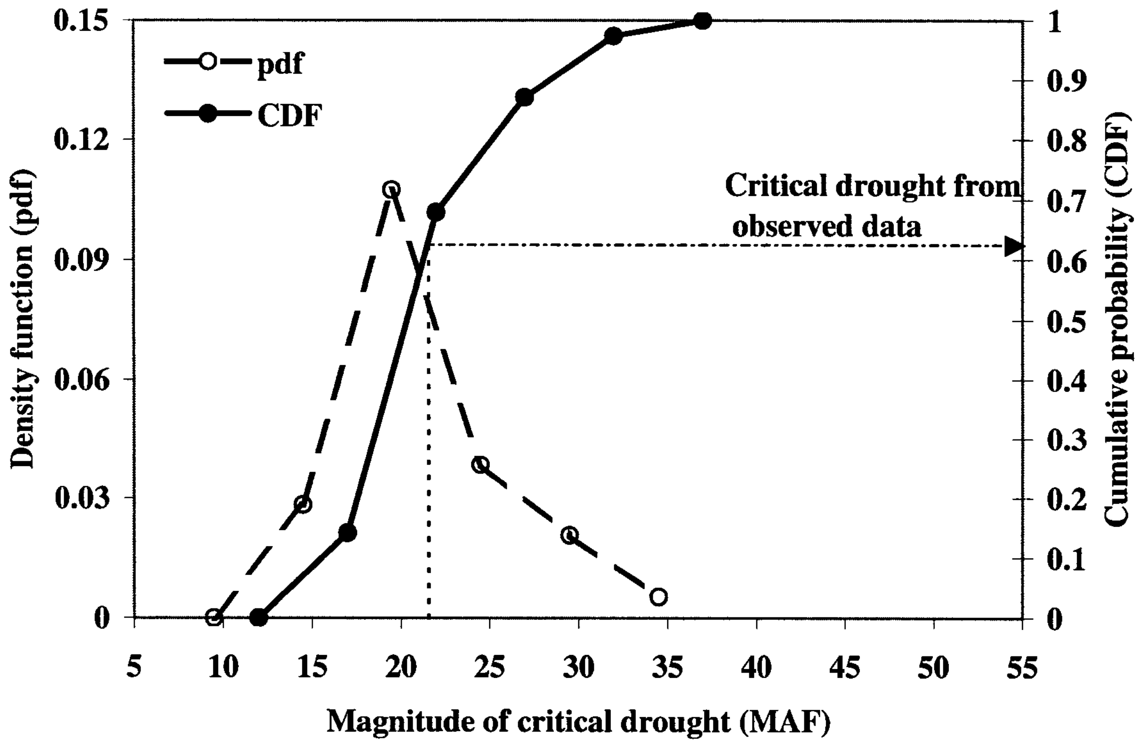

Fig 5.8 Probability distribution of the magnitude of the critical drought obtained from 417 sub-samples of length 97 years assembled from the combined annual cumulative flows based on the MLRLS model for site20 of the Colorado River. For comparison the critical drought obtained from the observed data is also shown.

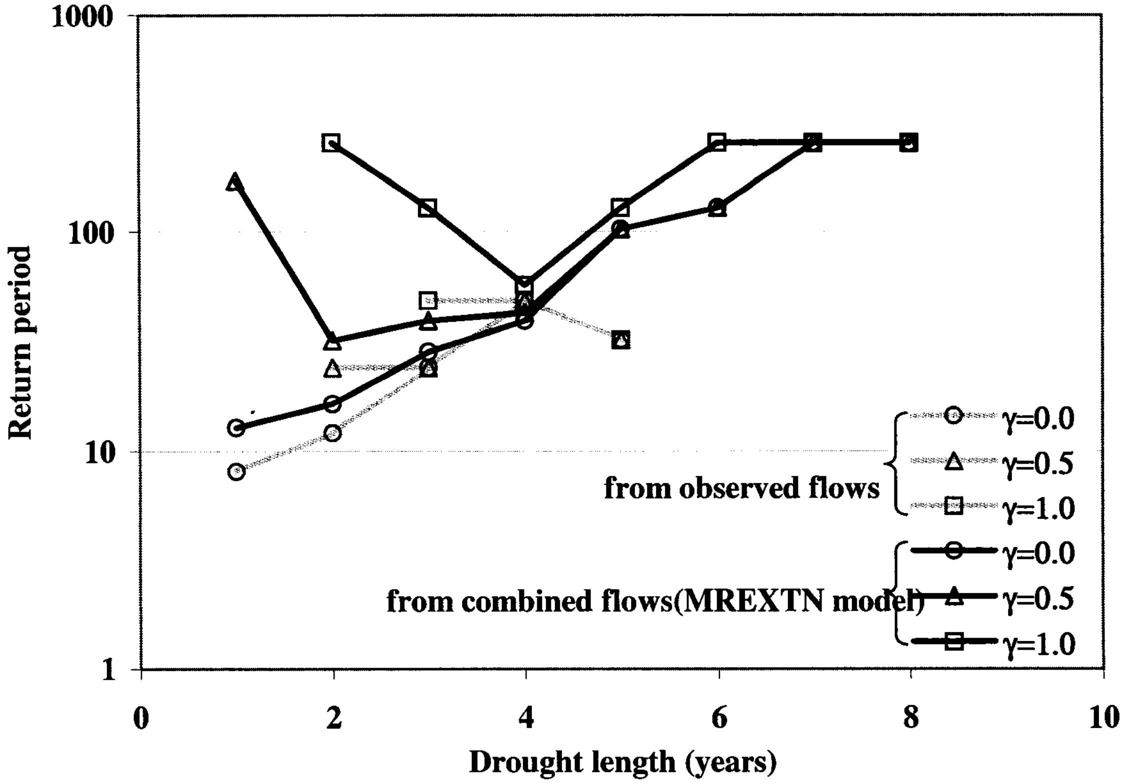

Fig 5.9 Return period of drought event ($D > D_o \cap L = l$) obtained from the observed annual cumulative flows, the combined annual cumulative flows based on MREXTN model for site20 of the Colorado River.

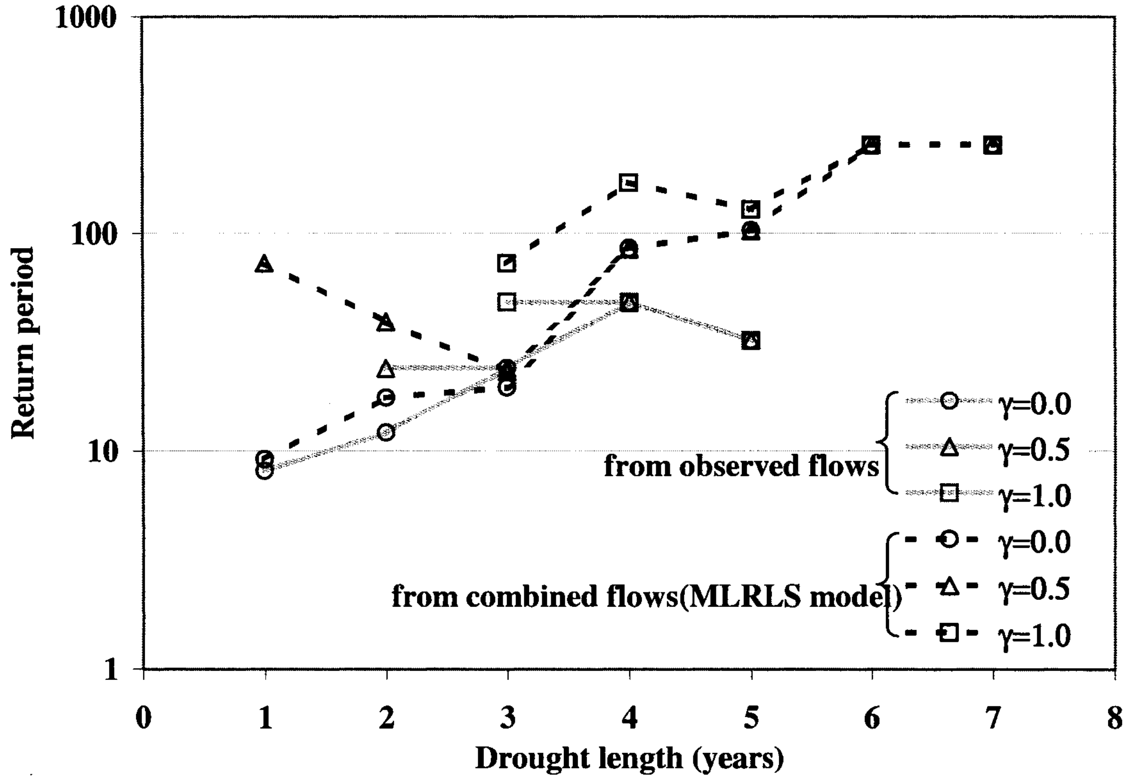

Fig. 5.10 Return period of drought event ($D > D_o \cap L = l$) obtained from the observed annual cumulative flows, the combined annual cumulative flows based on MLRLS model for site20 of the Colorado River.

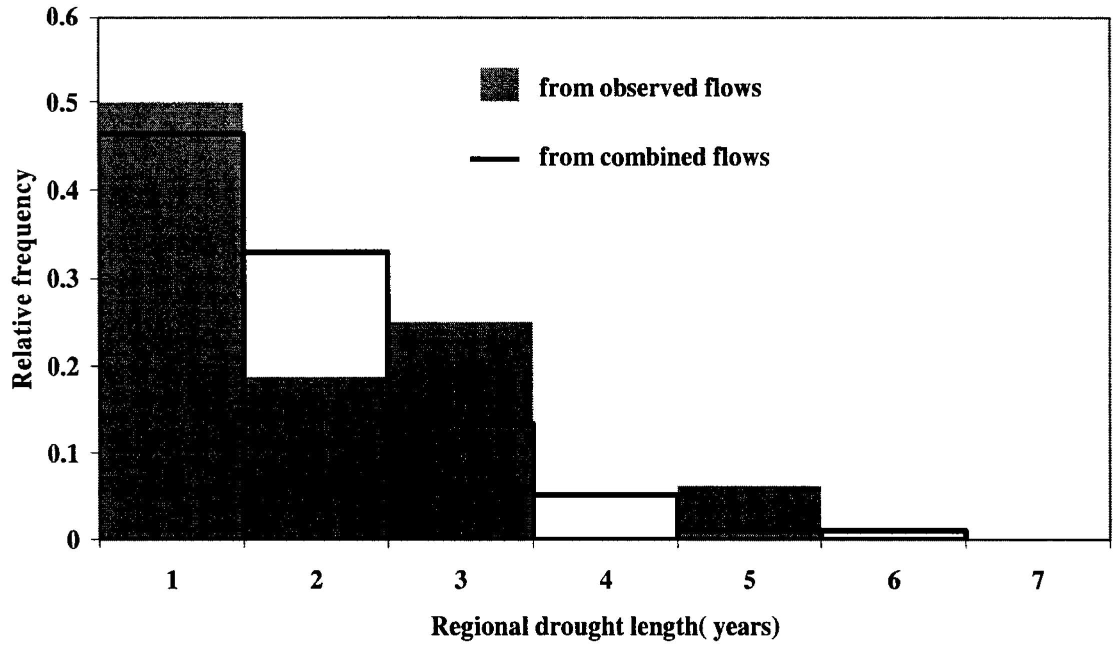

Fig 5.11 Frequency of regional drought duration obtained from the observed and the combined flows for sites 2, 6, 8, 9, 13, 16, 20, and 29 of the Colorado River. The combined flows are the observed (1906 – 2002) plus extendeded 1490 – 1905.

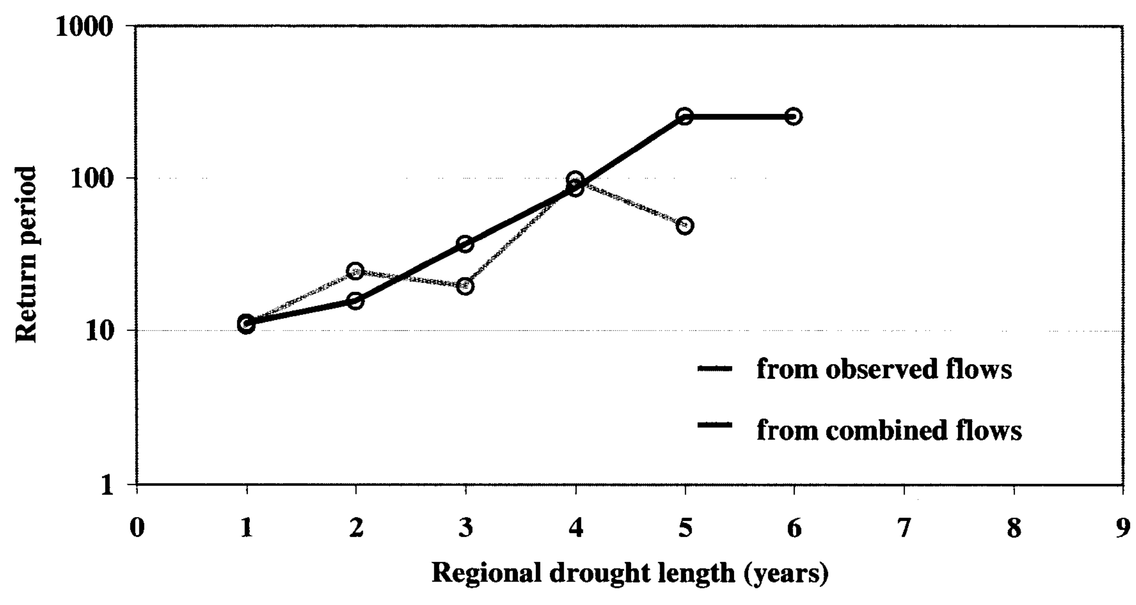

Fig 5.12 Return period of the regional drought event ($L = l$) obtained by analyzing the observed annual cumulative flows, and the combined annual cumulative flows for sites 2, 6, 8, 9, 13, 16, 20, and 29 of the Colorado River.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Planning and management studies of water resources systems consider temporal and spatial variability of hydrological processes like streamflows, particularly periods of extreme droughts. The general objective of this research is to characterize the severity of multiyear hydrologic droughts at one or more sites in a river basin using mathematical models that estimate the return period and risk of extreme droughts. Additional objective is to build a reliable record extension model to reconstruct longer streamflow records at many sites in a river system using tree ring indices for the purpose of improving drought characterization. In this study, a drought event length l occurs where consecutive deficits persist over the time period l.

Considering that drought deficits are beta distributed, mathematical models are developed to characterize the severity of hydrologic droughts. Specifically the occurrence probability, return period, and the risk of drought occurrence are determined assuming that the sequence of deficits and surplus are represented by low order discrete autoregressive moving average processes, i.e. DAR(1) or DARMA(1,1) models. The mathematical procedures have been verified using simulation experiments and later used to characterize droughts in the Poudre, the Colorado, and the Niger River basins.

Utilizing the new process Z_t that converts the joint occurrence of deficits in many sites to single process deficit occurrence, the developed mathematical procedures were extended to consider the characterization of regional hydrologic droughts. For bivariate case, mathematical procedures to evaluate $E[Z_t]$ and $V[Z_t]$ of the process Z_t are developed. The procedures have been applied to characterize regional hydrologic droughts in the Upper Colorado River system.

This study also developed a general record extension model that can be used to lengthen short records in one or more sites given longer records of tree ring indices. In particular, multivariate record extension model has been developed and verified using simulation experiments. Streamflows have been reconstructed using tree ring indices back to the year 1490 for many sites in the Colorado River. The reconstructed flows were used for the purpose of improving the characterization of hydrologic droughts in the Colorado River.

The general conclusion that emerged from this study is the applicability of the proposed mathematical procedures to estimate properties of multiyear droughts. The specific conclusions are:

1. Assuming the series of deficits and surpluses is represented by either DAR(1) or DARMA(1,1) process, analytical models were developed to estimate drought properties such as the occurrence probability, return period, and risk of hydrologic droughts. Drought properties obtained from analyzing the observed (historical) records are uncertain quantities especially for droughts of short length and considerable magnitude. In fact sometimes certain droughts properties cannot be calculated from the historical records because of the limited drought episodes that can be observed. The analytical models developed in this study were found to provide reliable drought properties. The results obtained from the analytical models were

found to agree with the empirical results obtained from analyzing the simulated flows. Therefore the mathematical models developed in this study can be used to predict extreme hydrologic drought events.

2. The Gamma distribution has been widely used to fit the distribution of drought deficits. However, sometimes drought properties obtained under such case may be unrealistic especially for droughts of short length (1-2 years) and high magnitude. On the other hand, the results of this study show that fitting the deficits using the Beta distribution function improves the characterization of extreme droughts when compared with other distributions (e.g. Gamma) and gives realistic results. These results and conclusions have been verified using simulation experiments.

3. Drought properties obtained utilizing the definition of the new process Z_t and the single site drought characterization procedures were found to be fairly good estimates of drought properties on a regional context when compared with properties obtained from analyzing the observed flows only. Considering the Upper Colorado Basin, the analysis of regional droughts growth patterns through the utilized procedure indicates direction of drought growth to be most probably north to south.

4. The multivariate record extension model MREXTN can be used to reconstruct reasonable streamflow records at many sites simultaneously using tree ring indices of longer records as proxy data. Based on several performance measures, the model MREXTN was found to perform well in terms of preserving the targeted key statistics compared to other existing record extension models. Drought properties obtained by analyzing the reconstructed streamflows obtained from MREXTN model show to be more reliable and much informative when compared with properties obtained by

analyzing the short observed flows. In particular, the results indicate that droughts of longer durations and greater magnitudes have happened in the past than droughts observed from analyzing the historical records.

Lastly, with the contributions of other related past studies, it is believed that this study provides analytical procedures and approaches that can be reliably used to characterize droughts and ultimately help in understanding the phenomenon of drought. However in the light of the assumptions made, it is recommended that further studies to be carried out beyond the scope of this study. It is recommended to:

(a) Extend drought characterization procedures to consider the case of droughts under periodic processes such as the case of monthly precipitation and monthly streamflows.

(b) Investigate and quantify the uncertainties in drought properties such as uncertainties in drought return period and risk.

(c) Extend the characterization of regional droughts considering alternative methods.

REFERENCES

Bayazit M. and Onoz B., 2005. Probabilities and return periods of multisite droughts. Hydrological Sciences. Vol. 50, No. 4, P: 605 – 615.

Biondi F., Kozubowski T., and Panorska A., 2005. A new model for quantifying climate episodes. International Journal of Climatology. Vol. 25, P: 1253 – 1264.

Bras R. and Rodriguez-Iturbe I., 1993. Random functions and hydrology. Dover publications, Revised edition 1993.

Buishand A., 1978. The binary DARMA(1,1) process as a model for wet-dry sequence. Technique Note 78-01, Dept of Math., Agricultural University, Wageningen, The Netherlands.

Byun Hi-Ryong and Wilhite D. A., 1999. Objective quantification of drought severity and duration. Journal of Climate. Vol. 12, P: 2747 – 2756.

Cancelliere A., Bonaccorso B., and Rossi G., 2003. On the probabilistic characterization of drought events. Proceedings of Hydrology Days, Colorado State University – Fort Collins, Co, P: 33 – 44.

Chang T. J., Kavvas M. L., and Delleur J. W., 1984. Daily precipitation modeling by discrete autoregressive moving average processes. Water Resources Research. Vol. 20, No. 5, P: 565 – 580.

Chung C., 1999. Probability distribution, risk, and return period of dependent hydrologic events. Ph. D dissertation, Colorado State University, Fort Collins, Co, USA.

Chung C. and Salas J. D., 2000. Drought occurrence probabilities and risk of dependent hydrologic processes. Journal of Hydrologic Engineering. Vol. 5, No. 3, P: 259 – 268.

Cleaveland M. K., 2000. A 963-year reconstruction of summer (JJA) streamflow in the white river, Arkansas, USA, from tree-rings. The Holocene. Vol. 10, No.1, P: 33 – 41.

Cramer H., and Leadbetter M., 1967. Stationary and related stochastic processes. Sample functions properties and their applications. Wiley, New York.

Douglas E. M., Vogel R. M. and Kroll C. N., 2002. Impact of streamflow persistence on hydrologic design. Journal of Hydrologic Engineering. Vol. 7, No. 3, P: 220 – 227.

Dracup A., Lee K. Seong and Paulson E. G., 1980. On the statistical characteristics of drought events. Water Resources Research. Vol. 16, No. 2, P: 289 – 296.

Earle C.J., 1993. Asynchronous droughts in California streamflow as reconstructed from tree rings. Quaternary Research. Vol. 39, P: 290 – 299.

Feller W., 1968. An introduction to probability theory and its application. Third edition, Vol.1, Wiley, New York.

Fernandez B. and Salas J. D., 1999. Return Period and Risk of Hydrologic Events: I-Mathematical Formulation. Journal of Hydrologic Engineering, Vol. 4, No.4, P: 297 – 307.

Fiering M.B., 1962. On the use of correlation to augment data. American Statistical Association Journal, March 1963.

Fiering M.B., 1963. Use of correlation to improve estimates of the mean and variance. Geological Survey Professional Paper 434-C.

Fritts H.C., Guiot J., and Gordon G.A., 1990. Section 4.3: Verification, *in* Cook E. R. and Kairiukstis L.A., eds. Methods of Dendrochronology, Applications in the Environmental Sciences. Kluwer Academic Publishers. P: 178 – 185.

Gedalof Z., Peterson D. L., and Mantua N. J., 2004. Columbia river flow and drought since 1750. Journal of the American Water Resources Association. Paper No. 03073, December 2004, P: 1 –14.

Genz A., 1992. Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics. Vol. 1, No. 2, P: 141 – 149.

Gilroy E. J.,1970. Reliability of a variance estimate obtained from a sample augmented by multivariate regression. Water Resources Research, Vol. 6, No.6, P: 1595 – 1600.

Gonzalez J. and Valdes J.B., 2003. Bivariate drought analysis using tree ring reconstruction. Journal of Hydrologic Engineering, Vol. 8, No.4, P: 247 – 257.

Grygier J. C., Stedinger J.R., 1989. A generalized maintenance of variance extension procedure for extending correlated series. Water Resources Research. Vol. 25, No. 3, P: 345 – 349.

Guerrero-Salazar P. and Yevjevich 1975. Analysis of Drought Characteristics by the Theory of Runs. Hydrology Paper 80. Colorado State University, Fort Collins, USA.

Heim Ricard R., 2002. A Review of Twentieth-Century Drought Indices Used in the United States. Bulletin of American Meteorological Society, August 2002, P: 1149 – 1165.

Hidalgo H. G., Piechota T. C., and Dracup J. A., 2000. Alternative principal components regression procedures for dendrohydrologic reconstructions. Water Resources Research. Vol. 36, No. 11, P: 3241 – 3249.

Hirsch R. M., 1982. A Comparison of four streamflow record extension techniques. Water Resources Research. Vol. 18, No. 4, P: 1081 – 1088.

Hisdal and Tallaksen L., 2000. Drought Event Definition. Assessment of the Regional Impact of Droughts in Europe, Technical Report No. 6.

Jacobs P., and Lewis P., 1977. A mixed autoregressive moving average exponential sequence and point process. Advanced Applied Probability. Vol. 9, P: 87 – 104.

Jacobs P., and Lewis P., 1978. Discrete time series generated by mixtures. I: correlational and runs properties. Journal of the Royal Statistical Society. Vol. 40, No. 1, P: 94 – 105.

Jacobs P., and Lewis P., 1978. Discrete time series generated by mixtures. II: asymptotic properties. Journal of the Royal Statistical Society. Vol. 40, No. 2, P: 222 – 228.

Johnson R. A. and Wichern D. W., 2005. Applied multivariate statistical analysis. Fifth edition, 2005.

Kingery, R. K., 1992. A stochastic analysis of spatial droughts in Colorado. Msc. Thesis, Colorado State University, Fort Collins, USA.

Lane W. L., 1982. Corrected parameter estimates for disaggregation schemes. Statistical Analysis of rainfall and runoff, edited by V. P. Singh, Water resources publications, Littleton, Colorado, 1982.

Loaiciga Hugo, Michaelsen J., Garver S., Haston L and Leipnik R, 1992. Droughts in river basins of the western United States. Geophysical Research Letters. Vol. 19, No. 20, P: 2051 – 2054.

Loaiciga H. A., Huston L., Michaelsen J, 1993. Dendrohydrology and long term hydrologic phenomena. Review of Geophysics. Vol. 31, No. 2, P: 151 –171.

Loaiciga Hugo and Leipnik R, 1996. Stochastic Renewal Model of Low-flow Streamflow Sequence. Stochastic Hydrology and Hydraulics. Vol. 10, P: 65 – 85.

Matalas N. C, Jacobs B., 1964. A correlation procedure for augmenting hydrological data. Geological Survey Professional Paper 434-E.

Meko D., Stockton C.W., Boggess W.R., 1995. The tree-ring record of severe sustained drought. Water Resources Bulletin. Vol. 31, No. 5, P: 789 – 801.

Meko D., Therrell M., Baisan C., and Hughes M., 2001. Sacramento river flow reconstructed to A.D. 869 from tree rings. Journal of the American Water Resources Association. Vol. 37, No. 4, P: 1029 – 1039.

Moran M. A., 1974. On estimators obtained from a sample augmented by multiple regression. Water Resources Research. Vol.10, No.1, P: 81 – 85.

Ponce V., Rajendra P., and Ercan S., 2000. Characterization of drought across climatic spectrum. Journal of Hydrologic Engineering. Vol. 5, No. 2, P: 222 – 224.

Rossi G. and Cancelliere A., 2003. At-site and regional drought identification by REDIM model. Tools For Drought Mitigation in Mediterranean Regions, edited by Rossi G. et al. 2003. Kluwer Academic Publishers, 2003.

Salas J. D., Delleur J. W., Yevjevich V., and Lane W. L., 1980. Applied modeling of hydrologic series. Water Resources Publications, Littleton Colorado, 1980.

Salas J. D., Fu C., Cancelliere A., Dustin D., Bode D., Pineda A., and Vincent E., 2005. Characterizing the severity and risk of drought in the Poudre River, Colorado. Journal of Water Resources Planning and Management, Vol. 131, No. 5, P: 383 – 393.

Salas J. D., Tarawneh Z., Woodhouse C., and Lukas J., 2006. Revisiting methods for extending short records and reconstructing flows from tree ring indices. Journal paper to be submitted.

Santos M. A., 1983. Regional droughts: a stochastic characterization. Journal of Hydrology. Vol. 66, P: 183 – 211.

Schwager Steven J., 1983. Run probabilities in sequences of Markov-dependent trials. Journal of American Statistical Association. Vol. 78, No. 381, P: 168 – 175.

Sen Z., 1976. Wet and dry periods of annual flows series. Journal of the Hydraulics Division. Vol. 102, No. HY10, P:1503 – 1514.

Sen Z., 1979. Wet and dry periods in regional water exchange. Inputs for risk analysis in water systems, edited by Edward B et al. 1979, Water Resources Publications, Fort Collins Co, USA.

Sen Z., 1980. Regional drought and flood frequency analysis: theoretical consideration. Journal of Hydrology. Vol. 46, P: 265 – 279.

Sen Z., 1980. Statistical analysis of hydrologic critical droughts. Journal of the Hydraulics Division. Vol. 106, No. HY1, P: 99 – 115.

Sen Z., 1998. Probabilistic formulation of spatio-temporal drought pattern. Theoretical and Applied Climatology. Vol. 61, P: 197 – 206.

Sen Z., Altunkaynak A. and Ozger M., 2003. Autorun persistence of hydrologic design. Journal of Hydrologic Engineering. Vol. 8, No. 6, P: 329 – 338.

Sharma T. C., 1997. A drought frequency formula. Hydrological Sciences, Vol. 42, No. 6, P: 803 – 814.

Sharma T. C., 1998. An analysis of non-normal Markovian extremal droughts. Hydrological Processes. Vol. 12, P: 597 – 611.

Shiau J., Shen H., 2001. Recurrence analysis of hydrologic droughts of different severity. Journal of Water Resources Planning & Management, Vol. 127, No. 1, P: 30 - 40.

Shin H., and Salas J., 2000. Regional drought analysis based on neural networks. Journal of Hydraulic Engineering. Vol. 5, No. 2, P: 145 – 155.

Tarboton G. D., 1994. The source hydrology of severe sustained drought in the southwestern United States. Journal of Hydrology. Vol. 161, P: 31 – 69.

Tase, Norio, 1976. Area-Deficit-Intensity characteristics of droughts. Ph.D. Dissertation. Colorado State University, Fort Collins, USA.

Taylor H., and Karlin S.,1998. An introduction to stochastic modeling. Third edition, Academic Press, USA.

Vogel R. M., Stedinger J.R., 1985. Minimum variance streamflow record augmentation procedures. Water Resources Research. Vol. 21, No. 5, P: 715 – 723.

Woodhouse C. A., 2000. Extending hydrologic records with tree rings. Water Resources Impact. Vol. 2, No. 4, P: 25 – 27.

Woodhouse C. A., 2001. A Tree-ring Reconstruction of streamflow for the Colorado Front Range. Journal of the American Water Resources Association. Vol. 37, No. 3, P: 561 – 569.

Yevjevich V,. 1967. An objective Approach to definitions and investigations of continental hydrologic droughts. Hydrology paper 23, Colorado State University, Fort Collins, Colorado.