DISSERTATION

STOCHASTIC CHARACTERIZATION OF DROUGHTS IN STATIONARY AND PERIODIC SERIES

Submitted by

Antonino Cancelliere

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

Spring 2008

UMI Number: 3321264

INFORMATION TO USERS

COLORADO STATE UNIVERSITY

March 14, 2008

WE HEREBY RECOMMEND THAT THE **DISSERTATION** PREPARED UNDER OUR SUPERVISION BY **ANTONINO CANCELLIERE** ENTITLED **STOCHASTIC CHARACTERIZATION OF DROUGHTS IN STATIONARY AND PERIODIC SERIES** BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Richard E. Eykholt

Richard Eykholt
Physics Department

Jorge A. Ramirez

Duane C Boes

Co-Adviser Duane C. Boes
Statistics Department

Jose D. Salas

Adviser Jose D. Salas

Luis Garcia

Department Head Luis Garcia

ABSTRACT OF DISSERTATION

STOCHASTIC CHARACTERIZATION OF DROUGHTS IN STATIONARY AND PERIODIC SERIES

Stochastic modelling of droughts is a topic of great interest in water resources management. For instance, estimating drought probabilities and return periods helps in implementing risk based management decisions of water supply systems and provide useful information for drafting drought management plans. Due to the limited number of droughts that can be generally observed in historical series, the inferential approach, e.g. fitting a probability distribution to drought characteristics from an observed hydrological sample, leads to unreliable results. Furthermore, the multiyear spanning of droughts, as well as their multivariate framework requires the development of concepts and tools that differ significantly from those generally adopted to analyze other hydrological extremes, such as floods.

The research described in this dissertation includes the stochastic modelling of drought characteristics, in cases where the underlying series is either periodic or with non-negligible autocorrelation. In addition, extension of the derived models to the regional case has been explored. Moments of drought characteristics, probability distribution functions and return period have been developed in all cases, thus providing tools to characterize droughts in a broad set of situations that may occur in practice.

Drought length, accumulated deficit, and intensity in stationary dependent processes are investigated by assuming an autoregressive model for the underlying hydrological series. A Discrete Autoregressive Moving Average (DARMA(1,1)) model is adopted to model the sequence of dry-wet years that is obtained when a continuous process is clipped by a constant demand level. Further, analysis of accumulated deficit indicates that although the underlying process is stationary, the sequence of deficits is not, a feature that apparently has been overlooked by previous studies. Thus, a truncated multivariate distribution is proposed to model the deficits. Approximate expressions of the bivariate distribution of accumulated deficit and length and drought intensity and length are employed to estimate the return period of various types of drought scenarios.

Drought length in periodic stochastic hydrological series is analyzed by assuming that the dependence structure of the underlying hydrological series is a periodic lag 1 Markov chain. Moments and probability mass function of drought length are derived as well as return period of droughts with length greater or equal to a fixed value. In addition, the analysis includes the joint characterization of drought length and accumulated deficit (or intensity) in periodic dependent series. Approximate expressions for the moments and pdf's of droughts and a new formulation that enables one estimating return periods of droughts starting at a given season is presented.

Regional droughts have been analyzed in order to estimate the corresponding probabilities of occurrence and return periods. Making use of copulas concepts, approximate analytical expressions are derived for the moments of areal coverage of deficit, of areal deficit and of accumulated deficit that improves previous formulations. The corresponding approximate pdf's are also derived and employed to compute return periods of regional droughts.

The developed models have been validated by applying them to several hydrological series, ranging from precipitation to streamflows and drought monitoring indices such as Standardized Precipition Index and Palmer Drought Hydrological Index from regions characterized by different climatological regimes.

The overall conclusion of the research is that exact and/or approximate analytical expressions of probability distributions of drought characteristics derived from the statistics of the underlying hydrological series enables one a more reliable probabilistic characterization than employing the inferential approach. In addition, such analytical derivations may be useful for checking approximations or results obtained for more complex cases. The examples using a variety of water supply series and climatologic and hydrological drought indices illustrate and confirm the applicability of the analytical derivations obtained for drought characteristics and associated return periods.

Antonino Cancelliere
Department of Civil and Environmental Engineering
Colorado State University
Fort Collins, Colorado 80523
Spring 2008

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor, Dr. Jose D. Salas and co-advisor, Dr. Duane C. Boes, for their guidance and support throughout these years. I would like to thank also Dr. Jorge Ramirez for his continuous encouragement, Dr. Richard Eykholt, for his valuable insights, and Dr. Peter Brockwell, for his helpful discussions in and out the golf courses.

This long trip would not have been possible without the continuous invaluable support and encouragements of my "Maestro" Pippo Rossi, to whom goes my deepest gratitude for being an example as a teacher, as a researcher, but most of all as a friend.

I would most like to thank Melania for her patience and help during all these wonderful years, as well as Francesco, Mariachiara and Nicoletta, for bearing with my long absences from home.

This American journey gave me the opportunity to meet outstanding people that have honored me with their friendship: Mike and Jackie Hobbins, Claudio and Margherita Meyer, Peter and Darcy Molnar, Kyra Ryan, and Oli Sveinsson. Without them, my years in Fort Collins would have been undoubtely more sober, but surely not as wonderful as they have been.

Finally my last thank goes to my mother and father, for their being with me and for me all the time.

And then we will act as guardians for your crops and your vines, so that neither drought afflict them, nor excessive wet weather.
-Aristophanes, Clouds

Very great distress was caused this year by a drought. Not only was there an absence of water from the heavens, but the earth, through lack of its natural moisture, barely sufficed to keep the rivers flowing. In some cases the want of water made the cattle die of thirst.
-Livius, Titus : The History of Rome, Vol. I

The year was marked by a drought and failure of the crops. It is recorded that no rain fell for six months.
-Livius, Titus : The History of Rome, Vol. VI

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Preface

Drought is a complex phenomenon that can be characterized by its initiation, duration, amount of water deficit (accumulated deficit or average deficit or intensity), termination, and spatial extent. These characteristics are usually determined by comparing a water supply series versus a water demand series. Due to the stochastic nature of water supply series such as precipitation or streamflows, the corresponding drought characteristics are random and therefore a correct approach to their analysis is by means of stochastic methodologies.

Drought identification and characterization represents an important step for correct water resource management. Probabilistic characterization of droughts is extremely important in those regions where an accurate water resources planning and management requires a detailed knowledge of water shortage (Frick et al., 1990; Shiau and Shen, 2001; Cancelliere and Rossi, 2003). According to the definition of drought based on run theory, introduced by Yevjevich (1967), a drought event is defined as a succession of consecutive intervals where the hydrological variable under investigation remains below a threshold level. Once such a threshold x_o is fixed, the main drought characteristics or properties are identified as drought length L (length of negative run or number of consecutive intervals for which $x_t < x_o$, followed and preceded by at least one interval where $x_t \geq x_o$, as accumulated deficit D (sum of the deficits $S_t = x_o - x_t$ over the drought duration L), and as average drought intensity, defined as the ratio of severity to the drought length $I = D/L$. Such drought definition, although simple, yet yields an objective identification and characterization once the threshold level x_o has been selected. It can also be extended to the regional case, by introducing another threshold which represents the percentage of areas affected by deficit within a region above which a drought can be considered affecting significantly the whole region.

Despite the limitations, the main advantage of using run method to identify droughts is its suitability for a theoretical probabilistic analysis of the drought characteristics. Indeed, the relative simplicity of the drought definition allows for the derivation of the probabilistic features of the characteristics assuming a given stochastic structure for the underlying hydrological series. Such theoretical derivation is particularly important in light of the relatively short length of available observations of the hydrological variable, which limits the applicability of an inferential approach to find the probability distributions of drought characteristics. Further, knowledge of the probabilistic structure of droughts allows to extrapolate, within reasonable limits, information about return period of such events, beyond the length of the observed series.

Probabilistic characterization and modeling of drought occurrences has been the subjet of several works since the pioneering papers by Yevjevich and his co-authors (Yevjevich, 1967; Downer et al., 1967; Saldarriaga and Yevjevich, 1970; Millan and Yevjevich, 1971). Along the years such a topic has received much attention both from a scientific as well as practical standpoint, due to its implications for water resources management. Nonetheless, despite the efforts, derivation of probability distributions of drought characteristics and of related return period from the stochastic structure of the underlying hydrological variable is still an open problem except for a few, simple cases.

This dissertation attempts to fill-in some of the missing gaps in drought modelling and as a such, its main contribution to stochastic hydrology is the development of new models or improvements of existing ones for the stochastic characterization of droughts, both at site and in a region, with particular reference to stationary or periodic hydrological series.

1.2 Objectives

The research aims at providing a framework for the modelling of drought occurrences under the hypothesis of stationary or periodic hydrological series. In particular, moments, probability and return period concepts are investigated to model drought occurrences both at site and in a region and to provide tools potentially useful for water resource managers. Emphasis is given to derive explicit expressions of moments, probabilities and return period of droughts as a function of the stochastic structure of the underlying series. Specific objectives are:

- To derive analytical expressions for the probability density functions of at site drought characteristics length, accumulated deficit and intensity, both marginally and jointly, in the cases of independent or dependent time series either stationary or periodic. Assuming a stochastic model for the underlying hydrological series, exact and approximate pdf's of drought characteristics are derived as a function of the parameters of the underlying model and of the threshold level.

- To derive analytical expressions for the probability density functions of regional drought characteristics both marginally and jointly, for spatially dependent time series. Exact and approximate pdf's of drought characteristics are derived as a function of the cross variance-covariance of the underlying series, of the deficit threshold level and of the areal threshold

- To develop procedures to estimate the return period of accumulated deficit, intensity and drought length in stationary and periodic series. The approach proposed by Shiau and Shen (2001) is extended in order to derive return period of severity for stationary series on the basis of the marginal distributions of severity. Extension to the bivariate periodic case, taking into account jointly either accumulated deficit and length or intensity and length, is also pursued by means of the joint bivariate distribution of the two drought characteristics.

- to extend the return period analysis to the case of regional droughts, thus enabling to better characterize the occurrence of droughts over a region

1.3 Chapters outline

This dissertation is organized in four main chapters in the form of papers, and is bookended at the start by this introductory chapter and by a general chapter on drought identification and characterization and at the end by a summary and conclusions chapter. Given such structure, there may be some repetions of concepts, especially with regard to common methodologies as they appear across the different chapters. This has been kept to a minimum within the constraints that broadly each chapter stands alone as a ready-to-submit or already published paper. However, the numbering of equations, tables, figures, and sections follows a dissertation format and the references for all chapters are listed at the end.

After a review of literature on drought identification and characterization, in Chapter 3 droughts in stationary dependent processes are investigated. First, an analysis of drought length is presented, which indicates that when the underlying hydrological series possesses strong persistence a DARMA(1,1) model is adequate to model the sequence of dry-wet years that is obtained when a continuous process is clipped by a constant demand level. A comparison with the Markov model shows that DARMA(1,1) models are better able to capture the persistence in the series and that in particular, the resulting pmf of drought length fits well with the corresponding sample pmf of droughts identified on long series generated by simulation of an autoregressive process AR(1). Further, analysis of accumulated deficit indicates an interesting feature that apparently has been overlooked in previous studies, namely that although the underlying process is stationary, the sequence of deficits is not. In order to overcome the difficulties related to modelling non i.i.d. variables, the proposed approach is to model the deficits by means of a truncated multivariate distribution. This enables to compute the moments of accumulated deficit for droughts of fixed lengths, which can be used to derive approximate expressions of the bivariate distribution of accumulated deficit and length and drought intensity and length. The derived pdf are then employed to estimate analytically the return period of various types of critical droughts. Application to four annual streamflow series shows the suitability of the derived methodology to model droughts and drought return periods in dependent hydrological series. The results illustrated in chapter 3 have been partially published in Cancelliere et al. (2003); Salas et al. (2005).

Chapter 4 analyzes drought length in periodic stochastic hydrological series. Moments and probability mass function of drought length are derived analytically as a function of the dependence structure of the underlying hydrological series. Then, return period of droughts with length greater or equal to a fixed value is also derived analytically. Finally the derived analytical expressions are applied to several series including precipitation, streamflow and drought monitoring indices such as Standardized Precipition Index and Palmer Drought Hydrological Index. The investigations presented in Chapter 4 have been published in Cancelliere and Salas (2004).

In Chapter 5 the accumulated deficit in periodic series is investigated. In particular, the moments of accumulated deficit of droughts starting in a given season or in any season are derived, and approximate expressions for the corresponding pdf are also proposed. The return period formulation proposed by Shiau and Shen (2001) is then extended to the case of droughts starting in any season, and a new formulation that enables to estimate return periods of droughts starting at a given season is presented. The applicability of the methodology is then tested with reference to seasonal precipitation series.

In Chapter 6 the methodology proposed by Santos (1983) to model regional droughts is revised and extended in order to estimate return periods of drought. In particular, new analytical expressions are derived for the moments of areal coverage of deficit, of areal deficit and of accumulated deficit, based on more realistic hypotheses about the distributions of the aforementioned characteristics. Also, 2nd order moments that in some cases were not derived by Santos (1983) are derived making use of copulas concepts. The derived moments, pdf and return period are applied for the characterization of regional droughts in two case studies, namely Sicily island in Italy and the Front Range region in Colorado. The results of the application indicates that the derived expressions are able to model fairly well the occurrences and return period of droughts.

CHAPTER II

DROUGHT IDENTIFICATION AND CHARACTERIZATION

2.1 Drought definition

Defining droughts, as opposed to other extreme hydrological phenomena, is a difficult issue, due to the widely different and often conflicting concepts used in literature related to the particular academic field of study (e.g., hydrology, meteorology, agriculture and economy). In addition, the definition of drought may sensibly change according to the climate of the investigated region and the specific water use. Therefore, a first important problem one has to face in the study of drought is the preliminary identification of the component of the hydrological cycle of primary interest and of the type of human activity affected by water deficit.

According to Dracup et al. (1980), an accurate selection of the drought identification method requires to properly define:

1. the nature of the water deficit under study;

2. the aggregation time scale of the time series to be studied (month, season or year);

3. the analytical criterion to distinguish drought events from other events in the time series;

4. the regionalization method, in order to take into account the regional features of the phenomenon.

Drought definition strongly depends on the nature of the water deficit to be studied. Water deficit may be related to precipitation, streamflow, soil moisture, or any combination of these hydrometeorogical variable. Precipitation deficit from conventionally normal conditions is usually considered of interest in determining the cause of drought events and is generally defined as meteorological drought. On the other hand, if one is interested in determining the effects and impacts of drought events, attention should be focused on streamflow or groundwater flow and soil moisture deficit, respectively defined as hydrologic and agricultural drought.

After selecting the nature of water deficit to be studied, the unit of time to be used as the aggregation time scale or averaging period for the variable under consideration has to be properly chosen. Obviously, the selection of the time scale depends on the purpose for which the study is intended. The choice of such time scale affects two main aspect of the analysis:

- the sample size of events to be studied;

- the presence of autocorrelation and in general of time dependence in the investigated series.

The sample size has to be large enough to guarantee the population moments are well approximated by the sample statistics. Obviously, as the aggregation time scale increases, the sample size gets smaller and so does generally the autocorrelation or time dependence. Therefore, a short time scale results in a larger sample size and larger dependence, while a long averaging period results in a smaller sample size and smaller time dependence.

The third point, that is the criterion for drought identification in a time series, will be discussed in detail in the next paragraph.

Drought is intrinsically a spatially distributed phenomenon, therefore a characterization of its areal features represents an important step for the assessment of its entity and effects (Rossi et al., 1992). Nevertheless, most of the methods, based on an at-site approach, do not take explicitly into account the spatial characterization of drought, which is usually carried out at a second stage by spatially interpolating drought characteristics previously computed at a given number of sites. Such interpolated values can then be represented graphically for example by isolines maps. A few methods attempt to incorporate spatial information within drought characteristics for example to derive the relationship between a fixed value of a drought index and a percentage of the area of the region interested by the chosen value (Rossi, 2003), to compute regional drought characteristics expressing areal coverage and areal weighted deficit as in the extended run method (Yevjevich, 1967; Tase, 1976).

2.2 Identification and characterization of drought events

Since the sixties, several methods have been proposed for drought identification and characterization, generally based on a specific drought definition. For instance, Foley (1957) and Herbst et al. (1966) used precipitation data to characterise meteorological drought, Palmer (1965) included also evapotraspiration, runoff and soil moisture to define agricultural drought, while Russel et al. (1971) applied a balance between available water resources and municipal demands to characterise urban water supply shortage. More recently new indices have been proposed to investigate drought at different time scales in order to provide adequate information to different users (i.e., Standard Precipitation Index (McKee et al., 1993)).

Drought indices can be successfully applied for drought monitoring since they provide either directly or indirectly information on current water resources conditions. However, the developed indices generally fail to provide a complete and objective description of drought events, and therefore they are not well suited when the purpose is to identify and characterize historical droughts (Rossi and Cancelliere, 2003).

One of the first work concerned with the identification of droughts is due to Yevjevich (1967), where the use of the *run* concept is proposed as a method to identify and characterize in an objective way drought events and their statistical properties. A run is generally defined as a sequence of observations of the same kind preceded and succeeded by one or more observations of a different kind (Yevjevich, 1972). A drought is then defined as a run of intervals where the hydrometeorological variable under investigation remains below a chosen truncation level or threshold.

More specifically, given a discrete time series X_t, $t = 1, 2, ...n$ and a truncation level x_0, it is possible to identify positive and negative deviations according to the sign of the difference between the observed values of the variable and the fixed threshold. Then, a drought is defined as a sequence of intervals (negative run) characterized by negative deviations preceded and followed by at least one interval with deviation of the opposite sign. Negative deviations are usually referred to as *deficit*, whereas the corresponding run is termed as drought. By analyzing a given drought or negative run s, it is possible to identify the following characteristics:

- drought duration $L(s)$, defined as the number of consecutive intervals where the variable remains below the threshold;
- drought severity or cumulative deficit $D_c(s)$, defined as the sum of the negative

deviations, extended to the whole drought duration;

- drought intensity or magnitude $I(s)$, defined as the ratio between drought severity and duration.

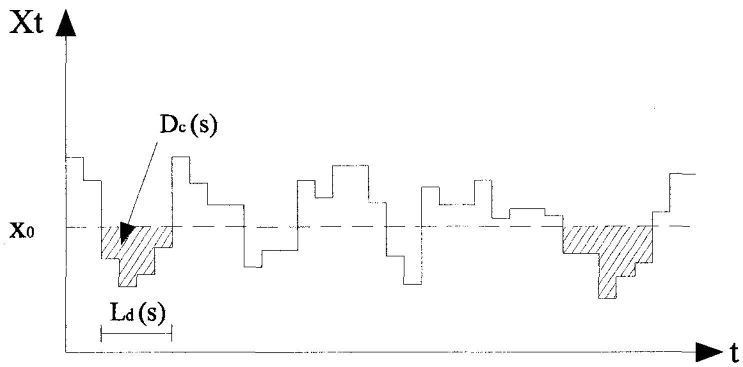


Figure 2.1: Droughts identified on hydrological series and their characteristics

More precisely, by letting S_t be the deficit in the interval t:

$$S_t = (x_0 - x_t) \cdot I[x_t] \tag{2.1}$$

where $I[x_t] = 1$ if $x_t < x_0$ and 0 otherwise.

The drought duration is given by:

$$L(s) = t_f - t_i + 1 \tag{2.2}$$

where t_f and t_i are such that $S_t > 0$ for $t_i \leq t \leq t_f$ and $D_{t_i - 1} = 0$, $D_{t_f + 1} = 0$.

The cumulated deficit or severity can be expressed by:

$$D_c(s) = \sum_{t=t_i}^{t=t_f} S_t = \sum_{t=1}^{L(s)} (x_0 - X_t) \quad for \quad X_t \geq x_0 \tag{2.3}$$

and finally the drought intensity is given by:

$$I(s) = \frac{D_c(s)}{L(s)} \tag{2.4}$$

Once the above characteristics have been computed for each drought, it is possible to characterise the whole series by considering the average, the maximum or the minimum of each characteristic and the total number of droughts.

The selection of the threshold is a function of the type of water deficit being studied. It may be a constant in the case of a non periodic (*e.g.* annual) time series or a seasonally varying truncation level in the case of a stationary periodic series.

In general, the threshold is chosen in order to be representative of the average demand related to the considered water resources (Yevjevich et al., 1983; Rossi et al., 1992). Usually, a measure of the central tendency of the hydrometeorological variable, such as the long-period mean or (or the median), is adopted. Theoretically, the use of the median as threshold level yields an identical number of drought and non-drought periods, while the use of the mean yields to the same total deviation from the threshold for droughts and non-droughts, which turns out in the same mean severities for both the two events in the series.

Other possible choices for the threshold include a fraction of the mean (Clausen and Pearson, 1995), a value corresponding to a given non-exceedence probability (Zelenhasic and Salvai, 1987), or a level defined as a function of the mean and the standard deviation (Ben-Zvi, 1987).

The method of run can be easily extended to regional drought analysis, by considering another truncation level A_0 representing a critical area, *i.e.* a value of the area affected by deficit above which a regional drought occurs. In this case, the area $A_d(t)$ describing the part of the region under study which, in a given time interval

t, is affected by a drought has determined by adding the areas of influence of single stations characterized by a deficit. If the sum of such areas is greater than the critical area, the drought properties are computing by considering only the successive intervals where $A_d(t) > A_0$.

The main advantages of defining drought in terms of run is the relative objectivity of the identification and characterization process, once the threshold level is defined. Thus the run method represents an ideal procedure for carrying out an analysis oriented to define best drought mitigation alternatives. Another advantage consists in the possibility of deriving the probabilistic features of drought characteristics analytically based on the stochastic properties of the underlying variable. This possibility is not limited to simple cases where time dependence can be neglected but also when a Markov chain structure is assumed for the variable of interest.

It may be worthwhile to note that other run definitions have been proposed in literature, depending on the specific objective for which definition is adopted. For instance, in defining a run of length r, Feller (1968) stated that "a sequence of n observations has as many runs of length r as there are non overlapping uninterrupted successions of exactly r events of the same kind". On the other hand, Schwager (1983) used a definition where the successions of events of the same kind may overlap. However he stated his analysis corresponds to that one carried out by Feller as the discussion centers on the first occurrence of the success run.

Therefore, with reference to the sequence $FSSSSF$, it contains:

- only one success run of length $r = 4$ according to Yevjevich;
- one success run of length $r = 4$ and $r = 3$ (ending at trial 5 and 4 respectively) and two success runs of length $r = 2$ (ending at trails 3 and 5) according to Feller;

- one success run of length $r = 4$, two of length $r = 3$ (ending at trials 4 and 5) and three success runs of length $r = 2$ (ending at trials 3, 4 and 5) according to Schwager.

2.3 Stochastic characterization of drought length

The properties of run-length and its application to drought have been widely investigated in literature. Indeed, the origins of the theory of runs probably trace back to the end of the eighteen century when there was a large interest in games of chance de Moivre (1967).

The earlier hydrologic studies concerned with such a concept were carried out by Downer et al. (1967), who studied the distribution and the statistical moments of positive and negative run-lengths for a sequence of independent identically distributed random variables, Llamas and Siddiqui (1969) who also considered the case of a two-state Markov process, and Saldarriaga and Yevjevich (1970) who derived analytical expressions to determine the probabilities of sequences of wet and dry years of specified lengths either for an independent series or a dependent stationary series of a variable which follows the first-order linear autoregressive model. In particular, for the latter case they found numerical solutions by computer integration of expansion equations defining such probabilities.

In order to derive the distribution of drought duration L, it is convenient to introduce a two state model indicating, for instance, a deficit by 0 and a surplus by 1. This approach enables to simplify the analytical formulation of the problem as well as to adopt discrete valued stochastic models, such as Markov chains.

According to the markovian hypothesis, the probability of occurrence of a given event at time t, depends on what happens in the previous $t - n$ time intervals. Such a dependence structure is suitable to various hydrological series with a certain approximation. In particular, if the outcome of the process at time t can be defined by using only the outcome at time $t-1$, the process is a first order Markov chain or a simple Markov chain.

Let y_t a boolean stationary variable such that:

$$p_0 = P[y_t = 0] = P[x_t < x_0] \qquad p_1 = P[y_t = 1] = P[x_t \geq x_0]$$

Further, let's consider the transition probabilities:

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \quad t = 1, .., \omega$$

where p_{ij} is given by:

$$p_{ij} = P[y_t = j | y_{t-1} = i] \quad i = 0, 1; \quad j = 0, 1$$

The probability of a negative run starting at time t, of length $L = l$ is:

$$P[L = l] = P[y_{t-1} = 1, y_t = 0, ..., y_{t+l-1} = 0, y_{t+l} = 1]$$

which, for a markovian process, can be written as:

$$P\left[L = l\right] = p_1 p_{10} \overbrace{p_{00} p_{00} \dots p_{00}}^{l-1 \text{ terms}} p_{01} = p_1 p_{10} \left(p_{00}\right)^{l-1} p_{01} \tag{2.5}$$

In particular, the probability that, once a negative run has just started, its length is equal to l, is expressed by:

$$f_L(l) = \frac{P[L=l]}{p_1 p_{10}} = (p_{00})^{l-1} p_{01} = (1-p_{01})^{l-1} p_{01} \qquad l = 1, 2, ... \tag{2.6}$$

where $p_1 p_{10}$ is the occurrence probability of a negative run. If x_t is an independent time series, it follows that $p_{00} = p_{10} = p_0$ and $p_{11} = p_{01} = p_1$, thus:

$$f_L(l) = {p_0}^{l-1} p_1 = (1-p_1)^{l-1} p_1 \qquad l = 1, 2, ... \tag{2.7}$$

It is easy to show the above probability mass functions are geometric with parameters respectively p_{01} and p_1, which can be estimated by applying a frequency analysis on y_t.

The equations related to the case of independent stationary series were first analytically derived by Feller (1968) on the basis of the sequential property of runs. Afterwards, Sen (1976), by using the same approach, derived explicit formulation for the case of lag-one Markov process. Such results have been widely applied by other authors in successive investigations showing that the suitability of the geometric distribution to model drought length.

2.4 Stochastic characterization of drought severity

Many studies have been carried out in order to analytically determine the probabilistic behaviour of drought severity, assuming a given stochastic structure of the underlying hydrological series. Although such an approach has found large application in literature, exact analytical derivation of probability distribution of drought severity and/or of the bivariate distribution of drought severity and drought duration, remain unsolved problems because of the difficulty to obtain a solution in closed form even assuming a simple stochastic structure for the underlying variable.

The first objective study related to drought severity has been performed by Downer et al. (1967), who derived analytical expressions for statistical moments of run-sum by making use of the moment and cumulative generating functions on the basis of normal independent processes.

Then, Sen (1977) has derived general expressions for the expected values and the variance of run sum by using the concept of conditional probability. According to the drought definition given by the run method, if x_t is the value assumed at time t by the hydrological variable, the drought severity D_c is the sum of the single deficits $S_t = x_0 - x_t$ over the drought duration L. It follows that drought severity D_c is a random variable function of the underlying hydrological process X_t and of the threshold x_0 and it can be formally defined as:

$$D_c = \sum_{t=1}^{L} S_t = \sum_{t=1}^{L} (x_0 - X_t) \qquad for \;\; X_t \leq x_0 \tag{2.8}$$

where S_t and L are both random variables.

Theoretically, the marginal pdf of D_c can be determined as:

$$f_{D_c} = \sum_{k=1}^{\infty} f_{D_c|L=k}(d_c) \cdot f_L(k) \tag{2.9}$$

where $f_{D_c|L=k}(d_c)$ is the pdf of drought severity conditioned by a fixed duration k and $f_L(k)$ is the pdf of drought duration L given by Equations (2.6) and (2.7) for two state Markov processes and independent processes respectively.

Following Sen (1977), based on equation 2.9, one can write:

$$\mathcal{E}\left[D_c\right] = \sum_{k=1}^{\infty} \mathcal{E}\left[D_c | L = k\right] \cdot f_L(k) \tag{2.10}$$

Taking into account equation 2.8, the conditional expected value of $D_c|L = k$ can be evaluated as:

$$\mathcal{E}\left[D_c|L=k\right] = \mathcal{E}\left[\sum_{t=1}^{L} S_t|L=k\right] = \mathcal{E}\left[\sum_{t=1}^{L=k} S_t\right] = k \cdot \mathcal{E}\left[S_t\right] \tag{2.11}$$

Substitution of equation 2.11 into equation 2.10 yields to:

$$\mathcal{E}\left[D_c\right] = \mathcal{E}\left[S_t\right] \sum_{k=1}^{\infty} k f_L(k) = \mathcal{E}\left[S_t\right] \cdot \mathcal{E}\left[L\right] \tag{2.12}$$

It can be noticed that the same result obtained by Sen (1977) can be derived directly by taking the expected value of $D_c = \sum_{t=1}^{L} S_t$ by means of the following property of conditional expectation:

$$\mathcal{E}\left[g(Y)\right] = \mathcal{E}\left[\mathcal{E}\left[g(Y)|X\right]\right]$$

Again by following the procedure adopted by Sen (1977), the variance of drought severity can be derived as:

$$\mathrm{Var}\left[D_c\right] = \mathcal{E}\left[{D_c}^2\right] - \mathcal{E}\left[D_c\right]^2 \tag{2.13}$$

where:

$$\mathcal{E}\left[{D_c}^2\right] = \sum_{k=1}^{\infty} \mathcal{E}\left[D_c^2|L=k\right] \cdot f_L(k) \tag{2.14}$$

Since:

$$\begin{aligned} \mathcal{E}\left[D_c^2|L=k\right] &= \mathrm{Var}\left[D_c|L=k\right] + \mathcal{E}\left[D_c|L=k\right]^2 = \\ &= \mathrm{Var}\left[\textstyle\sum_{t=1}^{L=k} S_t\right] + \mathcal{E}\left[D_c|L=k\right]^2 \end{aligned} \tag{2.15}$$

reminding the property of the variance of the sum of random variables and by substituting equation 2.11, equation 2.15 can be rewritten as:

$$\mathcal{E}\left[D_c^2|L=k\right] = k\mathrm{Var}\left[S_t\right] + 2\sum_{i<j}\sum \mathrm{Cov}\left[S_i, S_j\right] + k^2\mathcal{E}\left[S_t\right]^2 \tag{2.16}$$

Thus:

$$\begin{aligned}\mathcal{E}\left[D_c^2\right] = \mathrm{Var}\left[S_t\right]\mathcal{E}\left[L\right] + 2\sum_{i<j}\sum \mathrm{Cov}\left[S_i, S_j\right]\sum_{k=1}^{\infty} f_L(k) + \\ +\mathcal{E}\left[L^2\right]\mathcal{E}\left[S_t\right]^2\end{aligned} \tag{2.17}$$

the substitution of which in equation 2.13 after some algebra yields:

$$\begin{aligned}\mathrm{Var}\left[D_c\right] = \mathrm{Var}\left[S_t\right]\mathcal{E}\left[L\right] + 2\sum_{i<j}\sum \mathrm{Cov}\left[S_i, S_j\right]\sum_{k=1}^{\infty} f_L(k) + \\ +\mathrm{Var}\left[L\right]\mathcal{E}\left[S_t\right]^2\end{aligned} \tag{2.18}$$

Once again the same result can be derived directly by taking the variance of $D_c = \sum_{t=1}^{L} S_t$ by means of the following property:

$$\mathrm{Var}\left[Y\right] = \mathcal{E}\left[\mathrm{Var}\left[Y|X\right]\right] + \mathrm{Var}\left[\mathcal{E}\left[Y|X\right]\right]$$

It should be noted that for the case of independent drought events the covariance terms are clearly null. The expected value and variance of S_t can be computed once that the cumulative distribution function of single deficit is known. It can be shown that such a distribution is just the truncated distribution of the underlying random variable X_t with changed sign and shifted compared with the threshold level x_0.

As equation 2.9 points out, the probability distribution of drought severity D_c is a function of the distribution of $D_c|L$. Indeed, since $D_c|L = \sum_{t=1}^{L} S_t$, it should be possible, at least from a theoretical standpoint, to derive the distribution of $D_c|L$, provided the distribution of single deficit S_t is known. In practice, the above derivation can be carried out in closed form only in very few cases, depending on the distribution of S_t, which, as before mentioned, depends on the distribution of the underlying process X_t and on the threshold x_0. Therefore a traditional approach widely used to overcome this difficulty consists in assuming a parametric distribution for D_c, which can be fitted to a series of observed drought severities. Sen (1980b) has stated that, since the deficit sum D_c (drought severity) corresponding to a drought of a given duration L, is a summation of L random variables S_t, by virtue of the central limit theorem its distribution can be approximately given by a normal distribution. In particular, he applied this result in order to derive an analytical formulation of the probabilistic behaviour of maximum deficit sum for the case of stationary Markov processes on the basis of the theory of extremes of random number introduced into hydrology by Todorovich (1970). The same approach has been subsequently applied by Sharma (1995) to predict the expected value of the maximum deficit sum corresponding to a specific period T regarded equivalent to a return period, for different values of the threshold. It should be pointed out however that the Normal assumption for the severity is reasonable only for "large" L (again by virtue of the central limit theorem), which therefore casts some doubts on the above results when dealing with the relatively short length of droughts generally observed in hydrological series.

Making use of the theoretical results derived by Sen, Guven (1983) has derived approximate expressions for the probabilities of extreme droughts by assuming the distribution of drought severity approximately two parameter-gamma. Besides, he evaluated the parameters of such a distribution in terms of the threshold level x_0 and the first order autocorrelation coefficient of the underlying variable considered normally distributed and generated by a lag-one Markov process. Zelenhasic and Salvai (1987) and Mathier et al. (1992) have developed an approach to perform statistical frequency analysis of drought severity using the exponential distribution. Recently Shiau and Shen (2001) have applied the gamma distribution to model the conditional distribution of $D_c|L$ from generated data. More specifically, such distribution has been fitted to different categories of drought events with specific duration, determined by applying the run analysis to the generated data, then the distribution of drought severity is derived by equation 2.9.

2.5 Estimation of drought return period

A common approach used in hydrology and hydraulic engineering to define the risk of failure of a particular system, such as a reservoir, consists in deriving the return period of extreme hydrologic events.

The return period can be defined in different ways for different applications. Some authors (e.g., Lloyd, 1970; Loaicigica and Mariño, 1991; Shiau and Shen, 2001) have assumed the return period as the average elapsed time between occurrences of critical events (i.e., floods or drought events). This definition is useful when, provided that a critical event has just occurred, the interest is in the arrival time of the next one. An alternative definition of return period, widely used in flood analysis, is the average number of trials required to the first occurrence of a critical event (e.g., Vogel, 1987; Bras, 1990; Douglas et al., 2002). This other definition assumes that a finite time τ has elapsed since a critical event has occurred and the interest is in the remaining waiting time for the occurrence of the next critical event. These two definitions can be considered equivalent when $\tau = 0$ and in practice, they lead to the same result in simple cases, such as those related to independent events.

Fernandez and Salas (1999) have considered both definitions above to determine return period for different types of hydrological events. In particular, giving more emphasis to the second approach, they have proposed to estimate the return period of an event $D \geq D_0$ as:

$$T = \mathcal{E}\left[N\right] = \sum_{n=0}^{\infty} n f_{n,D_0} \tag{2.19}$$

where N is the number of trials (e.g. years) needed to the first occurrence of such an event and f_{n,D_0} is the probability that $D \geq D_0$ occurs for the first time at the $n th$ trial. More specifically, for this purpose they have used a recursive algorithm proposed by Schwager (1983), namely:

$$f_{n,r} = p^r \left(1 - S_{n-r,r} - \sum_{i=1}^{r-1} f_{n-i,r} p^{i-r}\right) \quad n > r \tag{2.20}$$

where r is equal to the drought critical duration, p is the exceedance probability of X_t compared to the threshold x_0 and $S_{n,r}$ is the probability that $D \geq r$ has occurred at or before trail n.

By solving equation 2.19 based on equation 2.20, they have also provided procedures for estimating return periods of (i) events consisting of runs of independent Bernoulli trials (i.e., meteorological droughts) and (ii) events consisting of runs in Markov dependent trials (i.e., hydrological droughts), with duration greater than or equal to a certain value. In particular, for the case of independent drought events, the authors have shown that the return period of a drought of length r can be evaluated both by equation 2.19 with the Schwager's solution and by the formula given by Feller (1968):

$$T = \mathcal{E}\left[N\right] = \frac{1 - p^r}{(1 - p)p^r} \tag{2.21}$$

which has been originally proposed to determine the expected interarrival time for independent Bernoulli events based on a slightly different definition of run compared with the one used by Schwager (cfr. Par. 2.4). It is worth noting that for $r = 1$ equation 2.21 reduces to the well known formula $T = 1/p$, widely used in flood analysis.

Chung and Salas (2000) estimated the return period of drought events with given duration for dependent hydrological drought events by using DARMA models, which are more adequate than Markovian models for modeling hydrological processes with a more persistent time dependence. Results related to periodic PDARMA models are presented in Chung (1999).

Shiau and Shen (2001) have derived the return period drought severity as the expected value of the elapsed time between two successive droughts with severity greater than or equal to a fixed value. Such an interarrival time can be formally written as:

$$T_{D_c} = \sum_{i=1}^{N_{d_c}} L_i \tag{2.22}$$

where L_i is the drought interarrival time between any two successive drought events (i.e. $L_i = L_{d_i} + L_{n_i}$) and N_{d_c} is the number of drought events until the occurrence of the next drought with severity greater than d_c (refer to Fig. 2.2 for further details). As above mentioned, the return period can be defined as the expected value of T_{D_c}, which through the previous equation can be expressed as:

$$\mathcal{E}\left[T_{D_c}\right] = \mathcal{E}\left[\sum_{i=1}^{N_{d_c}} L_i\right], \quad i = 1, ..., N_{d_c} \tag{2.23}$$

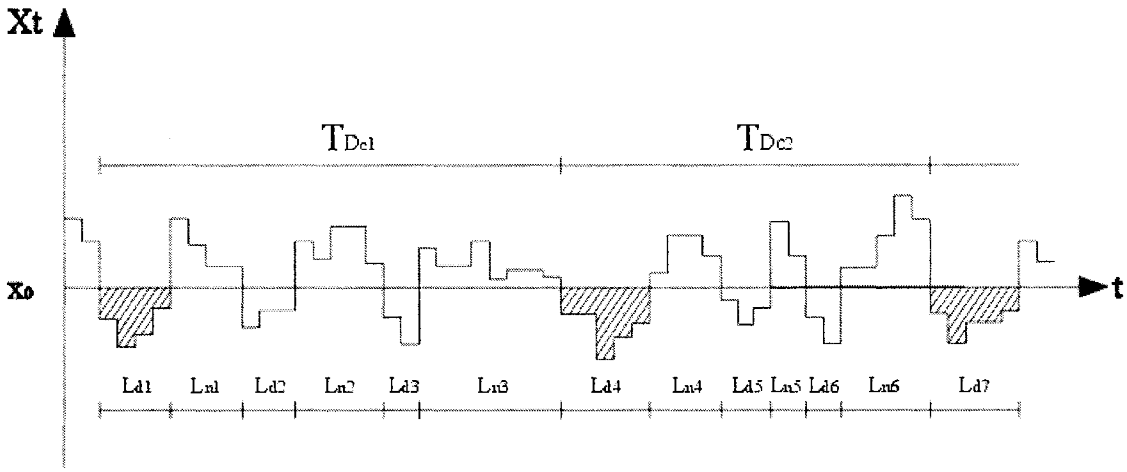


Figure 2.2: Interarrival time T_{D_c} between droughts with severity $\geq d_c$ (represented by hatched area)

Making use of a property of conditional expectation (Mood et al., 1974):

$$\mathcal{E}\left[\sum_{i=1}^{N_{d_c}} L_i\right] = \mathcal{E}\left[\mathcal{E}\left[\sum_{i=1}^{N_{d_c}} L_i | N_{d_c}\right]\right] \tag{2.24}$$

and reminding that for L_i identically distributed,

$$\mathcal{E}\left[\sum_{i=1}^{N_{d_c}} L_i | N_{d_c}\right] = N_{d_c}\mathcal{E}\left[L\right] \tag{2.25}$$

equation 2.23 becomes:

$$\mathcal{E}\left[T_{D_c}\right] = \mathcal{E}\left[N_{d_c}\right]\mathcal{E}\left[L\right] \tag{2.26}$$

Let $F_{D_c}(d_c)$ be the cumulative distribution function (cdf) of drought severity. Recalling the definition of N_{d_c}, it can be proven that it has a geometric distribution:

$$P(N_{d_c} = n) = F_{D_c}(d_c)^{n-1}[1 - F_{D_c}(d_c)] \tag{2.27}$$

with expected value:

$$\mathcal{E}\left[N_{d_c}\right] = \frac{1}{1 - F_{D_c}(d_c)} \tag{2.28}$$

Combining Equations (2.25) and (2.28), Shiau and Shen (2001) have found out that the return period can be evaluated by the following equation:

$$\mathcal{E}\left[T_{D_c}\right] = \frac{\mathcal{E}\left[L\right]}{1 - F_{D_c}(d_c)} \tag{2.29}$$

Hence, the evaluation of the return period of a drought event with severity D_c requires the knowledge of the expected value of interarrival time L and the cdf of drought severity D_c.

CHAPTER III

JOINT ANALYSIS OF DROUGHT LENGTH AND ACCUMULATED DEFICIT IN STATIONARY SERIES

3.1 General

When analyzing drought characteristics, the traditional inferential approach, based on fitting a probability distribution to drought characteristics identified on an observed hydrological series, is not generally feasible due to the limited number of droughts that can be identified even on relatively long observed series. This is particularly true when a multivariate analysis, for instance of accumulated deficit and length, is sought. In order to overcome such difficulty, some authors have attempted to derive analytically exact or approximate expressions for the probability distribution of drought characteristics, based on the stochastic properties of the underlying hydrological series and on the threshold level (Llamas and Siddiqui, 1969; Sen, 1976, 1977; Loaicigica and Leipnik, 1996; Bonaccorso et al., 2003; Salas et al., 2005). The above approaches however have usually neglected the time persistence observed in hydrological series, which generally has been modelled by means of a Markov chain model. In this chapter a general framework to derive probability distributions and return period of droughts in the case of serially dependent time series is investigated. More specifically, first the probability distribution of drought length arising from an autoregressive AR(1) model will be analyzed. It will be shown that when an AR(1) process is clipped by a constant threshold level, the resulting binary process does not follow a lag 1 Markov chain, and therefore the widely used geometric distribution to model drought length can only be considered as an approximation. In agreement with previous studies (Salas et al., 2001, 2005) results seem to indicate that a Discrete Autoregressive Moving Average DARMA(1,1) model represents a good approximation to model the sequence of deficit and surpluses.

Then, the moments of accumulated deficit conditioned on a fixed length are analyzed for the case of serially time dependent series. In particular it will be shown that due to the time dependence, the deficit series that is obtained by clipping a stationary series with a constant threshold is not identically and independently distributed (i.i.d). Thus, a fully multivariate approach, based on a truncated multivariate normal distribution, will be adopted to compute the moments of accumulated deficit. In order to overcome the numerical limitations due to the complexity of the approach, empirical relationships linking the moments of accumulated deficit for a fixed drought length to the lag-1 autocorrelation of the series and the threshold level will be developed. Such empirical relationships take into account different distributions for the underlying series, other than the normal, and therefore they are applicable also to skewed data.

Finally a formulation for the return period of droughts, taking into account jointly the length and either accumulated deficit or intensity, will be derived, that assumes accumulated deficit beta distributed. The applicability of the formulation has been tested by simulation, and applied to four streamflow series, namely the annual flows of the Poudre River at Mouth of the Canyon, of the Colorado River at Lee Ferry, of the Nile River at Aswan, and the St. Lawrence River at Cornwall. Results indicate that the proposed approach can be efficiently applied for estimating the return period of droughts in stochastic-dependent series, such as streamflows.

3.2 Pdf of Drought Length for Serially Dependent Hydrological Variables

In this section, the pdf of drought length is discussed in the case of serially dependent hydrological variables. In particular, is shown that the lag 1 Markov model, traditionally adopted to model the sequence of deficits and surpluses, is not adequate when the series presents significant serial dependence and therefore, following previous works (Chung and Salas, 2000), an approach based on Discrete Autoregressive Moving Average model is adopted.

Let's consider a stationary stochastic hydrologic variable denoted as X_t, $t = 1, 2, ...$, where t represents the year and a threshold water demand level denoted by x_0. Then, a drought event is defined as a succession (run) of consecutive intervals in which the water supply remains below the threshold level x_0. Following Yevjevich's drought definition (Yevjevich, 1967), the drought length L (length of negative run) is the number of consecutive time intervals in which $X_t < x_0$ preceded and followed by (at least one interval where) $X_t \geq x_0$. These conditions can be also represented using a Bernoulli variable Y_t as:

$$Y_t = 0 \text{ if } X_t < x_0 \tag{3.1}$$

$$Y_t = 1 \text{ if } X_t \geq x_0 \tag{3.2}$$

Obviously there is a one to one correspondence between the sequence of deficits identified using X_t and the sequence of zeroes in Y_t. Thus for analyzing the drought length properties we will restrict our attention to the latter process Y_t.

In general terms the stochastic properties of Y_t will depend on the stochastic structure of X_t, as well as on the threshold level x_0. Traditionally it has been assumed that after truncating the X_t series, the binary process Y_t follows a lag-1 Markov chain (Llamas and Siddiqui, 1969; Sen, 1976; Shiau and Shen, 2001). Although this may not be a bad assumption if the X_t exhibits a weak autocorrelation, however as the autocorrelation of X_t becomes larger, the corresponding time dependence of the binary process Y_t becomes increasingly not comparable with that of a lag-1 Markov chain.

In order to illustrate the above point, we will assume that the X_t process follows the well known AR(1) model, which takes the form (e.g. Salas (1993)):

$$X_t = \mu + \phi(X_{t-1} - \mu) + \epsilon_t \tag{3.3}$$

where μ is the mean of the process, ϕ is equal to the lag-1 autocorrelation ρ_1 and ϵ_t is a white noise process with variance σ_ϵ^2. Without loss of generality, we will assume $\mu = 0$ and $x_0 = \mu$. After truncation, in order for the Y_t process to be a lag-1 Markov chain, the folllowing condition must hold (Taylor and Karlin, 1998):

$$P[Y_t = i \mid Y_{t-1} = j_1, Y_{t-2} = j_2, Y_{t-3} = j_3, \ldots] = P[Y_t = i \mid Y_{t-1} = j_1] \qquad \forall t \tag{3.4}$$

where $i, j_1, j_2, j_3, \ldots$ take the values 0 or 1. Under the lag-1 Markov assumption, the process is completely specified by the transition probability matrix:

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \tag{3.5}$$

where :

$$p_{ij} = P\left[Y_t = j \mid Y_{t-1} = i\right] \quad , \quad i = 0,1;\ j = 0,1 \qquad \forall t \tag{3.6}$$

Making use of the conditional probability definition, the element p_{ij} can also be written as:

$$p_{ij} = \frac{P\left[Y_t = j, Y_{t-1} = i\right]}{P\left[Y_{t-1} = i\right]} \quad , \quad i = 0, 1; \; j = 0, 1 \qquad \forall t \tag{3.7}$$

On the other hand, with reference to two zero mean normal bivariate r.v.'s Z_1 and Z_2, Sheppard (1899) derived the following expression relating the joint probabilities to the correlation coefficient ρ between Z_1 and Z_2:

$$P[Z_1 \leq 0, Z_2 \leq 0] = P[Z_1 > 0, Z_2 > 0] = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho \tag{3.8}$$

Thus, assuming $Z_1 = X_t$ and $Z_2 = X_{t-1}$ in Eq.(3.8), and observing that $P\left[Y_{t-1} = 0\right] = P\left[Y_{t-1} = 1\right] = \frac{1}{2}$ and $p_{00} + p_{01} = p_{10} + p_{11} = 1$, the transition probabilities of the binary Y_t process can then be computed from eq.(3.7) as:

$$p_{00} = p_{11} = \frac{1}{2} + \frac{1}{\pi} \sin^{-1} \rho_1 \tag{3.9}$$

$$p_{01} = p_{10} = \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \rho_1 \tag{3.10}$$

where ρ_1 is the lag-1 autocorrelation coefficient of the X_t process.

To prove that Y_t is not a Markov chain, we will compare the two-step transition probability from zero to zero computed under the Markov chain assumption with the one obtained by means of Sheppard (1899) relationship. Under the Markov assumption, the two step transition probability matrix P_2 can be computed as $P_2 = PP$ (e.g. Taylor and Karlin (1998)). It follows:

$$P[Y_t = 0 \mid Y_{t-2} = 0] = p_{00}^2 + p_{10}^2 = (\frac{1}{2} + \frac{1}{\pi} \sin^{-1} \rho_1)^2 + (\frac{1}{2} - \frac{1}{\pi} \sin^{-1} \rho_1)^2 \tag{3.11}$$

$$P[Y_t = 0 \mid Y_{t-2} = 0] = \frac{1}{2} + 2(\frac{1}{\pi} \sin^{-1} \rho_1)^2 \tag{3.12}$$

Using Sheppard (1899) relationship, observing that the lag-2 autocorrelation for an AR(1) process is ρ_1^2 , one gets:

$$P[Y_t = 0 \mid Y_{t-2} = 0] = \frac{1}{2} + \frac{1}{\pi} \sin^{-1} \rho_1^2 \tag{3.13}$$

Note how the two transition probabilities in eqs.(3.12) and(3.13) differ. In figure 3.1 the comparison between the two transition probabilities given by eqs. (12) and(3.13) is shown, vs. the lag-1 autocorrelation coefficient ρ_1. Note that for small ρ_1 the difference between two-step transition probabilities is small. However, as ρ_1 increases, the difference becomes rather significant for a wide range of $\rho_1$values, while only for values of ρ_1 very close to 1 (unrealistic from an hydrological point of view), the two curves tend to merge. As a consequence, the Markov approximation does not appear applicable, since it is not capable to capture the dependence in the truncated process, for a large spectrum of ρ_1 values. The Figure confirms that, once the truncated AR(1) process is in deficit condition, it will have a stronger tendency to remain in deficit than the corresponding lag-1 Markov process and therefore a different method should be adopted to model the clipped process Y_t. Among the several models that can be adopted to model binary series, the Discrete Autoregressive Moving Average (DARMA) has received increased attention because of its capability of reproducing a wide range of autocorrelation functions. Low order DARMA models, originally introduced by Jacobs and Lewis (1977), have been applied to daily precipitation and streamflows (Chang et al., 1984a,b; Cheebane et al., 1995).

Here the DARMA(1,1) will be adopted to model the clipped process Y_t. Chung and Salas (2000) used the same model to derive the return period and risk of droughts

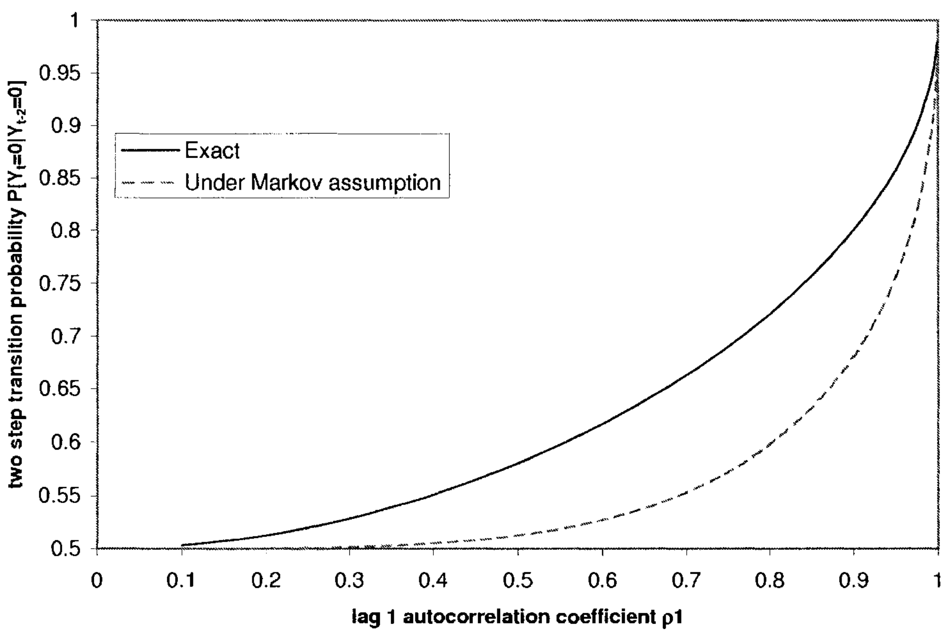


Figure 3.1: Two step transition probability $P[Y_t = 0|Y_{t-2} = 0]$ computed exactly (eq. 3.13) or by assuming a Markov chain dependence structure (eq. 3.12), as a function of the lag-1 autocorrelation coefficient of the continuous process X_t

of a given length or greater. In particular, they showed that a DARMA(1,1) is able to better model drought length probabilities of streamflows series exhibiting a strong autocorrelation, such as Niger River with respect to a DAR(1) model, which is a lag 1 Markov chain.

The binary DARMA(1,1) model Y_t takes only the values 0 and 1 and is defined as (Jacobs and Lewis, 1977):

$$Y_t = \begin{cases} Z_t & \text{with probability } \beta; \\ U_{t-1} & \text{with probability } (1-\beta); \end{cases} \tag{3.14}$$

and

$$U_t = \begin{cases} U_{t-1} & \text{with probability } \lambda; \\ Z_t & \text{with probability } (1-\lambda); \end{cases} \tag{3.15}$$

where Z_t is a sequence of independent Bernoulli variables that take the value 0 with probability π_0 and 1 with probability $\pi_1 = 1 - \pi_0$ and the variables are stationary and have the same marginal probability distribution π_0 and π_1. The autocorrelation function of Y_t is $\rho_k(Y) = c\lambda^{k-1}$, $k \geq 1$, where $c = (1-\beta)(\lambda + \beta - 2\lambda\beta)$.

Jacobs and Lewis (1978) showed that, even though the DARMA(1,1) process Y_t is not a Markov chain, the pair $\{Y_t, U_t\}$ forms a bivariate first-order Markov chain with transition probabilities $H_k(u,v) = P(Y_{t+1} = k), U_{t+1} = v | Y_t = m, U_t = u) = P(Y_{t+1} = k, U_{t+1} = v | U_t = u)$, since (Y_{t+1}, U_{t+1}) is indipendent of Y_t and u, v, k, m are $0, 1$ values. Then

$$H_0 = \begin{bmatrix} \lambda\,(1-\beta) + [1-\lambda\,(1-\beta)]\,\pi_0 & (1-\beta)\,(1-\lambda)\,\pi_1 \\ \beta\,(1-\lambda)\,\pi_0 & \beta\,\lambda\,\pi_0 \end{bmatrix} \tag{3.16}$$

$$H_1 = \begin{bmatrix} \beta\,\lambda\,\pi_1 & \beta\,(1-\lambda)\,\pi_1 \\ (1-\beta)\,(1-\lambda)\,\pi_0 & \lambda\,(1-\beta) + [1-\lambda\,(1-\beta)]\,\pi_1 \end{bmatrix} \tag{3.17}$$

The DARMA(1,1) process only takes values from {0,1} so that both of its probability distribution $\eth_0$ and $\eth_1$ sum to one, that is, $\eth_0 + \eth_1 = 1$. Estimates of $\eth_0$ and $\eth_1$ may be obtained from the run length property. Let $\overline{T_0}$ be the mean run length of state 0 and $\overline{T_1}$ be the mean run length of state 1, which are estimated from the sample 0-1 series. Then $\eth_0$ and $\eth_1$ are given by (Buishand, 1978):

$$\pi_0 = \frac{\overline{T_0}}{\overline{T_0} + \overline{T_1}} \tag{3.18}$$

and

$$\pi_1 = 1 - \pi_0 \tag{3.19}$$

respectively.

Chang et al. (1984a,b) noted that this estimator based on the occurrence probabilities of runs is conceptually reasonable since that the 0 and 1 spells govern the persistence characteristics of the series through the run lengths of 0 and 1.

Besides π_0 (or π_1), two more parameters λ and β,, need to be estimated for fitting a DARMA(1,1) model. For a DARMA(1,1) model, the method of moments may be used to obtain the estimate λ by matching the sample and model autocorrelation functions. For the DARMA(1,1) model, based on the autocorrelation function $\rho_k(Y) = c\lambda^{k-1}$, $k \geq 1$, where $c = (1-\beta)(\lambda + \beta - 2\lambda\beta)$ is the lag-1 autocorrelation coefficient, an estimator of λ may be the ratio of the second to the first sample autocorrelations, i.e.,

$$\hat{\lambda} = \frac{r(2)}{r(1)} \tag{3.20}$$

And, since the estimation of c is the lag-1 autocorrelation coefficient, parameter â may be estimated as

$$\hat{\beta} = \frac{(3\hat{\lambda}-1) - \sqrt{(3\hat{\lambda}-1)^2 - 4(2\hat{\lambda}-1)(\hat{\lambda}-\hat{c})}}{2(2\hat{\lambda}-1)} \tag{3.21}$$

The probability of the length L of a run of state i = 0 for a DARMA(1,1) is (Chang et al., 1984b):

$$P[L=n] = [H_1(0,0)\pi_0 + H_1(1,0)\pi_1][H_0^n(0) - H_0^{n+1}(0)] +$$
$$+[H_1(0,1)\pi_0 + H_1(1,1)\pi_1][H_0^n(1) - H_0^{n+1}(1)]/P[Y_0=1, Y_1=0] \tag{3.22}$$

where $H_0^n(j) = H_0^n(j,0) + H_0^n(j,1)$ and $j = 0,1$, in which $H_j^n(j,k)$ are the elements of the n-step transition matrix, and:

$$P(Y_0=1, Y_1=0) = \sum_{k=0}^{1}\sum_{j=0}^{1} H_0(j,k)\left[\pi_j - \sum_{l=0}^{1} H_0(l,j)\pi_l\right] \tag{3.23}$$

The expectation of L has also been derived by (Chang et al., 1984b) as:

$$E[L] = \frac{\pi_0[1-\beta\lambda+\beta(1-\lambda-\beta+2\beta\lambda)\{1-\pi_0\}]}{\{1-\pi_0\}[1-\lambda(1-\beta)\{1-\beta\pi_0\}-\beta\pi_0\{1-\beta(1-\lambda)\}]} \tag{3.24}$$

The applicability of the DARMA(1,1) to model drought length when the binary series is the result of clipping a continuous AR(1) process X_t has been tested here by comparing the drought length pdf obtained by simulating a long AR(1) process with the one resulting from a DARMA(1,1) process and a lag-1 Markov chain. In the latter case, the drought length probability is geometric with parameter p_{10} (Llamas and Siddiqui, 1969). In Figure 3.2, the comparison between the three pdf's is shown for different values of the autocorrelation coefficient ρ_1 of the continuous AR(1) process

X_t. The DARMA(1,1) parameters have been derived as a function of the AR(1) autocorrelation ρ_1 following Salas et al. (2001), while the parameter p_{10} of the drought pdf in the Markov chain case has been expressed by means of Eq.(3.10). The trivial case $\rho_1 = 0$ has not been plotted, since in this case the clipped Y_t process is just a sequence of i.i.d. Bernoulli variables and therefore the three distributions must coincide. From the figure it can be inferred an fairly good agreement between the drought length pdf related to the DARMA(1,1) model and the one obtained by simulating and clipping an AR(1) process. On the other hand, as larger values of ρ_1 are considered, the fitting of the pdf derived under the assumption of Markov chain sequence is worse and worse. This confirms that the DARMA(1,1) should be preferred to Markov chains when modelling the drought length pdf of a binary process obtained by clipping a continuous AR(1) process.

3.3 Moments of Deficits and Drought Amount

With reference to a stationary stochastic hydrological variable denoted as $X_t, t = 1, 2, ...$, and to a threshold water demand level denoted by x_0, the accumulated deficit D is defined as the sum of single deficits $S_t = x_0 - X_t$, for $x_o - X_t > 0$ extended to the duration L while the intensity I is defined as the ratio of accumulated deficit to drought length. Thus, the accumulated deficit of a drought starting at time t=t$_0$ can be expressed as:

$$D = \sum_{t=t_0}^{t_0+L-1} S_t = \sum_{t=t_0}^{t_0+L-1} (x_o - X_t) \quad \text{for} \quad X_t < x_o \tag{3.25}$$

where t_0 and L are such that $X_t < x_0$ for $t = t_0, ..., t_0 + L - 1$ and $X_{t_0-1} \geq x_0$ and $X_{t_0+L} \geq x_0$.

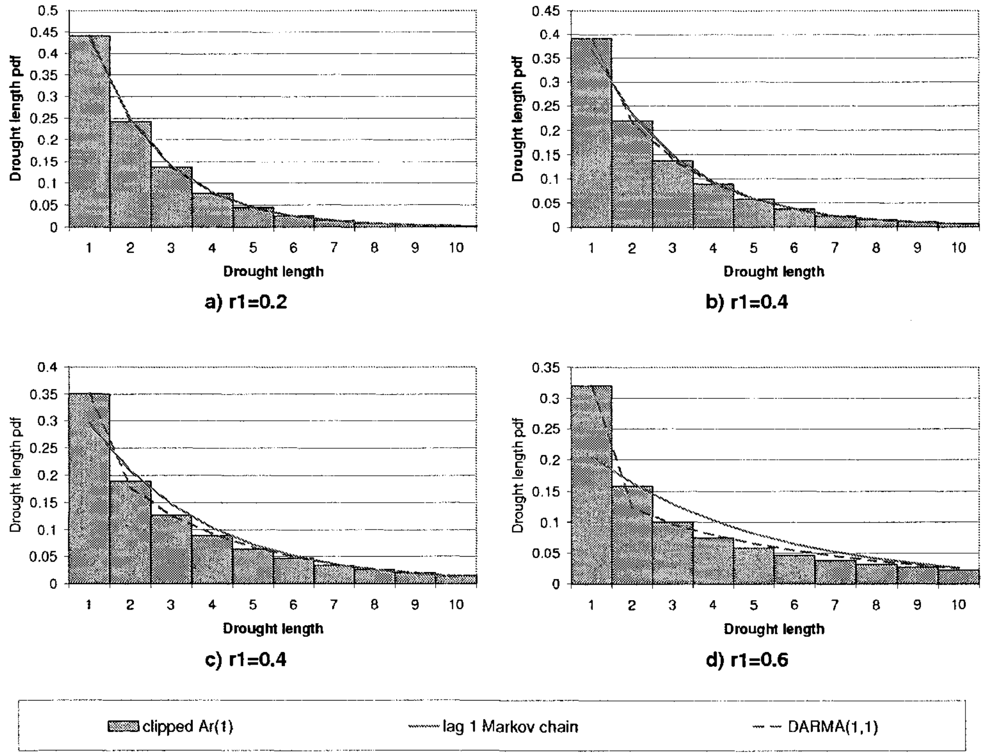


Figure 3.2: Drought length pdf corresponding to a clipped generated AR(1) process, to a DARMA(1,1) model and to a lag-1 Markov chain as a function of the lag-1 autocorrelation coefficient r_1

The intensity I is defined as:

$$I = \frac{D}{L} \tag{3.26}$$

The expected value of accumulated deficit conditioned on the fixed drought length $L = l$ can be derived by taking expectations on both sides of eq.(3.3):

$$E\left[D \mid L = l\right] = E\left[\sum_{t=t_0}^{t_0+L-1} S_t \mid L = l\right] = \sum_{t=t_0}^{t_0+L-1} E\left[S_t\right] \tag{3.27}$$

Similarly the variance:

$$Var\left[D \mid L = l\right] = Var\left[\sum_{t=t_0}^{t_0+L-1} S_t \mid L = l\right] = \sum_{j=0}^{l-1}\sum_{k=0}^{l-1} Cov\left[S_{t_0+j}, S_{t_0+k}\right] \tag{3.28}$$

Note that if the underlying series $X_t$is identically distributed, eq. (16) simplifies as:

$$E\left[D \mid L = l\right] = lE\left[S_t\right] \tag{3.29}$$

In addition, if the underlying series X_t, is serially independent, then $Cov\left[S_{t_0+j}, S_{t_0+k}\right] = 0$ for $j \neq k$, and the above expressions simplify as:

$$Var\left[D \mid L = l\right] = lVar\left[S_t\right] \tag{3.30}$$

since in this case the single deficits S_t will be i.i.d. and therefore they will have the same moments, and all the cross-covariances will be zero.

The moments of drought intensity I for fixed drought length $L = l$ can be easily computed as as a function of the corresponding moments of accumulated deficit as:

$$\mathbf{E}\left[I \mid L = l\right] = \mathbf{E}\left[D\right]/l \tag{3.31}$$

and

$$\mathbf{Var}\left[I \mid L = l\right] = \mathbf{Var}\left[D\right]/l^2 \tag{3.32}$$

as follows directly by taking expectations on both sides of $I = D/L$.

The expected value and the variance of the deficits S_t can be derived after truncation of the distribution of the underlying variable X_t. In particular, the r^{th} non-central moment of S_t can be expressed as a function of the pdf of X_t , $f_X(x)$ and of the threshold x_o as:

$$\mathbf{E}\left[S_t^k\right] = \frac{\int_{-\infty}^{x_o}(x_o - x)^r f_X(x)\mathbf{d}x}{\int_{-\infty}^{x_o} f_X(x)\mathbf{d}x} \tag{3.33}$$

Bonaccorso et al. (2003) derived the expectation and the variance of the single deficits S_t in the case of normal, log normal or gamma distributed X_t.

When the underlying hydrological series is not serially independent, eq.(3.19) no longer hold, since in this case the deficit S_t will exhibit temporal dependence. Consequently, $Cov\left[S_{t+j}, S_{t+k}\right] \neq 0$

Furthermore, the sequence of deficits will not be identically distributed in spite of the fact that the underlying process is stationary. Thus, the deficits are not independent and identically distributed (i.i.d.) and as stated above, eq.(3.30) will not hold.

In order to better illustrate the last point let's consider a drought event with duration l starting at time t_0. By definition, such an event is represented as:

$$\{X_{t_0-1} \geq x_o, X_{t_0} < x_o, \ldots, X_{t_0+k-1} < x_o, \ldots, X_{t_0+l-1} < x_o, X_{t_0+l} \geq x_o\} \tag{3.34}$$

The expected value of the k^{th} deficit will be the expectation of $S_{t_0+k-1} = x_o - X_{t_0+k-1}$

conditioned on the above sequence, i.e.:

$$\mathbf{E}\left[x_o - X_{t_0+k-1} \mid X_{t_0-1} \geq x_o, X_{t_0} < x_o, \ldots, X_{t_0+k-1} < x_o, \ldots, X_{t_0+l} \geq x_o\right] \quad k = 1, \ldots l-1 \tag{3.35}$$

To investigate numerically the behaviour of the above expectation, a 3,000,000 year autocorrelated series was generated using an AR(1) model considering different lag 1 autocorrelation coefficients ρ_1. The mean μ_x and the standard deviation σ_x were arbitrarily set equal to 700 mm and 200 mm, respectively, and the threshold was set equal to the mean, i.e. $x_0 = \mu_x$. After drought identification, the mean and standard deviation of the first, second, etc. deficits were computed for each drought length l.

In figure 3.3 the expected values of successive individual deficits for a drought of length 5 are plotted vs. the deficit position for different values of the lag 1 autocorrelation coefficient ρ_1. Some interesting considerations can be drawn from the inspection of figure 3.3. First of all the expectation of the k-th deficit depends on $k \leq l$. In other words, the expected value of a deficit depends on the position of the deficit within the drought, as well as on the drought length l. This confirms the non i.i.d. nature of the deficits when the underlying process X_t is serially dependent. In addition, figure 3.3 also reveals that the expected value of the k^{th} deficit will tend to increase as k increases. This is can be explained by observing that the surplus preceding the drought has an influence on the succeeding deficits, which will then tend to vanish for deficits at more distant time steps. The same figure 3.3 confirms that, because of symmetry, after the middle of the drought, the expectation will tend to decrease. Also the last deficit before the drought ends will have the same expectation

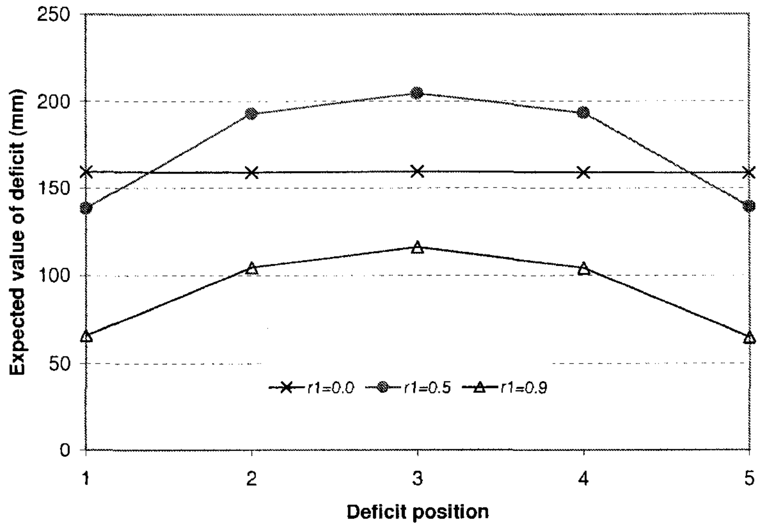


Figure 3.3: Expected value of deficits vs. the position within a drought of length 5 for different values of the lag 1 autocorrelation coefficient $r1$

as the first deficit, the second-last equal to the second, and so on. The practical significance of the above analysis is that for hydrological variables serially dependent, it is reasonable to expect that a multi-year drought will tend to get worse after the first year and the greatest deficits are to be expected in the central part of the drought.

A similar pattern is observed when a drought of length 9 is considered (fig. 3.4), also with respect to the standard deviation of the deficits (figs. 3.5 and 3.6). Because the deficits are not i.i.d. when the underlying X_t series is serially dependent, analytical or numerical derivation of the moments is more complex than in the independent case, since a in principle a full multivariate approach is needed. In general terms, the expected value of the k^{th} deficit for a drought of length l given by eq.(3.35), can be written as a function of the underlying multivariate pdf of X_t as:

$$\mathbf{E}\left[S_{t_0+k-1}\right] = \frac{\int_{x_o}^{\infty}\int_{-\infty}^{x_o}\cdots\int_{-\infty}^{x_o}\int_{x_o}^{\infty}(x_o - x_{t_0+k-1}) f(x_{t_0-1}, x_{t_0}, \ldots, x_{t_0+k-1}, \ldots, x_{t_0+l}, x_{t_0+l+1})\mathbf{dx}}{\int_{x_o}^{\infty}\int_{-\infty}^{x_o}\cdots\int_{-\infty}^{x_o}\int_{x_o}^{\infty} f(x_{t_0-1}, x_{t_0}, \ldots, x_{t_0+k-1}, \ldots, x_{t_0+l}, x_{t_0+l+1})\mathbf{dx}} \tag{3.36}$$

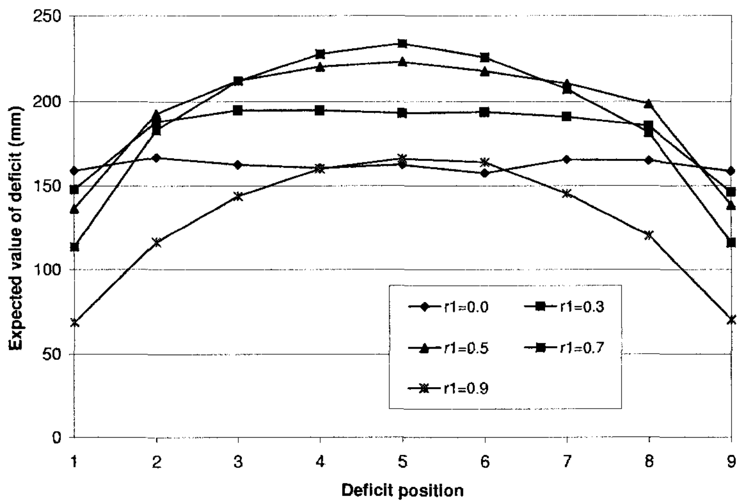


Figure 3.4: Expected value of deficits vs. the position within a drought of length 9 for different values of the lag 1 autocorrelation coefficient $r1$

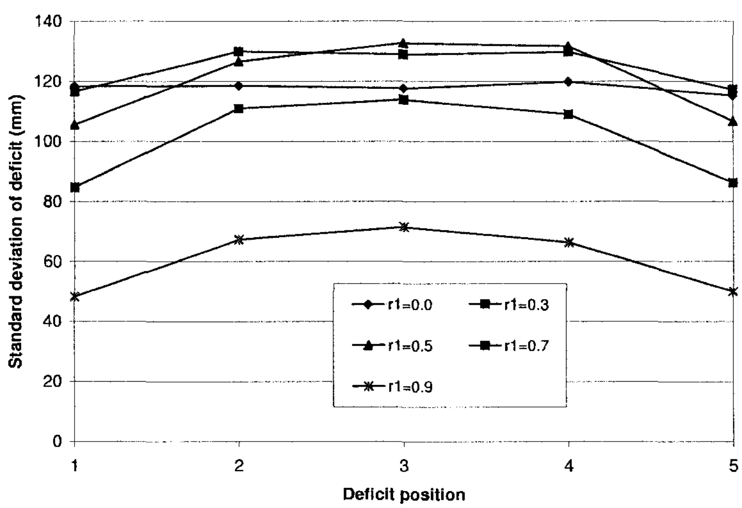


Figure 3.5: Standard deviation of deficits vs. the position within a drought of length 5 for different values of the lag 1 autocorrelation coefficient $r1$

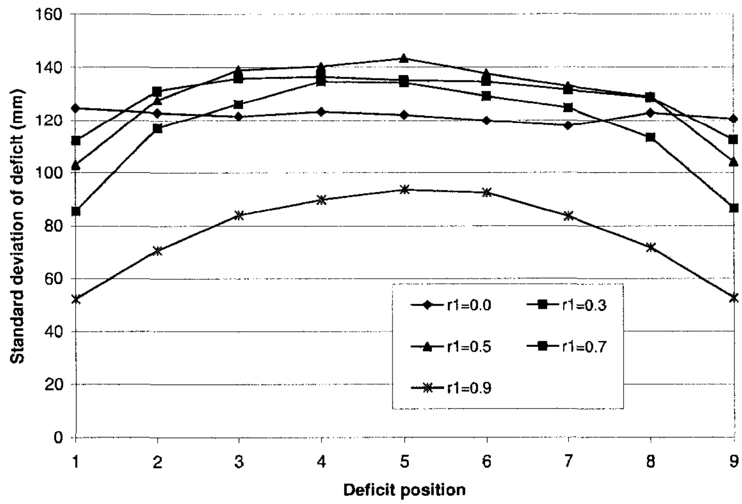


Figure 3.6: Standard deviation of deficits vs. the position within a drought of length 9 for different values of the lag 1 autocorrelation coefficient $r1$

where $f(.)$ denotes the multivariate distribution of the X_t process and $\mathbf{dx} = dx_{t_0-1}dx_{t_0} \ldots dx_{t_0+k-1} \ldots dx_{t_0+l}dx_{t_0+l+1}$. Computations of the integrals in eq.(3.36) require the knowledge of the multivariate distribution of the X_t process, which is rather difficult to define, except in a few cases. For example, if one assumes that X_t is multivariate normal, exact and numerical solution of the integrals are possible. The moment generating function and the cumulants of the truncated multivariate normal distribution have been investigated by Tallis (1961) and Finney (1962), respectively. Also, the algorithm MVNTRC which computes numerically expectations and covariances of truncated multivariate normals is available (Leppard and Tallis, 1989). Once the expectation $\mathbf{E}[S_{t+k-1}]$ is determined, the expected value of accumulated deficit follows directly from eq.(3.27) as:

$$\mathbf{E}[D \mid L = l] = \mathbf{E}\left[\sum_{k=1}^{l} S_{t+k-1}\right] \tag{3.37}$$

In order to compute the variance of D of eq.(3.28), the covariance $Cov[S_{t+j}, S_{t+k}]$ is needed. It can be computed as:

$$Cov\left[S_{t+j}, S_{t+k}\right] = \mathbf{E}\left[S_{t+j}S_{t+k}\right] - \mathbf{E}\left[S_{t+j}\right]\mathbf{E}\left[S_{t+k}\right] \tag{3.38}$$

where the expectation $\mathbf{E}\left[S_{t+j}S_{t+k}\right]$ can be computed by appropriate integration of the multivariate distribution of X_t as:

$$\mathbf{E}\left[S_{t+j}S_{t+k}\right] =$$

$$= \frac{\int_{x_o}^{\infty}\int_{-\infty}^{x_o}\cdots\int_{-\infty}^{x_o}\int_{x_o}^{\infty}(x_o - x_{t+k-1})(x_o - x_{t+j-1})f(x_{t-1},\ldots,x_{t+j-},\ldots,x_{t+k-1},\ldots,x_{t+l},x_{t+l+1})\mathbf{dx}}{\int_{x_o}^{\infty}\int_{-\infty}^{x_o}\cdots\int_{-\infty}^{x_o}\int_{x_o}^{\infty} f(x_{t-1},x_t,\ldots,x_{t+k-1},\ldots,x_{t+l},x_{t+l+1})\mathbf{dx}} \tag{3.39}$$

Once again, if the X_t are normal distributed, the above integration can be solved numerically making use of the algorithm MVNTRC.

In order to show numerically the applicability of the above integrations, simulations from the same normal AR(1) models adopted before have been carried out. The multivariate distribution of the X_t will be:

$$X_1, X_2, ..., X_n \tilde{} MVN(\mu, \Sigma) \tag{3.40}$$

where μ is the vector of (stationary) means and, due to the AR(1) nature of the process, the generic element of the variance covariance matrix Σ is given by:

$$\Sigma_{i,j} = \text{cov}[X_i, X_j] = \sigma^2 r_1^{|i-j|} \tag{3.41}$$

being $\sigma_x^2 = \text{var}[X_t]$.

In table 3.I the expectations and standard deviations of accumulated deficit for fixed drought durations computed by means of eqs.(3.27) and(3.28) respectively are reported for different values of the lag-1 autocorrelation coefficient r_1. The conditional moments of single deficits have been computed by means of the aforementioned algorithm MVNTRC. In the same table, the corresponding values obtained after simulating 3,000,000 years from an AR(1) process are shown in brackets. From the table a very good agreement between moments computed by numerical integration and by simulation can be inferred. The table also indicates that, as expected, both the mean and standard deviation of accumulated deficit increase with drought duration. More specifically, the mean increases proportionally with L for uncorrelated processes (r_1=0), in agreement with eq.(3.29), while it exhibits a non linear increase with L as the process becomes more and more serially dependent (r_1>0). Also it may be worthwhile to note how the mean of accumulated deficit decrease sensibly as the dependence of the process increases, in agreement with Figure 3.3. This is particularly significant for short durations since in this case, because of the high autocorrelation, the deficits will be strongly affected by the nearby surpluses, and therefore they will exhibit smaller values.

3.4 Empirical Approximations of the Moments of Accumulated Deficit for Autocorrelated and Skewed Data

The proposed procedure to compute the moments of drought characteristics, despite being exact, at least in the normal domain, still has some limitations due to the computational difficulties related to the numerical solutions of the expectations(3.36) and(3.39). Indeed, in the present version, the algorithm MVNTRC can compute expectations of up to 5 truncated multivariate variables. As a consequence, only droughts with length up to 3 can be considered, since in computing the expectation the surplus intervals preceeding and following the drought have to be taken into

Table 3.I: Expected value and standard deviation (mm) of accumulated deficit for fixed drought duration L, as a function of the lag-1 autocorrelation of the underlying series computed by means of eqs. (3.27) and (3.28), assuming a threshold $x_0 = \mu_x$. In brackets, the corresponding values obtained by simulating 3,000,000 years from an AR(1) process are shown.

lag 1 autocorrelation	$L = 1$		$L = 2$		$L = 3$	
	mean	std	mean	std	mean	std
$r_1 = 0.0$	159.6	120.6	319.2	170.5	478.7	208.8
	(159.6)	(120.6)	(319.0)	(169.9)	(478.0)	(209.6)
$r_1 = 0.2$	134.9	106.6	309.0	171.3	490.4	221.9
	(134.9)	(106.5)	(309.2)	(171.5)	(490.8)	(222.3)
$r_1 = 0.4$	108.7	88.9	277.5	159.5	471.3	222.4
	(108.4)	(88.7)	(276.9)	(159.3)	(472.8)	(222.5)
$r_1 = 0.6$	81.8)	68.7	225.5	133.5	408.0	200.1
	(81.5)	(68.5)	(225.1)	(132.7)	(406.9)	(198.5)
$r_1 = 0.8$	53.3	45.5	153.4	92.6	289.6	145.9
	(53.0)	(45.6)	(153.1)	(91.9)	(290.1)	(146.6)

account.

Furthermore, the numerical solution is limited to normally distributed hydrological variables, and therefore it does not apply to skewed data. Although in principle one might adopt a skewed multivariate distribution and proceed with the computation of the truncated moments, in practice this can be quite cumbersome, due to limited availability of skewed multivariate distributions simple enough to perform the complex integrations necessary to compute the conditional moments numerically, let alone analytically.

In order to overcome such limitations, an empirical approximation for computing the moments of accumulated deficit is developed here, based on the assumption that each moment can be approximately expressed as the product of two terms, the first of which is a function of the marginal distribution of the data (normal, log-normal, gamma, etc.) as well as of the threshold x_0 , while the second does not depend on the type of distribution but only on the lag-1 autocorrelation ρ_1 of the series, on the threshold x_o and on the drought length L. The first term should be able to capture the effect of the skewness of the distribution on the moments of accumulated deficit. A natural choice is to assume as the first terms the moments of accumulated deficit of a drought of length 1 under the hypothesis of non-autocorrelated process ($\rho_1 = 0$). Note that such moments are equal to the moments of the single deficits S_t , for which analytical expressions are available for several underlying distributions. Bonaccorso et al. (2003), among others, derived such moments for X_t normal, log-normal and gamma.

Thus, the proposed approximate expressions for the mean and the variance of accumulated deficit of fixed length l are:

$$\mathbf{E}\left[D_l(r_1)\right] = \mathbf{E}\left[D_1(0)\right]\Psi(r_1, x_o, l) \tag{3.42}$$

$$\mathbf{Var}\left[D_l(r_1)\right] = \mathbf{Var}\left[D_1(0)\right]\Upsilon(r_1, x_o, l) \tag{3.43}$$

where $\mathbf{E}\left[D_l(r_1)\right]$ and $\mathbf{Var}\left[D_l(r_1)\right]$ are the expected value and variance of accumulated deficit of droughts of length l of a series with lag 1 autocorrelation equal to r_1. Then,$\mathbf{E}\left[D_1(0)\right]$ and $\mathbf{Var}\left[D_1(0)\right]$ are function exclusively of the underlying marginal distribution (either skewed or not) and of the threshold x_0, whereas $\Psi(r_1, x_o, l)$ and $\Upsilon(r_1, x_o, l)$ are the dimensionless moments function of the autocorrelation of the process r_1, of the threshold x_o and of the fixed drought length l.

The advantage of the approximations given by eqs.(3.42) and(3.43) is evident, since the first terms can be virtually computed for any marginal distribution, as they involve a simple integration, whereas the two functions $\Psi(r_1, x_o, l)$ and $\Upsilon(r_1, x_o, l)$ can be derived from the corresponding moments of accumulated deficit computed assuming a normal multivariate distribution for the data, as shown previously. However, since the numerical approach to compute the moments of accumulated deficit can currently be employed only for droughts up to $L = 3$, empirical approximations to the functions $\Psi(r_1, x_o, l)$ and $\Upsilon(r_1, x_o, l)$ are derived here by means of a simulation approach.

More specifically, at first it has been assumed that the dependence of $\Psi(r_1, x_o, l)$ and $\Upsilon(r_1, x_o, l)$ on drought length l can be modeled by means of regression equations of the type:

$$\Psi(r_1, l) = a_m l^{b_m} \tag{3.44}$$

$$\Upsilon(r_1, l) = a_v l^{b_v} \tag{3.45}$$

where the parameters a_m , b_m , a_v and b_v have an implicit relationship with ρ_1 and x_o. Note that for $\rho_1 = 0$, $a_m = b_m = a_v = b_v = 1$ as follows directly from eqs.(3.18) and(3.19) since in this case $\mathbf{E}\left[D_l(0)\right] = \mathbf{E}\left[S_l\right]$ and $\mathbf{Var}\left[D_l(0)\right] = \mathbf{Var}\left[S_l\right]$.

In order to estimate the parameters of the regressions(3.44) and(3.45) by least squares, series of 3,000,000 years have been generated by means of an AR(1) for fixed lag-1 autocorrelations. After generation, droughts have been identified assuming a threshold of the type $x_o = \mu_X - \alpha\sigma_x$, assuming different α, $(\alpha = -.6, -.55, \ldots, .6)$, and the empirical values of $\Psi(r_1, x_o, l)$ and $\Upsilon(r_1, x_o, l)$ have been computed by dividing the accumulated deficit of each identified drought by the mean and the variance of accumulated deficit of length 1 when $\rho_1 = 0$. The parameters a_m , b_m , a_v and b_v have then been estimated by least squares. The procedure has been repeated for several ρ_1 values ranging from 0 up to .9 in .05 steps.

At first, 2nd degree polynomials have been fitted to a_m , b_m , a_v and b_v as a function of r_1 for fixed α. Plots of the estimated parameters of such 2nd degree polynomials vs. α indicated a fairly linear relationship, and therefore, linear regressions have been subsequentially applied to express each of the parameters of the fitted 2nd degree polynomial to α. This resulted in the following expressions for a_m , b_m , a_v and b_v as a function of ρ_1 and α.

$$\begin{aligned}
a_m &= (-6.8341\alpha - 3.7491)\,\rho_1^2 + (12.469\alpha - 6.8581)\,\rho_1 + 1.0 \\
b_m &= (0.18128\alpha + 0.081687)\,\rho_1^2 + (-0.17899\alpha + 0.59243)\,\rho_1 + 1.0 \\
a_v &= (-64.512\alpha - 3.7329)\,\rho_1^2 + (114.29\alpha - 84.279)\,\rho_1 + 1.0 \\
b_v &= (0.35109\alpha + 0.55935)\,\rho_1^2 + (-0.37248\alpha + 1.0930)\,\rho_1 + 1.0
\end{aligned} \tag{3.46}$$

The validity of the derived empirical approximations has been verified by comparing the moments of accumulated deficit computed by simulating 3,000,000 years from an AR(1) process with those computed by eqs. (3.42) and(3.43). In particular, either normal or log-normal (with skewness $\gamma = .5, 1$) series have been generated assuming different lag 1 autocorrelation coefficients ($\rho_1 = 0, .4, .8$). After generation, droughts have been identified using different thresholds and the expected value and variance of accumulated deficit have been computed for different drought durations ($L = 1, 3, 5, 7$) and for each combination of skewness γ and lag 1 autocorrelation r. Then the normalized differences between the moments of accumulated deficit computed from generated samples and the ones computed by means of eqs.(3.42) and(3.43) have been computed.

The results of such comparison related to the expected value and variance of accumulated deficit expressed in percentage is reported in Tables 3.II and 3.III for different combinations of drought length L, lag-1 autocorrelation r_1, truncation level α and skewness of the underlying series.

From Table 3.II it can be inferred how the errors related to the expected values of accumulated deficit are relatively small for a wide range of L, r_1, γ, and α values, and in general below 5% for most cases. In particular, an apparent increase of the errors is observed as increasing drought lengths L and lag-1 autocorrelations r_1 are considered. On the other hand no significant, nor consistent differences can be inferred as different threshold levels α and skewness coefficients γ are considered.

Similar considerations can be drawn from the inspection of Table 3.III, where the errors related to the variance of accumulated deficit are shown. The same pattern of increasing errors as increasing drought lengths L and lag-1 autocorrelations r_1 are considered can be observed, although in this case the errors tend to become relatively large, occasionally above 30%. However it should be pointed that for a range of autocorrelation usually observed in annual hydrological series ($r_1 \leq .4$) and for the range of drought lengths of interest for practical purposes, the errors remain acceptable and generally far below 10%.

3.5 Joint Probability Distribution of Drought Characteristics

Since the three characteristics L, D and I are not mutually independent, one must use a multivariate formulation to model their joint occurrence. From a theoretical standpoint, the bivariate distribution of any two drought characteristic follows from the multivariate distribution of the underlying series X_t. However, its direct analytical derivation is rather complex, if not impossible, except in a few simplified cases. Here a different approach will be followed, that capitalizes on the particular link between D and L , and I and L. Indeed the bivariate pdf of D and L can be written as the product of the conditional distribution of the two variables as (e.g. Mood et al. (1974)):

$$f_{D,L}(d,l) = f_{D|L=l}(d) f_L(l) \tag{3.47}$$

Table 3.II: Difference (%) between expected value of accumulated deficit of fixed length L computed by the empirical approximations given by eq. (3.42) and the one obtained by simulation. α is the threshold coefficient ($x_0 = \mu_x - \alpha\sigma_X$), r_1 is the autocorrelation coefficient of the generated AR(1) process and γ is the skewness.

		$\alpha = .0$			$\alpha = .2$			$\alpha = .5$		
	L	$\gamma = 0$	$\gamma = .5$	$\gamma = 1.0$	$\gamma = 0$	$\gamma = .5$	$\gamma = 1.0$	$\gamma = 0$	$\gamma = .5$	$\gamma = 1.0$
$r_1 = .0$	1	0.01	0.26	-0.04	0.15	-0.28	0.08	0.09	-0.04	-0.26
	2	0.07	0.21	-0.03	-0.15	-0.24	0.21	0.48	-0.00	-0.09
	3	-0.03	0.07	-0.26	0.01	0.04	-0.05	-0.18	0.11	0.24
	4	-0.33	-0.22	0.10	-0.25	0.31	0.17	-0.48	0.29	0.06
	5	0.33	-0.38	-0.21	-0.67	0.14	0.01	-1.59	-0.51	0.22
	6	0.18	-0.00	-0.47	-0.08	0.41	0.13	0.19	0.62	0.30
	7	-0.19	-0.06	0.06	-0.09	1.11	0.09	-1.12	-0.82	-2.67
	8	-0.56	-0.06	-0.19	-0.13	-0.52	0.23	-4.07	3.29	-0.16
	9	0.48	-0.06	0.57	-0.16	1.84	0.71	-	-	3.58
	10	0.47	-0.14	0.26	-1.24	1.41	-0.19	-	-	-
$r_1 = .2$	1	-3.24	-3.13	-3.71	-3.59	-3.25	-3.64	-4.16	-3.83	-4.18
	2	1.77	1.36	0.41	1.87	1.05	0.33	1.83	1.14	0.38
	3	2.69	1.12	0.67	2.27	1.46	0.67	2.63	2.10	1.13
	4	1.94	0.74	-0.05	2.18	0.85	0.21	2.32	1.50	0.41
	5	0.69	-0.68	-1.29	1.32	-0.22	-1.21	1.56	0.37	-0.30
	6	-0.29	-1.48	-2.62	-0.19	-1.16	-2.09	-0.28	0.36	-1.92
	7	-0.92	-2.65	-3.81	-0.57	-2.06	-3.65	0.57	-0.79	-1.72
	8	-2.29	-4.15	-4.99	-2.74	-3.68	-4.21	-2.83	-2.93	-3.24
	9	-3.13	-2.86	-5.56	-2.74	-5.08	-5.64	-6.73	-2.96	-5.04
	10	-5.19	-6.80	-6.62	-4.29	-3.61	-7.23	-4.36	-2.57	-8.33
$r_1 = .4$	1	-5.98	-5.75	-5.89	-6.19	-5.47	-5.65	-5.87	-5.86	-6.03
	2	1.37	0.61	-0.30	1.58	0.73	-0.29	1.51	1.35	0.46
	3	3.54	2.06	0.91	4.00	2.42	1.06	4.43	2.64	1.69
	4	3.40	1.36	-0.25	3.58	1.37	0.71	4.31	2.17	1.14
	5	1.67	-0.08	-2.04	2.75	0.24	-1.31	3.30	1.48	-0.13
	6	0.71	-2.05	-3.31	0.38	-1.47	-3.44	1.28	-1.05	-2.12
	7	-1.84	-3.51	-5.92	-1.27	-3.39	-5.86	-0.20	-3.20	-4.09
	8	-3.60	-6.13	-7.51	-2.06	-5.87	-7.38	-1.79	-3.17	-6.63
	9	-5.61	-8.35	-9.96	-5.60	-7.08	-9.38	-3.82	-6.88	-7.50
	10	-8.10	-10.66	-11.91	-6.06	-8.64	-11.25	-5.33	-8.18	-10.19
$r_1 = .6$	1	-5.56	-4.12	-3.72	-5.41	-4.84	-3.84	-6.01	-4.95	-4.06
	2	-0.26	-0.34	-0.66	0.27	0.47	-0.96	0.49	0.99	0.32
	3	2.55	2.18	0.38	3.56	2.52	0.38	4.52	3.42	1.83
	4	3.98	2.26	0.39	4.96	2.39	0.78	5.41	3.53	1.60
	5	3.23	1.49	-0.63	3.96	1.85	-0.77	4.70	2.30	0.90
	6	2.19	-0.57	-2.90	2.75	-0.44	-3.18	2.72	1.04	-0.53
	7	0.31	-2.14	-5.11	0.59	-2.31	-4.77	1.27	-0.80	-3.60
	8	-1.58	-4.86	-7.64	-1.23	-4.96	-7.24	0.04	-3.35	-6.09
	9	-3.14	-7.86	-10.04	-3.37	-7.24	-9.92	-2.52	-5.32	-8.14
	10	-5.72	-9.52	-11.82	-6.03	-9.67	-12.17	-4.08	-7.96	-9.94
$r_1 = .8$	1	0.61	2.58	4.14	-0.02	2.63	3.48	0.20	2.03	3.89
	2	-0.00	1.48	1.71	0.25	1.48	1.77	1.36	2.67	2.72
	3	1.32	2.26	1.36	2.49	2.57	2.46	3.16	3.39	3.55
	4	2.70	2.30	1.73	3.03	2.82	1.37	3.90	3.89	3.10
	5	3.30	1.53	1.22	3.35	2.89	0.31	5.43	3.95	1.56
	6	3.62	1.72	-0.69	3.42	1.64	-0.63	5.06	3.34	1.87
	7	1.83	0.30	-2.20	2.63	-0.25	-1.68	3.49	2.10	0.24
	8	2.62	-0.48	-3.41	2.12	-0.94	-3.10	2.98	0.11	-1.95
	9	0.57	-2.11	-4.89	1.12	-2.42	-5.58	1.79	-0.22	-3.62
	10	-0.90	-4.19	-7.55	-0.11	-4.67	-7.69	-0.88	-3.38	-5.03

Table 3.III: Difference (%) between variance of accumulated deficit of fixed length L computed by the empirical approximations given by eq. (3.43) and the one obtained by simulation. α is the threshold coefficient ($x_0 = \mu_x - \alpha\sigma_X$), r_1 is the autocorrelation coefficient of the generated AR(1) process and γ is the skewness.

		$\alpha = .0$			$\alpha = .2$			$\alpha = .5$		
	L	$\gamma = 0$	$\gamma = .5$	$\gamma = 1.0$	$\gamma = 0$	$\gamma = .5$	$\gamma = 1.0$	$\gamma = 0$	$\gamma = .5$	$\gamma = 1.0$
	1	0.20	0.26	0.17	0.34	-0.04	0.01	0.22	-0.17	-0.04
	2	0.12	-0.04	0.27	-0.47	-0.52	0.05	2.24	0.14	-0.09
	3	-0.26	-0.27	0.02	-1.69	-0.26	-0.03	-1.23	0.37	0.51
	4	-0.53	1.32	-0.28	0.76	0.98	0.23	-0.10	-1.07	-2.39
$r_1 = .0$	5	-1.08	-0.98	2.12	-2.84	-2.34	-2.67	-3.49	1.40	-1.77
	6	-1.41	-0.72	0.10	-0.28	-0.61	-1.61	3.28	5.67	-8.33
	7	0.25	2.10	-1.06	-4.00	1.44	3.64	-4.93	-26.07	-4.59
	8	2.62	-2.09	-3.06	-0.40	2.23	-4.73	2.83	-4.06	9.29
	9	2.18	3.67	1.14	-16.84	1.94	0.42	-	-	13.74
	10	-9.85	-3.51	-0.27	-5.25	7.98	14.66	-	-	-
	1	-1.35	2.26	4.60	-0.53	1.56	3.00	-1.64	0.63	1.42
	2	8.15	8.33	7.92	8.76	7.87	7.33	8.83	8.31	7.40
	3	9.76	8.08	6.90	9.57	6.80	7.56	11.85	9.54	8.30
	4	8.57	5.90	5.25	8.40	5.96	4.31	10.87	8.52	7.47
$r_1 = .2$	5	3.92	3.34	4.04	7.76	4.23	3.92	13.75	5.40	5.71
	6	4.51	1.95	1.43	5.74	0.84	2.25	9.64	4.09	3.41
	7	5.80	0.56	-1.97	4.45	-2.29	0.78	7.69	6.44	6.19
	8	-1.71	-4.69	-2.85	0.56	-6.50	-7.67	4.05	2.79	0.24
	9	-4.42	-0.58	-7.44	-10.02	-2.99	-0.45	-11.31	-4.20	-11.92
	10	-11.43	-13.47	-3.51	-5.02	-0.71	-10.91	16.15	-3.72	-11.12
	1	-4.67	1.01	5.48	-5.49	0.43	3.81	-4.26	-0.33	3.11
	2	6.57	7.94	9.83	8.23	7.51	8.34	8.78	9.13	9.82
	3	10.38	8.87	10.22	12.44	9.89	8.94	12.93	11.52	9.79
	4	10.47	7.45	6.32	10.47	8.29	7.15	13.72	10.27	9.00
$r_1 = .4$	5	7.88	5.04	2.07	7.38	5.71	5.00	10.77	8.83	6.01
	6	7.55	1.22	-0.20	5.46	4.66	-1.11	9.03	5.61	1.86
	7	-0.27	-2.78	-7.61	1.73	-4.73	-8.00	2.77	0.34	-1.15
	8	-2.76	-6.01	-9.30	0.01	-6.18	-8.45	0.31	3.91	-9.20
	9	-5.64	-12.93	-15.97	-7.56	-9.36	-9.94	4.93	-11.89	-15.04
	10	-10.16	-19.34	-19.78	-12.11	-8.96	-15.94	-4.98	-8.36	-25.24
	1	-6.26	3.45	12.76	-6.79	1.57	9.59	-7.52	-0.13	6.64
	2	-4.03	3.54	8.27	-0.48	3.29	5.49	0.08	4.44	6.48
	3	1.79	5.13	7.09	4.49	5.89	5.07	7.50	6.36	4.98
	4	5.61	5.86	6.54	8.20	4.45	3.82	9.92	7.28	4.39
$r_1 = .6$	5	6.21	3.82	1.36	7.24	4.96	0.57	11.54	7.02	2.96
	6	5.10	1.13	-3.05	5.48	2.28	-5.20	5.26	5.62	-0.74
	7	2.31	-1.44	-4.79	2.76	-3.12	-4.57	7.14	-2.08	-6.29
	8	-0.64	-11.00	-11.83	-1.62	-8.25	-18.21	4.81	-1.90	-11.66
	9	-7.43	-16.16	-14.98	-4.15	-9.36	-18.13	-1.45	-7.36	-19.16
	10	-12.35	-17.67	-22.12	-11.21	-21.04	-22.86	-17.24	-19.19	-20.57
	1	19.99	32.24	40.23	17.33	29.23	36.43	15.28	24.55	32.83
	2	9.60	20.40	27.86	6.84	16.58	22.73	5.97	14.16	19.59
	3	7.46	17.98	22.62	8.24	14.56	19.26	8.42	14.05	17.86
	4	12.32	16.71	22.36	11.89	15.26	17.59	10.71	14.96	16.04
$r_1 = .8$	5	13.66	15.29	19.08	10.52	14.91	12.96	14.82	14.13	12.25
	6	15.56	17.71	16.03	15.05	14.64	15.16	16.34	13.39	14.07
	7	17.59	14.33	15.39	16.66	14.02	12.65	14.76	15.28	8.99
	8	19.28	15.32	14.18	15.19	10.73	10.06	15.48	13.69	7.75
	9	17.98	16.77	10.41	16.92	11.02	6.28	18.82	12.40	0.42
	10	14.91	12.35	6.93	16.55	4.78	3.44	10.30	10.97	2.58

where $f_{D,L}$ is the bivariate pdf, $f_{D|L=l}$ is the distribution of D conditioned on a fixed L, and $f_L(l)$ is the marginal pdf of L. A similar expression can be written for the bivariate distribution of I and L as:

$$f_{I,L}(i,l) = f_{I|L=l}(d) f_L(l) \tag{3.48}$$

Thus, the joint distributions $f_{D,L}(d,l)$and $f_{I,L}(i,l)$ can be determined once the conditional distributions $f_{D|L=l}(d)$ and $f_{I|L=l}(d)$ and the distribution $f_L(l)$ are known. Assuming that the underlying hydrological variable follows an AR(1) model, it has already been shown that a DARMA(1,1) is able to reproduce well the pdf of drought length $f_L(l)$.

In general, for fixed L=l, the accumulated deficit is the sum of single deficits, extended to the drought length L=l. Therefore in principle, its exact distribution could be derived s a function of the multivariate distribution of the single deficits. However, such analytical derivation can be quite difficult to express in closed form. To overcome such difficulty, some authors have assumed a parametric distribution for $D \mid L$, and have estimated the parameters from observed droughts (Guven, 1983; Sharma, 1995; Shiau and Shen, 2001). In some cases, due to the limited number of droughts that can be observed in the usually available records, synthetic generation (Shiau and Shen, 2001) or long series reconstructed from tree rings records (Gonzalez and Valdes, 2003) have been used.

Here a different approach for estimating the parameters of the distribution of $D \mid L$ will be employed, that capitalizes on the expressions of the moments of accumulated deficits derived previously. More specifically, the approximate moments of accumulated deficit computed on the basis of the distribution of the underlying process X_t and of the threshold are used to estimate the parameters by method of moments.

This enables to fully exploit the available information from the whole observed series, thus allowing for a reliable estimation of the distribution of drought characteristics even on relatively short series.

Several authors have adopted a gamma distribution to fit observed accumulated deficit (Guven, 1983; Shiau and Shen, 2001; Bonaccorso et al., 2003; Gonzalez and Valdes, 2003; Salas et al., 2005). Here, in order to select an appropriate distribution for accumulated deficit conditioned on a fixed drought length when the underlying series is autocorrelated, three two parameters distributions (namely , log-normal, gamma, and beta) have been compared by means of simulation, and their goodness of fit has been evaluated by means of the Probability Plot Correlation Coefficient statistic (PPCC) proposed by Filliben (1975). The PPCC statistic is defined as the correlation coefficient between the observation sorted in ascending order and the quantiles computed by means of the selected distribution corresponding to a nonexceeding probability equal to the sample frequency of the observation. In particular, method of moments has been applied to estimate the parameters of each distribution, making use of the approximate moments derived in eqs.(3.42) and(3.43), as a function of the underlying variable X_t , of the threshold parameter α and of the lag-1 autocorrelation r_1. Also, Weibull formula has been adopted to compute the sample frequency.

The PPCC statistic has been computed for each of the three parametric cdf, for different values of L , α and r_1 , assuming a normal or a log normal distribution with skewness coefficient $\gamma = .5$ and $\gamma = 1$ for the underlying generated AR(1) process respectively. Higher values of PPCC generally indicate a better fit. The computed PPCC statistics are reported in Table 3.5, from which it can be inferred that the beta distribution leads to values of PPCC generally higher that those corresponding to the other two distributions, for all combinations of L , α and r_1. This seems to confirm that the beta seems is to be preferred to model accumulated deficit in serially

Table 3.IV: Probability Plot Correlation Coefficient (PPCC) for Beta, Gamma e Log Normal distribution fitted to accumulated deficit of fixed length L using the moments computed by the empirical approximations given by eqs. (3.42) and (3.43). Threshold $x_0 = \mu_x$. ρ_1 is the autocorrelation coefficient of the generated normal AR(1) process.

		γ=0			γ=0.5			γ=1		
	L	Beta	Gamma	LogN	Beta	Gamma	LogN	Beta	Gamma	LogN
	1	0.999	0.993	0.961	0.995	0.988	0.953	0.995	0.980	0.943
	3	0.999	0.997	0.985	0.998	0.996	0.984	0.998	0.994	0.981
ρ_1=0	5	1.000	0.999	0.992	0.998	0.997	0.989	0.999	0.996	0.989
	7	0.999	0.998	0.993	0.998	0.997	0.991	1.000	0.999	0.994
	9	0.996	0.995	0.989	0.998	0.998	0.995	0.999	0.997	0.992
	1	0.999	0.996	0.967	0.997	0.994	0.963	0.998	0.991	0.960
	3	1.000	0.999	0.990	0.999	0.998	0.987	0.999	0.996	0.984
ρ_1=.4	5	1.000	1.000	0.994	1.000	0.999	0.993	1.000	0.998	0.991
	7	1.000	1.000	0.996	1.000	0.999	0.995	1.000	0.998	0.993
	9	1.000	0.999	0.994	0.999	0.999	0.995	1.000	0.998	0.994
	1	0.999	0.998	0.975	0.998	0.997	0.977	0.998	0.997	0.979
	3	1.000	0.999	0.990	1.000	0.999	0.989	1.000	0.998	0.989
ρ_1=.8	5	1.000	1.000	0.994	1.000	0.999	0.994	1.000	0.999	0.992
	7	1.000	1.000	0.996	1.000	0.999	0.994	1.000	0.999	0.994
	9	0.998	0.999	0.999	0.999	1.000	0.998	1.000	0.999	0.996

dependent series. The beta pdf takes the form (Johnson et al., 1994):

$$f_{D|L=l}(d) = \frac{1}{B(p,q)} \frac{(d-a)^{p-1}(b-d)^{q-1}}{(b-a)^{p+q-1}} \quad (a \leq d \leq b) \tag{3.49}$$

where $B(p,q)$ is the complete beta function $\int_0^\infty t^{p-1}(1-t)^{q-1}dt$, and a and b are the lower and upper bound respectively. In our case, $a = 0$ and $b = lx_o$, since a drought of length l cannot have accumulated deficit greater than lx_o , and the parameters p, q can be estimated as a function of the first two moments of accumulated deficit $\mu_D = \mathbf{E}[D]$ and $\sigma_D^2 = \mathbf{Var}[D]$ as (Johnson et al., 1994):

$$p = \left(\frac{\mu_D}{lx_o}\right)^2 \left(1 - \frac{\mu_D}{lx_o}\right) \left(\frac{\sigma_D^2}{(lx_o)^2}\right)^{-1} - \frac{\mu_D}{lx_o} \tag{3.50}$$

$$q = \frac{\mu_D}{lx_o} \frac{1 - \frac{\mu_D}{lx_o}}{\frac{\sigma_D^2}{(lx_o)^2}} - 1 - p \tag{3.51}$$

Then, the bivariate pdf of accumulated deficit and length takes the following form:

$$f_{D,L}(d,l) = \frac{1}{B(p,q)} \frac{(d)^{p-1}(lx_o - d)^{q-1}}{(lx_o)^{p+q-1}} f_L(l) \quad (0 \leq d \leq lx_o) \tag{3.52}$$

where the parameters q and p can be estimated by combining eqs.(3.42) and 3.43) with(3.50) and(3.51) and $f_L(l)$ is the pdf of drought length assuming a DARMA(1,1) process.

Furthermore, from the conditional distribution of accumulated deficit given drought length (eq. 3.49), the conditional distribution of drought intensity I given length L can be also derived. Indeed, since the drought intensity is the ratio of accumulated deficit to drought length, i.e. I=D/l, the conditional pdf of intensity I given a fixed length L=l can be derived from eq.(3.49)as

$$f_{I|L}(i) = \frac{1}{B(p_i,q_i)} \frac{(i)^{p_i-1}(x_o - i)^{q_i-1}}{(x_o)^{p_i+q_i-1}} \quad (0 \leq i \leq x_o) \tag{3.53}$$

where p_i and q_i can be computed again using eqs.(3.50) and(3.51), using the moments given by eqs.(3.38) and(3.39).

Thus, the bivariate pdf of intensity and length can be found in a similar fashion as in(3.52) above. It follows:

$$f_{I,L}(i,l) = \frac{1}{B(p_i,q_i)} \frac{(i)^{p_i-1}(x_o - i)^{q_i-1}}{(x_o)^{p_i+q_i-1}} f_L(l) \quad (0 \leq i \leq x_o) \tag{3.54}$$

By integrating appropriately the bivariate pdf's, the occurrence probability of various drought events can be found. In particular, with reference to different types of critical drought events, the following expressions can be adopted:

(1) for drought event E = {D>D_0 and L= l_0 (l_0 =1,2,)}:

$$P[D > D_o, L = l_o] = \int_{D_0}^{l_o x_o} f_{D,L}(z, l_0) dz = \int_{D_0}^{l_o x_o} \frac{1}{B(p,q)} \frac{(z)^{p-1}(l_o x_o - z)^{q-1}}{(l_o x_o)^{p+q-1}} f_L(l_o) dz \tag{3.55}$$

(2) for drought event E = {D>D_0 and L $\geq$ l_0 (l_0 =1,2,)}:

$$P[D > D_o, L \geq l_o] = \sum_{l=l_0}^{\infty} \int_{D_0}^{l x_o} f_{D,L}(z, l) dz = \sum_{l=l_0}^{\infty} \int_{D_0}^{l x_o} \frac{1}{B(p,q)} \frac{(z)^{p-1}(l x_o - z)^{q-1}}{(l x_o)^{p+q-1}} f_L(l) dz \tag{3.56}$$

(3) for drought event E = {I > I_0 and L = l_0 (l_0 =1,2,)}::

$$P[I > I_o, L = l_o] = \int_{I_0}^{x_o} f_{I,L}(z, l_0) dz = \int_{I_0}^{x_o} \frac{1}{B(p_i,q_i)} \frac{(z)^{p_i-1}(x_o - z)^{q_i-1}}{(x_o)^{p_i+q_i-1}} f_L(l_o) dz \tag{3.57}$$

(4) for drought event E = {I > I_0 and L $\geq$ l_0 (l_0 =1,2,)}:

$$P[I > I_o, L \geq l_o] = \sum_{l=l_0}^{\infty} \int_{I_0}^{x_o} f_{I,L}(z, l) dz = \sum_{l=l_0}^{\infty} \int_{I_0}^{x_o} \frac{1}{B(p_i,q_i)} \frac{(z)^{p_i-1}(x_o - z)^{q_i-1}}{(x_o)^{p_i+q_i-1}} f_L(l) dz \tag{3.58}$$

Furthermore, the marginal probability of droughts events E={D > D_0 } or E={I > I_0 } can be obtained from eqs.(43) or (45), respectively, by letting l_0 =1. It should be noted that despite the apparent complexity of the above expressions, the integrations can be carried out efficiently making use of numerical tools for the beta pdf that are available in most mathematical and statistical software.

3.6 Return Period of Critical Drought Events E

A common approach to investigate hydrological events with a certain magnitude or greater is based on the concept of return period or recurrence interval. Such a concept has been widely used in engineering design of structures subject to extreme hydrological events such as floods. Recently some authors (Fernandez and Salas, 1999; Shiau and Shen, 2001; Bonaccorso et al., 2003; Gonzalez and Valdes, 2003; Salas et al., 2005) extended the concept in order to determine the return period related to drought events and derived some analytical expressions for its estimation. Gonzalez and Valdes (2006) derived a new Drought Frequency Index, (DFI) based on the concept of mean frequency of recurrence of extreme persistent events.

Following a traditional definition, the return period of a generic event with a certain magnitude of greater is defined as the average elapsed time, or mean inter-arrival time, between two of such events (Lloyd, 1970; Loaicigica and Mariño, 1991; Fernandez and Salas, 1999). When the event under analysis is extreme precipitation or streamflow, the traditional approach consists in analyzing the series X_t of annual maxima of the hydrological variable and to assume time independence for such observations. With reference to a certain magnitude x_o , the sequence of events $\{X_t > x_o\}, \{X_t \leq x_o\}$ can be seen as a sequence of i.i.d. Bernoulli r.v. (Bernoulli trials), with probability of success $P[X_t > x_o]$. It follows that the interarrival times between any two events $\{X_t > x_o\}$ is distributed according to a geometric distribution (Mood et al., 1974), and therefore the expected value of such interarrival time (return period) is given by the well known expression:

$$T = \frac{1}{P[X_t > x_o]} \tag{3.59}$$

The above expression however cannot be applied to droughts, since they usually span several years and therefore in general it is not possible to identify the time unit necessary to model the sequence of events $\{X_t > x_o\}, \{X_t \leq x_o\}$. In other words, since droughts have variable lengths, there is no "trial" that can be adopted as time unit.

Shiau and Shen (2001) have developed a formulation for return period of droughts with a fixed accumulated deficit or greater. Cancelliere and Salas (2004) extended such a formulation to drought length in periodic series, while Cancelliere et al. (2003) and Gonzalez and Valdes (2003) included other characteristics like drought length and intensity, under the assumption of lag-1 Markov dependence. Here the latter formulation will be extended also to the case of AR(1) dependence for the X_t process, making use of the approximation to the bivariate distributions of accumulated deficit and length or intensity and length previously derived.

For analyzing the severity of droughts and the associated risks we need to specify the drought event under consideration. Let us denote by $\{E\}$ a critical drought event like the ones previously defined and by $\{\bar{E}\}$ the corresponding non critical drought events. For example, if $\{E\} = \{D > D_0, L \geq l_0\}$, the corresponding non drought events $\{\bar{E}\}$ will be $\{D \leq D_0, L \geq l_0\} \cup \{D > D_0, L < l_0\} \cup \{D \leq D_0, L < l_0\}$.

With reference to Figure 3.7 , between any two critical droughts (filled circles), there must be 1 critical drought, namely the first, and $N_l - 1$ non critical droughts. It follows that the inter-arrival time between critical droughts T_l is the sum of the inter-arrival times between N_l droughts of which 1 is critical (the first) and the remaining N_l are non critical i.e.

$$T_l = L \mid \{E\} + Ln + \sum_{j=2}^{N_l} \left(L_j \mid \{\bar{E}\} + Ln_j\right) \tag{3.60}$$

where $L \mid \{E\}$ is the duration of a critical drought event $\{E\}$, Ln the duration of a non-drought event and $L \mid \{\bar{E}\}$ is the duration of non-critical drought events $\{\bar{E}\}$. Note that all of the above quantities are random variables. By definition, the return period T is the expected value of T_l , which can be computed by taking expectation

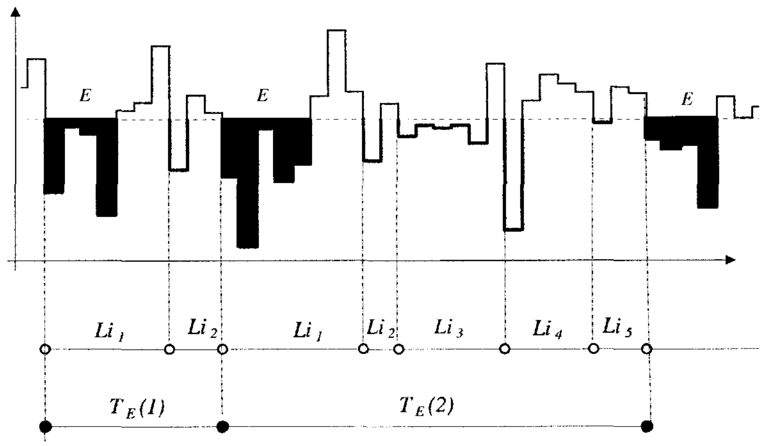


Figure 3.7: Definition of inter-arrival times of critical droughts E (e.g. $\{ D > D_0$ and $L \geq 4 \}$). The open circles denote the initiation of any droughts while the filled circles denote the initiation of critical droughts E.

on both side of eq.(3.60) as:

$$\mathbf{E}[T_l] \quad = \mathbf{E}[L \mid \{E\}] + \mathbf{E}[Ln] + \mathbf{E}\left[\sum_{j=2}^{N_l} (L_j \mid \{\bar{E}\} + Ln_j)\right]$$

$$\mathbf{E}[T_l] \quad = \mathbf{E}[L \mid \{E\}] + \mathbf{E}[Ln] + \mathbf{E}\left[\sum_{j=2}^{N_l} L_j \mid \{\bar{E}\}\right] + \mathbf{E}\left[\sum_{j=2}^{N_l} Ln_j\right] \tag{3.61}$$

Solving the expectations involving sums by means of conditional expectation concepts one gets:

$$\mathbf{E}[T_l] = \mathbf{E}[L \mid \{E\}] + \mathbf{E}[Ln] + [\mathbf{E}[N_l] - 1]\,\mathbf{E}\left[L \mid \{\bar{E}\}\right] + [\mathbf{E}[N_l] - 1]\,\mathbf{E}[Ln] \tag{3.62}$$

Recalling the definition of N_l , for $N_l = n_l$ droughts, there must be $n_l - 1$ non critical and only one critical drought E. Assuming independence between drought and/or non-drought events, the probability $P[N_l = n_l]$ can be determined as:

$$P[N_l = n_l] = (1 - P[E])^{n_l - 1} P[E] = (1 - P[E])^{n_l - 1} P[E] \tag{3.63}$$

which suggests that N_l is geometrically distributed with parameter $P[E]$. Then its expected value is (Mood et al., 1973):

$$\mathbf{E}\left[N_l\right] = \frac{1}{P[E]} \tag{3.64}$$

Therefore, from Eq.(3.64) we get

$$T = \mathbf{E}\left[T_l\right] = \mathbf{E}\left[L \mid \{E\}\right] + \mathbf{E}\left[Ln\right] + \left[\frac{1}{P[E]} - 1\right] \mathbf{E}\left[L \mid \{\bar{E}\}\right] + \left[\frac{1}{P[E]} - 1\right] \mathbf{E}\left[Ln\right] \tag{3.65}$$

Rearranging terms, observing that $P[E] + P[\bar{E}] = 1$, and making use of the following identity:

$$\mathbf{E}\left[L \mid \{\bar{E}\}\right] P[\bar{E}] + \mathbf{E}\left[L \mid \{E\}\right] P[E] = \mathbf{E}\left[L\right] \tag{3.66}$$

eq.(3.65) simplifies as:

$$T = \frac{\mathbf{E}\left[L\right] + \mathbf{E}\left[Ln\right]}{P[E]} \tag{3.67}$$

Note that in deriving eq.(3.67) only the independence between drought events has been assumed.

Equation(3.67) allows one to compute the return period of any critical drought $\{E\}$ once the probability $P[E]$ is defined and the expected value of drought and non-drought durations are known. The former probability can be computed by making use of the joint pdf's derived in the previous paragraph, while the expected value of drought and non-drought can be computed assuming a DARMA(1,1) process as an approximation for the clipped series Y_t.

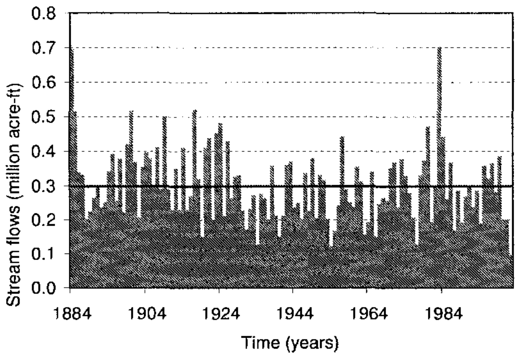
3.7 Applications

The proposed methodology to compute probabilities and return periods of droughts in stochastic series has been applied to four annual streamflow series, namely the Poudre River at Mouth of the Canyon, the Colorado River at Lee Ferry, the Nile River at Aswan, and the St. Lawrence River at Cornwall.

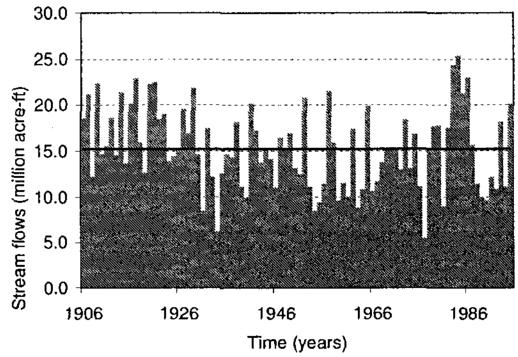
In figure 3.8 the time series of the investigated rivers are shown, along with the respective long term means. In all cases, several severe droughts can be identified on the series. For example, the series of Poudre flows indicates a wet period of about 40-45 years in the first part of the record and a drier period in the rest of the record; this wet/dry flow pattern has been characteristic in the region. In addition, the time series plot indicates the various drought episodes that have occurred in the Poudre River throughout the historical record, such as those of the 1930's and 1950's, as well as the recent three-year drought. Similarly, the other series indicate the presence of serious droughts that occurred in the different periods. In particular, the very severe drought that occurred in the 30's is clearly indicated by the St. Lawrence river plot.

From the comparison of the plots, the different degree of persistence of the four series can be inferred. For example the St. Lawrence river and the Nile river seem to exhibit some long term cycles which is characteristic of highly autocorrelated processes whereas this is less evident for the other two investigated series.

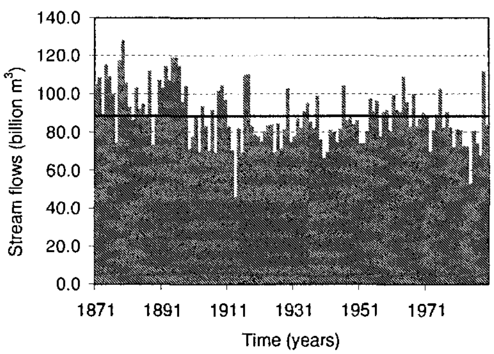
Table 3.V summarizes a number of basic flow statistics that were determined from the four annual flow records, which allow to compare the different hydrological regimes. From the table it can be inferred the different skewness of the investigated series, which is significantly different than zero for the Poudre whereas it is very low for the other three series. Also the table indicates the different lag 1 autocorrelation of



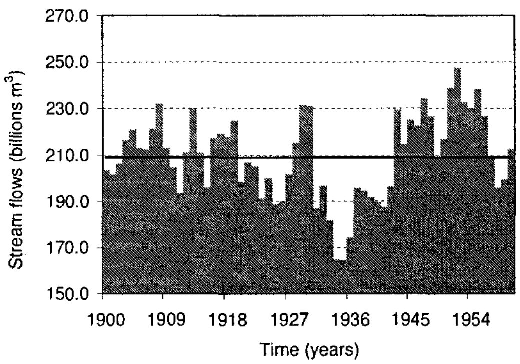
(a) Poudre River at Mouth of the Canyon annual flows



(b) Colorado at Lees Ferry annual flows



(c) Nile River at Aswan annual flows



(d) St. Lawrence river at Cornwall annual flows

Figure 3.8: Annual flow records of (a)the Poudre River (1888-2002), (b) Colorado at Lees Ferry (1906-1995), (c) Nile at Aswan (1871-1989) and (d) St. Lawrence at Cornwall (1900-1960). The figure shows some extreme drought episodes on all series as well as the different degrees of serial time dependence in the four series

.

Table 3.V: General description of the investigated streamflow series

Station and units	Period of record	Mean	Standard deviation	Skewness	Lag-1 autocorrelation
Poudre River at Mouth of the Canyon [acf*10^6]	1888-2002	.299	.11	.99	.153
Colorado River at Lees Ferry [acf*10^6]	1906-1995	15.23	4.39	.175	.277
Nile River at Aswan [m^3*10^9]	1871-1989	88.28	14.60	.25	.397
St. Lawrence River at Cornwall [$m^3 * 10^9$]	1900-1960	209.09	18.10	-.27	.712

the four series, ranging from an almost negligible value for the Poudre River (r_1=.153), to a relatively high value of $r_1 = .712$ for the St. Lawrence River.

In figure 3.9 the number of droughts of different lengths identified in the four investigated series is plotted. In all cases a threshold x_0 equal to the long term mean has been assumed. From the figure it can be inferred how the number of identified droughts is relatively low, even when short lengths are considered, and in any case certainly not enough for a reliable inference analysis about the probabilistic behaviour of droughts. For instance, with reference to the Poudre river, despite the fact that a fairly long record of observations is available (118 years), only 15 droughts of length 1, 5 of length 2 and 4 of length 3 are observed. Thus, any attempt to fit different pdf to accumulated deficit for each fixed length would lead to unreliable estimates of the parameters. This confirms the need of expressing the parameters of the distribution of accumulated deficit for fixed drought length as a function of the statistics of the underlying series, as proposed in the present chapter. Due to the limited number of droughts that can be identified in the investigated flow records, in addition to observed series, synthetically generated data has been used in order to verify the proposed analytical expressions to compute return period of droughts, However it may be worthwhile to mention that all the parameters of the analytical expressions have

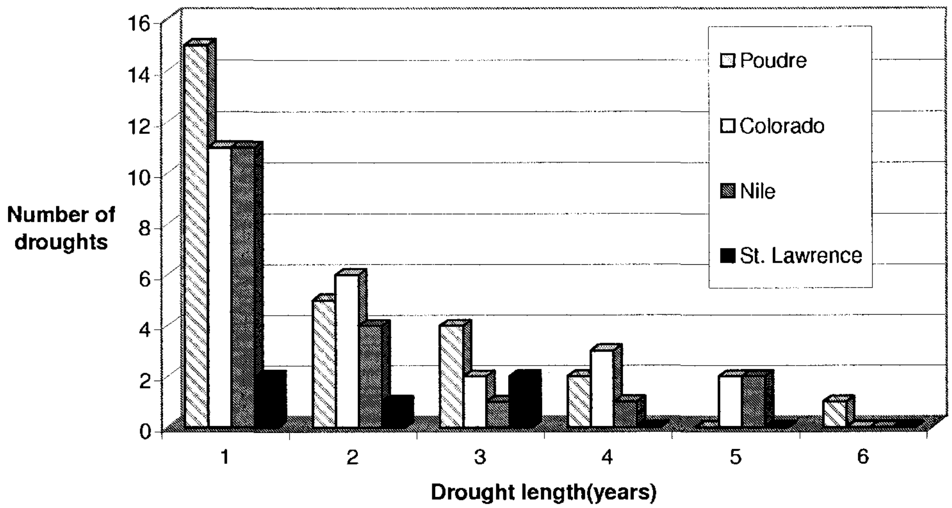


Figure 3.9: Number of droughts of different lengths identified in the four analyzed streamflow series

been fitted.making use only of the observations. In particular, AR(1) models have been fitted to the Poudre, Colorado and St. Lawrence series, while an ARMA(1,1) has been adopted to model the Nile series. Because the Poudre annual flows are significantly skewed, a log-normal AR(1) model has been adopted. Therefore, original data have been first transformed using a logarithmic transformation:

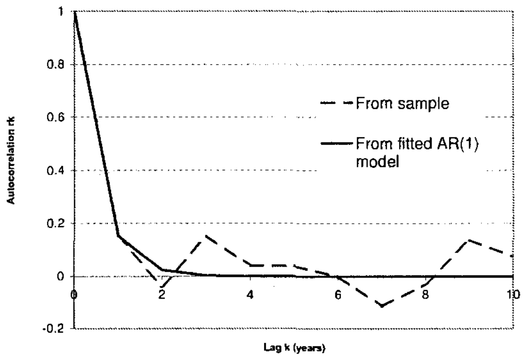
$$y_t = \ln X_t \tag{3.68}$$

that gave an approximately normally distributed flow series. After generation, the transformed series have been inverted into the original flow domain by using $X_t = e^{Y_t}$.

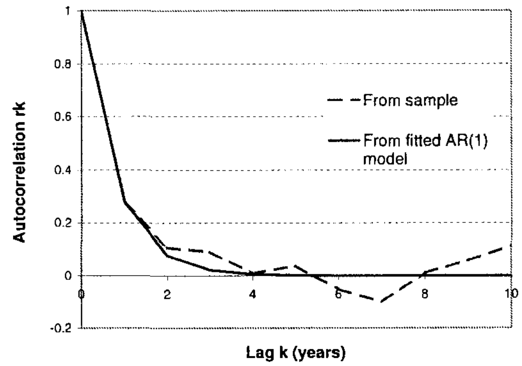
Since the other three series do not exhibit significant skewness, no transformation has been carried out.

In Figures 3.10, the autocorrelation function r_k of the sample series and of the fitted models are shown for lags k 0-10 years.

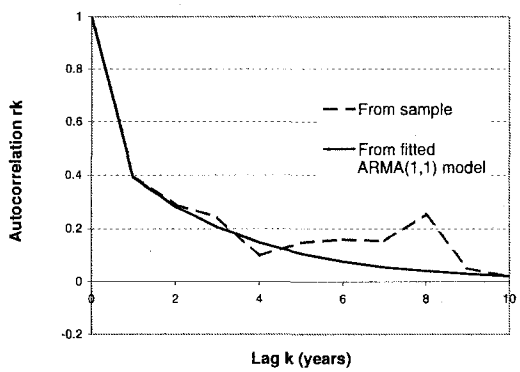
The fitted models have been used for simulating 3,000,000-year samples from



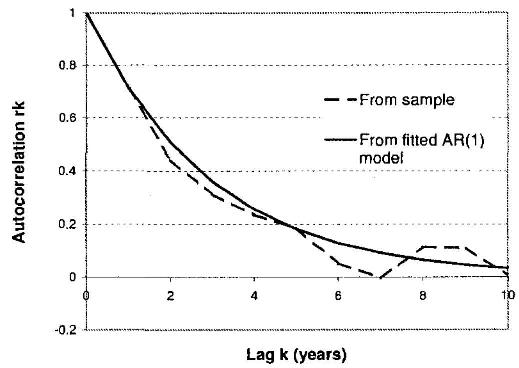
(a) Poudre River at Mouth of the Canyon annual flows



(b) Colorado at Lees Ferry annual flows



(c) Nile River at Aswan annual flows



(d) St. Lawrence river at Cornwall annual flows

Figure 3.10: Sample autocorrelation function of the annual flows of (a)the Poudre River (1888-2002), (b) Colorado at Lees Ferry (1906-1995), (c) Nile at Aswan (1871-1989) and (d) St. Lawrence at Cornwall and autocorrelation function of fitted AR(1) models

.

which the return period of different types of drought events have been determined. The drought analysis has been performed by assuming a parameterization for the deficit threshold D $_0$ as a fraction of the threshold x_o, by means of a deficit coefficient δ, i.e.

$$D_0 = \delta x_o \delta = 0.0, 0.25, ..., 2.00 \tag{3.69}$$

Likewise, the threshold drought intensity I_0 is defined as a fraction of x_o , i.e. $I_0 = \zeta x_o$, ζ=0.0, 0.10, ...,0.40 , and ζ is called the drought intensity coefficient. Then one can determine the magnitude of various droughts by simply varying the coefficients δ and ζ as the case may be. The water demand threshold x_0 has been assumed in all cases equal to the long term mean μ_x.

Then, DARMA(1,1) models have been fitted to the binary series obtained by clipping the observations by the respective thresholds. In Figure 3.11, the drought length pdf corresponding to each fitted DARMA(1,1) model are compared with the corresponding drought length frequencies computed from the observations and from the generated series. In the plot, approximate 90% confidence intervals for the drought length frequencies from the observations are also shown. These confidence intervals have been obtained by generating 1000 series of length equal to each observed sample using the corresponding DARMA(1,1) model, and by computing the drought length frequencies. Then, for each drought length, 5% and 95% percentiles have been computed from the 1000 values obtained by generation. The relative width of the confidence intervals is indicative of the sampling variability of the drought length frequencies computed from the observed droughts in a sample. From the plot, a general good agreement between the drought length pdf obtained from the DARMA(1,1) model and the observed frequencies can be inferred, showing that a DARMA(1,1) model is able to model drought length probabilities if the underlying series exhibit

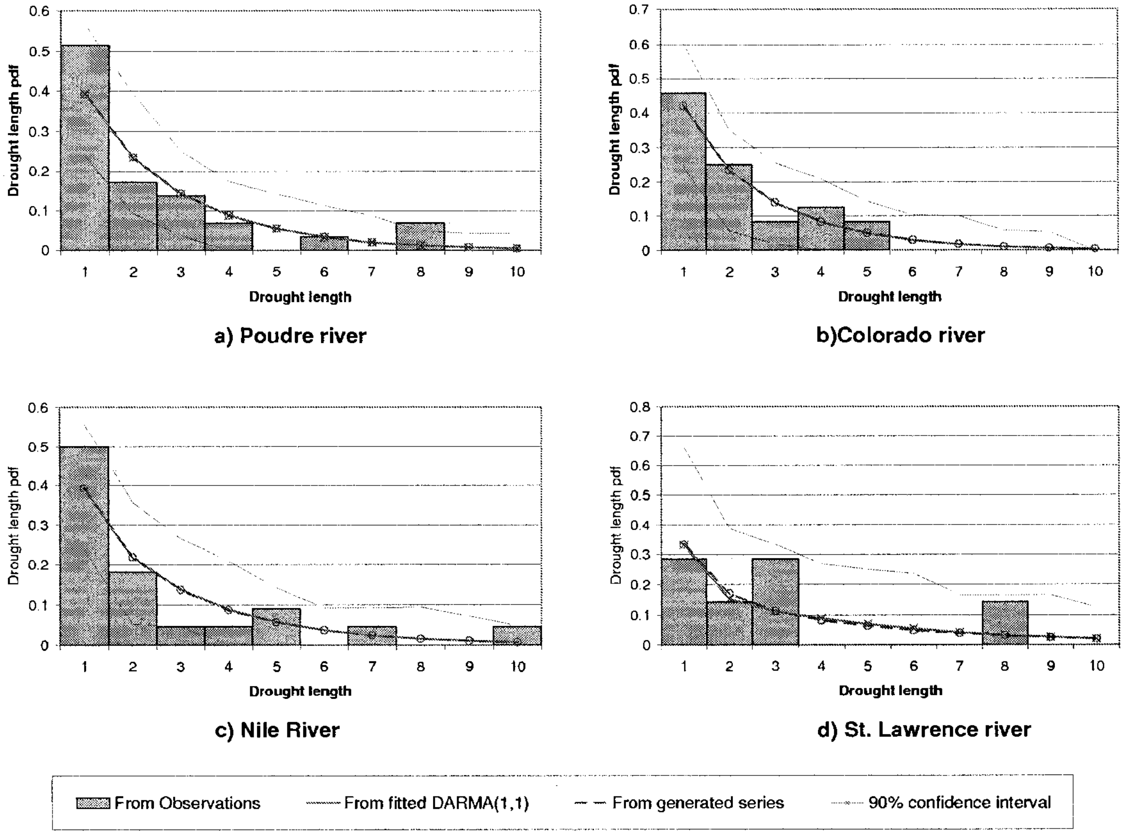


Figure 3.11: Drought length pdf obtained from the observed series, from fitted DARMA(1,1) model and from clipped generated series for the four investigated rivers. Threshold $x_0 = \mu_x$. The approximate 90% confidence intervals for the sample pdf are also shown.

auocorrelation, as is the case for AR(1) or ARMA(1,1) models. In order to model accumulated deficit for fixed drought length, the parameters of the beta pdf have been computed from the basic statistics of the analyzed series. In particular, first the moments of accumulated deficits have been computed by means of the proposed approximations given by eqs.(3.42) and(3.43), as a function of the expected value and variance of single deficits, of the skewness coefficient and of the lag-1 autocorrelation coefficient. For Colorado, Nile and St. Lawrence (normal case), the expected value and variance of single deficits for the serially independent case $\mathbf{E}[D_1(0)]$ and

$\mathbf{Var}[D_1(0)]$ have been computed as (Bonaccorso et al., 2003):

$$\mathbf{E}[D_1(0)] = x_0 - \mu_x + \frac{\sigma_x}{p_0}\phi\left(\frac{x_0 - \mu_x}{\sigma_x}\right) \tag{3.70}$$

and

$$\mathbf{Var}[D_1(0)] = \frac{\sigma_x}{p_0}\phi\left(\frac{x_0 - \mu_x}{\sigma_x}\right) \cdot \left[(\mu_x - x_0) - \frac{\sigma_x}{p_0}\phi\left(\frac{x_0 - \mu_x}{\sigma_x}\right)\right] + \sigma_x^2 \tag{3.71}$$

where $p_0 = \Phi\left(\frac{x_0 - \mu_x}{\sigma_x}\right)$.

For the Poudre river, similar expressions have been adopted, valid for log-normal series with parameters μ_y and σ_y (Bonaccorso et al., 2003):

$$\mathbf{E}[D_1(0)] = x_0 - \frac{\mu_x}{p_0} \cdot \Delta \tag{3.72}$$

and

$$\mathbf{Var}[D_1(0)] = \frac{\mu_x^2 e^{\sigma_y^2}}{p_0} \cdot \Psi - \frac{\mu_x^2}{p_0^2} \cdot \Delta^2 \tag{3.73}$$

where $\mu_x = e^{\mu_y + \frac{\sigma_y^2}{2}}$, $\Delta = \Phi\left(\frac{\ln x_0 - \mu_y}{\sigma_y} - \sigma_y\right)$ and $\Psi = \Phi\left(\frac{\ln x_0 - \mu_y}{\sigma_y} - 2\sigma_y\right)$.

Then the parameters p and q of the joint distribution of accumulated deficit and length or intensity and length have been estimated by means of eqs.(3.50) and(3.51).

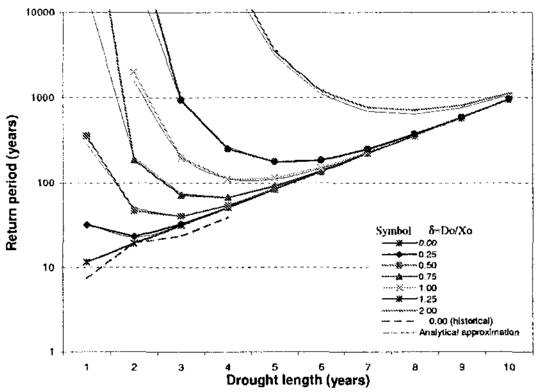
Finally, return period has been computed by means of Eq.(3.67), where the probability $P[E]$ of a critical drought E has been computed using the Eqs.(3.55),(3.56),(3.57), and(3.58), according to the different types of critical events considered. Figure 3.12 shows the return periods of droughts defined as in case(1) above, i.e. E = $\{\ D > D_0$ and $L = l_0\ (\ l_0 = 1, 2, \ldots\ )\}$, which were obtained from the generated samples and from Eq.(3.67) for various values of the deficit coefficient δ and for a threshold $x_o = \mu_x$. Several interesting features may be observed. First of all, for a given drought duration

l_0 , the return period $T \to \infty$ as $\delta \to \infty$, which means that estimating the return period T for large values of δ may require a very long sample. On the other hand, Fig. 3.12 shows that using Eq.(3.67) one can always find the results for any δ and any l_0. Figure 3.12 also shows the close results that are obtained from the generated samples and from Eq.(3.67). This is also true in the case of Poudre river, thus confirming the validitiy of the procedure when skewed series are analyzed. However, one may note that as δ increases there are some differences between them, although such differences cannot be considered particularly significant, since they are generally related to droughts with return period exceeding 1000 years, thus for a range of return period of little interest from a practical standpoint. The figure also shows that for all values of the deficit coefficient δ , as the drought duration l_0 increases, the return period T converges to the return period curve for $\delta = 0$. The comparison between the results obtained for the four rivers reveals also some features related to the presence of persistence in the series. In particular, the curve related to $\delta = 0$, which represents critical droughts of fixed length, and of any accumulated deficit, is less steep in the case of the Nile and St. Lawrence rivers than in the other cases. This is consistent with the fact that such series, being more persistent, will tend to exhibit longer droughts than the other, less persistant series, and therefore return period of longer droughts will be higher, whereas short droughts will have higher return periods. Figure3.12 shows also the return periods obtained from the historical streamflows for δ =0 and for some l_0. In order to somewhat increase the reliability of the estimation, return period has been computed by considering, besides interarrival times between critical droughts, also the first arrival time (beginning from the first year of the series), as well as the time from the last critical drought to the last year of the series. Nonetheless, not many drought episodes can be observed from the historical records, and therefore the return periods estimated from the samples have a high variability, and cannot be considered reliable for $\delta > 0$.

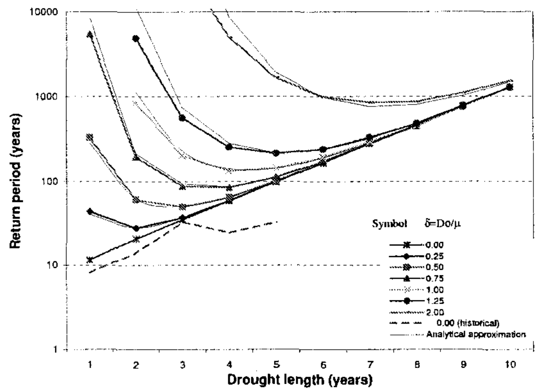
Figure 3.13 shows the return periods of droughts defined as in case(2) above, i.e. E $= \{ D > D_0$ and $L \geq l_0$ ($l_0 = 1, 2, ...$)}, which were obtained from the generated samples and from Eq.(3.67) for various values of δ. The figure shows that as in case(1) above, as l_0 increases all return period curves converge to a single curve that is independent of δ. As in the previous case, the plot confirms the close results that are obtained from the generated samples and from Eq.(3.67).
Also, Fig. 3.13 indicates that return periods of critical droughts generally decrease as the persistence of the series increases. This is particularly evident by comparing the curves related to the Poudre river (Fig.3.13(a)) with the ones related to the St. Lawrence river (Fig. 3.13(d)).

In addition, Fig. 3.13 shows the return periods based on the historical record for some values of l_0 and δ. Unlike the case(1) droughts where only a few estimates of return periods can be made for very small values of l_0 and δ (Fig. 3.12), for case(2) droughts, estimates of return periods based on the historical records can be made for several values of l_0 and δ as shown in Fig. 3.13. This allows comparing the return periods obtained from the generated sample (or from Eq.(3.67)) versus those obtained from the historical records so that further evaluation of the simulation model and the analytical formulation be made. However, also in this case the results obtained from the historical sample must be used with caution because as l_0 increases and as δ increases the return periods estimated from the historical sample becomes unreliable. For instance, in Fig. 3.13 the results are not reliable for $l_0 > 4$ or $\delta > 1$.

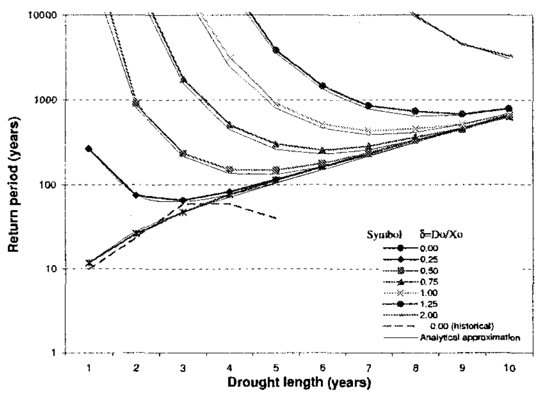
Figure 3.14 shows the return periods for droughts defined as in case(3) above, i.e. E $= \{ I > I_0$ and $L = l_0$ ($l_0 = 1, 2, ...$)}, which were obtained from the generated sample and from Eq.(3.67) for various values of the drought intensity coefficient ζ. The return period curves are increasing functions of both l_0 and ψ. The figure also



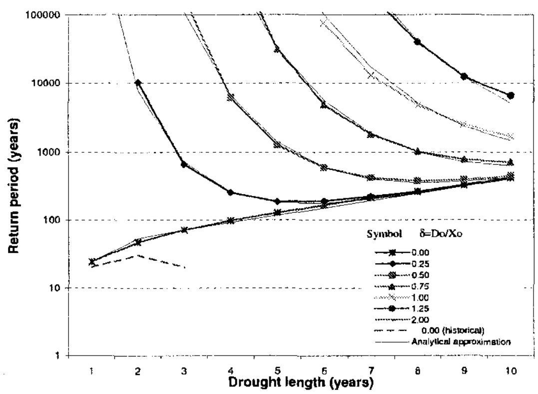
(a) Poudre River annual flows



(b) Colorado at Lee Ferry annual flows



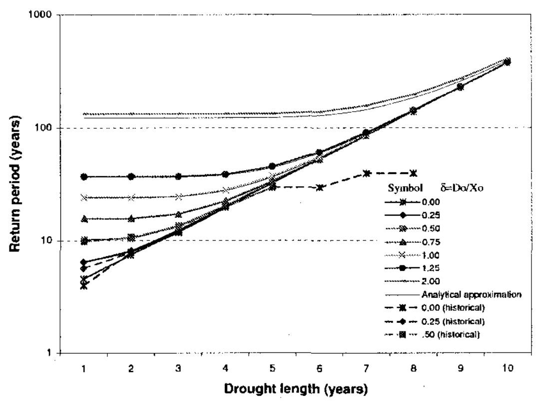
(c) Nile at Aswan annual flows



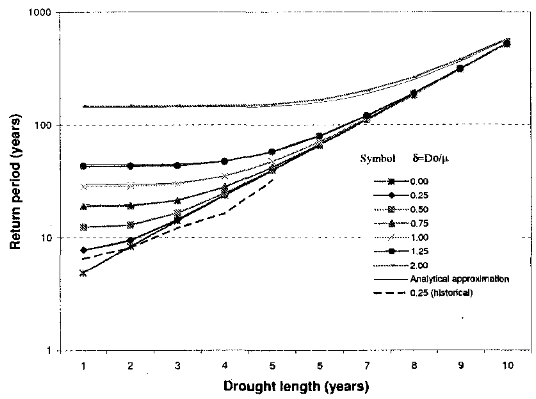
(d) St. Lawrence River annual flows

Figure 3.12: Return period of drought events defined by $\{ D > D_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$ obtained from the generated synthetic annual flows of the investigated rivers and from Eq.(3.67) for various values of the deficit coefficient δ ($D_0 = \delta x_o$). Threshold $x_o = \mu_x$. The return period curves for δ=0 obtained from the historical samples are also shown

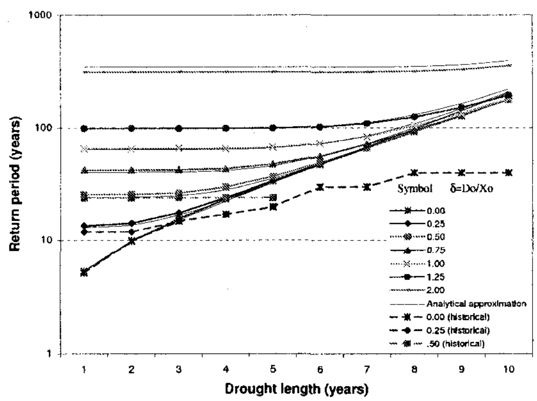
.



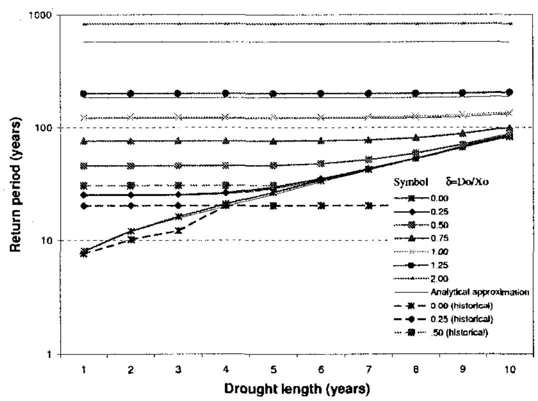
(a) Poudre River annual flows



(b) Colorado at Lee Ferry annual flows



(c) Nile at Aswan annual flows

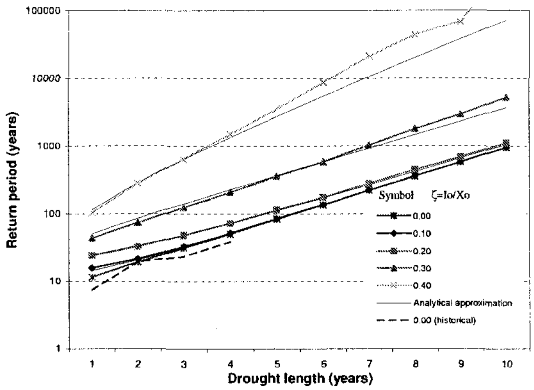


(d) St. Lawrence River annual flows

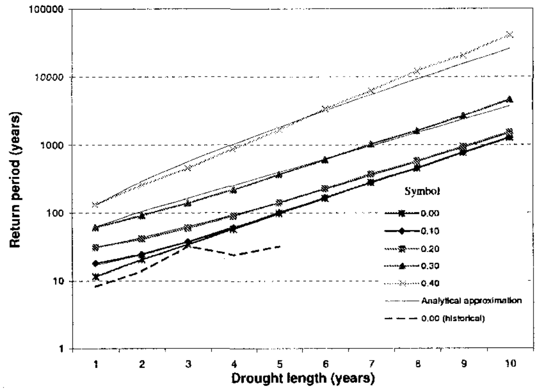
Figure 3.13: Return period of drought events defined by $\{\ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$ obtained from the generated synthetic annual flows of the investigated rivers and from Eq.(3.67) for various values of the deficit coefficient δ ($D_0 = \delta x_o$). Threshold $x_o = \mu_x$. The return period curves for a few δ values obtained from the historical samples are also shown

shows results obtained from the historical sample for a few values of l_0 and $\zeta = 0$. Again in this case the results are not reliable for $l_0 > 2$ for the same reasons already discussed above. In addition, Fig. 3.14 shows that as the persistence of the series increases, the steepness of the curves decreases, which is somewhat in agreement with case(1) above. This is particularly evident for the St. Lawrence river (Fig. 3.14(d)), for which the return periods corresponding to relatively high intensity coefficients ($\zeta > .1$) does not change significantly with drought length. Furthermore, the results obtained from the generated samples and from Eq.(3.67) are quite comparable with some exceptions, as is the case of the Poudre river (Fig. 3.14(a)) for $\zeta = .40$. As in case(1) above, it should be noted however that the largest differences are observed for very high return periods (>1000).

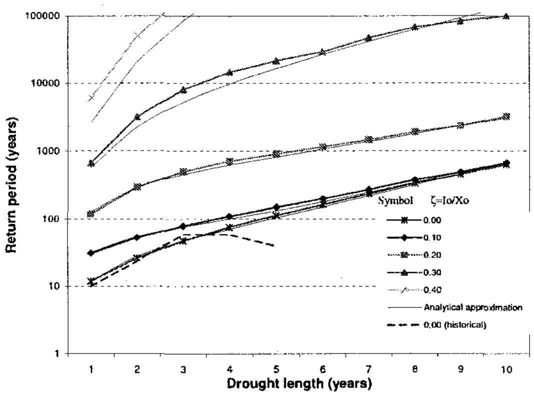
The return periods for drought events defined as in case(4) above, i.e. $E = \{ I > I_0 \text{ and } L \geq l_0\}$ are shown in Fig. 3.15 for various values of l_0 and ζ. The return period curves are increasing functions of both l_0 and ζ. From Fig. 3.15 a similar behaviour as in cases(1) and(3) above can be inferred, namely that as the persistence of the series increases, the steepness of the curves decreases. The figure also shows that return periods obtained from the generated samples and from Eq.(3.67) are quite comparable for low values of ζ , while the differences generally increase with l_0 and ζ. Likewise, Fig. 3.15 shows that the return periods obtained from the historical record for $\zeta = 0$ and $l_0 < 5$ correspond quite well with those obtained from the generated samples and from Eq.(3.67) although some underestimation can be observed, especially when drought length increases. This is not surprising, since, the number of observed droughts decreases with l_0 , and therefore the historical return period estimates become more and more unreliable.



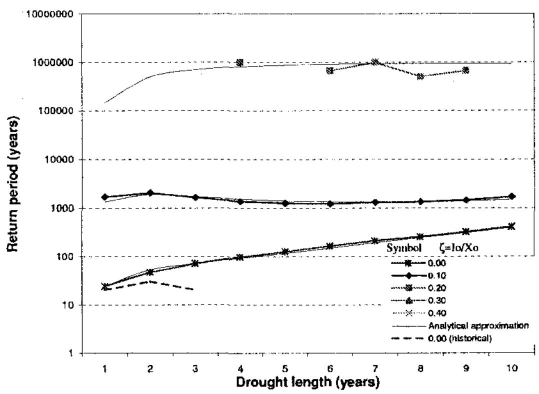
(a) Poudre River annual flows



(b) Colorado at Lee Ferry annual flows

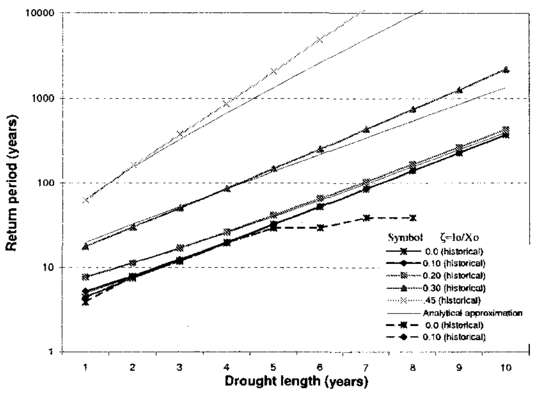


(c) Nile at Aswan annual flows

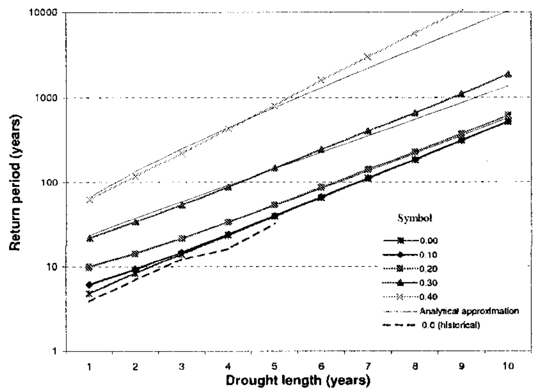


(d) St. Lawrence River annual flows

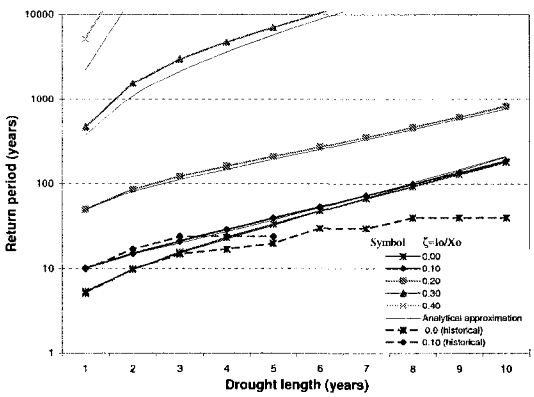
Figure 3.14: Return period of drought events defined by $\{\ I > I_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$ obtained from the generated synthetic annual flows of the investigated rivers and from Eq.(3.67) for various values of the intensity coefficient ζ ($I_0 = \zeta x_o$). Threshold $x_o = \mu_x$. The return period curves for ζ=0 obtained from the historical samples are also shown



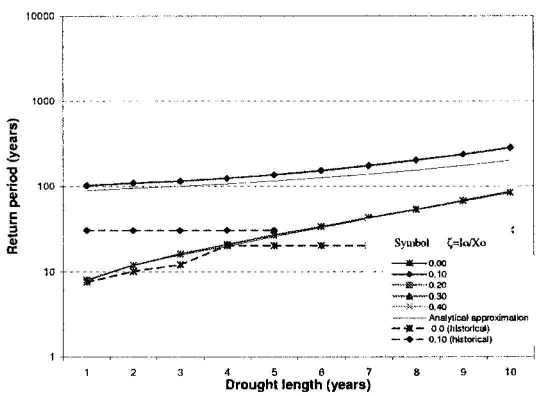
(a) Poudre River annual flows



(b) Colorado at Lee Ferry annual flows



(c) Nile at Aswan annual flows



(d) St. Lawrence River annual flows

Figure 3.15: Return period of drought events defined by $\{\ I > I_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$ obtained from the generated synthetic annual flows of the investigated rivers and from Eq.(3.67) for various values of the intensity coefficient ζ ($I_0 = \zeta x_o$). Threshold $x_o = \mu_x$. The return period curves for a few ζ values obtained from the historical samples are also shown

3.8 Conclusions

An accurate planning and management of water resources systems must take into account the occurrence of droughts. For example, the implementation of water use restrictions in a given water district is very much related to the degree of severity that an evolving drought reaches at a given point in time. While this question has been typically answered using drought indices such as the Palmer Drought Severity Index (Palmer, 1965) or the Standardized Precipitation Index (McKee et al., 1993), unfortunately they do not provide any information on the frequency or recurrence time of the drought.

In this chapter the concept of return period (mean interarrival time) has been applied to characterize the severity of extreme droughts. In particular, a formulation that enables the estimation of return period of critical droughts occurring on stochastic time-dependent hydrological variables has been presented. The formulation allows to specify drought events taking into account not only the duration of the drought but drought deficit or drought intensity, etc. Although previous studies exist where droughts are analyzed in a bivariate fashion, (e.g. citetGonzalez2003; Cancelliere et al. (2003)), however they have usually neglected the time-dependence in the hydrological series under investigation. Indeed, it has been shown that when time dependence is considered, even computing the first two moments of drought characteristics becomes difficult, since in this case the series of the deficits resulting after the hydrological variable is truncated by means of a water demand threshold is not stationary. Thus a fully multivariate approach is needed. Here a truncated multivariate normal has been proposed to model the deficits, that enables to compute by numerical integration the first two moments of accumulated deficit or intensity, for droughts of fixed length.

Furthermore, when a time dependent process, such as an AR(1) is clipped by a constant threshold level, the resulting binary process does not follow a lag 1 Markov chain, and therefore the widely used geometric distribution to model drought length may become not appropriate. Thus, drought length has been modeled asssuming as an approximation that the clipped binary process follows a DARMA(1,1) model, for which several theoretical and practical results are available (Jacobs and Lewis, 1977; Chang et al., 1984a; Chung and Salas, 2000).

Also, empirical approximations of the moments of accumulated deficit, function of the lag-1 autocorrelation of the underlying hydrological variables and of the threshold level are proposed , that allow to derive the pdf of drought characteristics without resorting to numerical algorithms to solve the multivariate integrals. The empirical approximations are such that they can be applied also to skewed data.

Such moments have then been used to define the distribution of the accumulated deficit, assumed beta. The result is a bivariate formulation for drought characteristics length and accumulated deficit or intensity pdf, that allows to compute the probability of occurrence of several types of critical droughts as a function of the basic statistics (mean, standard deviation, lag-1 autocorrelation coefficient and skewness coefficient) of the underlying hydrological variable, as well as of the threshold level. The derived pdf have been applied to compute the return period of critical droughts, extending the methodology proposed by Shiau and Shen (2001) in order to consider also drought length.

The procedure has been applied to four streamflow annual series, namely the annual flows of the Poudre River at Mouth of the Canyon, of the Colorado River at Lee Ferry, of the Nile River at Aswan, and the St. Lawrence River at Cornwall. The four series exhibit different level of persistence thus enabling to test the methodology when different hydrological regimes are considered. After AR(1) and ARMA(1,1) models have been fitted to the series and synthetic series have been generated, the comparison between return periods obtained by means of the analytical approximations and those computed on the observed and generated samples has been carried out. The results of such comparison confirm the applicability of the proposed methodology, since it is able to capture reasonably well the stochastic variability of droughts in streamflow series with different degrees of persistence.

The overall conclusion of the present chapter is that even when time dependence in the underlying hydrological series is taken into account, drought occurrences and return periods can still be modeled by means of an analytical approach, although is some cases, empirical approximations are necessary to overcome the difficulties related to solving complex integrals. Such an approach is particularly useful in light of the generally limited sample length of observed hydrologic records, which hinders the applicability of an inferential approach to model the occurrence of droughts.

CHAPTER IV

PROBABILISTIC MODELLING OF DROUGHT LENGTH IN PERIODIC-STOCHASTIC SERIES

4.1 General

Drought length characteristics, such as the probability mass function and moments, are derived herein assuming that the underlying water supply series after is clipped by a periodic water demand threshold level results in a two-state periodic Markov chain. For this purpose we use the drought definition as in Yevjevich (1967). Let us consider a periodic stochastic water supply series denoted as $X_{\nu,\tau}$, $\nu = 1, 2, ...$, $\tau = 1, ..., \omega$ where ν represents the year, τ represents the season (e.g. a month), and ω is the number of seasons (e.g. $\omega = 12$ for monthly data). In addition, consider a variable water demand series denoted by $x_{o\tau}$. Then a drought event is defined as a succession of consecutive periods (run) in which the water supply remains below the threshold level $x_{o\tau}$. Thus, the drought length L (length of negative run) is the number of consecutive time intervals (seasons) in which $X_{\nu,\tau} < x_{o\tau}$ preceded and followed by (at least one season where) $X_{\nu,\tau} \geq x_{o\tau}$. This drought definition is schematically shown in 4.1.

These conditions can be also represented using a Bernoulli variable $Y_{\nu,\tau}$ as

$$
\begin{cases} Y_{\nu,\tau} = 0 & if \;\; X_{\nu,\tau} < x_{o\tau} \\ Y_{\nu,\tau} = 1 & if \;\; X_{\nu,\tau} \geq x_{o\tau} \end{cases} \tag{4.1}
$$

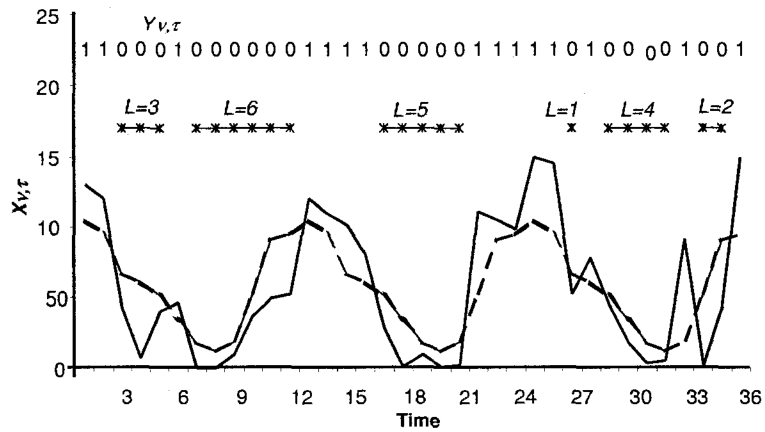


Figure 4.1: Definition of drought episodes for a periodic-stochastic water supply series $X_{\nu,\tau}$ (thick line) and a periodic water demand $x_{0\tau}$ (line of segments). The resulting discrete process $Y_{\nu,\tau}$ is also shown.

Obviously there is a one to one correspondence between the sequence of deficits identified using $X_{\nu,\tau}$ and the sequence of zeroes in $Y_{\nu,\tau}$. Thus for analyzing the drought length properties we will restrict our attention to the latter process $Y_{\nu,\tau}$.

The probability mass function (pmf) of drought length can be derived assuming that the variable $Y_{\nu,\tau}$ follows a two-state periodic Markov chain, i.e.:

$$
\begin{aligned}
\mathrm{P}\left[Y_{\nu,\tau} = j_\tau \mid Y_{\nu,\tau-1} = j_{\tau-1}, Y_{\nu,\tau-2} = j_{\tau-2}, Y_{\nu,\tau-3} = j_{\tau-3}, \ldots\right] = \\
\mathrm{P}\left[Y_{\nu,\tau} = j_\tau \mid Y_{\nu,\tau-1} = j_{\tau-1}\right]
\end{aligned} \tag{4.2}
$$

Equation 4.2 states that the probability of a value of the Markov chain at a given time interval ν, τ conditioned on past values depends only on the value of the chain at time $\nu, \tau - 1$. The Markovian assumption for the discrete process $Y_{\nu,\tau}$ may be useful where the original process representing the water supply, $X_{\nu,\tau}$, follows the well-known periodic lag-l autoregressive process (Salas, 1993). Two-state periodic Markov models have been applied, among others, by Fernandez and Vergara (1998) to determine the return period of droughts for monthly precipitation and streamflow series and by Katz and Parlange (1995) to model hourly precipitation.

Because $Y_{\nu,\tau}$ has a Markov chain structure, its multivariate distribution is determined uniquely once the periodic transition probability matrices are fixed. They are given by

$$P_\tau = \begin{pmatrix} p_{00\tau} & p_{01\tau} \\ p_{10\tau} & p_{11\tau} \end{pmatrix} \quad , \quad \tau = 1, \ldots, \omega \tag{4.3}$$

such that $P_\tau = P_{\tau+k\omega}$, $k = ..., -1, 0, 1, ...$ and the elements $p_{ij\tau}$ are given by

$$p_{ij\tau} = P\left(Y_{\nu,\tau} = j \mid Y_{\nu,\tau-1} = i\right) \quad , \quad i = 0, 1; \; j = 0, 1 \tag{4.4}$$

In addition,

$$p_{i\tau} = P(Y_{\nu,\tau} = i), \;\; i = 0, 1 \tag{4.5}$$

denote the unconditional probabilities. Furthermore, the foregoing probabilities must satisfy:

$$p_{i0_\tau} + p_{i1_\tau} = 1 \;\; , \;\; i = 0, 1 \tag{4.6}$$

and

$$p_{1_\tau} = p_{1_{\tau-1}}\left(1 - p_{10_\tau}\right) + p_{01_{\tau-1}}\left(1 - p_{1_{\tau-1}}\right) \tag{4.7}$$

for $\tau = 1, ..., \omega$. Equation 4.7 is a direct consequence of the theorem of total probabilities.

4.2 Probability mass function of drought lengths

The conditional probability that a drought has length l given that it starts at season τ may be derived as

$$\begin{aligned} f_{L\tau}(l) &= P\left[L = l\,|drought\ starts\ at\ time\ \tau\right] = \frac{P[L = l \cap drought\ starts\ at\ time\ \tau]}{P(drought\ starts\ at\ time\ \tau)} \\ &= \frac{P[L_\tau l]}{P(drought\ starts\ at\ time\ \tau)} \end{aligned} \tag{4.8}$$

where $L_\tau = l$ denotes a drought of length l that starts at season τ. The joint probability in the numerator denoted as $P[L_\tau = l]$ is equal to

$$\mathrm{P}\left[L_\tau = l\right] = \mathrm{P}\left[Y_{\nu,\tau-1} = 1, Y_{\nu,\tau} = 0, ..., Y_{\nu,\tau+l-1} = 0, Y_{\nu,\tau+l} = 1\right] \tag{4.9}$$

Because of the Markov chain structure of $Y_{\nu,\tau}$ 4.9 may be expressed as

$$P\left[L_\tau = l\right] = p_{1_{\tau-1}}p_{10_\tau}\overbrace{p_{00_{\tau+1}}p_{00_{\tau+2}}\cdots p_{00_{\tau+l-1}}}^{l-1\text{ terms}}p_{01_{\tau+l}} = p_{1_{\tau-1}}p_{10_\tau}\left(\prod_{i=1}^{l-1}p_{00_{\tau+i}}\right)p_{01_{\tau+l}} \tag{4.10}$$

Equation 4.10 represents the probability that a drought (run) lasts l time intervals when it begins at season τ. In addition, since $P(drought\ starts\ at\ time\ \tau) = p_{1_{\tau-1}}p_{10_\tau}$, then 4.8 simplifies to

$$f_{L_\tau}(l) = \frac{P\left[L_\tau = l\right]}{p_{1_{\tau-1}}p_{10_\tau}} = \left(\prod_{i=1}^{l-1}p_{00_{\tau+i}}\right)p_{01_{\tau+l}} = \left(\prod_{i=1}^{l-1}\left(1 - p_{01_{\tau+i}}\right)\right)p_{01_{\tau+l}} \qquad l = 1, 2, ... \tag{4.11}$$

Equation 4.11 represents the pmf of drought length given that the drought begins at season τ.

Furthermore, the probability that a drought of length L starts at any season can be derived as follows. Denote $DS = drought\ starts\ at\ time\ \tau$, then

$$f_L(l) = P\left[L = l \,|DS\right] = \frac{P[L = l \cap DS]}{P(DS)} \tag{4.12}$$

The probability in the numerator can be determined using the theorem of total probability as:

$$P\left[L = l \cap DS\right] =$$

$$= \sum_{\tau=1}^{\omega} P\left[L = l \cap DS | drought\ starts\ at\ time\ \tau\right] P\left(drought\ starts\ at\ time\ \tau\right)$$

$$= \sum_{\tau=1}^{\omega} f_{L\tau}(l) P\left(drought\ starts\ at\ time\ \tau\right) \tag{4.13}$$

Since $P(drought\ starts\ at\ season\ \tau) = p_{1\tau-1}p_{10\tau}$ the foregoing expression can be rewritten as:

$$P[L = l \cap DS] = \sum_{\tau=1}^{\omega} f_{L\tau}(l) p_{1\tau-1} p_{10\tau} \tag{4.14}$$

In addition, the probability in the denominator of 4.12 can be determined as

$$\begin{aligned} P\left(DS\right) &= \sum_{\tau=1}^{\omega} P\left(DS \cap season\ \tau\right) = \sum_{\tau=1}^{\omega} P\left(drought\ starts\ at\ time\ \tau\right) && (4.15) \\ &= \sum_{\tau=1}^{\omega} p_{1\tau-1} p_{10\tau} && (4.16) \end{aligned}$$

Substituting (4.14) and (4.16) into (4.12) gives

$$f_L(l) = P\left[L = l \,|\, DS\right] = \frac{\sum_{\tau=1}^{\omega} f_{L\tau}(l) p_{1\tau-1} p_{10\tau}}{\sum_{\tau=1}^{\omega} p_{1\tau-1} p_{10\tau}} \tag{4.17}$$

Therefore the probability that a drought lasts l time periods given that it begins at any season is a weighted average of the seasonal probabilities with weighting coefficients $p_{1\tau-1}p_{10\tau}$

The expression (4.17) represents the pmf of drought length when the underlying (0,1) process is periodic first-order Markov. As a particular case, the pmf of drought length for periodic independent processes can be obtained using

$$p_{01_\tau} = p_{1_\tau} p_{00_\tau} = p_{10_\tau} = p_{0_\tau} = 1 - p_{1_\tau} \; \tau = 1, 2, \ldots, \omega \tag{4.18}$$

Therefore, (4.17) reduces to

$$f_L(l) = \frac{\sum_{\tau=1}^{\omega} \left(\prod_{i=1}^{l-1} \left(1 - p_{1_{\tau+i}}\right)\right) \cdot p_{1_{\tau+l}} p_{1_{\tau-1}} p_{0_\tau}}{\sum_{\tau=1}^{\omega} p_{1_{\tau-1}} p_{0_\tau}} \tag{4.19}$$

Clearly for the stationary case and a constant threshold, (4.11), (4.17), and (4.19) are applicable by making $\omega = 1$. This yields the following expressions

$$f_L(l) = p_{00}^{l-1} p_{01} = (1 - p_{01})^{l-1} p_{01} \tag{4.20}$$

$$f_L(l) = p_0^{l-1} p_1 = p_1 (1 - p_1)^{l-1} \tag{4.21}$$

which are valid for the serially dependent and independent cases, respectively. The expressions (4.20) and (4.21) were previously derived by Sen (1976) following a different approach.

4.3 Moments of drought lengths

It may be useful deriving the moments of drought lengths in closed form. The r-th moment about the origin of drought length for periodic first-order Markov process is given by:

$$E\left[L^r\right] = \sum_{l=1}^{\infty} l^r f_L(l) = \sum_{l=1}^{\infty} l^r \frac{\sum\limits_{\tau=1}^{\omega} \left(\prod_{i=1}^{l-1} \left(1 - p_{01_{\tau+i}}\right) p_{01_{\tau+l}} p_{1_{\tau-1}} p_{10_\tau} \right)}{\sum\limits_{\tau=1}^{\omega} p_{1_{\tau-1}} p_{10_\tau}} \tag{4.22}$$

which may be rearranged as

$$E\left[L^r\right] = \frac{\sum\limits_{\tau=1}^{\omega} \left\{ p_{1_{\tau-1}} p_{10_\tau} \left[\sum\limits_{l=1}^{\infty} l^r \prod_{i=1}^{l-1} \left(1 - p_{01_{\tau+i}}\right) p_{01_{\tau+l}} \right] \right\}}{\sum\limits_{\tau=1}^{\omega} p_{1_{\tau-1}} p_{10_\tau}} \tag{4.23}$$

The closed form expression can be obtained using the following proposition.

Proposition 1 *Let a_k, b_i be periodic sequences with period ω and c_i be a non-periodic sequence. Then:*

$$\sum_{i=2}^{\infty} c_i b_i \prod_{k=1}^{i-1} a_k = \sum_{t=1}^{\omega} b_t \left(\prod_{k=1}^{t-1} a_k \right) \sum_{i=0}^{\infty} \left(\prod_{k=1}^{\omega} a_k \right)^i c_{i\omega+t} \tag{4.24}$$

Proof: Let us start by observing that the infinite summation over a non-periodic sequence f_i can be split in a double sum as follows:

$$\sum_{i=1}^{\infty} f_i = \sum_{i=0}^{\infty} \sum_{\tau=1}^{\omega} f_{i\omega+\tau} = \sum_{\tau=1}^{\omega} \sum_{i=0}^{\infty} f_{i\omega+\tau}$$

Then, the left hand side of (4.24) can be rewritten as:

$$\sum_{i=1}^{\infty} c_i b_i \prod_{k=1}^{i-1} a_k = \sum_{i=0}^{\infty} \sum_{\tau=1}^{\omega} c_{i\omega+\tau} b_{i\omega+\tau} \prod_{k=1}^{i\omega+\tau-1} a_k$$

Since, a_k and b_τ are periodic it follows $b_{i\omega+\tau} = b_\tau$ and

$$\prod_{k=1}^{i\omega+\tau-1} a_k = \left(\prod_{k=1}^{\tau-1} a_k\right) \left(\prod_{k=1}^{\tau-1} a_k\right)^i$$

which completes the proof.

By virtue of the foregoing proposition and letting $a_k = (1 - p_{01\tau+k})$, $b_\tau = p_{01\tau+l}$ and $c_i = i^r$, the term in square brackets in 4.23 can be rewritten as

$$\left[\sum_{l=1}^{\infty} l^r \prod_{i=1}^{l-1} \left(1 - p_{01_{\tau+i}}\right) p_{01_{\tau+l}}\right] = \tag{4.25}$$

$$\sum_{t=1}^{\omega} \left\{ p_{01_{\tau+t}} \prod_{k=1}^{t-1} \left(1 - p_{01_{t+k}}\right) \sum_{i=0}^{\infty} \left[\prod_{k=1}^{\omega} \left(1 - p_{01_{t+k}}\right)^i (i\omega + t)^r\right]\right\} \tag{4.26}$$

In addition, the infinite summation in (4.26) can be determined using results from the geometric series:

$$\sum_{i=0}^{\infty} (a + bi)\, q^i = \frac{a}{1-q} + \frac{bq}{(1-q)^2} \tag{4.27}$$

Then, the expected value of L follows from (4.23)-(4.27) by letting $r = 1$ which gives

$$\mathcal{E}[L] = \frac{\sum_{\tau=1}^{\omega} \left\{ p_{1_{\tau-1}} p_{10_\tau} \left[\sum_{t=1}^{\omega} \left(p_{01_{\tau+t}} \prod_{k=1}^{t-1} \left(1 - p_{01_{\tau+k}}\right) \frac{t - \Delta(t+\omega)}{(1-\Delta)^2}\right)\right]\right\}}{\sum_{\tau=1}^{\omega} p_{1_{\tau-1}} p_{10_\tau}} \tag{4.28}$$

where $\Delta = \prod_{k=1}^{\omega} \left(1 - p_{01_{\tau+k}}\right)$.

Equation (4.28) represents the expected value of drought length regardless of the initial season for a first-order periodic Markov process. The expected value for the stationary process results by letting $\omega = 1$ and (4.28) becomes

$$\mathcal{E}[L] = \frac{1}{1 - p_{00}} \tag{4.29}$$

and

$$\mathcal{E}[L] = \frac{1}{1 - p_0} \tag{4.30}$$

for the dependent and independent cases, respectively. Obviously for these cases the same results follow directly by determining the expectations from (17) and (18).

4.4 Return period of drought lengths

It is of interest here to derive analytical expressions to estimate the return period of droughts with length L equal to or greater than a fixed value for the case of periodic series. Here, the return period of a drought of length L=l is defined as the mean interarrival time between droughts with lengths that are equal or greater than l. Assuming that drought events are independent Shiau and Shen (2001) developed a procedure to derive the return period related to a drought with accumulated deficit greater or equal to a given value. Here such procedure will be extended for deriving the return period of drought lengths for both stationary and periodic dependent processes.

Referring to 4.2 the inter-arrival time T_l is the sum of the inter-arrival times between droughts of any length, i.e.

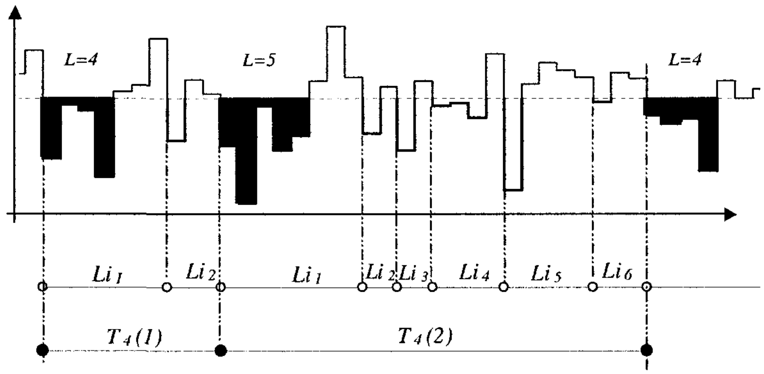


Figure 4.2: Definition of inter-arrival times of droughts with lengths $L \geq 4$. The open circles denote the initiation of droughts of any length while the filled circles denote the initiation of droughts of length equal or greater than 4.

$$T_l = \sum_{j=1}^{N_l} Li_j \tag{4.31}$$

where Li_j is the inter-arrival time between any two successive drought events (regardless of the length) and N_l is the number of drought events until the occurrence of the next drought event with length equal to or greater than l.

Note that both N_l and Li_j are random variables. Thus the return period T is the expected value of T_l, which under the assumption that Li_j is independent gives

$$T = E(T_l) = E\left(\sum_{j=1}^{N_l} Li_j\right) = E(N_l)\, E(Li_j) \tag{4.32}$$

Recalling the definition of N_l, for $N_l = n_l$ droughts, there must be $n_l - 1$ droughts with length less than l and only one drought with length equal or greater than l. Then, the probability $P(N_l = n_l)$ can be determined as:

$$P(N_l = n_l) = [P(L < l)]^{n_l - 1} P(L \geq l) = [P(L < l)]^{n_l - 1}[1 - P(L < l)] \tag{4.33}$$

where $P(L \leq l)$ is the CDF of L given by (4.28). In addition, because L is a discrete variable one can write $P(L < l) = F_L(l-1)$, it follows

$$P(N_l = n_l) = [F_L(l-1)]^{n_l - 1}[1 - F_L(l-1)] \tag{4.34}$$

which suggests that N_l is geometrically distributed with parameter $F_L(l-1)$. Then its expected value is Mood et al. (1974)

$$E\left(N_l\right) = \frac{1}{1 - F_L(l-1)} \tag{4.35}$$

Therefore, from (4.32) and (4.35) we get

$$T = E(T_l) = \frac{E(Li_j)}{1 - F_L(l-1)} \tag{4.36}$$

Note that in deriving (4.36) only the independence between the lengths of drought events has been assumed. Such condition holds even if the underlying process is a periodic Markov chain since by definition, a drought must be preceded and succeeded by at least one surplus intervals and the lag-1 Markov model assumes independence at lags greater than one. As a consequence, drought characteristics do not depend on those of preceeding droughts.

Furthermore, to apply (4.36) we must obtain the expected value $E(Li_j)$. This term can be determined by observing that the inter-arrival time $Li_j$is the sum of the drought length L and the surplus (non-drought) length S. Accordingly, one can write, $E(Li_j) = E(L) + E(S)$. The first term $E(L)$ is given by (4.28) and $E(S)$ can be obtained from the same equation by simply switching the deficit probabilities with surplus probabilities, i.e.

$$E(S)=\frac{\sum_{\tau=1}^{\omega}\left\{p_{0_{\tau-1}}p_{01_{\tau}}\left[\sum_{t=1}^{\omega}\left(p_{10_{\tau+t}}\prod_{k=1}^{t-1}\left(1-p_{10_{\tau+k}}\right)\frac{t-\Delta^{*}(t+\omega)}{(1-\Delta^{*})^{2}}\right)\right]\right\}}{\sum_{\tau=1}^{\omega}p_{0_{\tau-1}}p_{01_{\tau}}} \tag{4.37}$$

where $\Delta^{*}=\prod_{k=1}^{\omega}\left(1-p_{10_{\tau+k}}\right)$.

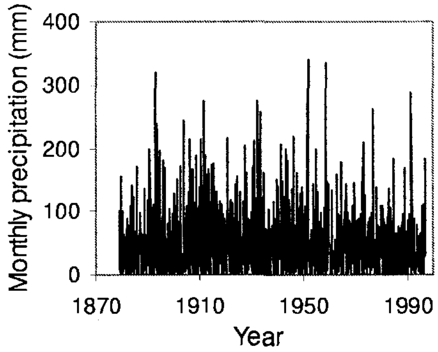
4.5 Drought length in periodic series

The procedures for estimating the pmf, low order moments, and return period of drought length as outlined in the previous sections of this paper have been applied to a variety of data sets, namely: (1) monthly precipitation for Caltanissetta, Italy (source: Regional Hydrographic Service of Sicily, Italy), (2) weekly precipitation for Fort Collins at Colorado State University (CSU) weather station (source: http:// climate.atmos.colostate.edu/ dataaccess.shtml), (3) monthly streamflows of the Niger River at Koulikoro (source: The Global Runoff Data Centre, D-56068 Koblenz, Germany), (4) the three-month Standardized Precipitation Index (SPI) based on the monthly precipitation series for the Northeastern Climate Division of Colorado (source: http://lwf.ncdc.noaa.gov/oa/climate/ onlineprod/drought/ftppage.html; NCDC, NOAA), and (5) Palmer Hydrologic Drought Index (PHDI) for the Northeastern Climate Division of Colorado (source: http://lwf.ncdc.noaa.gov/oa/climate/ onlineprod/drought/ftppage.html; NCDC, NOAA). Table 4.I summarizes the main characteristics of the referred data and Fig. 4.3 gives the time series plots of the basic data utilized in the study.

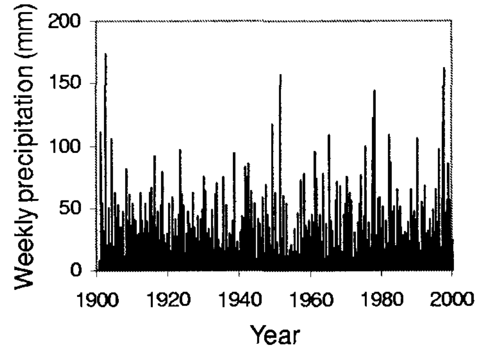
The SPI (McKee et al., 1993) and PHDI (Palmer, 1965; Karl and Knight, 1985) are widely used indices for monitoring drought severity in the United States and some

Table 4.1: General description of the periodic-stochastic data utilized in characterizing droughts lengths

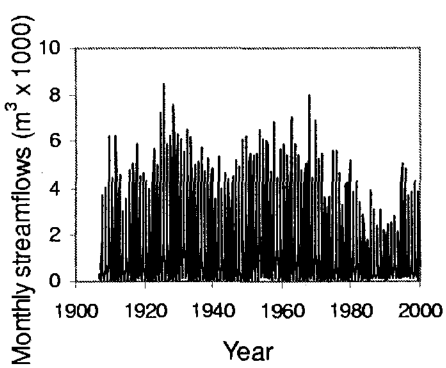
Station/region Name	Type of Data	Period of record	Mean	Standard deviation
Caltanissetta, Italy	Monthly precipitation (mm)	1879-1996	47.1	48.35
Fort Collins, Colorado at CSU Weather Station	Weekly precipitation (mm)	1901-1999	7.42	13.5
Niger River at Koulikoro	Monthly streamflows (m^3/s)	1907-1999	1,373	1,726
Northeastern Colorado Climate Division	Standard Precipitation Index (SPI) for k=3	1896-2000	-0.0047	1
Northeastern Colorado Climate Division	Palmer Hydrologic Drought Index (PHDI) for 6-month time interval	1895-2001	1.485	3.227



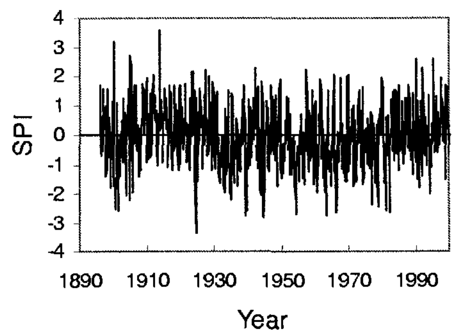
(a) Monthly precipitation for Caltanissetta



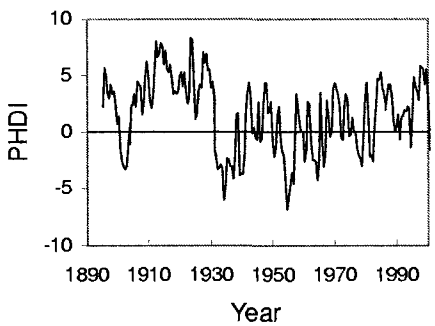
(b) Weekly precipitation for Fort Collins



(c) Monthly flows for the Niger River



(d) 3 month - SPI



(e) 6 month average - PHDI

Figure 4.3: Time series plots of the five data sets investigated: (a) monthly precipitation for Caltanissetta, (b) weekly precipitation for Fort Collins, (c) monthly streamflows for the Niger River, (d) 3-month SPI for Northeastern Colorado, and (e) 6-month average PHDI for Northeastern Colorado.

other countries. Recently Heim [2002] reviewed the use of the SPI, PHDI, and some other indices in the United States. The SPI is a meteorological index originally introduced by McKee et al. (1993) that takes into account the different time scales at which the drought phenomenon affects different water uses. In practice, computation of the index involves fitting the gamma probability distribution to the cumulative precipitation series over a given number of months (e.g. three-months), computing the corresponding non-exceedence probabilities, and relating them to the standardized normal quantiles that are called the SPI. The PHDI is a slight modification of the Palmer Drought Severity Index (PDSI), which was originally developed by Palmer (1965). It takes into account precipitation, evapotranspiration, and soil water content in an area, and as such, is more oriented towards a meteorological/agricultural drought characterization. The index is based on a simple water balance of the unsaturated soil and classifies droughts on the basis of the comparison of the precipitation in a given month and that (referred to as climatically appropriate) needed to maintain average soil water content. The index is standardized and can take positive values (wet conditions) or negative values (drought conditions).

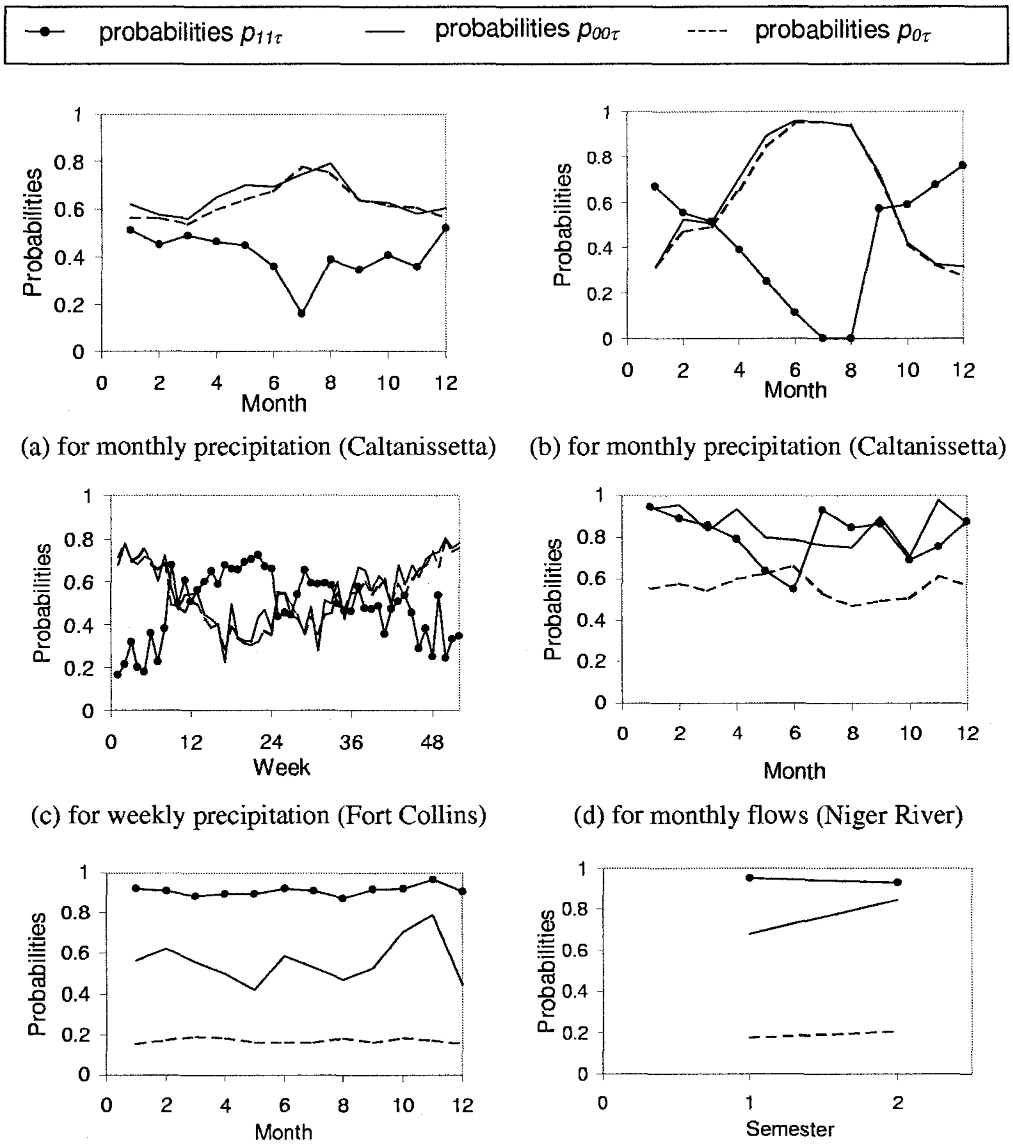
In our study we used weekly and monthly data for characterizing droughts at short and medium time scales while the SPI and PHDI series were utilized for characterizing droughts at longer (3-month and 6-month) time scales. The SPI series was derived from the monthly precipitation series based on a 3-month time span. On the other hand, the monthly PHDI time series were obtained from the web site as referred to above from which the 6-month average (seasonal) time series was obtained.

The drought analysis based on the referred data was performed using both constant and periodic water demands (thresholds) depending on the case. For the drought analysis regarding monthly precipitation and monthly streamflows we used periodic thresholds of the form $x_{0\tau} = \hat{\mu}_\tau - \alpha\hat{\sigma}_\tau$ ($\hat{\mu}_\tau$ and $\hat{\sigma}_\tau$ represent the monthly mean and monthly standard deviation of the water supply series, respectively and $\alpha \geq 0$ is a coefficient) and constant thresholds, i.e. $x_{0\tau} = x_0$ (where x_0 remains the same across the year). On the other hand, for the weekly precipitation series we used $x_{0\tau} = 2.54mm(0.1in)$ because our interest in this case was characterizing periods of weeks with no significant precipitation. Likewise, for the applications regarding the drought indices we used constant thresholds: -1 for the SPI and –2 for the PHDI.

Firstly, the transition probabilities $p_{00\tau}$, $p_{01\tau}$, $p_{10\tau}$, and $p_{11\tau}$, and the marginal probabilities $p_{0\tau}$ and $p_{1\tau}$ were estimated for $\tau = 1, ..., \omega$ (ω varied depending on the case) using the procedure outlined in Appendix A, e.g. using (A.2) and (A.3) for estimating $p_{00\tau}$ and $p_{11\tau}$. Figure 4 illustrates the probabilities $p_{00\tau}$, $p_{11\tau}$, and $p_{0\tau}$ that were obtained for the five data sets considered for demand thresholds given by: $x_{0\tau} = \hat{\mu}_\tau$ for the data of Caltanissetta and the Niger River, $x_{0\tau} = 2.54mm$; for the weekly precipitation of Fort Collins, and $x_{0\tau}$ =–1, and –2 for the SPI and the PHDI data, respectively. For Caltanissetta, a constant demand threshold equal to the sample mean, i.e. $x_{0\tau}$ =47.1 mm, was also analyzed to illustrate a case where there is a major difference between the water supply and the demand across the year. Some of the referred probabilities show a marked seasonality, e.g. $p_{11\tau}$ in Figures 4.4(a) and 4.4(b), while some others appear to be constant throughout the year, e.g. $p_{0\tau}$ in Figure 4.4(e). However, in all cases at least one of the referred probabilities appears to be seasonal, which justifies using a periodic model and equations for drought characterization that include periodic probabilities.

Clearly, the seasonality in the transition probabilities must arise from the seasonality of the variance-covariance structure of the underlying series.

The empirical probability mass function (pmf) of drought length was obtained from the historical sample data and from (4.17) using the estimated transition and marginal probabilities as referred to above. The pmf's illustrated in Fig. 4.5(a)-(f) correspond



(a) for monthly precipitation (Caltanissetta) (b) for monthly precipitation (Caltanissetta)

(c) for weekly precipitation (Fort Collins) (d) for monthly flows (Niger River)

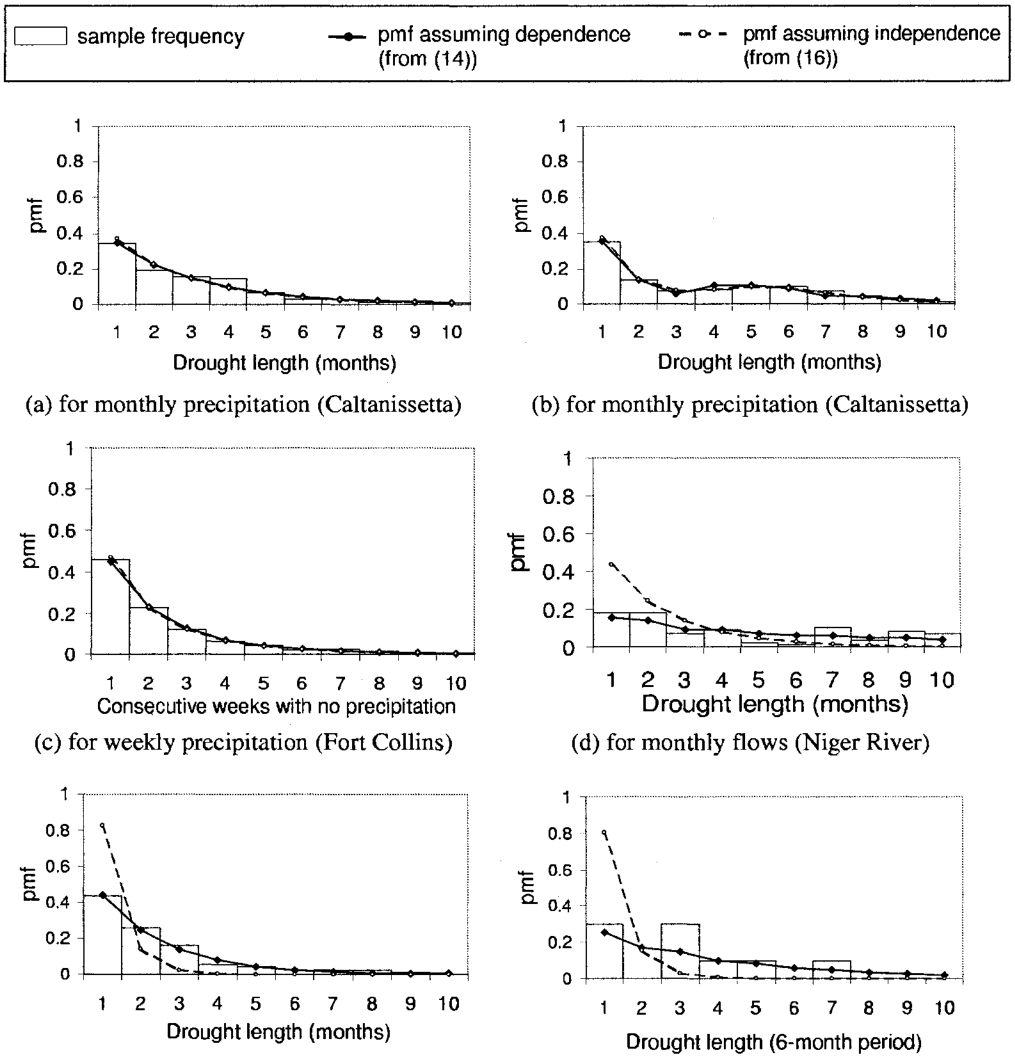
(e) for the 3-month SPI (f) for the 6-month average PHDI

Figure 4.4: Month-to-month, week-to-week, or season-to-season probabilities $p_{11\tau}, p_{00\tau}$, and $p_{0\tau}$ calculated for the five data sets and demand thresholds: (a)$x_{0\tau} = \hat{\mu}_\tau$ for monthly precipitation at Caltanissetta, (b) $x_{0\tau} = 47.1$ mm for monthly precipitation at Caltanissetta, (c) $x_{0\tau} = 2.54$ for weekly precipitation at Fort Collins, (d) $x_{0\tau} = \hat{\mu}_\tau$ for monthly streamflows of the Niger River, (e) $x_{0\tau} = -1$ for the 3-month SPI for Northeastern Colorado, and (f) $x_{0\tau} = -2$ for the 6-month average PHDI for Northeastern Colorado.

to the five data sets considered and the same demand thresholds used for estimating the probabilities shown in Figure 4.4 (refer to the previous paragraph). Generally the empirical and derived pmf's show a pretty good agreement between them. The agreement between the pmf's appear to be closer and better for the monthly data of Caltanissetta, the weekly data of Fort Collins, and the SPI data for which the pmf's show well defined exponential decay as the drought length increases (refer to Figures 4.5(a), (c), (e)). Also for Caltanissetta when a constant demand threshold is adopted, Fig.4.5(b) shows that the pmf's obtained from the sample and from (4.17) are quite similar. In fact, because of a consecutive dry period generally occurring for about 6 months every year the resulting sample pmf of drought length is bimodal and such a feature is also reproduced by the pmf derived from (4.17). For the Niger River flows and the PHDI data the sample pmf's show a slow decay as the drought length increases, which is an indication of the significant frequency of long drought lengths. For these two cases the pmf's derived from (4.17) reasonably follow the decay pattern shown by the sample pmf's but the overall agreement is not as clear as for the other four cases. One may note that for these two cases the sample pmf's appear somewhat complex not only because of the long tails but also because of zero probabilities for some drought lengths. However, these features should not be surprising for significantly dependent processes.

In order to illustrate and examine the effect of dependence on the distribution of drought length we also plotted on Fig. (4.5) the pmf's obtained from (4.19), which assumes independence. Clearly for the monthly and weekly precipitation where the temporal dependence is small, there is little difference between the pmf's obtained assuming independence and dependence. On the other hand, for the other three cases the differences (between the pmf's) are remarkable.

The mean, standard deviation, and skewness coefficient of drought lengths for the



(a) for monthly precipitation (Caltanissetta)

(b) for monthly precipitation (Caltanissetta)

(c) for weekly precipitation (Fort Collins)

(d) for monthly flows (Niger River)

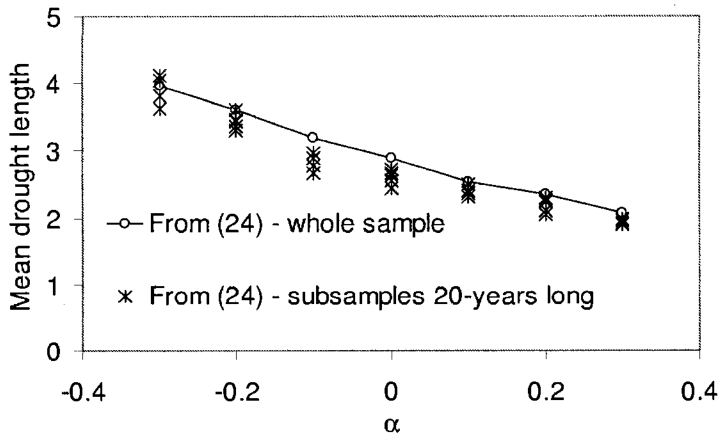
(e) for the 3-month SPI

(f) for the 6-month average PHDI

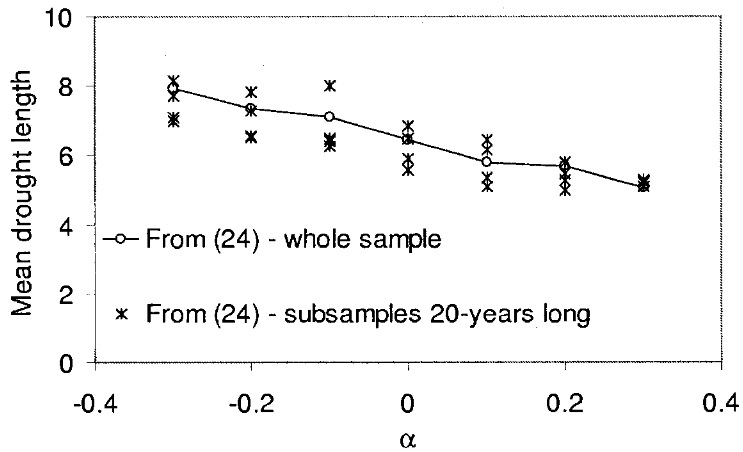
Figure 4.5: Probability mass function (pmf) of drought length estimated for the five data sets assuming demand thresholds: (a) $x_{0\tau} = \hat{\mu}_\tau$ for monthly precipitation at Caltanissetta, (b) $x_{0\tau} = 47.1$ mm for monthly precipitation at Caltanissetta, (c) $x_{0\tau} = 2.54$ for weekly precipitation at Fort Collins, (d) $x_{0\tau} = \hat{\mu}_\tau$ for monthly streamflows of the Niger River, (e) $x_{0\tau} = -1$ for the 3-month SPI for Northeastern Colorado, and (f) $x_{0\tau} = -2$ for the 6-month average PHDI for Northeastern Colorado.

five data sets considered and the same demand thresholds as in the previous analysis were estimated from the historical sample and from (4.23) and (4.28). The results shown in Table 2 suggest that (4.23) and (4.28) provide quite good estimates of the low order moments of drought length. Although some differences are noted regarding the standard deviation and skewness for the Niger River and the PHDI series, overall there is a fairly good correspondence between the analytical and historical estimates. In addition, the expected drought lengths for the monthly precipitation at Caltanissetta and the monthly flows of the Niger River were determined from (4.28) using variable demand thresholds with α varying between -0.3 and $+0.3$. To see the effect of the sample size on the estimated mean drought length the estimates were based on the entire sample sizes (1,416 for Caltanissetta and 1,116 for Niger) and based on non-overlapping subsamples of size 20-years (240 months). The results in Figure 4.6 show that (as expected) because the thresholds decrease as α increases, the mean drought length becomes smaller. Overall the estimated drought lengths fall near those obtained from the entire sample size. However, some differences are observed. For Caltanissetta the smaller subsamples appear to underestimate the mean drought length. The underestimation could be of the order of 20 % based on the subsamples analyzed. For the Niger the underestimation is somewhat less noticeable (than for Caltanissetta) but the differences between the estimated drought lengths based on the subsamples relative to the mean drought length based on the entire record could be of the order of 30 %. It may be worthwhile to note that over the range of values of α examined the sample variability of the mean drought length does not vary significantly.

The return period of droughts lengths for the five data sets investigated were estimated from (4.36) using the analytical expressions for the expected inter-arrival times of drought events and the distribution of drought lengths assuming the periodic



(a) for monthly precipitation (Caltanissetta)



(b) for monthly streamflows (Niger River)

Figure 4.6: Comparison of the mean drought lengths obtained from (24) based on estimates derived from the entire record (full line) versus those derived from subsamples 20-years long. The demand threshold levels have been defined by $x_{0\tau} = \hat{\mu}_\tau - \alpha\hat{\sigma}_\tau$.

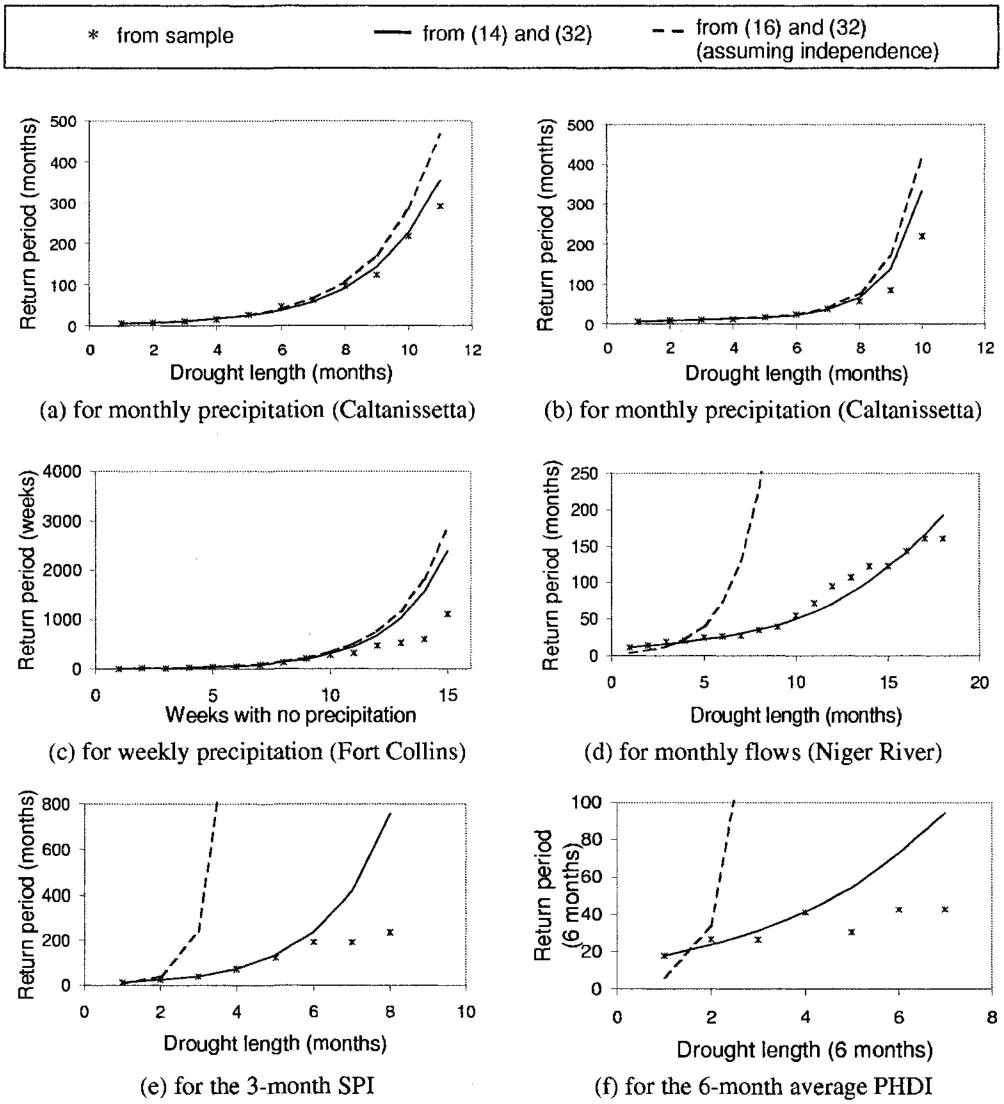
Markov chain as described above. Likewise, the return period was estimated from the historical sample as the average inter-arrival time between droughts with lengths equal or greater than a given value. The estimates were made for the same demand thresholds as indicated previously in relation to Figures 4.4 and 4.5. The results are shown in Figure 4.7(a)-(f). In all cases the general pattern is similar, i.e. there is an excellent agreement between the return periods T estimated from (4.36) and those estimated from the historical sample. This is especially so for the first few values of drought lengths, thereafter the sample estimates appear to drift below the analytical estimates. For example, Figure 4.7(e) for the SPI data shows that after the drought length of 5 months the sample estimates drift to smaller values than those obtained from (4.36). This can be explained because as the drought length increases there are smaller numbers of drought episodes that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time become unreliable and underestimated. In fact, beyond a certain drought length no droughts may be observable from the historical sample and no return period can be determined.

Furthermore, the degree of persistence is certainly a factor in determining the return period versus drought length function. Figure 4.7 also compares the return periods obtained from (4.36) under the assumptions of independence and simple Markov chain dependence. As expected, the differences can be notable as Fig. 4.7 (d)-(f) show for the Niger, SPI, and PHDI series for which the temporal dependencies are more significant than for the two cases of precipitation series. The general pattern is that for a given drought length the return period T for an independent series is bigger than that for a dependent series, conversely for the same return period the drought length is larger for a dependent series than for an independent series. These results may suggest that the reason why the sample estimates of T fall below those estimated from a model after a certain drought length is because the model may not have the required persistence. While this is a plausible possibility, it must be weighted against two other factors, namely, the sample variability of the estimates of T from a short sample and the fact that as the drought length increases the sample estimates of T either cannot be estimated or are underestimated as already discussed above.

4.6 Final remarks

An extreme drought is a complex phenomenon that evolves through time and space in a random fashion. The reviewed literature revealed that the major emphasis in analyzing drought properties using analytical methods has been for cases where the underlying stochastic processes are stationary. Because annual hydrological processes, such as annual streamflow, have been often assumed to be stationary, most analytical developments and applications on drought characteristics and drought frequency have been centered about annual processes. Research on drought analysis where the underlying water supply process is periodic-stochastic is lacking.

In this chapter, the probability mass function (pmf) of drought length, its associated low order moments, and the return period of drought events have been derived assuming a periodic two-state simple Markov chain and the drought definition as in Yevjevich (1967). The derived pmf's allow estimating the occurrence probabilities of droughts of a given length where either the drought begins in a given season or regardless of the initial season. The applicability of the drought formulations has been illustrated using a variety of water supply series (e.g. monthly and weekly precipitation for Caltanissetta and Fort Collins, respectively, monthly streamflow series for the Niger River, and the 3-month SPI and the PHDI for the Northeastern Climate Region of Colorado), and using both constant and periodic water demand thresholds



(a) for monthly precipitation (Caltanissetta)

(b) for monthly precipitation (Caltanissetta)

(c) for weekly precipitation (Fort Collins)

(d) for monthly flows (Niger River)

(e) for the 3-month SPI

(f) for the 6-month average PHDI

Figure 4.7: Comparison between the return periods of drought events, with lengths equal to or greater than l obtained from the historical samples and from (32) assuming dependence and independence and demand thresholds: (a) $x_{0\tau} = \hat{\mu}_\tau$ for monthly precipitation at Caltanissetta, (b) $x_{0\tau} = 47.1$ mm for monthly precipitation at Caltanissetta, (c) $x_{0\tau} = 2.54$ for weekly precipitation at Fort Collins, (d) $x_{0\tau} = \hat{\mu}_\tau$ for monthly streamflows of the Niger River, (e) $x_{0\tau} = -1$ for the 3-month SPI for Northeastern Colorado, and (f) $x_{0\tau} = -2$ for the 6-month average PHDI for Northeastern Colorado.

depending on the case (e.g. $x_{0\tau} = \hat{\mu}_\tau$) for drought analysis regarding monthly precipitation and monthly streamflows and constant demand thresholds for the monthly precipitation at Caltanissetta, the weekly precipitation series at Fort Collins, and the SPI and PHDI drought indices).

The comparison of the empirical probability mass function (pmf) of drought length obtained from the historical sample and from (4.17) generally showed a very good agreement between them. The model pmf appears to represent quite well the empirical pmf for the monthly and weekly precipitation data and for the SPI data. Furthermore, when a constant threshold is adopted for the monthly precipitation series, the resulting pmf of drought length appears bimodal and such a feature is also reproduced by the pmf derived from (4.17). For the Niger River flows and the PHDI data the pmf's obtained from the historical sample are more complex with a slow decay as the drought length increases and with a few zero frequencies for some drought lengths. Nevertheless, the pmf's obtained from (4.17) resemble closely such long tail of the frequency distribution. For the data analyzed in this study it is not clear whether a model with more persistence than the periodic Markov chain would give better results. In addition, the mean, standard deviation, and skewness coefficient of drought lengths for the five data sets and demand thresholds considered were estimated from the historical sample and from (4.23) and (4.28). The results suggest that (4.23) and (4.28) provide quite good estimates of the low order moments of drought lengths. Furthermore, the return period of droughts lengths for the five data sets investigated were estimated from (4.36) and from the historical sample. In all cases the results showed an excellent agreement between the two estimates (analytical and historical sample) for the first few values of drought lengths, thereafter the sample estimates appear to drift below the analytical estimates. This can be explained because as the drought length increases there are smaller numbers of droughts that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time become unreliable and generally underestimated. The comparison between the return periods obtained assuming independence and simple Markov dependence illustrate clearly the effect of dependence in the behavior of the return period versus drought length, i.e. for a given drought length the return period is bigger for an independent process. This effect should not be confused with that caused by the underestimation arising from estimates of the return period because of a small sample as discussed above.

The overall conclusion of the present chapter is that simple concepts and definitions may enable one characterizing droughts using probability and stochastic approaches. In fact, in some cases, it may be possible to derive drought characteristics analytically. Such analytical derivations are particularly useful in cases of short hydrologic records for which the applicability of an inferential approach to find the probability distributions of drought characteristics is either impractical or not feasible. In addition, they may be useful for checking approximations or results obtained for more complex cases. The examples using a variety of water supply series (e.g. monthly and weekly precipitation and streamflows) and climatologic and hydrological drought indices illustrate and confirm the applicability of the analytical derivations obtained for drought lengths and associated return periods.

CHAPTER V

JOINT ANALYSIS OF DROUGHT CHARACTERISTICS IN PERIODIC SERIES

5.1 Drought characterization on periodic series

Probabilistic features of drought characteristics length, accumulated deficit and intensity will be derived assuming periodicity for both the underlying water supply series and the demand series.

Let us consider a periodic stochastic water supply series denoted as $X_{\nu,\tau}$, $\nu = 1, 2, ..., \tau = 1, ..., \omega$ where ν represents the year, τ represents the season (e.g. a month), and ω is the number of seasons (e.g. $\omega = 12$ for monthly data). In addition, consider a variable water demand series denoted by $x_{o\tau}$. Following the drought definition given by Yevjevich (1967), a drought event is defined as a succession of consecutive periods (run) in which the water supply remains below the threshold level $x_{o\tau}$. With reference to a drought starting at season τ, we can define the drought length L_τ (length of negative run), i.e. the number of consecutive time intervals (seasons) in which $X_{\nu,\tau} < x_{o\tau}$, preceded and followed by (at least one season where) $X_{\nu,\tau} \geq x_{o\tau}$, the drought magnitude D_τ (often referred to as accumulated deficit), defined as the sum of the deficits $S_{\nu,\tau} = x_{o\tau} - X_{\nu,\tau}$, extended to the drought duration L_τ and the

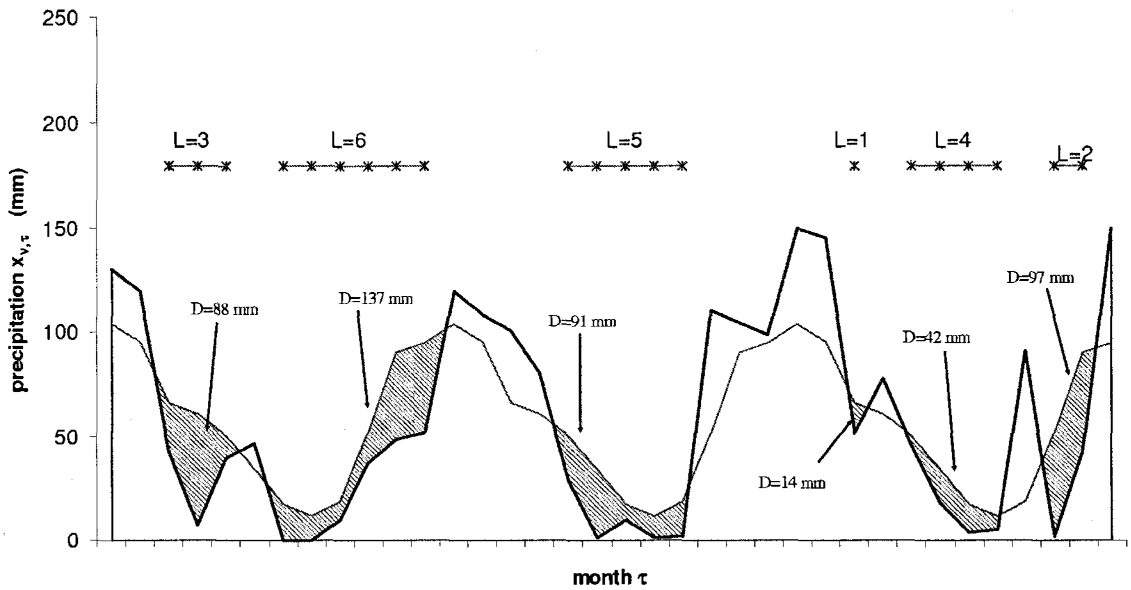


Figure 5.1: Definition of drought episodes for a periodic-stochastic water supply series

intensity (or average deficit) I_τ defined as the ratio of accumulated deficit to drought duration. This drought definition is schematically shown in Figure 5.1 where the drought shaded areas indicate the period where the hydrological series is below the periodic threshold.

The above defined characteristics are related to droughts starting at season τ. It may also be of some interest to define the drought characteristics duration L, accumulated deficit D and intensity I regardless of the initial season τ.

In general terms, drought characteristics are not mutually independent and therefore, in order to investigate their probabilistic behaviour, it is appropriate to follow a bivariate approach, i.e. to consider jointly duration L and accumulated deficit D or duration L and intensity I. Depending on the combination of drought characteristics we can specify different critical drought events. For instance, one may consider only the duration of a drought regardless of the deficit, or duration with a certain degree of deficit, or duration and a given intensity, etc. Furthermore, we can also distinguish events on the basis of their beginning season.

The probabilistic characterization of drought events requires the definition of the marginal and/or of the joint bivariate pdfs of drought characteristics. The moments and distribution of drought length have already been derived in Chapter 4. In what follows first the moments and the marginal distribution of accumulated deficit will be derived for the periodic case. The periodic bivariate distributions of L_τ and D_τ or L_τ and I_τ for a drought whose onset is in season τ will be derived. Then the corresponding bivariate pdfs regardless of the initial season will be derived as a mixture of such pdf's. Finally, analytical expressions for the return period of the considered events will be derived and some examples of applications are illustrated.

5.2 Probability distributions of drought characteristics

5.2.1 Moments and probability distribution of accumulated deficit (D)

In order to determine an approximate expression for accumulated deficit for periodic dependent series, the proposed approach is to assume drought accumulated deficit to be gamma distributed with parameters (r, β), i.e.:

$$f_D(d) = \frac{1}{\beta\Gamma(r)} e^{-\frac{d}{\beta}} \left(-\frac{d}{\beta}\right)^{r-1} \tag{5.1}$$

where the parameters r and β will depend on the two first moments of D. The latter moments can be expressed as a function of the moments of the deficits and of drought length, thus allowing for a more reliable estimate since on historical records the number of observed deficits is obviously consistently larger than the observed droughts. Thus, the derived pdf for accumulated deficit will be exact from a second order point of view, since the first two moments of accumulated deficit will be preserved.

More specifically, let $\mathcal{E}[D]$ and $\text{Var}[D]$ be the expected value and the variance of D respectively. If D is gamma distributed with parameters (r, β) the following relationships will hold (Mood et al., 1974):

$$\mathcal{E}[D] = r\beta \tag{5.2}$$

$$\text{Var}[D] = r\beta^2 \tag{5.3}$$

which solved for (r, β) yield:

$$r = \frac{\mathcal{E}^2[D]}{\text{Var}[D]} \tag{5.4}$$

$$\beta = \frac{\text{Var}[D]}{\mathcal{E}[D]} \tag{5.5}$$

The proposed approach is to determine $\mathcal{E}[D]$ and $\text{Var}[D]$ as a function of the stochastic structure of the underlying hydrological variable. More specifically, let's start by observing that the above moments are related to the conditional moments of accumulated deficit of droughts starting in season τ $\mathcal{E}[D_\tau]$ and $\text{Var}[D_\tau]$. To derive the expected value and variance of the accumulated deficit of droughts regardless of the initial season, a procedure similar to that already described in Chapter 4 regarding drought length can be applied. Indeed, the expected value and variance of accumulated deficit will be simply the weighted average of the expected value and variance conditioned on the initial season, i.e.:

$$\mathcal{E}[D] = \frac{\sum_{\tau=1}^{\omega} \mathcal{E}[D_\tau] \, p_{1_{\tau-1}} p_{10_\tau}}{\sum_{\tau=1}^{\omega} p_{1_{\tau-1}} p_{10_\tau}} \tag{5.6}$$

$$\text{Var}[D] = \frac{\sum_{\tau=1}^{\omega} \text{Var}[D_\tau] \, p_{1_{\tau-1}} p_{10_\tau}}{\sum_{\tau=1}^{\omega} p_{1_{\tau-1}} p_{10_\tau}} \tag{5.7}$$

Assuming that a drought starts at season τ and lasts L_τ, the accumulated deficit is by definition:

$$D_\tau = \sum_{j=0}^{L_\tau - 1} S_{\tau+j} \tag{5.8}$$

where S_τ represents the deficit at interval τ and L_τ is a random variable. The expected value of D_τ will be:

$$\mathcal{E}\left[D_\tau\right] = \mathcal{E}\left[\sum_{j=0}^{L_\tau - 1} S_{\tau+j}\right] \tag{5.9}$$

The above expectation can be worked out by means of conditional expectation concepts as:

$$\mathcal{E}\left[\sum_{j=0}^{L_\tau - 1} S_{\tau+j}\right] = \mathcal{E}\left[\mathcal{E}\left[\sum_{j=0}^{L_\tau - 1} S_{\tau+j} \middle| L_\tau\right]\right] = \mathcal{E}\left[\sum_{j=0}^{L_\tau - 1} \mathcal{E}\left[S_{\tau+j}\right]\right] \tag{5.10}$$

The expectation can be solved following the definition:

$$\mathcal{E}\left[D_\tau\right] = \sum_{l=1}^{\infty}\sum_{j=0}^{l-1} \mathcal{E}\left[S_{\tau+j}\right] f_{L_\tau}(l) \tag{5.11}$$

which, by substituting the expression for $f_{L_\tau}(l)$ given by Eq. (4.17), becomes:

$$\mathcal{E}\left[D_\tau\right] = \sum_{l=1}^{\infty}\sum_{j=0}^{l-1} \mathcal{E}\left[S_{\tau+j}\right] \left(\prod_{i=1}^{l-1} p_{00_{\tau+i}}\right) p_{01_{\tau+l}} \tag{5.12}$$

In order to express 5.12 in closed form, the following proposition is useful:

Proposition 2 *Let a_j, b_i and c_l be periodic sequences with periodicity ω. Then:*

$$\sum_{l=1}^{\infty}\left[\left(\sum_{j=0}^{l-1} a_j\right)\left(\prod_{i=1}^{l-1} b_i\right) c_l\right] = \tag{5.13}$$

$$\sum_{k=0}^{\infty}\sum_{h=0}^{\omega-1}\left[\left(k\sum_{j=0}^{\omega-1} a_j + \sum_{j=0}^{h} a_j\right)\left(\prod_{i=1}^{\omega} b_i\right)^k \prod_{i=1}^{h} b_i \;\; c_{k\omega+h+1}\right] = \tag{5.14}$$

$$\sum_{h=0}^{\omega-1}\left\{c_{h+1}\prod_{i=1}^{h} b_i\left[\sum_{j=0}^{\omega-1} a_j\left(\sum_{k=0}^{\infty} k\left(\prod_{i=1}^{\omega} b_i\right)^k\right) + \sum_{j=0}^{h} a_j \sum_{k=0}^{\infty}\left(\prod_{i=1}^{\omega} b_i\right)^k\right]\right\} \tag{5.15}$$

Again, making use of the result $\sum_{i=0}^{\infty}(a+bi)q^i = \frac{a}{1-q} + \frac{bq}{(1-q)^2}$, the infinite sums can be solved, thus obtaining:

$$\sum_{h=0}^{\omega-1}\left\{c_{h+1}\prod_{i=1}^{h} b_i\left[\frac{\prod_{i=1}^{\omega} b_i}{(1-\prod_{i=1}^{\omega} b_i)^2}\sum_{j=0}^{\omega-1} a_j + \frac{\sum_{j=0}^{h} a_j}{1-\prod_{i=1}^{\omega} b_i}\right]\right\} \tag{5.16}$$

Applying the above proposition to Eq. (5.12) with the following positions:

$$a_j = \mathcal{E}\left[S_{\tau+j}\right] \tag{5.17}$$

$$b_i = p_{00_{\tau+i}} \tag{5.18}$$

$$c_h = p_{01_{\tau+h}} \tag{5.19}$$

the expectation $\mathcal{E}\left[D_\tau\right]$ takes the following closed form:

$$\mathcal{E}\left[D_\tau\right] = \sum_{h=0}^{\omega-1} p_{01_{\tau+h+1}} \prod_{i=1}^{h} p_{00_{\tau+i}}\left[\frac{\prod_{i=1}^{\omega} p_{00_{\tau+i}}}{(1-\prod_{i=1}^{\omega} p_{00_{\tau+i}})^2}\sum_{j=0}^{\omega-1}\mathcal{E}\left[S_{\tau+j}\right] + \frac{\sum_{j=0}^{h}\mathcal{E}\left[S_{\tau+j}\right]}{1-\prod_{i=1}^{\omega} p_{00_{\tau+i}}}\right] \tag{5.20}$$

Derivation of the variance of D_τ is more involving than finding its expectation.

Making use of conditional expectation concepts, the variance can be written as:

$$
\begin{aligned}
\mathrm{Var}\left[D_\tau\right] &= \mathrm{Var}\left[\sum_{j=0}^{L_\tau-1} S_{\tau+j}\right] && (5.21)\\
&= \mathcal{E}\left[\mathrm{Var}\left[\sum_{j=0}^{L_\tau-1} S_{\tau+j}\middle| L_\tau\right]\right] + \mathrm{Var}\left[\mathcal{E}\left[\sum_{j=0}^{L_\tau-1} S_{\tau+j}\middle| L_\tau\right]\right] = && (5.22)\\
&= \mathcal{E}\left[\sum_{i=0}^{L_\tau-1}\sum_{j=0}^{L_\tau-1} \mathrm{Cov}\left[S_{\tau+i}, S_{\tau+j}\right]\right] + \mathrm{Var}\left[\sum_{j=0}^{L_\tau-1} \mathcal{E}\left[S_{\tau+j}\right]\right] && (5.23)
\end{aligned}
$$

The first term represent the expected value of the sum of all the elements of the variance-covariance matrix of the vector $\{S_\tau, S_{\tau+1}, \ldots, S_{\tau+L_\tau-1}\}$. Assuming the process uncorrelated, it will reduce to:

$$
\mathcal{E}\left[\sum_{i=0}^{L_\tau-1}\sum_{j=0}^{L_\tau-1} \mathrm{Cov}\left[S_{\tau+i}, S_{\tau+j}\right]\right] = \mathcal{E}\left[\sum_{i=0}^{L_\tau-1} \mathrm{Var}\left[S_{\tau+i}\right]\right] \qquad (5.24)
$$

Making use of Proposition 2, the above term can be expressed in closed form as:

$$
\mathcal{E}\left[\sum_{i=0}^{L_\tau-1} \mathrm{Var}\left[S_{\tau+i}\right]\right] =
$$

$$
\sum_{h=0}^{\omega-1}\left\{p01_{\tau+h+1}\prod_{i=1}^{h} p00_{\tau+i}\left[\frac{\prod_{i=1}^{\omega} p00_{\tau+i}}{(1-\prod_{i=1}^{\omega} p00_{\tau+i})^2}\sum_{j=0}^{\omega-1}\mathrm{Var}\left[S_{\tau+j}\right] + \frac{\sum_{j=0}^{h}\mathrm{Var}[S_{\tau+j}]}{1-\prod_{i=1}^{\omega} p00_{\tau+i}}\right]\right\} \qquad (5.25)
$$

The last term of Eq. (5.23) can be written as:

$$
\mathrm{Var}\left[\sum_{j=0}^{L_\tau-1}\mathcal{E}\left[S_{\tau+j}\right]\right] = \mathcal{E}\left[\left(\sum_{j=0}^{L_\tau-1}\mathcal{E}\left[S_{\tau+j}\right]\right)^2\right] - \mathcal{E}^2\left[\sum_{j=0}^{L_\tau-1}\mathcal{E}\left[S_{\tau+j}\right]\right] \qquad (5.26)
$$

The first expectation can be computed numerically as:

$$\mathcal{E}\left[\left(\sum_{j=0}^{L_\tau-1}\mathcal{E}\left[S_{\tau+j}\right]\right)^2\right]=\sum_{l=1}^{\infty}\left(\sum_{j=0}^{l-1}\mathcal{E}\left[S_{\tau+j}\right]\right)^2 f_{L_\tau}(l) \tag{5.27}$$

where $f_{L_\tau}(l)$ is given by Eq. (4.11), while the last term is just the rhs of Eq. (5.20) squared.

Finally, combining Equations (5.23), (5.25), (5.26), and (5.27), an expression for the variance of accumulated deficit of droughts starting in season τ can be derived.

5.2.2 Joint distribution of drought length (L) and accumulated deficit (D) or intensity (I)

The derivation of the bivariate probability distribution function (pdf) of accumulated deficit and length can be pursued by considering the distribution of accumulated deficit conditioned on a fixed drought length $L_\tau=l$. Then the bivariate pdf of accumulated deficit D_τ and length L_τ can be found as:

$$f_{D_\tau,L_\tau}(d,l)=f_{D_\tau|L_\tau}(d)f_{L_\tau}(l) \tag{5.28}$$

where $f_{D_\tau|L_\tau}(d)$ is the conditional pdf of D_τ given $L_\tau=l$ and $f_{L_\tau}(l)$is the marginal pdf of L_τ . Deriving the exact analytical expression of $f_{D_\tau|L_\tau}(d)$ considering all the features of the underlying process is not an easy task. As an approximation, we will assume accumulated deficit $D_{\tau|L}$ to be gamma distributed with parameters (r_τ,β_τ), namely:

$$f_{D_\tau|L_\tau}(d)=\frac{1}{\beta_\tau\Gamma(r_\tau)}\left(\frac{d}{\beta_\tau}\right)^{r_\tau-1}\exp\left(-\frac{d}{\beta_\tau}\right) \tag{5.29}$$

As will be shown later, in the above conditional pdf the parameters have an implicit dependence on L_τ. In addition, if we assume that the sequence of surpluses and deficits follows a periodic lag 1 Markov chain, the marginal pdf $f_{L_\tau}(l)$ of drought duration can be expressed as (Cancelliere and Salas, 2004) as:

$$f_{L_\tau}(l) = \left(\prod_{i=1}^{l-1}\left(1 - p_{01_{\tau+i}}\right)\right) p_{01_{\tau+l}} l = 1, 2, ... \tag{5.30}$$

where p_{01_τ} represents the transition probability of observing a surplus in season τ given a deficit occurred in season τ *-1*.

Combining these results, the bivariate pdf of accumulated deficit and duration, for drought starting in season τ, becomes:

$$f_{D_\tau, L_\tau}(d, l) = \frac{1}{\beta_\tau \Gamma(r_\tau)}\left(\frac{d}{\beta_\tau}\right)^{r_\tau - 1} \exp\left(-\frac{d}{\beta_\tau}\right)\left(\prod_{i=1}^{l-1}\left(1 - p_{01_{\tau+i}}\right)\right) p_{01_{\tau+l}} \quad l = 1, 2, ... \tag{5.31}$$

Furthermore, from the conditional distribution of accumulated deficit given drought duration given by Eq. (5.29), the conditional distribution of drought intensity I_τ given duration L_τ can be also derived. Indeed, since the drought intensity is the ratio of accumulated deficit to drought length, i.e. $I_\tau = D_\tau / l$, the conditional pdf of intensity I_τ given a fixed length $L_\tau = l$ can be derived from Eq. (5.29) as:

$$f_{I_\tau | L_\tau}(i) = \frac{1}{\beta_\tau \Gamma(r_\tau)}\left(\frac{li}{\beta_\tau}\right)^{r_\tau - 1} \exp\left(-\frac{li}{\beta_\tau}\right) \tag{5.32}$$

which is also gamma distributed, i.e. G($l\ r_\tau$, β_τ / l). Thus, the bivariate pdf of intensity and length can be found in a similar fashion as in Eq. (5.30) above. It follows:

$$f_{I_\tau,L\tau}(d,l) = \frac{1}{\beta_\tau \Gamma(r_\tau)} \left(\frac{ld}{\beta_\tau}\right)^{r_\tau - 1} \exp\left(-\frac{ld}{\beta_\tau}\right) \left(\prod_{i=1}^{l-1} \left(1 - p_{01_{\tau+i}}\right)\right) p_{01_{\tau+l}} \tag{5.33}$$

The periodic bivariate pdf's of accumulated deficit and duration and of intensity and duration given by eqs. (5.31) and (5.33) respectively are related to droughts starting at season τ. It may be of some interest to derive the bivariate pdf's of drought characteristics, regardless of the initial season which will find application to compute return periods of droughts, as shown later. Such stationary pdf's can be seen as a mixture of the periodic pdf's, with weights given by the ratio of the probability of observing a drought starting in a given season to the probability of having a drought at any seasons (Cancelliere and Salas, 2004). The probability of having a drought in season τ is given by (Cancelliere and Salas, 2004):

$$P(drought starts at season \tau) = p_{1\tau-1} p_{10\tau} \tag{5.34}$$

where $p_{1\tau}$is the probability of having a surplus in season τ and p_{10_τ} represents the transition probability of observing a deficit in season τ given a surplus occurred in season τ *-1.*

Then the bivariate pdf of accumulated deficit D and duration L, regardless of the initial season will be:

$$f_{D,L}(d,l) = \frac{\sum\limits_{\tau=1}^{\omega} f_{D_\tau,L_\tau}(d,l) p_{1\tau-1} p_{10\tau}}{\sum\limits_{\tau=1}^{\omega} p_{1\tau-1} p_{10\tau}} \tag{5.35}$$

Similarly the bivariate pdf of intensity I and duration L, regardless of the initial season will be:

$$f_{I,L}(i,l) = \frac{\sum\limits_{\tau=1}^{\omega} f_{I_\tau,L_\tau}(i,l) p_{1\tau-1} p_{10\tau}}{\sum\limits_{\tau=1}^{\omega} p_{1\tau-1} p_{10\tau}} \tag{5.36}$$

By integrating appropriately the above bivariate pdf's the occurrence probability of droughts with different characteristics (i.e. combination of cumulative deficit and duration or intensity and duration) can be computed. In particular, the following droughts events are considered here:

- Drought events E with duration L equal to a given number of months l_0 and accumulated deficit D greater than a specified quantity D_0, i.e. $E = \{D > D_0$ and $L = l_0$ (l_0=1,2,...)$\}$:

$$P[D > D_0, L = l_0] = \int_{D_0}^{\infty} f_{D,L}(z, l_0) dz \tag{5.37}$$

- Drought events E with duration L equal or greater than a given number of months l_0 and accumulated deficit D greater than a specified quantity D_0, i.e. $E = \{D > D_0$ and $L \geq l_0$ (l_0=1,2,...)$\}$:

  P$[D > D_0, L \geq l_0]$=

$$P[D > D_0, L \geq l_0] = \int_{D_0}^{\infty} \sum_{l=l_0}^{\infty} f_{D,L}(z, l) dz \tag{5.38}$$

- Drought events E with duration L equal to a given number of months l_0 and intensity I greater than a specified quantity I_0, i.e. $E = \{I > I_0$ and $L = l_0$ (l_0=1,2,...)$\}$:

$$P[I > I_0, L = l_0] = \int_{I_0}^{\infty} f_{I,L}(z, l_0)dz \qquad (5.39)$$

- Drought events E with duration L greater or equal to a given number of months l_0 and intensity I greater than a specified quantity I_0, i.e. $E = \{I > I_0$ and $L \geq l_0$ (l_0=1,2,...)$\}$:

$$P[I > I_0, L \geq l_0] = \int_{I_0}^{\infty} \sum_{l=l_0}^{\infty} f_{I,L}(i, l)di \qquad (5.40)$$

Droughts where $D > D_0$ and $L \geq 1$ represent droughts regardless of the drought duration; conversely droughts where $D > 0$ (i.e. $D_0 = 0$), and $L \geq l_0$ represent droughts regardless of the amount of deficit. Furthermore, the marginal probability of droughts events $D > D_0$ or $I > I_0$ can be obtained from Eqs.(5.38) or (5.40), respectively, by letting l=1. All the above events can also be particularized for droughts beginning at a given season τ by replacing the bivariate pdf's in Eqs. (5.37), (5.38), (5.39) and (5.40) by the corresponding pdf's given by Eqs. (5.31) and (5.33).

It should be noted that despite the apparent complexity of the above expressions, the integrations can be carried out efficiently making use of numerical tools for the gamma pdf that are available in most mathematical and statistical software.

Application of the derived pdf's involves the estimation of the parameters, namely the parameters of the gamma distribution r_τ, β_τ and the transition probabilities p_{01_τ}. Estimation of the latter has been discussed in Chapter 4, thus we will concentrate our attention to the estimation of the parameters of the gamma.

In principle, such parameters could be estimated from the observed droughts, by means of traditional estimation techniques such as method of moment or maximum likelihood. However, given the relatively limited sample length of the available records, the number of observed droughts is seldom sufficient to yield reliable estimates for these parameters. This is particularly true in light of the fact that r_τ, β_τ have an implicit dependence on L_τ and therefore they must be estimated separately for each drought duration and for each month τ. To illustrate the point, in Table 5.I , the number of droughts identified on the Caltanissetta 3-months precipitation series (1921-2000) assuming a constant threshold equal to the 3-months mean are reported for various drought duration. It can be inferred that, despite the relatively length of the available sample (80 years), the identified droughts are not enough to estimate reliably the parameters, since, with few exceptions, they are generally considerably less than 10, if not none.

A different approach will be employed here, namely to derive analytical expressions for the parameters r_τ, β_τ of the gamma distribution based on underlying distribution of $X_{\nu,\tau}$ and of the threshold level $x_{o\tau}$. This will enable to fully exploit the available information from the whole observed series, thus allowing for a reliable estimation of the joint distribution of drought characteristics even on relatively short series. In the following, first the moments of drought characteristics will be derived as a function of the underlying distribution of $X_{\nu,\tau}$ and of the threshold level $x_{o\tau}$. Then, the parameters of the conditional distribution will be expressed as a function of such moments, thus providing and approximate, yet reliable expression for the joint distribution of drought characteristics.

Let $E\left[D_{\tau|L}\right]$ and var $\left[D_{\tau|L}\right]$ be the expected value and the variance of $D_{\tau|L}$ respectively.

Assuming $D_{\tau|L}$ gamma distributed the following relationships will hold between the moments and the parameters (r_τ,β_τ) (Mood et al., 1974):

$$E\left[D_{\tau|L}\right] = r_\tau \beta_\tau \tag{5.41}$$

$$\text{var}\left[D_{\tau|L}\right] = r_\tau \beta_\tau^2 \tag{5.42}$$

which solved for (r_τ, β_τ) yields:

$$r_\tau = \frac{E^2\left[D_{\tau|L}\right]}{\text{Var}\left[\left[D_{\tau|L}\right]\right]} \tag{5.43}$$

(18)

$$\beta_\tau = \frac{\text{var}\left[D_{\tau|L}\right]}{E\left[D_{\tau|L}\right]} \tag{5.44}$$

The proposed approach is to determine $E\left[D_{\tau|L}\right]$ and var $\left[D_{\tau|L}\right]$ as a function of the stochastic structure of the underlying hydrological variable. More specifically, let's start by observing that the above moments are related to the moments of the single deficits which, in turn depend on the distribution of the underlying variable and on the threshold level. The $D_{\tau|L}$ is the sum of the deficits, extended to a fixed duration L_τ

$$D_{\tau|L} = \sum_{i=0}^{L_\tau - 1} S_{\nu,\tau+i} \tag{5.45}$$

Taking expectation of both sides of Eq. (5.45) and remembering that L_τ is to be considered fixed:

$$E\left[D_{\tau|L}\right] = E\left[\sum_{i=0}^{L_\tau - 1} S_{\nu,\tau+i}\right] = \sum_{i=0}^{L_\tau - 1} E\left[S_{\nu,\tau+i}\right] \tag{5.46}$$

Similarly, taking variance of both sides of Eq. (5.45):

$$\mathrm{Var}\left[\left[D_{\tau|L}\right]\right] = \mathrm{Var}\left[\left[\sum_{i=0}^{L_\tau - 1} S_{\nu,\tau}\right]\right] = \sum_{i=0}^{L_\tau - 1}\sum_{j=0}^{L_\tau - 1} \mathrm{Cov}\left[\left[S_{\nu,\tau+i}, S_{\nu,\tau+j}\right]\right] \tag{5.47}$$

If we assume that the deficits are uncorrelated, eq. (5.47) simplifies as:

$$\mathrm{Var}\left[\left[D_{\tau|L}\right]\right] = \sum_{i=0}^{L_\tau - 1} \mathrm{Var}\left[\left[S_{\nu,\tau+i}\right]\right] \tag{5.48}$$

Eqs. (5.46) and (5.47) allow to determine the moments of $D_{\tau|L}$ as a function of the moments of the deficits $S_{\nu,\tau}$, under the assumption of uncorrelated deficits. In general terms, the pdf of deficits $S_{\nu,\tau}$ is the pdf of $X_{\nu,\tau}$ truncated, i.e.:

$$\mathrm{f}_{S_{\nu,\tau}}(\mathrm{s}) = \frac{1}{\mathrm{p}} \cdot \mathrm{f}_{\mathrm{X}_{\nu,\tau}}\left(\mathrm{x}_{O\tau} - \mathrm{s}\right) \cdot \mathrm{I}(\mathrm{s})_{(0,\mathrm{x}_{o\tau})} \tag{5.49}$$

where p is a rescaling term, necessary for $\mathrm{f}_{S_{\nu,\tau}}(\mathrm{s})$ to be a pdf. It can be computed as:

$$p = \int_0^{x_{0\tau}} \mathrm{f}_{\mathrm{X}_{\nu,\tau}}(x) \cdot d\mathrm{x} \tag{5.50}$$

Thus the k^{th} moment of the deficits $S_{\nu,\tau}$ is given by:

$$\mathcal{E}\left[D_t^k\right] = \frac{1}{\mathrm{p}_0} \int_0^{\infty} \mathrm{s}^{\mathrm{k}} \cdot \mathrm{f}_{\mathrm{X}_{\nu,\tau}}\left(\mathrm{x}_{0\tau}\text{-}\mathrm{s}\right) \cdot ds \tag{5.51}$$

The moments of the deficits have been derived, among others, by Bonaccorso et al. (2003), assuming the underlying variable normal, or log-normal or gamma distributed. Then, the parameters of the distribution of cumulative deficit conditioned on a fixed length and on an initial season can be computed by means of eqs. (5.43) and (5.44).

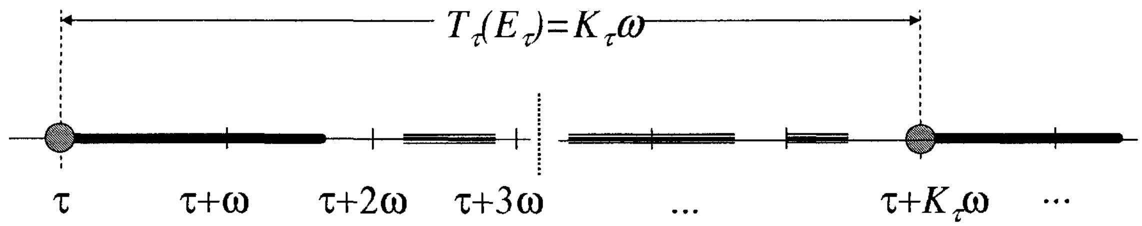


Figure 5.2: Definition of interarrival time $T_\tau(E_\tau)$ of critical droughts starting at a given season τ, with characteristics E_τ, e.g. $E_\tau = \{D_\tau > D_0$ and $L_\tau = l_0$ (l_0=1,2,...)$\}$. Critical events are depicted by solid thick lines, while non critical events (i.e. events that either don't start on season τ or that do not have characteristics E_τ) are depicted by striped lines.

5.3 Return period of droughts

The return period T of a drought events with characteristics E, e.g. $E = \{L = l_0$and $D > D_0\}$, will be derived for both droughts starting at a given season τ and for droughts starting at any season. For both cases return period is defined as the mean interarrival time between droughts with characteristics E.

5.3.1 Droughts starting at a given season τ

To derive the return period of critical droughts starting at a given season τ, let's consider critical drought events with characteristics E_τ, e.g. $E_\tau = \{D_\tau > D_0$ and $L_\tau = l_0$ (l_0=1,2,...)$\}$. In Figure 3 such critical events are depicted by solid thick lines, while non critical events (i.e. events that either don't start on season τ or that do not have characteristics E_τ) are depicted by striped lines.

Since all critical events must start at season τ, the interarrival time $T_\tau(E_\tau)$ between any two of such events must be a multiple of the number of seasons ω, i.e.:

$$T_\tau(E_\tau) = K_\tau\omega \quad K_\tau = 1, 2, \ldots \tag{5.52}$$

where K_τ is a discrete random variable that can take only integer values. Thus

K_τ represents the number of intervals in season τ before a second critical drought occurs.

Although in general terms K_τ starts at 1, however its lower bound will depend on the duration of the critical events we are considering. For example, with reference to the droughts depicted in Figure 5.2, since the first critical drought has duration greater than ω but less than 2ω K_τ must be greater than 1. This is just a consequence of the fact that the interarrival time between critical drought must obviously exceed the duration of such critical droughts. In a more formal way, if the duration is such that $j\omega \leq L_\tau < (j+1)\omega - 1, j = 0, 1, 2, 3, \ldots$ then $T_\tau(E_\tau) \geq (j+1)\omega$ and consequently $K_\tau \geq (j+1)$. The above relationship holds for $\omega > 1$. For practical purposes, j can be computed as the integer part of the ratio of L_τ to ω.

Due to the one-to-one correspondence between $T_\tau(E_\tau)$ and K_τ , the pdf of $T_\tau(E_\tau)$ can be pursued by deriving the pdf of K_τ. To do this, we will concentrate our attention to critical events with fixed duration , e.g. $E_\tau = \{D_\tau > D_0$ and $L_\tau = l_0, (l_0 = 1, 2, \ldots)\}$ or $E_\tau = \{I_\tau > I_0$ and $L_\tau = l_0, (l_0 = 1, 2, \ldots)\}$.

Also, we define the integer j such that $j\omega \leq l_0, (j+1)\omega - 1$. To derive the pdf of K_τ a Bernoulli trial approach will be followed. Let's denote by success an interval in season τ where a critical event E_τ starts (indicated by a solid circle in Figure 5.2, and by failure an interval in season τ where a critical event E_τ does not start. Then, K_τ represents the number of failures until the first success. The probability of a success (critical drought at season τ), here indicated as Ps_τ, will be the product of the probability that a drought occurs starting at season τ times the probability that it has characteristics E_τ:

$$Ps_\tau = P[\text{ a drought occurs starting at season } \tau] \cdot P[E_\tau] = p_{1_{\tau-1}} p_{10_\tau} \cdot P[E_\tau] \quad (5.53)$$

where $p_{1_{\tau-1}} p_{10_\tau}$ represents the probability that a drought occurs starting at season τ (Cancelliere and Salas, 2004). Following the Bernoulli trial scheme for the sequence of failures and success, we observe that if $K_\tau = k$ there must be *k-1* failures before 1 success. Thus the pdf of K_τ must be geometric with parameter Ps_τ. However, since we have the condition $K_\tau \geq (j+1)$, which essentially takes into account the fact that the first drought must terminate before the next one starts, such pdf becomes:

$$P\left[K_\tau = k \mid K_\tau \geq j+1\right] = Ps_\tau \left(1 - Ps_\tau\right)^{k-j-1} \quad k = j+1, j+2, \ldots \tag{5.54}$$

In deriving Eq. (5.54) we capitalize on the lack of memory property of the geometric distribution [Mood et al., 1974].

Finally, the probability

$$P\left[T_\tau(E_\tau) = t\right] \tag{5.55}$$

can be computed by observing that:

$$P\left[T_\tau\left(E_\tau\right) = t\right] = P\left[K_\tau \omega = t\right] = P\left[K_\tau = t/\omega\right] \tag{5.56}$$

where:

$$t = \omega\left(j+1\right), \omega\left(j+2\right), \ldots \tag{5.57}$$

Combining Eqs. (5.54) and (5.57), the following expression for the pdf of interarrival time is obtained:

$$P\left[T_\tau\left(E_\tau\right) = t\right] = Ps_\tau \left(1 - Ps_\tau\right)^{t/\omega-j-1} \quad t = \omega\left(j+1\right), \omega\left(j+2\right), \ldots \tag{5.58}$$

To compute the return period, first let's compute the expected value of K_τ:

$$E\left[K_\tau\right] = \sum_{k=j+1}^{\infty} kP\left[K_\tau = k \mid K_\tau \geq j+1\right] = \sum_{k=j+1}^{\infty} kPs_\tau \left(1 - Ps_\tau\right)^{k-j-1} \tag{5.59}$$

The infinite sum can be solved by a change of variable i=k-j, thus:

$$E\left[K_\tau\right] = \sum_{i=1}^{\infty} (i+j)\, Ps_\tau \left(1 - Ps_\tau\right)^{i-1} = j + \frac{1}{Ps_\tau} \tag{5.60}$$

as follows directly from a property of the geometric sum. Then, the return period will be:

$$E\left[T_\tau(E_\tau)\right] = E\left[K_\tau \omega\right] = \omega \left(j + \frac{1}{Ps_\tau}\right) \tag{5.61}$$

5.3.2 Return period of droughts beginning at any season

It is of interest to derive the return period T of droughts events beginning at any season, identified on periodic series and defined (jointly) in terms of duration and either accumulated deficit or intensity. However, since the derivation follows basically the same approach already adopted to derive return period for stationary series (see Par. 3.6), most of the details will be omitted in order to avoid repetitions.

In general terms the return period of drought event with characteristics E, is given by eq. (3.67), here repeated for convenience:

$$T = \frac{\mathcal{E}\left[L\right] + \mathcal{E}\left[Ln\right]}{P[E]} \tag{5.62}$$

where $\mathcal{E}\left[L\right]$ is the expected value of drought length, $\mathcal{E}\left[Ln\right]$ is the expected value of non drought periods length (surplus) and $P[E]$ is the probability of a drought event with characteristics E.

Both expectations have been derived in closed form in Chapter 4 for the general periodic Markov case as:

$$E(Ln) = \frac{\sum_{\tau=1}^{\omega}\left\{p_{0_{\tau-1}}p_{01_{\tau}}\left[\sum_{t=1}^{\omega}\left(p_{10_{\tau+t}}\prod_{k=1}^{t-1}\left(1-p_{10_{\tau+k}}\right)\frac{t-\Delta^{*}(t-\omega)}{(1-\Delta^{*})^{2}}\right)\right]\right\}}{\sum_{\tau=1}^{\omega}p_{0_{\tau-1}}p_{01_{\tau}}} \tag{5.63}$$

$$E(L) = \frac{\sum_{\tau=1}^{\omega}\left\{p_{1_{\tau-1}}p_{10_{\tau}}\left[\sum_{t=1}^{\omega}\left(p_{01_{\tau+t}}\prod_{k=1}^{t-1}\left(1-p_{01_{\tau+k}}\right)\frac{t-\Delta(t-\omega)}{(1-\Delta)^{2}}\right)\right]\right\}}{\sum_{\tau=1}^{\omega}p_{1_{\tau-1}}p_{10_{\tau}}} \tag{5.64}$$

where $\Delta = \prod_{k=1}^{\omega}\left(1-p_{01_{\tau+k}}\right)$ and $\Delta^{*} = \prod_{k=1}^{\omega}\left(1-p_{10_{\tau+k}}\right)$

Equation (5.62) can be easily particularized for the different drought events considered previously. For example, the return period of a drought with accumulated deficit D greater than a specified quantity D_0 can be expressed as:

$$T = \frac{E(L)+E(Ln)}{\mathrm{P}\left[D > D_0\right]} \tag{5.65}$$

where the probability P[ ] can be computed using the Gamma distribution in Eq. (5.28).

As another example, the return period of a drought with accumulated deficit D greater than a specified quantity D_0 and duration L equal to a given number of years l_0, i.e. $E = \{D > D_0$ and $L = l_0$ $(l_0=1,2,\ldots)\}$ can be expressed as:

$$T = \frac{E(L)+E(Ln)}{\mathrm{P}\left[D > D_0,\ L = l_o\right]} \tag{5.66}$$

where the probability P[ ] can be computed from Eq. (5.37). Similar expressions can be derived for the other considered cases.

5.4 Applications

5.4.1 Marginal distribution and return period of accumulated deficit

The methodology to derive the pdf of accumulated deficit and to estimate return period of droughts in periodic series has been tested with reference to 6 monthly precipitation series in Italy, namely Agrigento, Termini Imerese, Caltanissetta, Linguaglossa, Trapani and Milano Brera. The record lengths range from 104 years for Linguaglossa station to 234 years for Milano Brera.

First, the analytical pdf of accumulated deficit has been computed for each series as in Eq. (5.1) assuming as a threshold value $x_{0\tau} = \mu_{X_\tau} - \alpha\sigma_{X_\tau}$, where μ_{X_τ} and σ_{X_τ} are the monthly mean and monthly standard deviation standard deviations of each precipitation series respectively and α is a threshold coefficient. Two values of α have been considered corresponding to two different demand levels, namely $\alpha = 0$ and $\alpha = .2$

In Figure 5.3-5.9 the qq-plots between the sample quantiles of accumulated deficit from the samples and the corresponding theoretical quantiles computed by means of Eq. (5.1) are shown for the two investigated threshold levels $\alpha = 0.0$ and $\alpha = 0.2$. In order to make the results comparable among the different series, accumulated deficit has been divided by the corresponding mean annual precipitation. From the plots a remarkable goodness of fit of the theoretical distribution to the sample one for all investigated series and for the two thresholds can be inferred, as the points tends to align around the 45 degree line. Deviations from such a line for large values of accumulated deficit can be generally ascribed to sample variability of the estimated quantiles from the observed samples. The plots confirm the validity of the derived expression that seems suitable to estimate drought probabilities for a wide range of accumulated deficits.

The return period of accumulated deficit for the six precipitation series investigated were estimated from Eq. (5.65) using the analytical expressions for the expected inter-arrival times of drought events and the distribution of accumulated deficit as described above. Likewise, the return period was estimated from the historical samples as the average inter-arrival time between droughts with accumulated deficit greater than a given value. The estimates were made for the same demand thresholds as indicated previously in relation to Figures 5.3-5.9. The results are shown in Figures 5.10-5.16. Again, in order to make the results comparable among the different series, accumulated deficit has been divided by the corresponding mean annual precipitation.

In all cases the general pattern is similar, i.e. there is an excellent agreement between the return periods T estimated from Eq. (5.65) and those estimated from the historical sample. This is especially so for smaller values of accumulated deficit, thereafter the sample estimates generally appear to drift above or below the analytical estimates. For example, Figure 5.5(b) for Caltanissetta precipitation data shows that for dimensionless accumulated deficits greater than .15 the sample estimates drift to smaller values than those obtained from Eq. (5.65). On the other hand Figure 5.9(b) for milano Brera precipitation data shows that for dimensionless accumulated deficits greater than .18 the sample estimates drift to larger values than those obtained from Eq. (5.65) The fluctuations of the sample return period can be explained because as larger values of accumulated deficit are considered, there are smaller numbers of drought episodes that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time becomes unreliable. In fact, beyond a certain accumulated deficit no droughts may be observable from the historical sample and no return period can be determined.

The general agreement between analytical and observed return period confirms the validity of the proposed methodology thus indicating its suitability to extrapolate

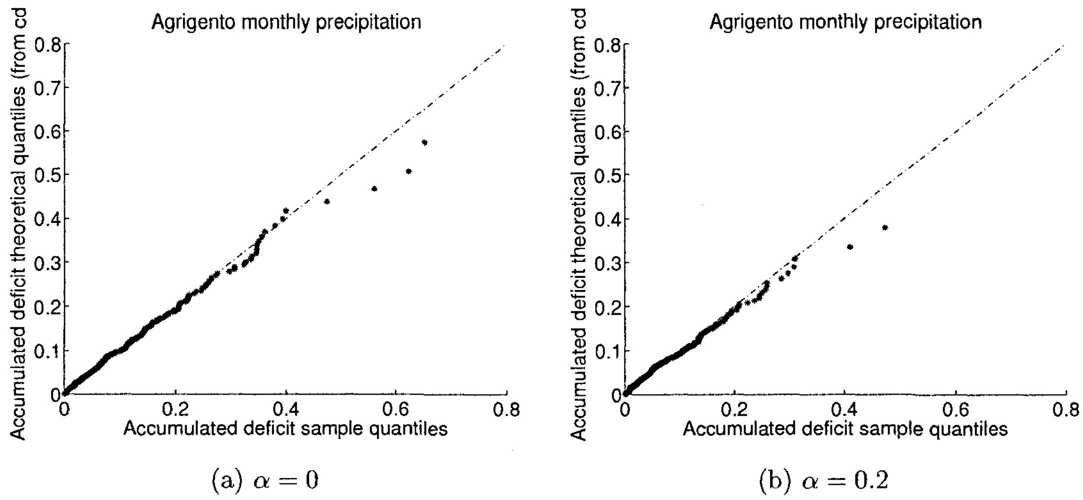
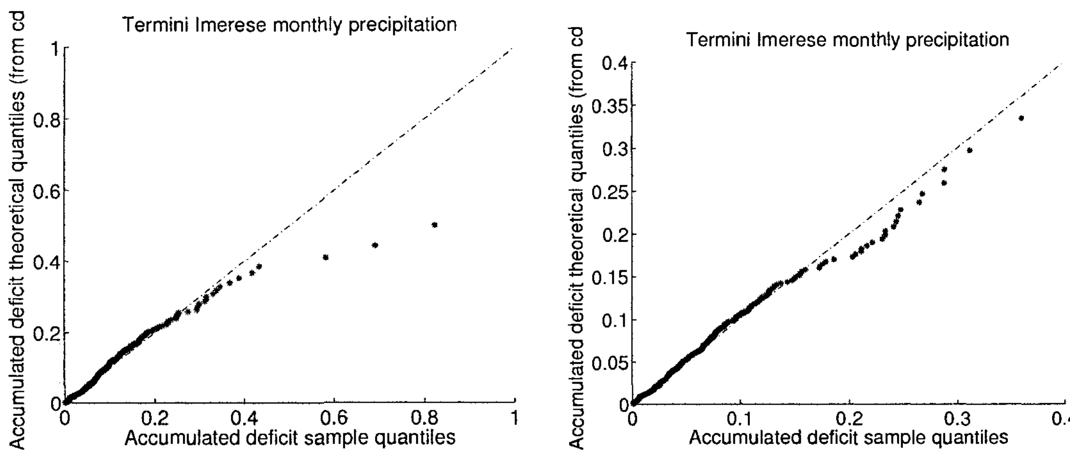


Figure 5.3: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Agrigento monthly precipitation series, for different values of α (threshold level $x_{0\tau} = \mu_{X_\tau} - \alpha\sigma_{X_\tau}$)

within reasonable limits information about return period of drought events, beyond the length of the observed series.

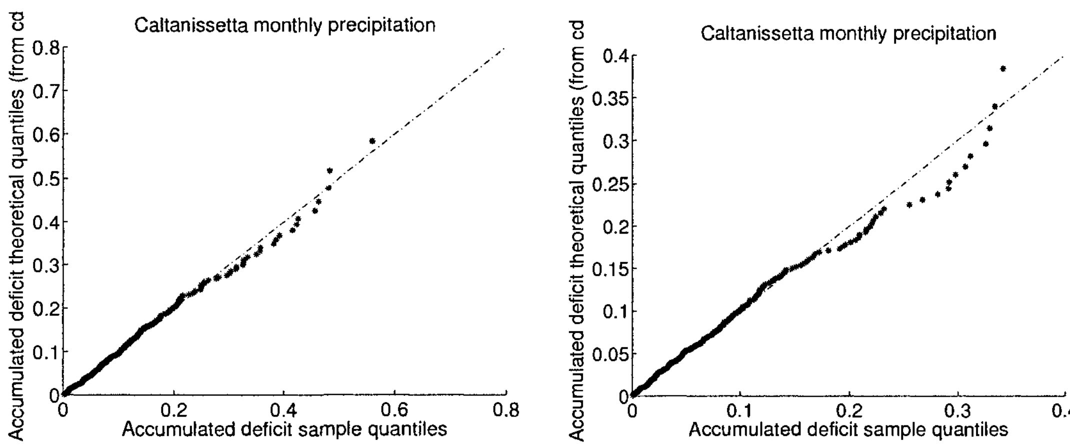
5.4.2 Return period of droughts starting at a given season

The derived expressions to estimate probabilities and return period of drought events starting at a given season as outlined in previous sections of this chapter have been applied to the 3-months precipitation for Caltanissetta, Italy (source: Regional Hydrographic Service of Sicily) for which 119 years of observations are available. Validation of the derived expressions has been carried out by comparing the values of expected values of deficit return periods of droughts events of different severity with the ones estimated on the series as the average of the interarrival times between any two of such droughts observed. In order to overcome the difficulties related to the limited number of observed droughts, that hinders the possibility of obtaining reliable estimates of return periods the proposed expressions have also been validated with



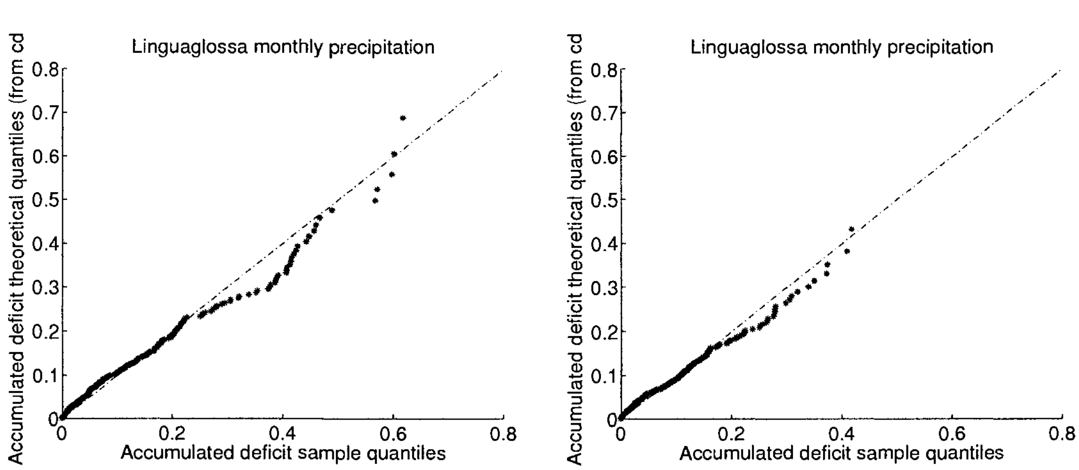
(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.4: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Termini Imerese monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$ (b) $\alpha = 0.2$

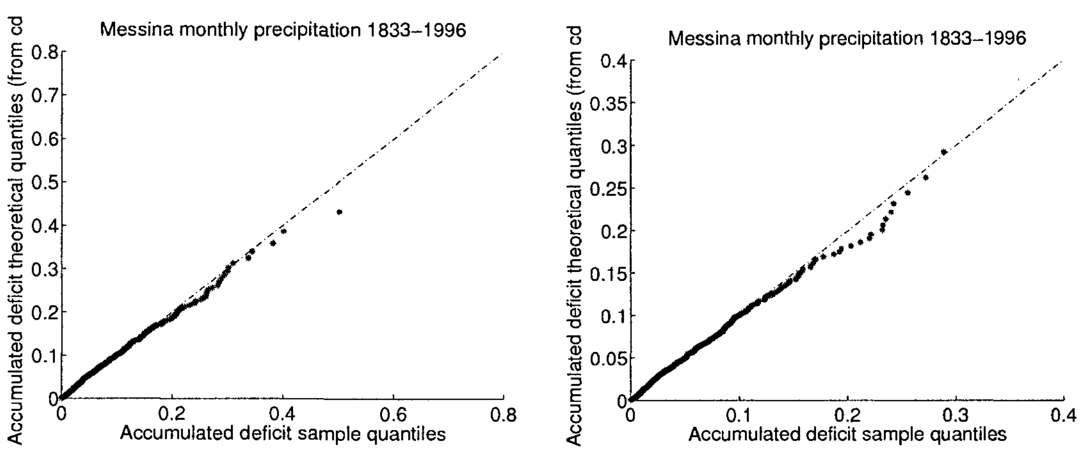
Figure 5.5: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Caltanissetta monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$

(b) $\alpha = 0.2$

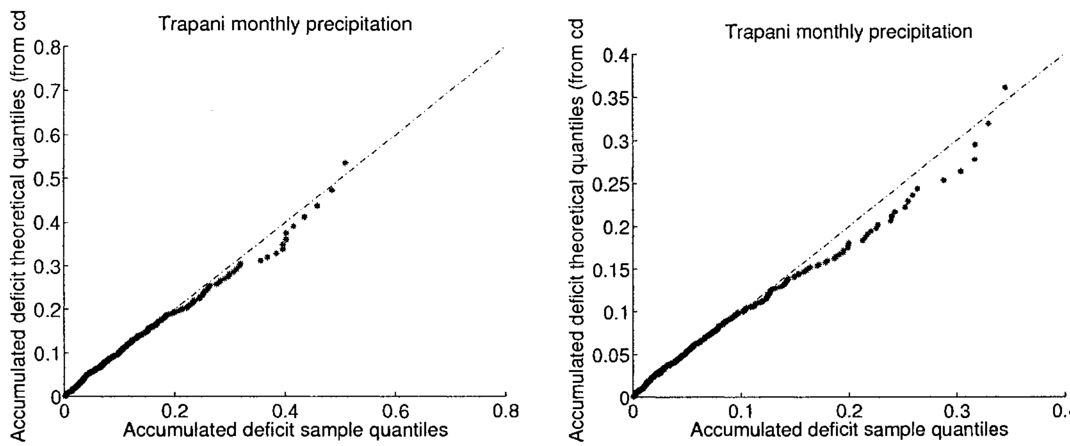
Figure 5.6: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Linguaglossa monthly precipitation series, for different values of α (threshold level $x_{0_\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$

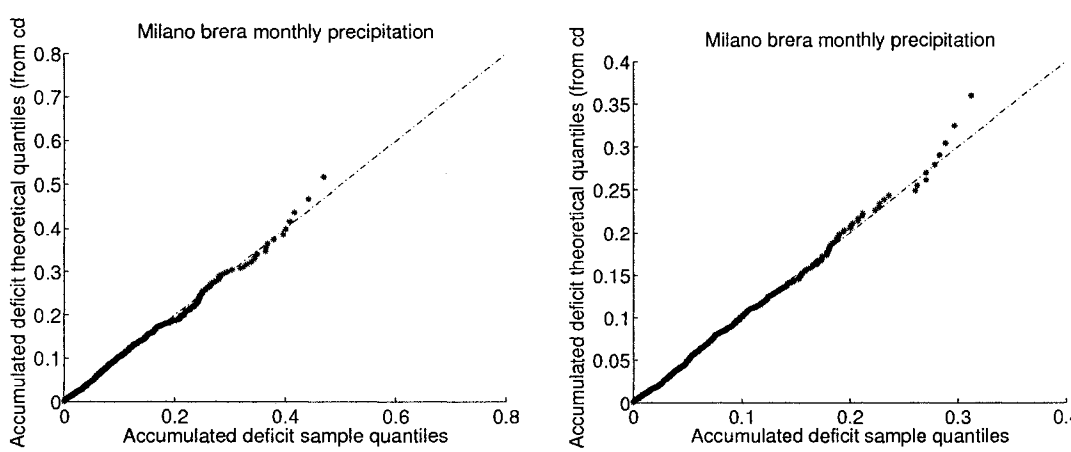
(b) $\alpha = 0.2$

Figure 5.7: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Messina monthly precipitation series, for different values of α (threshold level $x_{0_\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



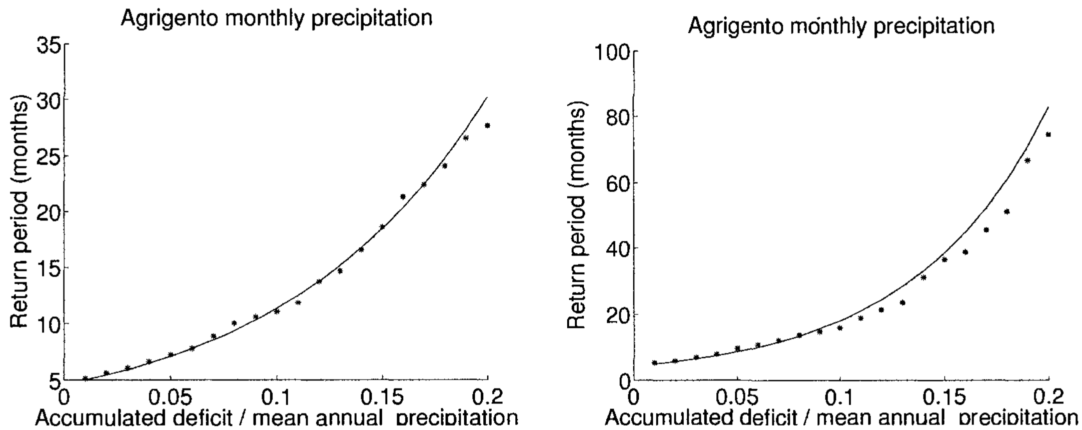
(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.8: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Trapani monthly precipitation series, for different values of α (threshold level $x_{0_\tau} = \mu_{X_\tau} - \alpha\sigma_{X_\tau}$)



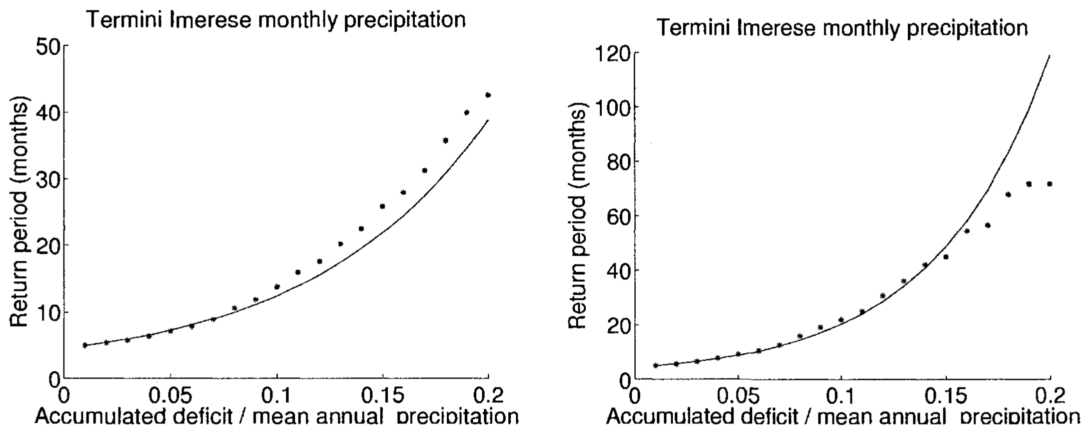
(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.9: Q-Q plot of drought accumulated deficit (divided the mean annual precipitation) for Milano Brera monthly precipitation series, for different values of α (threshold level $x_{0_\tau} = \mu_{X_\tau} - \alpha\sigma_{X_\tau}$)



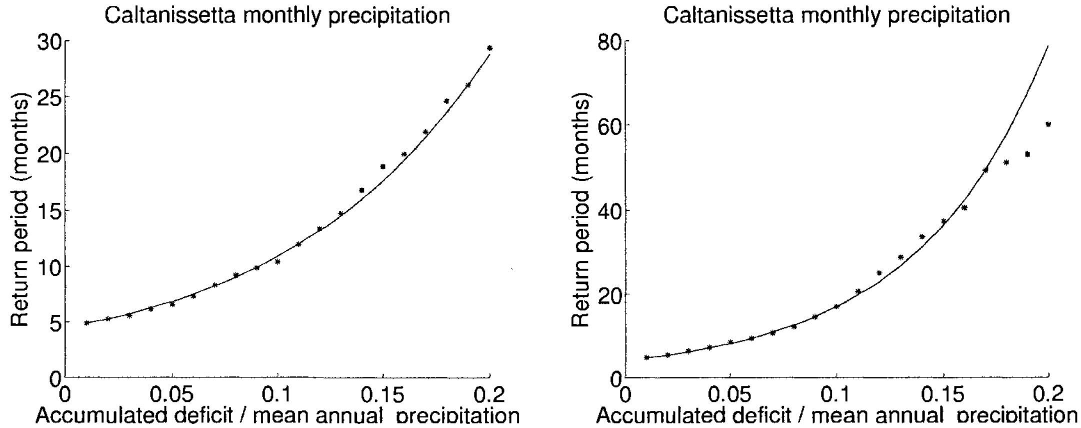
(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.10: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Agrigento monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$ (b) $\alpha = 0.2$

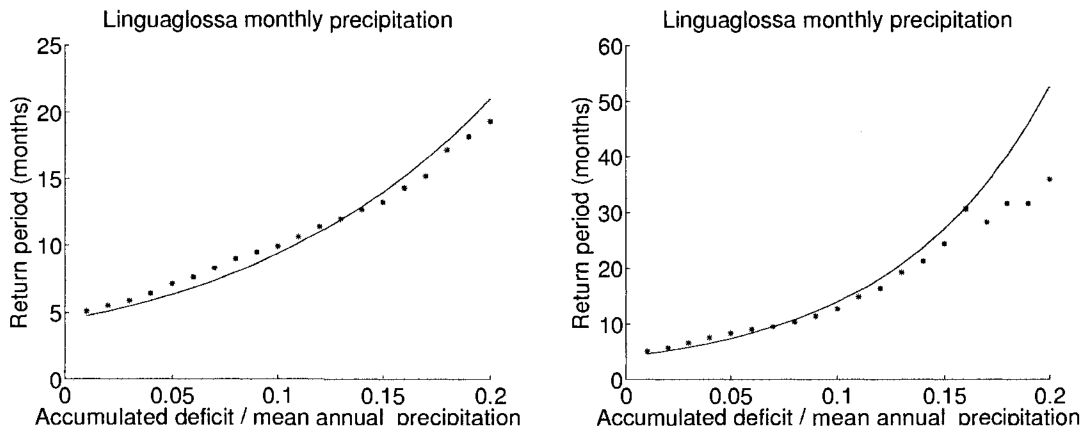
Figure 5.11: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Termini Imerese monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$

(b) $\alpha = 0.2$

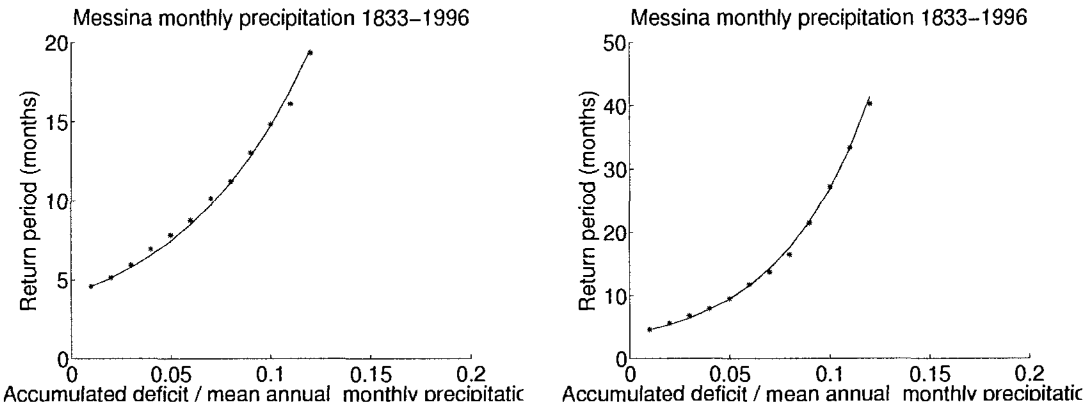
Figure 5.12: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Caltanissetta monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$

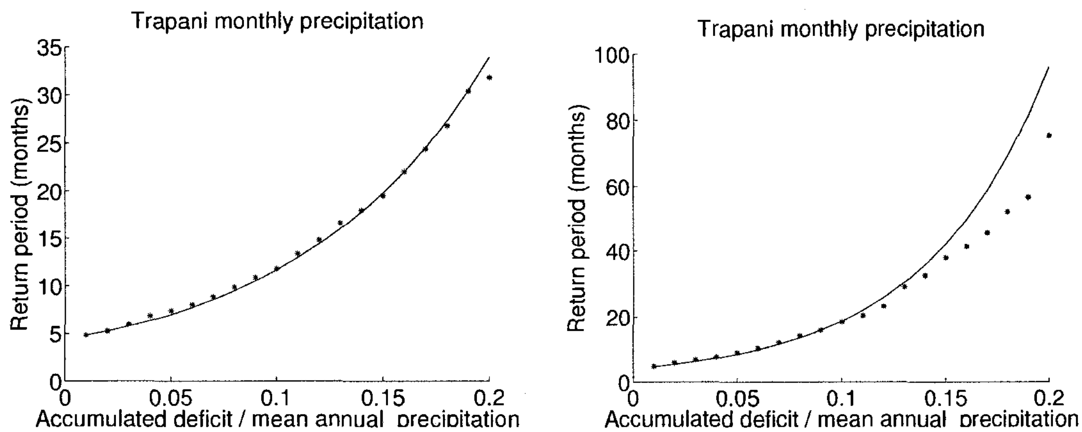
(b) $\alpha = 0.2$

Figure 5.13: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Linguaglossa monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



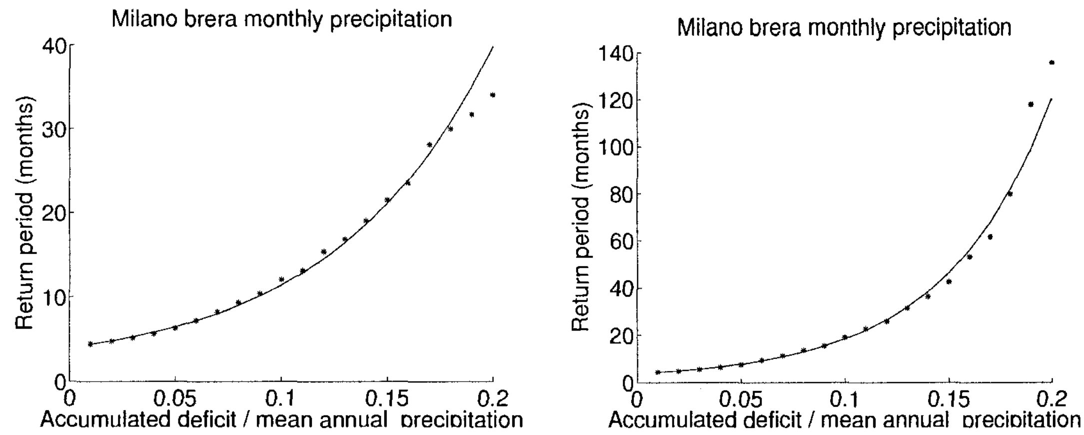
(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.14: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Messina monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.15: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Trapani monthly precipitation series, for different values of α (threshold level $x_{0\mathcal{T}} = \mu_{X_\mathcal{T}} - \alpha\sigma_{X_\mathcal{T}}$)



(a) $\alpha = 0$ (b) $\alpha = 0.2$

Figure 5.16: Observed (*) and theoretical (-) return periods of severity computed by Eq. (5.65) for Milano Brera monthly precipitation series, for different values of α (threshold level $x_{0\tau} = \mu_{X_\tau} - \alpha\sigma_{X_\tau}$)

reference to synthetically generated series, using lag 1 periodic autoregressive models (Salas, 1993) estimated on the basis of the aforementioned data. In particular, logarithmic transformations have been applied to all data prior to model estimation, in order to take into account the presence of skewed data. In all cases, the fitted models are able to reproduce well the main periodic statistics such as means, standard deviations and lag 1 autocorrelation coefficient.

First the proposed expressions to compute return period of drought events beginning at a given season τ have been validated using Caltanissetta three-months precipitation series. In particular, a constant threshold, equal to the average of the four three-months means has been selected, thus leading to a significant periodicity in the corresponding drought characteristics. In Figure 5.17 the expected value of accumulated deficit, estimated by means of Eq. (5.46) and the corresponding mean value estimated numerically from the generated precipitation series is shown as a function of drought length and initial season $\tau = 1$ and $\tau = 3$. In the same figure, the corresponding mean value computed from the observed precipitation series is also shown, only for the cases in which at least two droughts have been observed in the record. From the figure it can be inferred a fairly good agreement between theoretical, observed and simulated results. Figure 5.18 shows the analogous results related to the standard deviation of accumulated deficit. Again, a good agreement between analytical (computed by means of Eq. (5.48)), observed and simulate standard deviation can be inferred. It may be worthwhile to note that some of the discrepancies between theoretical and observed values can be explained in terms of the very low number of droughts identified on the observed record, as can be inferred from Table 5.I. This is for example the case of $\tau = 1$, $L = 1$, for which only two droughts have been identified on the observed record.

Then, return period of droughts of different severities have been computed by fixing different drought durations and accumulated deficit or drought intensity. For easy of reference, we will introduce a *deficit coefficient* λ, defined as the ratio of accumulated deficit to the average demand threshold:

$$\lambda = \frac{D_{o_\tau}}{\bar{x}_o} \tag{5.67}$$

where $\bar{x}_o = \sum_{\tau=1}^{\omega} x_{o\tau}/\omega$. Likewise, we can also introduce a drought intensity coefficient ψ as:

$$\psi = \frac{I_{o_\tau}}{\bar{x}_o} \tag{5.68}$$

By varying λ or ψ droughts of different severities can be defined. In particular, deficit coefficients λ=0,.25,.5,.75,1 and intensity coefficients ψ=0,.25,.5,.75,1 have been selected. Also different durations l_0 varying between 1 and 8 have been fixed.

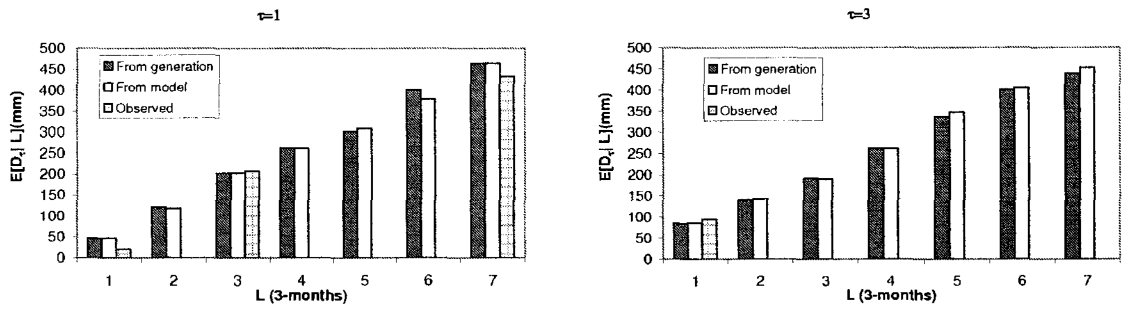


Figure 5.17: Expected value of accumulated deficit for Caltanissetta 3-months precipitation estimated on generated series and by means of Eq. (5.46) as a function of drought duration for initial season τ=1 and τ=3. The corresponding mean value computed on droughts identified on the observed precipitation series is also shown, only for the cases in which at least two droughts have been observed in the record.

Table 5.I: Number of droughts of different duration L_τ starting at season τ identified on Caltanissetta 3-months precipitation record.

	L=1	L=2	L=3	L=4	L=5	L=6	L=7
τ=1	2	1	20	1	0	0	2
τ=2	6	59	8	0	0	2	2
τ=3	9	0	0	0	0	0	0
τ=4	2	0	0	0	0	0	0

In figure 5.19 the comparison of return periods of events $E_\tau = \{D_\tau > \lambda\bar{x}_o$ and $L_\tau = l_0$ (l_0=1,2,...)$\}$ computed by averaging the interarrival times of corresponding droughts on the generated Caltanissetta 3-months precipitation series are compared with the theoretical values computed by means of Eq. (32). Similarly, in figure 7 the return periods of events $E_\tau = \{I_\tau > \psi\bar{x}_o$ and $L_\tau = l_0$ (l_0=1,2,...)$\}$ is reported. Several interesting features can be observed. First of all, return periods, although generally increasing with drought duration, seem to follow periodic patterns. Such periodicity descends directly from the fact that the original periodic series has been clipped by a constant threshold, thus leading to marked periodic features in the corresponding droughts. Furthermore, as drought duration increases, return periods tend to depend less and less on λ. This indicates that as drought duration increases, the probability of occurrence of a drought is mostly function of drought duration.

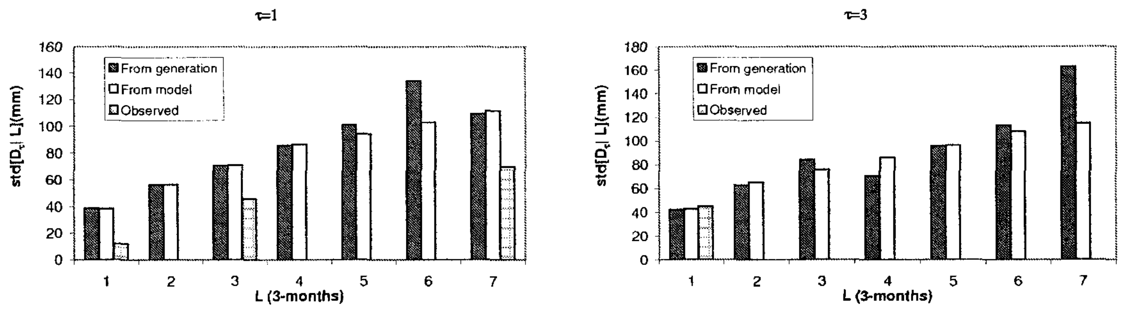


Figure 5.18: Standard deviation of accumulated deficit for Caltanissetta 3-months precipitation estimated on generated series and by means of Eq. (5.46) as a function of drought duration for initial season τ=1 and τ=3. The corresponding standard deviation computed on droughts identified on the observed precipitation series is also shown, only for the cases in which at least two droughts have been observed in the record.

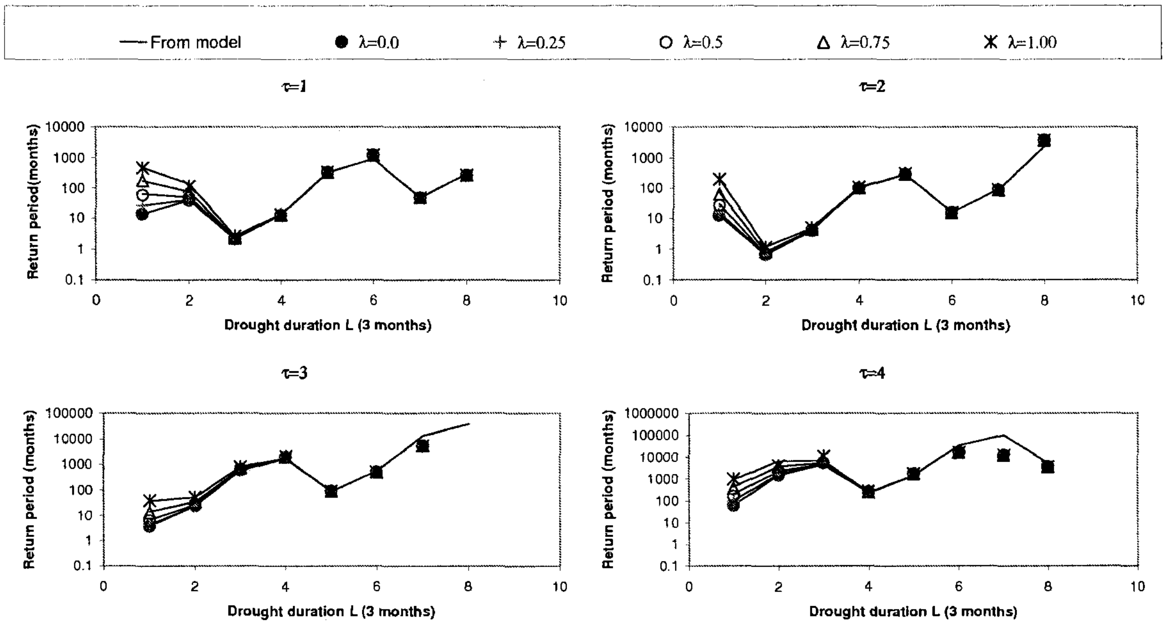


Figure 5.19: Return period of drought events defined by $E_\tau = \{D_\tau > \lambda \bar{x}_o$ and $L_\tau = l_0$ (l_0=1,2,...)$\}$ obtained from the generated synthetic 3-months Caltanissetta precipitation and from Eq. (32) for various values of th deficit coefficient λ and for fixed beginning season τ.

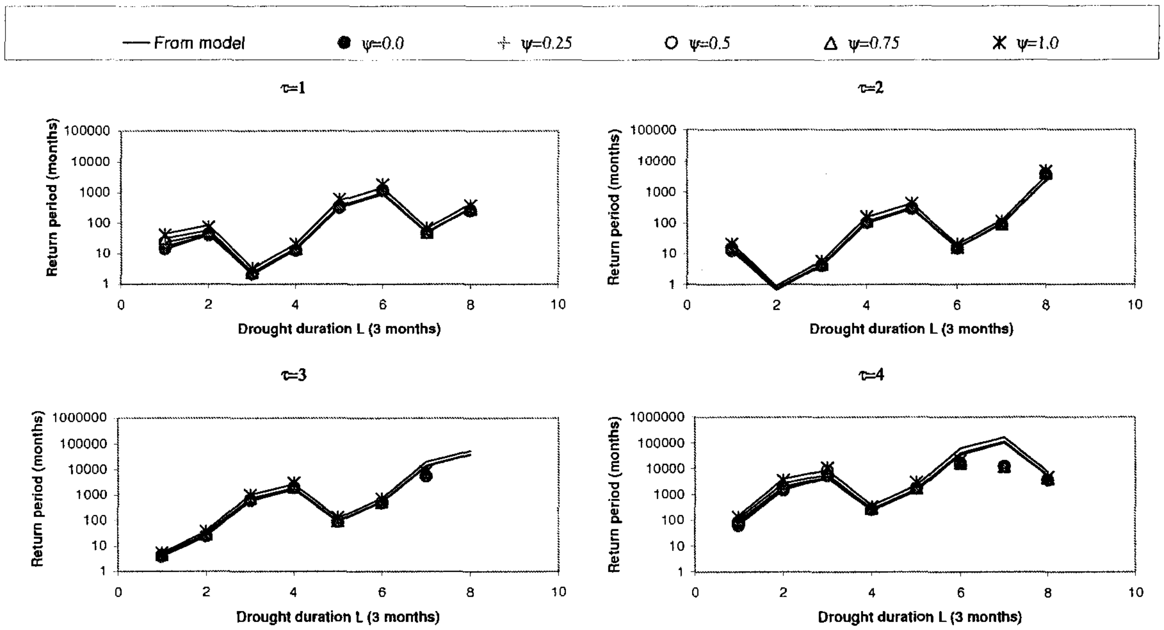


Figure 5.20: Return period of drought events defined by $E_\tau = \{I_\tau > \psi\bar{x}_o$and $L_\tau = l_0$ $(l_{mathit0}=1,2,\ldots)\}$ obtained from the generated synthetic 3-months Caltanissetta precipitation and from Eq. (32) for various values of the intensity coefficient ψ and for fixed beginning season τ.

5.5 Final remarks

Traditionally, major emphasis in analyzing drought properties using analytical methods has been for cases where the underlying stochastic processes are stationary and therefore the annual time scale has been used. For drought management purposes however, and especially for drought monitoring, the annual time scale may reveal to be inadequate as it does not allow to capture in sufficient details the feature of an ongoing droughts. Furthermore, planning and management of water supply systems often requires time scales smaller than the year. Nonetheless, research on drought analysis where the underlying water supply process is periodic-stochastic is lacking.

In this chapter, probability distributions of drought characteristics in periodic series have been investigated both marginally and jointly. In particular, probability density functions of drought characteristics, their associated low order moments, and the return period of drought events have been derived based on the drought definition as in Yevjevich (1967). The derived pdf's allow estimating the occurrence probabilities and return period of droughts where either the drought begins in a given season or regardless of the initial season. The applicability of the drought formulations has been illustrated using six precipitation series assumed as supply series, and using periodic water demand thresholds. The comparison of the empirical probability mass function (pdf) of accumulated deficit obtained from the historical sample and from Eq. (5.28) generally showed a very good agreement between them. The model pdf appears to represent quite well the empirical pdf for all series. Also the comparison between accumulated deficit moments for droughts starting at a given season computed by the model and estimated both on observed and generated sample indicates the general adequacy of the developed expressions.

Furthermore, the return period of droughts with accumulated deficit D greater than a fixed value d for the six data sets investigated were estimated from Eq. (5.65) and from the historical sample. In all cases the results showed an excellent agreement between the two estimates (analytical and historical sample) for smaller values of accumulated deficit, thereafter the sample estimates appear to drift above or below the analytical estimates. This can be explained because as the drought accumulated deficit increases there are smaller numbers of droughts that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time become unreliable. Also the derived expressions for return period of droughts with accumulated deficit or intensity greater than a given value and length equal to a given duration have been applied to the 3-months precipitation series of Caltanissetta for droughts starting at a given season. The results indicate the validity of the proposed expressions, with a good agreement between return periods computed from analytical expressions and those estimated from generated samples.

The overall conclusion of the present chapter is that analytical formulations can be adopted to model drought characteristics in periodic series. Such analytical derivations are particularly useful in cases of short hydrologic records for which the applicability of an inferential approach to find the probability distributions of drought characteristics is either impractical or not feasible. Furthermore, they can be used to extrapolate within reasonable limits information about return period of drought events, beyond the length of the observed series. The examples using six monthly precipitation series illustrate and confirm the applicability of the analytical derivations obtained for accumulated deficit or intensity and associated return periods.

CHAPTER VI

PROBABILISTIC MODELLING OF REGIONAL DROUGHTS

6.1 General

Drought identification and characterization in a region is an important step in planning for drought mitigation and management. While at-site analysis can provide useful information on drought occurrences in a limited area or in a river, regional analysis enables to identify droughts that affect significantly a large region by considering, besides length and accumulated deficit, also the areal extension of the drought (Tase, 1976; Rossi, 2003; Santos, 1983; Rossi and Cancelliere, 2003).

Regional drought analysis differs from the regional flood counterpart since the latter deals with improving the parameter's estimates, while the former has the objective to describe regional features peculiar to droughts such as areal coverage.

Tase (1976) investigated the relationship between area and deficit characteristic of drought events by simulating precipitation in a grid assuming an decaying exponential dependence function with distance. Probabilities of areal coverage of droughts were further investigated analyzing the effects of the size and the shape of the area.

Rossi (1988) investigated regional drought characteristics in Sicily by means of run method and concluded that due to the contemporaneous occurrence of droughts in different area of the island, interbasin water transfer cannot be adopted as a mitigation measure. Features of regional droughts in Colorado have been investigated by Kingery (1992) making use of a Montecarlo simulation approach for generating precipitation and to analyze the corresponding regional drought characteristics. Sen (1980a) Derived the probability distributions of regional drought characteristics assuming independence in space for the precipitation field. Santos (1983) analyzed the probabilistic features of regional drought characteristics, assuming precipitation distributed according to a multivariate normal.

Bayazit and Onoz (2005) derived the pdf and return period of regional drought length assuming a Markovian spatial dependence structure for streamflows in Turkey.

In the present chapter, an attempt to revise and improve Santos (1983) model is proposed. More specifically, some of Santos assumptions are reviewed and alternatives are proposed that lead to an improved formulation. In particular, Santos model relies heavily on the multivariate normal. Although this is not a bad assumption in many cases, however the result presented in this chapter show that the distribution of some drought characteristics is too skewed to be assumed normal. Furthermore, in some cases the drought characteristics are bounded (Tase, 1976). Thus, alternative distributions are proposed in order to improve the results. Also, here return period of alternative critical regional droughts will be derived, using the derived distributions of drought characteristics as building blocks.

As in Santos (1983) throughout the development of the model we will try as close as possible to derive the pdf of drought characteristics based on the statistics of the underlying precipitation.

The chapter is organized as follows. First a review of regional drought characteristics is presented. Then moments and pdf of regional deficit and regional droughts characteristics are derived and the return period formulation originally proposed by Shiau and Shen (2001) is extended to the regional case. The methodology is then applied to two case study, namely Sicily region in the Mediterranean Sea and Eastern Colorado, and results obtained by the model are validated by simulation.

6.2 Regional drought characteristics

The method of run, originally devised for at-site drought identification (Yevjevich, 1967), has been extended by Tase (1976) to the case of regional droughts, by considering series of the variable of interest (typically precipitation) observed at several stations and selecting, besides the truncation level at each site, an additional threshold, which represents the value of the area affected by deficit above which a regional drought is considered to occur. Thus, for a given interval t, two indices are computed, representative of the areal extension and amount of the deficits over the investigated region.

More specifically, let's assume that in a given area of interest there are m precipitation stations. Each station k can be characterized by the corresponding area of the Thiessen's polygon s_k, which can also be expressed as a fraction a_k of the total area as:

$$a_k = \frac{s_k}{\sum_{j=1}^{m} s_j} \tag{6.1}$$

In what follows, the following notation will be adopted: $X_{k,t}$ will indicate the precipitation at station k at time t, x_{ok} will indicate the threshold level at station k. Then the following indicator variable can be defined for station k:

$$I_{k,t} = 0 \;\; if \;\; X_{k,t} \leq x_{ok}$$
$$I_{k,t} = 1 \;\; if \;\; X_{k,t} > x_{ok} \tag{6.2}$$

Obviously $I_{k,t}$ is a Bernoulli r.v. with probability $p_k = P[X_{k,t} \leq x_{ok}]$ of being 1. Furthermore, $I_{k,t} = 1$ indicates that a deficit occurs at time t at site k, i.e. $X_{k,t} \leq x_{ok}$.

The fraction of the area affected by deficit in a given interval t will be termed here as areal coverage of deficit Ad_t and is defined as:

$$Ad_t = \sum_{j=1}^{m} a_j I_{j,t} \tag{6.3}$$

The regional extension of the run method assumes that in interval t a regional deficit occurs if the areal coverage of deficit exceeds a fixed threshold a_{crit} (Tase, 1976). Accordingly, the areal deficit D_t can be computed as:

$$D_t = \sum_{j=1}^{m} a_k I_{k,t}(x_{ok} - X_{k,t}) \;\; \text{if} \;\; Ad_t \geq a_{crit} \tag{6.4}$$

The Ad_t index is a measure of the area affected by deficit, expressed as a fraction of the total area and it ranges between 0 and 1. The D_t index provides some insight on the total amount of the deficit in the area being basically a sum of the deficits at each sites, weighted by the corresponding influence areas, computed for example by Thiessen polygons.

Then, regional drought is defined as an uninterrupted sequence of regional deficits preceded and succeeded by surpluses. A regional drought can be characterized by its duration L:

$$L = t_f - t_i + 1 \tag{6.5}$$

where t_i and t_f are such that $D_t > 0$ for $t_i \leq t \leq t_f$ and $D_{t_i-1} = 0$, $D_{t_f+1} = 0$.

Regional accumulated deficit is computed as:

$$DC = \sum_{t=t_i}^{t_f} D_t \tag{6.6}$$

while the regional drought intensity is given by:

$$ID = \frac{DC}{L} \tag{6.7}$$

Finally the mean areal coverage of drought can be computed by:

$$AD = \frac{1}{L} \sum_{t=t_i}^{t_f} Ad_t \tag{6.8}$$

It should be noted that extension of the definitions to other hydrological variables besides precipitation such as streamflows is straightforward. For instance, one may be interested in analyzing droughts occurring at several streamflow sites, whose water supply a given water system. The only required modification to the methodology is in the definition of the weights a_c assigned to each series, since in this case they may for example represent weighting factors assigned to each streamflow source, as a function of the senior/junior water rights structure.

6.3 Moments and pdf's of regional drought characteristics

It is of interest here to derive the moments and the pdf of the regional drought characteristics based on the distribution of precipitation $X_{k,t}$. To better illustrate the assumptions made during the derivation of such moments and pdf, we will occasionally refer to results related to the case studies presented in details later in the chapter.

More specifically we will assume that at a given time interval t the $X_{k,t}$ will be jointly distributed according to a multivariate normal with vector mean $\underline{\hat{\mu}}$ and variance-covariance matrix Σ (Kotz et al., 2000):

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,m} \\ \gamma_{2,1} & \sigma_2^2 & \gamma_{2,3} & \cdots & \gamma_{2,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{m,1} & \gamma_{m,2} & \gamma_{m,3} & \cdots & \sigma_m^2 \end{pmatrix} \tag{6.9}$$

where $\sigma_k^2 = \mathrm{Var}\left[X_{k,t}\right]$ and $\gamma_{j,k} = \mathrm{Cov}\left[X_{j,t}, X_{k,t}\right]$.

For easy of notation we introduce the r.v. $Y_{k,t} = a_k(x_{ok} - X_{k,t})$ which represents the deviation of areal precipitation at station k and time t from the threshold. Since $X_{k,t}$ is multivariate normal it it follows $Y_{k,t} \sim \mathrm{N}(a_k(x_{ok} - \mu_k), a_k^2\sigma_k^2)$. Further, we can introduce the r.v.:

$$Z_t = \sum_{k=1}^{m} Y_{k,t} I_{k,t} = \sum_{k=1}^{m} a_k(x_{ok} - X_{k,t}) I_{k,t} \tag{6.10}$$

which represents the sum of precipitation deficits at time t, weighted by the influence areas a_k. The expected value of Z_t has been derived by Santos (1983) as:

$$\mathcal{E}\left[Z_t\right] = \sum_{k=1}^{m} \left[x_{ok} - \left[\mu_k - \frac{\sigma_k}{p_k}\phi\left(\frac{x_{ok} - \mu_k}{\sigma_k}\right)\right]\right] a_k p_k \tag{6.11}$$

where in usual notation $\phi()$ stands for the standard normal pdf. Also, given the normality of $X_{k,t}$, the probability p_k can be computed as $p_k = \Phi\left(\frac{x_{ok}-\mu_k}{\sigma_k}\right)$.

The variance has also been derived by Santos (1983) as:

$$\text{Var}\left[Z_t\right] = \sum_{k=1}^{m}\sum_{j=1}^{m} \text{Cov}\left[Y_{k,t}, Y_{j,t}|Y_{k,t} > 0, Y_{j,t} > 0\right] p_{k,j} \tag{6.12}$$

where $p_{k,j} = P[Y_{k,t} > 0, Y_{j,t} > 0]$.

The conditional covariance in Eq. (6.12) is the covariance of truncated bivariate normal r.v.. The moment generating function and the cumulants of the truncated multivariate normal distribution have been investigated by Tallis (1961) and Finney (1962) respectively. Kotz et al. (2000) provide analytical expressions to compute the first two moments of a truncated bivariate normal distribution. Further, the algorithm MVNTRC (Leppard and Tallis, 1989) is available which computes numerically expectations and covariances of truncated multivariate normals. On the other hand, the joint bivariate probability in Eq. (6.12)can be computed numerically for example by means of the algorithm MULNOR (Shervish, 1984).

Exact derivation of the cdf of Z_t is not a simple task since Z_t is the sum of truncated normal variables. Bonaccorso et al. (2003), approximated the truncated normal by means of a gamma distribution. Using the same approximation, Z_t will be the sum of gamma variables,and therefore gamma itself, with cdf:

$$f_{Z_t}(z) = \frac{1}{\beta_z \Gamma(r_z)} e^{-\frac{z}{\beta_z}} \left(-\frac{z}{\beta_z}\right)^{r_z - 1} \tag{6.13}$$

where the parameters r_z, β_z can be expressed in terms of the moments of Z_t given in Eqs. (6.11) and (6.12) as:

$$r_z = \frac{\mathcal{E}\left[Z_t\right]^2}{\text{Var}\left[Z_t\right]} \tag{6.14}$$

$$\beta_z = \frac{\mathrm{Var}\left[Z_t\right]}{\mathcal{E}\left[Z_t\right]} \tag{6.15}$$

The first two moments of areal coverage of deficits have alredy been derived by Santos (1983) as:

$$\mathcal{E}\left[Ad_t\right] = \mathcal{E}\left[\sum_{k=1}^{m} a_k I_{k,t}\right] = \sum_{k=1}^{m} a_k p_k \tag{6.16}$$

and:

$$\mathrm{Var}\left[Ad_t\right] = \sum_{j=1}^{m} a_k p_k (1 - p_k) + 2 \sum_{k=1}^{m} \sum_{j=k+1}^{m} a_k a_j (p_{k,j} - p_k p_j) \tag{6.17}$$

Derivation of the exact distribution of the areal coverage of deficit is cumbersome except for simple cases. For example, Santos (1983) assumed Ad_t normally distributed since it is the sum of r.v.'s therefore by virtue of the central limit theorem its asymptotic distribution will be normal. Such normality assumption however is not realistic, since the areal coverage of deficit Ad_t is bounded between 0 and 1 (Tase, 1976). Following Tase (1976), here a beta distribution is adopted (Johnson et al., 1994):

$$f_{Ad_t}(a) = \frac{1}{B(p,q)} a^{p-1} (1-a)^{q-1} \quad (0 \leq a \leq 1) \tag{6.18}$$

where B(p,q)is the complete beta function $\int_0^{\infty} w^{p-1}(1-w)^{q-1} dw$. In our case, the parameters p, q can be estimated as a function of the first two moments $\mu_A = \mathcal{E}\left[Ad_t\right]$ and $\sigma_A^2 = \mathrm{Var}\left[Ad_t\right]$ given by eqs. 6.16 and 6.17 as (Johnson et al., 1994):

$$p = \mu_A^2 - \frac{1 - \mu_A}{\sigma_A^2} - \mu_A \tag{6.19}$$

$$q = \mu_A - \frac{1 - \mu_A}{\sigma_A^2} - 1 - p \tag{6.20}$$

Santos (1983) derived an expression for the expectation of the areal deficit D_t under the assumtpion that Z_t and Ad_t are jointly normally distributed but she did not derive an explicit expression for the variance. Nonetheless, the hypothesis of Z_t and Ad_t distributed according to a bivariate normal is not realistic, due to the skewness of Z_t and to the fact that Ad_t is bounded. Therefore, a different approach will be followed here to derive the distribution of areal deficit D_t and its moments. More specifically, we will start by observing that by definition the Cumulative Distribution Function (cdf) of D_t is given by the following conditional cdf:

$$F_{D_t}(d) = F_{Z_t|Ad_t>a_c}(d) \tag{6.21}$$

Derivation of the conditional cdf on the l.h.s. of eq. (6.21) requires the definition of the bivariate cdf of Z_t and Ad_t and of the marginal of Ad_t However, exact derivation of such bivariate distribution is quite complicated therefore an approach based on copulas will be followed here. Copulas are mathematical tools that enable to "build" multivariate distributions starting from known marginals.

Several copulas have been proposed by different authors. Among the several copulas proposed, Frank's copula has been adopted here (Frank, 1979), after the Ali-Mikhail-Haq copula did not give satisfactory results. With reference to two generic r.v. U and V with marginal cdf's $F_U(u)$ and $F_V(v)$ respectively, Frank's bivariate copula assumes that a bivariate cdf can be built as:

$$F_{U,V}(u,v) = \alpha^{-1} ln[1 + (e^{\alpha F_U(u)} - 1)(e^{\alpha F_V(v)} - 1)(e^{\alpha} - 1)^{-1}] \tag{6.22}$$

where α is a parameter that models the dependence between U and V. In our case, a bivariate cdf of Z_t and Ad_t can be written as:

$$F_{Z_t,Ad_t}(z,a) = \alpha^{-1} ln[1 + (e^{\alpha F_{Z_t}(z)} - 1)(e^{\alpha F_{Ad_t}(a)} - 1)(e^{\alpha} - 1)^{-1}] \tag{6.23}$$

Although in principle the dependence parameter α could be estimated from an available sample of Z_t and Ad_t, it may be preferable to estimate it by method of moments making use of the covariance between Z_t and Ad_t which in turn can be derived as a function of the underlying precipitation statistics. More specifically, the covariance between Z_t and Ad_t can be computed as:

$$\mathrm{Cov}\left[Z_t, Ad_t\right] = \mathrm{Cov}\left[\sum_{k=1}^{m} Y_{k,t} I_{k,t}, \sum_{j=1}^{m} a_j I_{j,t}\right] = \sum_{k=1}^{m}\sum_{j=1}^{m} a_j \mathrm{Cov}\left[Y_{k,t} I_{k,t}, I_{j,t}\right] \tag{6.24}$$

The covariance in the r.h.s. can be rewritten as:

$$\begin{aligned} \mathrm{Cov}\left[Y_{k,t} I_{k,t}, I_{j,t}\right] &= \mathcal{E}\left[Y_{k,t} I_{k,t} I_{j,t}\right] - \mathcal{E}\left[Y_{k,t} I_{k,t}\right] \mathcal{E}\left[I_{j,t}\right] \\ &= \mathcal{E}\left[Y_{k,t} | I_{k,t} = 1, I_{j,t} = 1\right] p_{k,j} - \mathcal{E}\left[Y_{k,t} | I_{k,t=1}\right] p_k p_j \end{aligned} \tag{6.25}$$

where the conditional expectations can be computed as before.

Once the the bivariate cdf of Z_t and Ad_t is known, the cdf of D_t follows as:

$$F_{D_t}(d) = F_{Z_t|Ad_t > a_c}(d) = \frac{F_{Z_t}(d) - F_{Z_t,Ad_t}(d, a_c)}{1 - F_{Ad_t}(a_c)} \tag{6.26}$$

where $F_{Ad_t}()$ is the marginal cdf of Ad_t

Then, the first two moments of areal deficit D_t can be computed. Instead of taking derivatives of Eq. (6.26) and carrying out the necessary integrations, we will make use of an alternative definition of expected value and variance based only on the cdf (Mood et al., 1974):

$$\mathcal{E}\left[D_t\right] = \int_0^{\infty} [1 - F_{D_t}(z)]dz \tag{6.27}$$

$$\operatorname{Var}\left[D_t\right] = \int_0^{\infty} 2z[1 - F_{D_t}(z)]dz \tag{6.28}$$

Although in principle eq. 6.26 completely defines the distribution of D_t, for what follows it is convenient to approximate such distribution by a gamma distribution with pdf:

$$f_{D_t}(d) = \frac{1}{\beta\Gamma(r)} e^{-\frac{d}{\beta}} \left(-\frac{d}{\beta}\right)^{r-1} \tag{6.29}$$

where the parameters r, β can be expressed in terms of the moments of D_t given in Eqs. (6.27) and (6.27) as:

$$r = \frac{\mathcal{E}\left[D_t\right]^2}{\operatorname{Var}\left[D_t\right]} \tag{6.30}$$

$$\beta = \frac{\operatorname{Var}\left[D_t\right]}{\mathcal{E}\left[D_t\right]} \tag{6.31}$$

Such gamma approximation will enable to derive in a straighforward fashion the cdf of drought characteristics accumulated deficit and intensity.

The adequacy of the above approximations has been assessed by comparing the copula approximated cdf given by Eqs. (6.26), the gamma approximation given by eq. (6.29) and the empirical cdf (estimated by means of Weibull Plotting Position formula) of D_t computed from generated data. The details of such generation will be illustrated in the application paragraph.

In Figure 6.1, the results of such a comparison are shown. From the figure it can be inferred that the copula approximation fits well the lower and upper part of the empirical cdf, while in the central part, the differences are noticeable. Overall

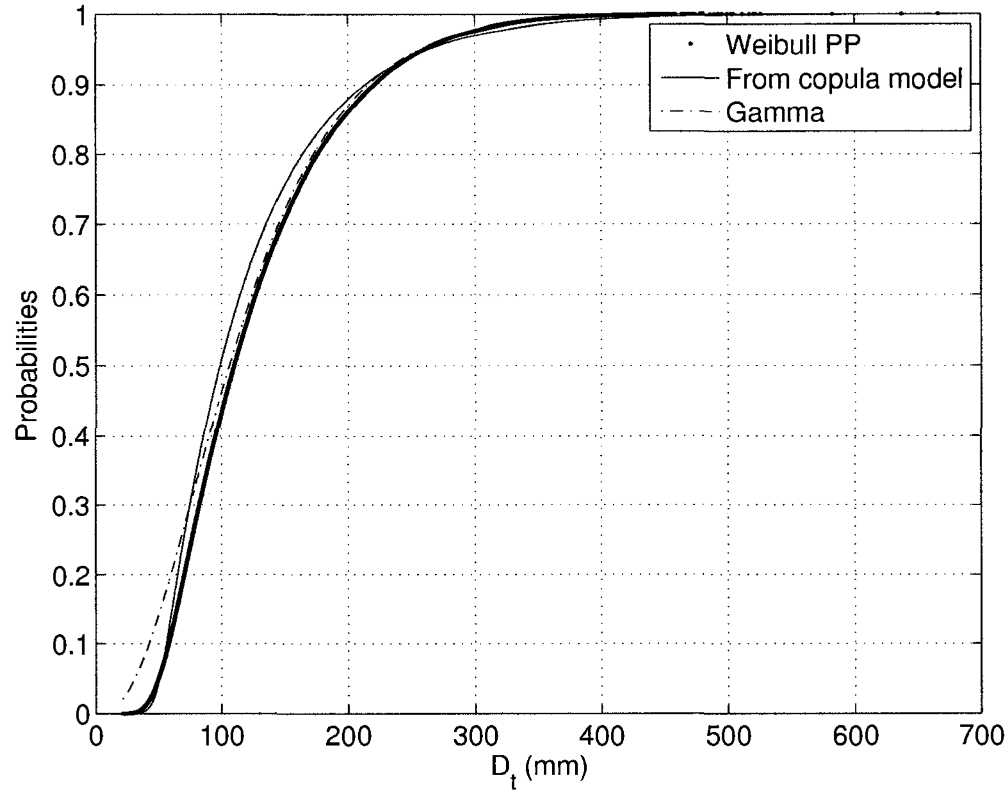


Figure 6.1: Comparison among cdf approximated by copula given by Eqs. (6.26), the gamma approximation given by eq. (6.29) and the empirical cdf (estimated by means of Weibull Plotting Position formula) of D_t computed from generated data.

however, the gamma approximation seems to better fit the empirical cdf, especially for quantiles above 50%.

6.4 Moments and pdf of regional drought characteristics

From a formal standpoint, a regional drought differs from an at-site drought only in what concerns the definition of deficit. At-site deficit occurs whenever the variable of interest is below the threshold (Yevjevich, 1967), whereas, in the regional case, a regional deficit occurs whenever the areal coverage of deficit Ad_t exceeds a critical threshold a_{crit} (Tase, 1976). Once a regional deficit occurs, the regional characteristics are computed in a very similar way as in the at site analysis. Therefore, probabilistic formulation of regional drought characteristics can inherit most of the results of the at-site case. By definition, a regional deficit occurs when the areal coverage of deficit Ad_t exceeds the areal threshold a_c. Thus, the occurrence probability pd of a regional deficit is given by:

$$pd = P[Ad_t > a_{crit}] \tag{6.32}$$

The above probability can be computed by integrating the beta pdf of Ad_t, given in equation 6.18.

As in the at site case, under the hypothesis of independence in time for $X_{k,t}$ the regional drought duration will be distributed according to a geometric distribution with parameter $1 - pd$ (Downer et al., 1967), i.e.:

$$f_L(l) = P[L = l] = (1 - pd)pd^{l-1} \qquad l = 1, 2, \ldots \tag{6.33}$$

since the occurrence of regional deficits is a sequence of Bernoulli trials.

The moments of L follow as:

$$\mathcal{E}\,[L] = \frac{1}{1 - pd} \tag{6.34}$$

and

$$\mathrm{Var}\,[L] = \frac{pd}{(1 - pd)^2} \tag{6.35}$$

The expected value of the regional accumulated deficit DC can be derived from eq. (6.6) as:

$$\mathcal{E}\,[DC] = \mathcal{E}\,[D_t]\,\mathcal{E}\,[L] = \frac{\mathcal{E}\,[D_t]}{1 - pd} \tag{6.36}$$

while the variance takes the following form:

$$\operatorname{Var}[DC] = \operatorname{Var}[D_t]\,\mathcal{E}[L] + \operatorname{Var}[L]\,\mathcal{E}[D_t]^2 = \frac{\operatorname{Var}[D_t]}{1-pd} + \frac{pd\;\mathcal{E}[D_t]^2}{(1-pd)^2} \tag{6.37}$$

where the expected value and variance of D_t have already been derived in Eqs. (6.27) and (6.28).

Before tackling the marginal distribution of DC, it may be of some interest to derive an approximate expression for the joint pdf of regional drought length and accumulated deficit.

The regional accumulated deficit DC given by Eq. (6.6) is a sum of a random number of random variables, i.e. DC is the sum of L successive regional deficits D_t, where the regional drought duration L is a random variable. To derive its distribution, a scheme proposed by Salas et al. (2005) for the at site drought case will be extended to the regional case. More specifically, first, the derivation of the bivariate probability distribution function (pdf) of deficit and length can be pursued by considering the distribution of accumulated deficit conditioned on a fixed drought length $L = l$. Then the bivariate pdf of accumulated deficit D and length L can be expressed as:

$$f_{DC,L}(d,l) = f_{DC|L=l}(d) f_L(l) \tag{6.38}$$

where $f_{DC|L=l}$ is the conditional pdf of DC given $L = l$ and is the marginal pdf of L. The r.v. $DC|L = l$ is the sum of exactly l regional deficits D_t, which have been assumed i.i.d gamma distributed with parameters r, β given by eqs. (6.30) and (6.31). Since $DC|L = l$, is the sum of l gamma variates, it will also be gamma distributed, with parameters (lr, b)(Mood et al. (1974); Guven (1983); Shiau and Shen (2001)). On the other hand, it has already been shown that the pmf of regional drought length is geometric (see Eq. (6.33)). Combining these results, the following expression for the bivariate distribution of DC and L is obtained:

$$f_{DC,L}(d,l) = \frac{1}{\beta\Gamma(l\ r)} e^{-\frac{d}{\beta}} \left(-\frac{d}{\beta}\right)^{l\ r-1} (1-pd)pd^{l-1} \tag{6.39}$$

Furthermore, from the conditional distribution of regional accumulated deficit given drought length (Eq. 6.39), the conditional distribution of regional drought intensity I given length L can be also derived. Indeed, since the drought intensity is the ratio of accumulated deficit to drought length, i.e. $I{=}D/l$, the conditional pdf of intensity I given a fixed length $L{=}L$ can be derived from Eq. 6.39 as

$$f_{I|L}(i) = \frac{l}{\beta\Gamma(r)} \left(\frac{l\ i}{\beta}\right)^{r-1} \exp\left(-\frac{li}{\beta}\right) \tag{6.40}$$

which is also gamma distributed, i.e. $G(r,\beta)$. Thus, the bivariate pdf of intensity and length can be found in a similar fashion as in (6.38) above. It follows:

$$f_{I,L}(i,l) = \frac{l}{\beta\Gamma(r)} \left(\frac{l\ i}{\beta}\right)^{r-1} \exp\left(-\frac{li}{\beta}\right) f_L(l) \tag{6.41}$$

Therefore by integrating appropriately the bivariate pdf's, the occurrence probability of various drought events can be found. In particular, with reference to different types of critical drought events, the following expressions can be found (Salas et al., 2005):

(1) for drought event $E = \{DC{>}D_0$ and $L{=}\ l_0\ (l_0{=}1{,}2{,}\ldots)\}$:

$$P[DC > D_o, L = l_o] = \int_{D_0}^{\infty} f_{DC,L}(z,l_0)dz = \int_{D_0}^{\infty} \frac{1}{\beta\Gamma(l_0 r)} \left(\frac{z}{\beta}\right)^{l_0 r-1} e^{-\frac{z}{\beta}} f_L(l)dz \tag{6.42}$$

(2) for drought event $E = \{DC{>}D_0$ and $L \geq l_0$ (l_0=1,2,...)$\}$:

$$P[DC{>}D_0, L \geq l_0] = \int_{D_0}^{\infty} \sum_{l=l_0}^{\infty} f_{DC,L}(z,l)dz = \int_{D_0}^{\infty} \sum_{l=l_0}^{\infty} \frac{1}{\beta\Gamma(lr)} \left(\frac{z}{\beta}\right)^{lr-1} e^{-\frac{z}{\beta}} f_L(l)dz \tag{6.43}$$

(3) for drought event $E = \{I > I_0$ and $L = l_0$ (l_0=1,2,...)$\}$:

$$P[I{>}I_0, L = l_0] = \int_{I_0}^{\infty} f_{I,L}(z,l_0)dz = \int_{I_0}^{\infty} \frac{l}{\beta\Gamma(l_0 r)} \left(\frac{l_0\, z}{\beta}\right)^{l_0 r-1} e^{-\frac{l_0\, z}{\beta}} f_L(l)dz \tag{6.44}$$

(4) for drought event $E = \{I > I_0$ and $L \geq l_0$ (l_0=1,2,...)$\}$:

$$P[I{>}I_0, L \geq L_0] = \int_{I_0}^{\infty} \sum_{l=l_0}^{\infty} f_{I,L}(z,l)dz = \int_{I_0}^{\infty} \sum_{l=l_0}^{\infty} \frac{l}{\beta\Gamma(lr)} \left(\frac{l\, z}{\beta}\right)^{lr-1} e^{-\frac{l\, z}{\beta}} f_L(l)dz \tag{6.45}$$

Furthermore, the marginal probability of droughts events $DC > D_0$ or $I > I_0$ can be obtained from Eqs.(6.43) or (6.45), respectively, by letting l_0=1, thus yielding:

$$P[D{>}D_0, L \geq l_0] = \int_{D_0}^{\infty} \sum_{l=1}^{\infty} f_{D,L}(z,l)dz = \int_{D_0}^{\infty} \sum_{l=1}^{\infty} \frac{1}{\beta\Gamma(lr)} \left(\frac{z}{\beta}\right)^{lr-1} e^{-\frac{z}{\beta}} f_L(l)dz \tag{6.46}$$

$$P[I{>}I_0, L \geq 1] = \int_{I_0}^{\infty} \sum_{l=1}^{\infty} f_{I,L}(z,l)dz = \int_{I_0}^{\infty} \sum_{l=1}^{\infty} \frac{l}{\beta\Gamma(lr)} \left(\frac{l\, z}{\beta}\right)^{lr-1} e^{-\frac{l\, z}{\beta}} f_L(l)dz \tag{6.47}$$

Alternatively, one can assume regional accumulated deficit DC to be marginally gamma distributed with parameters r_D, β_D. Such parameters can be linked to the moments of DC as (Mood et al., 1974):

$$r_D = \frac{\mathcal{E}\left[DC\right]^2}{\mathrm{Var}\left[DC\right]} \tag{6.48}$$

$$\beta_D = \frac{\mathrm{Var}\left[DC\right]}{\mathcal{E}\left[DC\right]} \tag{6.49}$$

where the moments $\mathcal{E}\left[DC\right]$ and $\mathrm{Var}\left[DC\right]$ are given by Eqs. (6.36) and (6.37) respectively. The pdf of regional accumulated deficit is therefore:

$$f_{DC}(d) = \frac{1}{\beta\Gamma(r_D)} e^{-\frac{d}{\beta_D}} \left(-\frac{d}{\beta_D}\right)^{r_D - 1} \tag{6.50}$$

As already mentioned, the regional drought definition differs substantially from the corresponding at-site one only in what regards the definition of the regional deficit. It follows that, in order to derive an expression for the return period of critical regional droughts, the scheme similar to the one developed for the at-site case can be applied (Shiau and Shen, 2001; Bonaccorso et al., 2003; Gonzalez and Valdes, 2003; Salas et al., 2005).

For analyzing the return period of droughts we need to specify the regional drought event under consideration. In what follows we will refer to a generic regional critical drought E, i.e. a drought that is critical with respect to one or more of its characteristics. For example E might indicate a drought with regional accumulated deficit DC greater or equal to a fixed amount or a regional drought that exceed a given duration. Likewise, a critical regional drought E might be a drought whose areal extent exceed a fixed percentage of the whole region.

Then, adopting the same scheme of Salas et al. (2005), let us denote by $\{E\}$ a critical drought event and by $\{\bar{E}\}$ the corresponding non critical drought events. For example, if $\{E\} = \{DC > DC_0, L \geq l_0\}$, the corresponding non drought events $\{\bar{E}\}$ will be $\{DC \leq D_0, L \geq l_0\}$, $\{DC > D_0, L < l_0\}$, and $\{DC \leq D_0, L < l_0\}$.

Assuming independence between drought events, the following expression for return period of droughts with characteristics E can be derived (Salas et al., 2005):

$$T(E) = \frac{\mathcal{E}\left[L\right] + \mathcal{E}\left[Ln\right]}{P[E]} \tag{6.51}$$

where $\mathcal{E}\left[L\right]$ is the average duration of a critical drought event $\{E\}$, and $\mathcal{E}\left[Ln\right]$ is the duration of a non-drought event.

Equation (6.51) allows one to compute the return period of any critical drought $\{E\}$ once the probability $P(E)$ is defined and the expected value of drought and non-drought durations are known. The former probability can be computed by making use of the joint pdf's of regional drought characteristics derived previously. The expected value of regional drought length L is given by Eq. (6.34) whereas it is easy to see that the non-drought length is distributed according to a geometric distribution with parameter pd, and therefore its expected value will be:

$$\mathcal{E}\left[Ln\right] = \frac{1}{pd} \tag{6.52}$$

Combining Eq. (6.51) with Eqs. (6.34) and (6.52) the return period of regional critical drought finally becomes:

$$T = \frac{1}{pd(1 - pd)P[E]} \tag{6.53}$$

6.5 Application of the regional model

Application of the regional model has been carried out with reference to two regions with different areal extension and climatic characteristics, namely Sicily region, Italy and Eastern Colorado, USA.

Sicily is one of the largest islands in the Mediterranean sea and with a surface of approx. 25,000 km^2. The climate is semiarid, with mean annual precipitation around 700 mm and an elevated intra-annual variability as well as from year to year. Drought analysis in Sicily has been carried out with reference to 43 precipitation stations located in the island. Available precipitation data include annual precipitation covering the period 1921-2000 with occasional missing years that have been filled by means of linear regression technique.

The investigated region in Colorado is the plain located East of the Front Range and covers an area of 114,000 km^2. The climate is semiarid with an mean annual precipitation in the order ot 375 mm. Colorado precipitation data used in this study is part of the National Weather Service cooperative network weather station database (source http:// cdo.ncdc.noaa.gov/ CDO/ dataproduct). The original dataset for the whole state included more than 500 stations covering the period 1885-2003, with very few stations with available observations before 1931. Out of the original dataset, 72 stations have been selected with the criteria of being located east of the Front Range in Colorado, and of having at least 30 years of observations after 1931. The resulting dataset has been completed by filling occasionally missing yearly observations by linear regression.

The location of the stations along with the related Thiessen's polygons are shown in Figure 6.2 and 6.3 for Sicily and Colorado respectively whereas the main statistical characteristics of annual precipitation are reported in Tables 6.I and 6.II.

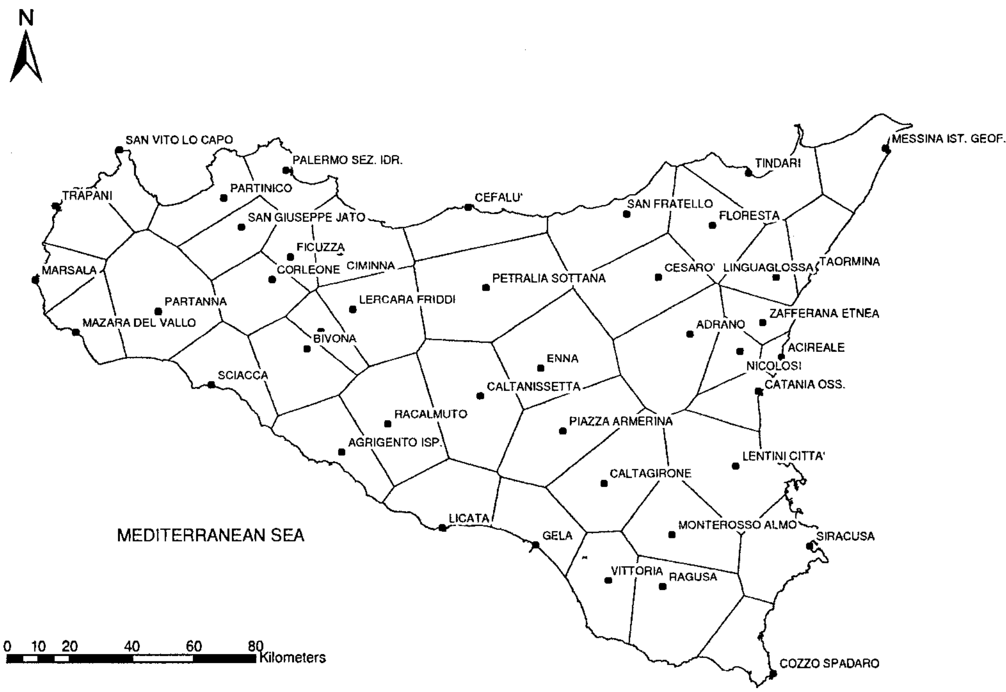


Figure 6.2: Localization of the investigated 43 precipitation stations in Sicily and related Thiessen's polygons

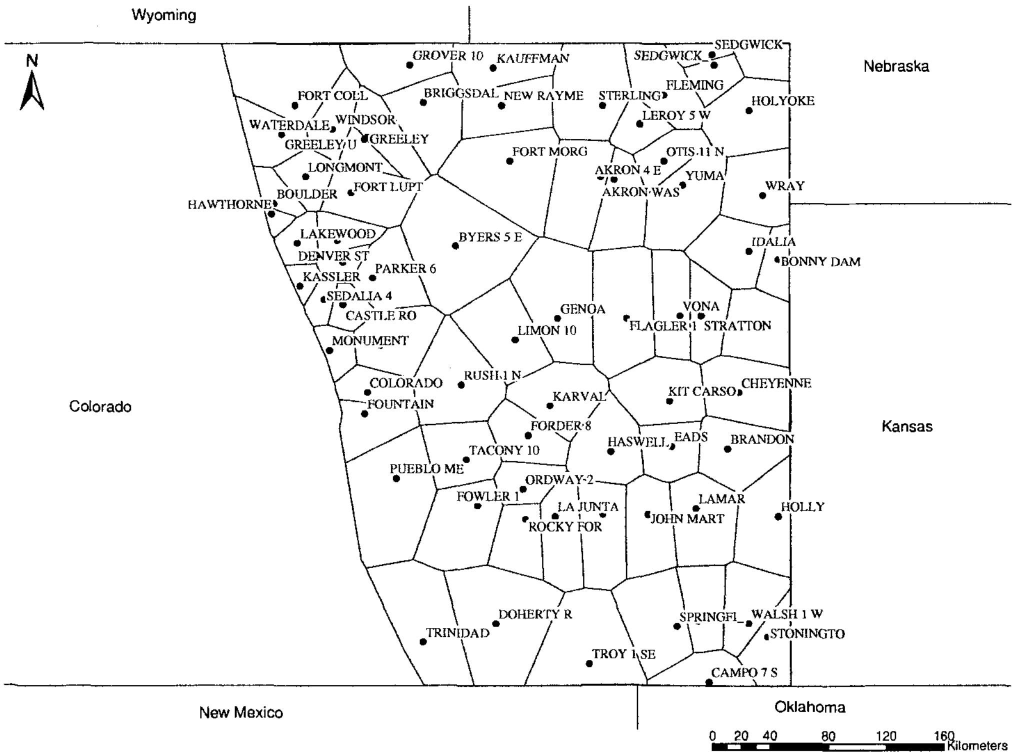


Figure 6.3: Localization of the investigated 72 precipitation stations in Eastern Colorado and related Thiessen's polygons

Table 6.I: Main statistical characteristics of investigated stations in Sicily

Station name	Elevation(m)	Mean(mm)	Std(mm)	Skewness
Tindari	280	693.4	135.0	0.63
San Fratello	690	928.6	185.4	0.56
Cefalu'	30	684.1	156.8	0.75
Ciminna	525	635.2	157.1	0.51
Ficuzza	681	830.8	203.4	0.46
Partinico	189	690.4	162.5	0.14
Palermo Sez. Idr.	19	628.8	152.1	1.39
San Giuseppe Jato	450	777.6	174.1	0.26
Trapani	2	469.6	123.1	0.32
San Vito Lo Capo	6	491.0	122.8	0.43
Marsala	12	498.6	123.3	0.05
Mazara Del Vallo	8	518.2	152.1	0.98
Partanna	407	680.0	184.1	0.62
Corleone	588	731.7	173.6	0.30
Sciacca	56	555.8	148.1	0.90
Piano Del Leone	831	826.2	183.8	0.27
Bivona	521	815.6	200.5	1.25
Lercara Friddi	658	593.9	137.1	0.66
Racalmuto	450	675.7	164.1	0.08
Agrigento Isp.	175	498.0	141.3	0.61
Petralia Sottana	930	805.4	181.5	0.81
Enna	950	842.1	358.8	1.59
Caltanissetta	375	541.3	144.6	0.66
Licata	70	453.2	165.4	1.47
Gela	30	408.4	118.0	0.24
Piazza Armerina	721	684.8	183.1	0.15
Monterosso Almo	691	706.4	187.8	0.10
Vittoria	168	502.2	133.5	0.35
Ragusa	515	723.3	208.3	0.51
Cozzo Spadaro	50	410.3	165.0	0.85
Siracusa	23	557.0	194.0	0.35
Lentini Citta'	43	644.4	237.5	0.52
Cesaro'	1100	755.3	156.5	0.06
Adrano	589	497.2	145.1	0.52
Caltagirone	513	542.0	152.4	0.47
Nicolosi	695	1121.9	464.4	1.55
Zafferana Etnea	590	1336.2	476.7	0.55
Linguaglossa	530	1119.5	371.6	0.87
Acireale	194	813.9	254.4	0.28
Catania Oss.	75	682.4	300.3	1.56
Floresta	1270	1236.3	277.3	0.44
Taormina	248	759.6	216.5	0.56
Messina Ist. Geof.	54	764.2	147.6	0.40

Table 6.II: Main statistical characteristics of investigated stations in Eastern Colorado

Station name	Coop ID	Elevation (m)	Mean(in)	Std(in)	Skewness
AKRON 4 E	50109	1384	16.47	3.04	0.052
AKRON WAS	50114	1421	16.61	3.39	0.005
BONNY DAM	50834	1133	17.06	3.82	-0.209
BOULDER	50848	1672	18.90	4.55	0.449
BRANDON	50895	1196	13.57	3.38	0.221
BRIGGSDAL	50945	1473	13.06	3.26	0.480
BURLINGTON	51121	1291	16.12	4.42	0.323
BYERS 5 E	51179	1555	14.88	3.87	-0.005
CAMPO 7 S	51268	1255	16.55	3.75	0.244
CASTLE ROCK	51401	1936	17.13	3.35	0.879
CHERRY CREEK	51547	1721	17.15	4.51	0.246
CHEYENNE	51564	1295	15.39	4.11	0.347
COLORADO	51778	1884	16.18	4.17	0.409
DENVER ST	52220	1611	15.59	4.29	0.271
DOHERTY R	52312	1565	13.64	3.24	0.474
EADS	52446	1285	14.72	3.94	-0.097
EASTONVILLE	52494	2198	18.60	4.48	0.474
FLAGLER 1	52932	1500	15.69	3.54	0.116
FLEMING	52944	1292	17.76	3.53	-0.138
FORDER 8	52997	1458	11.66	3.20	0.484
FORT COLLINS	53005	1525	14.96	4.21	0.617
FORT LUPTON	53027	1531	12.39	3.14	0.191
FORT MORGAN	53038	1320	12.92	2.88	-0.166
FOUNTAIN	53063	1695	16.06	5.23	0.666
FOWLER 1	53079	1320	11.00	3.47	0.445
GENOA	53258	1709	15.70	3.67	0.306
GREELEY	53546	1418	11.68	2.66	0.408
GREELEY U	53553	1437	13.76	2.93	0.467
GROVER 10	53643	1549	13.80	3.47	0.367
HASWELL	53828	1379	13.29	3.84	0.154
HAWTHORNE	53850	1806	21.06	4.49	0.213
HOLLY	54076	1033	15.70	4.70	0.451
HOLYOKE	54082	1137	17.73	4.27	0.014
IDALIA	54242	1209	18.26	4.06	0.292
JOHN MART	54388	1163	12.29	3.35	0.507
JULESBURG	54413	1057	16.82	3.88	-0.017
KARVAL	54444	1547	13.67	4.02	0.581
KASSLER	54452	1703	17.63	4.10	0.427
KAUFFMAN	54460	1600	13.64	2.91	0.193

Table 6.II: (cont.) Main statistical characteristics of investigated stations in Eastern Colorado

Station name	Coop ID	Elevation (m)	Mean(in)	Std(in)	Skewness
KIT CARSON	54603	1317	14.50	4.19	0.543
LA JUNTA	54720	1278	11.47	3.23	0.181
LAKEWOOD	54762	1719	16.36	3.88	-0.040
LAMAR	54770	1106	14.85	3.90	0.315
LAS ANIMAS	54834	1186	12.47	3.47	0.511
LEROY 5 W	54945	1363	17.66	3.61	0.167
LIMON 10	55015	1696	14.76	3.69	0.610
LONGMONT	55116	1509	13.09	3.59	0.235
MONUMENT	55730	2239	18.89	3.59	0.515
NEW RAYME	55922	1458	15.18	4.45	0.538
ORDWAY 2	56131	1315	11.06	3.55	0.709
OTIS 11 N	56192	1274	14.91	3.62	0.026
PARKER 6	56326	1923	14.04	3.55	-0.103
PUEBLO ME	56740	1439	11.85	3.15	0.424
ROCKY FORT	57167	1271	11.79	3.57	0.804
RUSH 1 N	57287	1845	13.72	4.02	0.589
SEDALIA 4	57510	1821	17.18	4.86	0.293
SEDGWICK	57513	1092	16.82	3.45	0.102
SEDGWICK	57515	1216	17.66	3.16	-0.312
SPRINGFIELD	57862	1345	15.66	3.83	0.262
SPRINGFIELD	57866	1409	15.93	3.76	0.319
STERLING	57950	1200	14.97	3.46	-0.024
STONINGTON	57992	1159	16.65	4.24	0.478
STRATTON	58008	1342	17.45	4.73	0.594
TACONY 10	58157	1512	11.15	2.88	-0.101
TRINIDAD	58434	1750	13.07	3.22	0.513
TROY 1 SE	58468	1709	14.47	3.22	0.544
VONA	58722	1373	16.27	3.76	0.125
WALSH 1 W	58793	1213	17.93	4.06	0.550
WATERDALE	58839	1594	15.87	4.29	0.538
WINDSOR	59147	1457	12.44	2.82	0.363
WRAY	59243	1122	17.51	4.44	0.442
YUMA	59295	1262	16.98	3.98	0.093

As a first step, regional droughts that occurred in the two investigated regions have been identified and characterized using the procedure outlined in previous paragraphs. In particular, for each region, influence areas a_k at each station have been computed by Eq. (6.1) where s_k is the area of the Thiessen's polygon corresponding to station

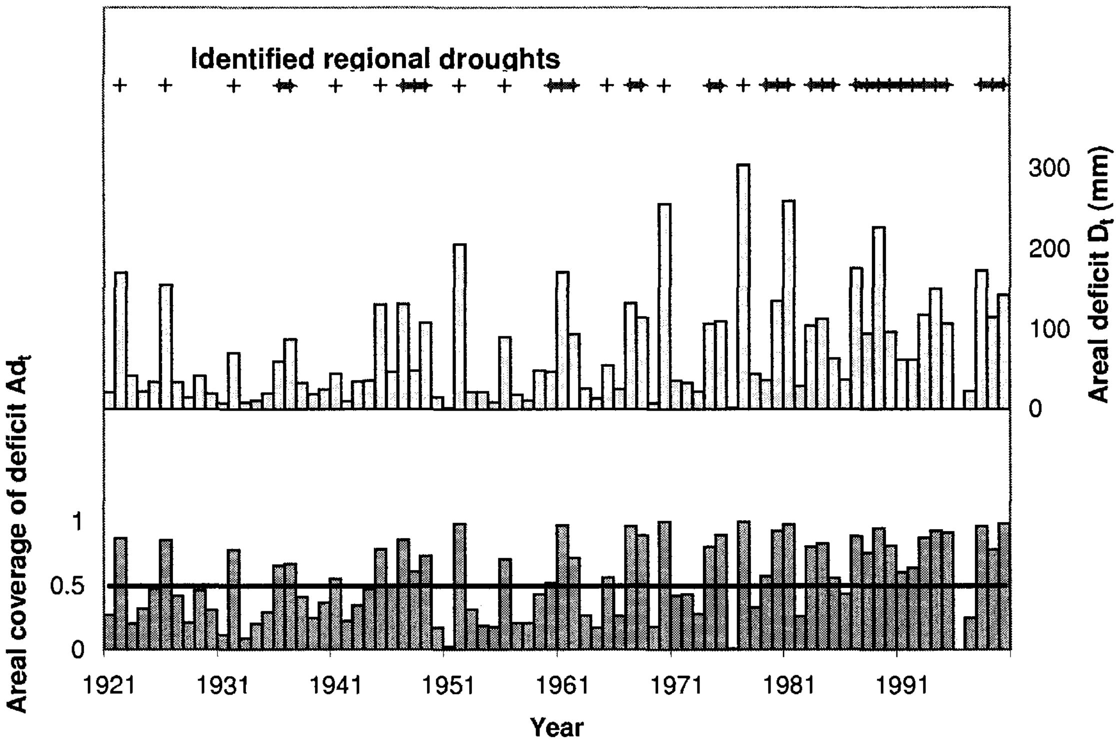


Figure 6.4: Areal coverage of deficit Ad_t, regional deficits D_t and identified regional droughts in Sicily during the 1921-2000 period.

k. Also, for each station k a threshold corresponding to the observed sample mean has been assumed, i.e. $x_{ok} = \overline{X}_k$ for $k = 1, 2, \ldots$, and fixed critical area thresholds $a_c = .5$ and $a_c = .7$. In Figure 6.4 the areal coverage of deficit Ad_t and the regional deficits D_t observed in Sicily during the period 1921-2000 are shown for $a_c = .5$. In addition, the plot indicates the various regional drought episodes that have occurred in Sicily throughout the historical record, corresponding to periods during which the areal coverage of deficit is above the threshold. For instance, the severe drought that occurred between the 1980's and 1990's is clearly shown. The figure confirms the intensification of droughts in the latter half of the century, with a tendency to longer and more severe droughts.

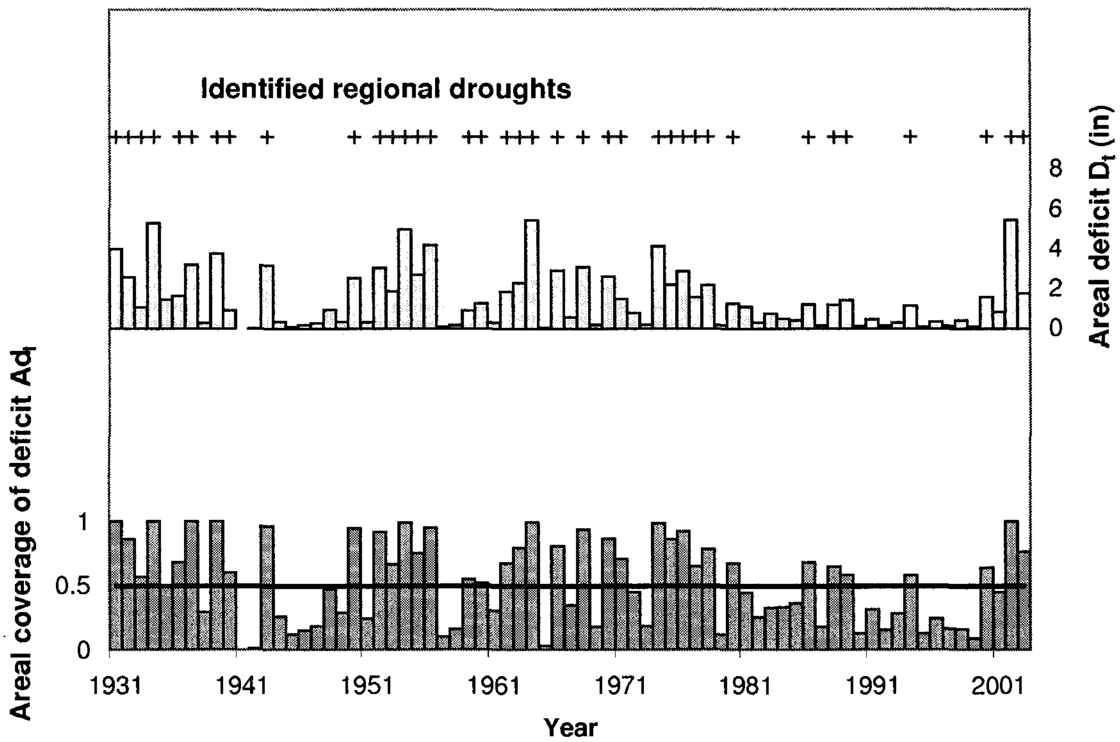


Figure 6.5: Areal coverage of deficit, regional deficits and identified regional droughts in Eastern colorado during the 1931-2003 period.

Similarly in Figure 6.5 the areal coverage of deficit Ad_t and the regional deficits D_t observed during the 1931-2003 period in Eastern Colorado are shown. The plot indicates the various regional drought episodes that have occurred in the region throughout the historical record, such as that in the 1930's, 1950's and 1970's.

From Figures 6.4 and 6.5, it can be inferred that 19 and 18 regional droughts can be identified from the available precipitation samples in Sicily and Colorado, respectively. Such number of droughts is not adequate to perform statistical inference on drought probabilities especially if a joint analysis of drought characteristics such as accumulated deficit and length is sought. To better illustrate the point, if the 19 regional droughts identified in Sicily are disaggregated according to the drought length, 10 droughts have length 1, 3 length 2, 5 length 3 and 1 length 9. This means that it is not possible to obtain reliable estimates of the probability of droughts by fitting a parametric distribution to the observed droughts of fixed length, let alone estimates of return period. Thus, in order to verify the analytical model developed previously, two records of 200000 years precipitation at the stations in Sicily and Colorado have been generated assuming a normal distribution. Such normality assumption has not been rejected by means of a skewness test in 22 out of 43 stations in Sicily and in 64 out of 72 stations in Colorado with a significant level 5%. In particular, generation of normal vector $\underline{X}_t = \{X_{1,t}, X_{2,t}, \ldots, X_{m,t}\}^T$ with $m = 43$ for Sicily and $m = 72$ for Eastern Colorado has been carried out by first estimating the cross variance-covariance matrix of observed annual precipitation at the m stations $\hat{\Sigma}$. The generic element $\hat{\gamma_{i,j}}$ of $\hat{\Sigma}$ represents the covariance between contemporaneous observations at station i and j and has been computed as:

$$\hat{\gamma_{i,j}} = \frac{\sum_{t=1}^{m}(x_{it} - \overline{x}_i)(x_{jt} - \overline{x}_j)}{\left(\sum_{t=1}^{m}(x_{it} - \overline{x}_i)^2 \sum_{t=1}^{m}(x_{jt} - \overline{x}_j)^2\right)^{1/2}} \tag{6.54}$$

where $\overline{x}_i$ is the sample mean of annual precipitation at station i.

Then, a random vector $\mathbf{X}_t$ has been generated as (Salas, 1993):

$$\underline{X}_t = \hat{\mathbf{B}}\underline{\varepsilon}_t + \hat{\underline{\mu}} \tag{6.55}$$

where $\hat{\mathbf{B}}$ is the Cholesky decomposition of $\hat{\mathbf{\Sigma}}$ (i.e a the solution of the matrix equation $\hat{\mathbf{B}}\hat{\mathbf{B}}^T = \hat{\mathbf{\Sigma}}$), $\hat{\underline{\mu}} = \{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m\}$ is the vector of the sample means of annual precipitation and $\underline{\varepsilon}_t$ is a vector of i.i.d. standard normal random variables.

After generation, regional drought identification in each investigated region has been carried out fixing for each station a threshold equal to the long term mean and two areal thresholds, namely $a_{crit} = 0.5$ and $a_{crit} = 0.7$. Then regional deficit and drought characteristics Ad_t, Dd_t, L and DC have been estimated on the generated samples. Corresponding moments have also been computed by means of the aforementioned expressions.

As a first step, the comparison between computed, generated and observed moments (i.e. computed from the observed samples) of Ad_t, Dd_t has been carried out. The results of such comparison are reported in Tables 6.III and 6.IV for Sicily and Colorado respectively, for the two fixed areal threshold values $a_c = .5$ and $a_c = .7$. It can be inferred from the tables a general good agreement between the derived moments of deficit characteristics and corresponding moments computed on regional deficits identified on generated series, especially with respect to expected values. Also the moments computed from the observed sample generally match well the derived and generated ones for both regions.

Tables 6.V and 6.VI shows similar results related to the moments of regional drought duration L and accumulated deficit DC for Sicily and Eastern Colorado respectively. Again a good agreement between moments computed by the proposed

Table 6.III: Comparison of expected values and standard deviations of areal coverage of deficit Ad_t and areal deficit D_t (mm) in Sicily for various combination of threshold values x_{ok} and a_{crit}. Moments of Ad_t from models computed by eqs. 6.16 and 6.17, and of D_t computed by eqs. 6.27 and 6.28. Moments from simulation have been estimated on a 200000 years generated sample, observed moments have been computed on 80 years of observed precipitation

			$\mathcal{E}[Ad_t]$	$\sqrt{\mathrm{Var}[Ad_t]}$	$\mathcal{E}[D_t]$	$\sqrt{\mathrm{Var}[D_t]}$
		From model	0.500	0.290	120.11	71.55
$x_{ok} = \overline{X}_k$	$a_{crit} = .5$	From simulation	0.499	0.291	126.1	66.6
		Observed	0.540	0.300	123.7	62.7
		From model	0.500	0.290	155.5	72.55
$x_{ok} = \overline{X}_k$	$a_{crit} = .7$	From simulation	0.499	0.291	160.2	63.8
		Observed	0.540	0.300	145.9	56.3

approximations and from generated or sample series can be inferred.

Next, a comparison between the pdf's of drought characteristics derived analytically and the corresponding sample pdf computed from the generated series has been performed. In Figure 6.6 the comparison between the sample pmf of regional drought length L in Sicily computed from the generated series and the corresponding analytical pmf given by Eq. (6.33) is shown. The plot indicates how the analytical pmf follows closely the sample pmf computed from the generated series for both values of the critical area a_c. The plot also shows the sample pmf of regional drought length computed form the sample series (80 years). It can be inferred how the latter pmf, being computed from a limited number of droughts, exhibits some sample variability. On the other hand, the pmf derived analytically shows a very good agreement with the sample pmf computed from the generated sample. Similar arguments can be made with reference to Figure 6.7, which shows the same comparison among drought length pmf's related to Eastern Colorado. It may be worthwhile to stress once again

Table 6.IV: Comparison of expected values and standard deviations of areal coverage of deficit Ad_t and areal deficit D_t (in) in Eastern Colorado for various combination of threshold values x_{ok} and a_{crit}. Moments of Ad_t from model computed by eqs. 6.16 and 6.17, and moments of D_t computed by eqs. 6.27 and 6.28. Moments from simulation have been estimated on a 200000 years generated sample, observed moments have been computed on 72 years of observed precipitation

			$\mathcal{E}[Ad_t]$	$\sqrt{\mathrm{Var}[Ad_t]}$	$\mathcal{E}[Dd_t]$	$\sqrt{\mathrm{Var}[Dd_t]}$
$x_{ok} = \overline{X}_k$	$a_{crit} = .5$	From model	0.500	0.311	2.532	1.523
		From simulation	0.500	0.310	2.520	1.403
		Observed	0.520	0.319	2.543	1.336
$x_{ok} = \overline{X}_k$	$a_{crit} = .7$	From model	0.500	0.311	3.101	0.963
		From simulation	0.500	0.310	3.191	1.341
		Observed	0.520	0.319	3.280	1.166

Table 6.V: Comparison of expected values and standard deviations of regional drought duration L and regional cumulated deficit DC (mm) in Sicily for various combination of threshold values x_{ok} and a_{crit}. Moments of L from model computed by eqs. 6.34 and 6.35, and moments of DC computed by eqs. 6.36 and 6.37. Generated moments have been estimated on a 200000 years generated sample, observed moments computed on 80 years of observed precipitation

			$\mathcal{E}[L]$	$\sqrt{\mathrm{Var}[L]}$	$\mathcal{E}[DC]$	$\sqrt{\mathrm{Var}[DC]}$
$x_{ok} = \overline{X}_k$	$a_{crit} = .5$	From model	1.998	1.412	241.7	199.2
		From simulation	1.972	1.358	248.6	193.6
		Observed	2.105	1.883	260.4	232.3
$x_{ok} = \overline{X}_k$	$a_{crit} = .7$	From model	1.429	0.782	222.4	149.7
		From simulation	1.411	0.765	226.1	145.3
		Observed	1.667	0.907	243.2	136.5

Table 6.VI: Comparison of expected values and standard deviations of regional drought duration L and regional cumulated deficit DC (in) in Eastern Colorado for various combination of threshold values x_{ok} and a_{crit}. Moments of L from model computed by eqs. 6.34 and 6.35, and moments of DC computed by eqs. 6.36 and 6.37. Generated moments have been estimated on a 200000 years generated sample, observed moments computed on 72 years of observed precipitation

			$\mathcal{E}[L]$	$\sqrt{\mathrm{Var}[L]}$	$\mathcal{E}[DC]$	$\sqrt{\mathrm{Var}[DC]}$
		From model	2.000	1.414	5.075	4.194
$x_{ok} = \overline{X}_k$	$a_{crit} = .5$	From simulation	2.018	1.469	5.085	4.181
		Observed	2.056	1.349	5.227	4.688
		From model	1.429	0.782	4.686	3.246
$x_{ok} = \overline{X}_k$	$a_{crit} = .7$	From simulation	1.472	0.847	4.697	3.189
		Observed	1.533	0.743	5.030	2.848

that the analytical distributions have not been fitted to drought lengths but have been derived as a function of the statistics of the underlying precipitation and of the thresholds x_{ok} and a_c.

In Figure 6.8 and 6.9 the quantiles of regional accumulated deficit in Sicily and Eastern Colorado respectively computed from the generated sample are plotted versus the corresponding analytical counterpart computed by means of the pdf derived in Eq. (6.50). For easy of reference, in the plots the dimensionless accumulated deficit $\delta = DC/\overline{\mu}$ is plotted, where $\overline{\mu}$ is the average of the means of annual precipitation at the investigated stations in each region. In the same plot, the analogous comparison with the dimensionless quantiles computed on the observed sample is also shown by circle symbols. Figure 6.8(a) indicates that there is a fairly good agreement between the pdf of regional accumulated deficit in Sicily derived analytically and the ones computed from the generated or observed sample for $a_c = .5$, since the points generally tend to align along the 45 degrees line. Similar behaviour can be inferred for the

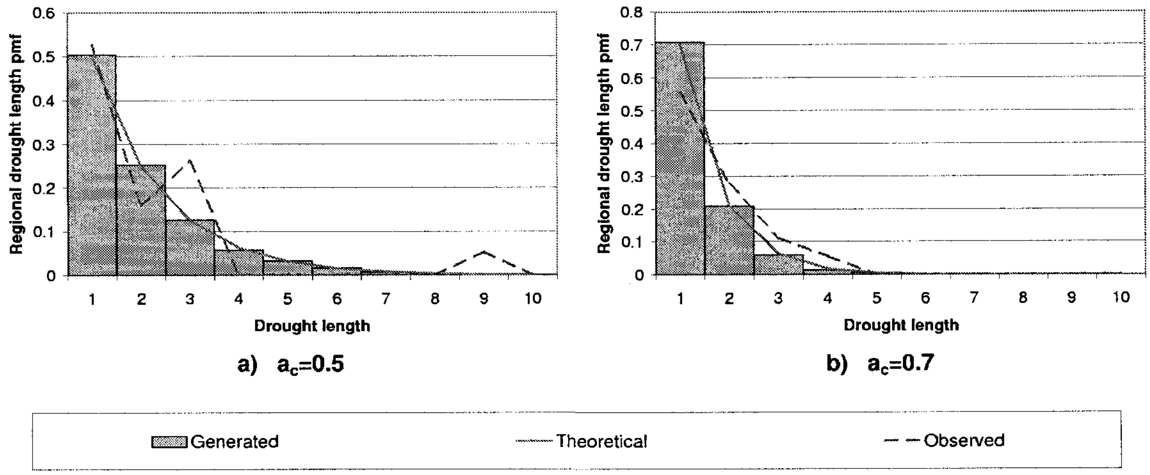


Figure 6.6: Regional drought length pmf in Sicily estimated from generated sample (200000 years) and from observed sample (80 years) and computed analytically by eq. (6.33) for critical areal threshold $a_c = .5$ and $a_c = .7$.

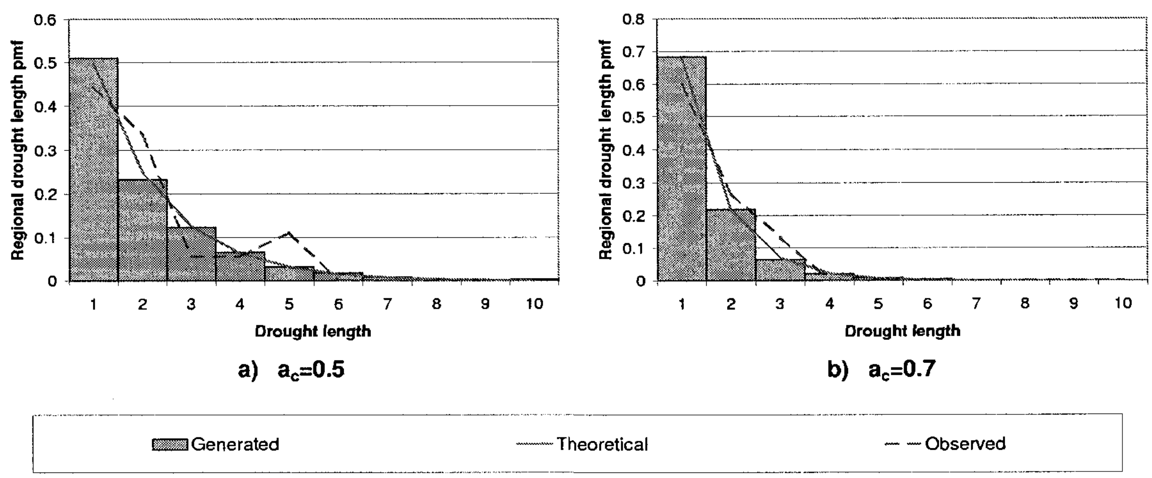
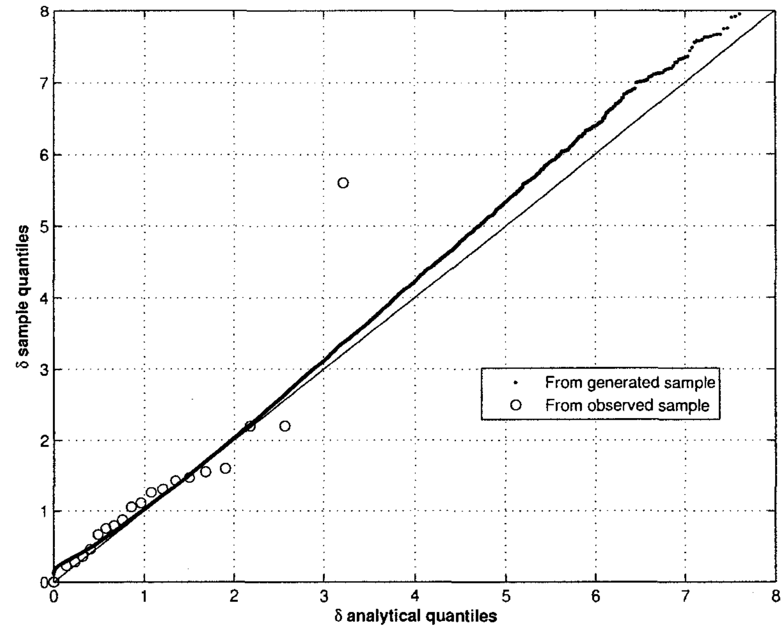


Figure 6.7: Regional drought length pmf in Eastern Colorado estimated from generated sample (200000 years) and from observed sample (72 years) and computed analytically by eq. (6.33) for critical areal threshold $a_c = .5$ and $a_c = .7$.

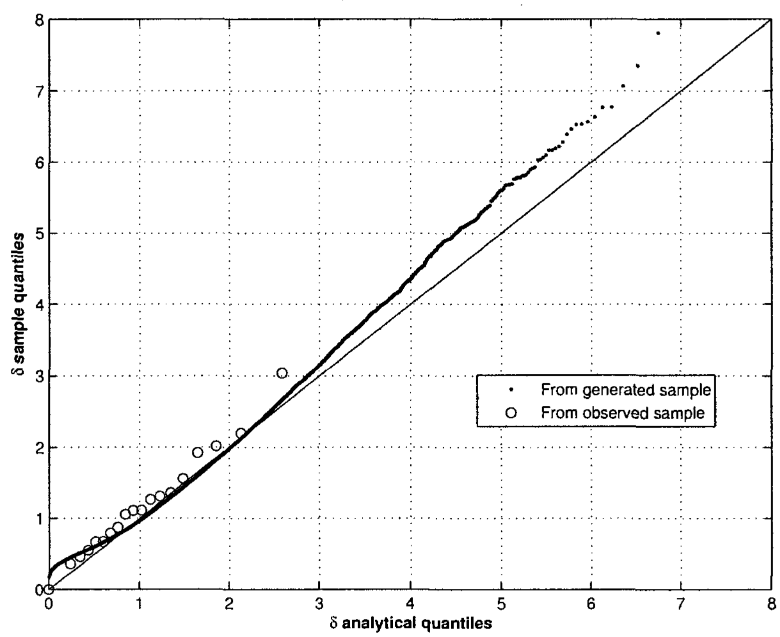
Colorado case from Figure 6.9(a). Both plots show also that the points deviate somewhat from the line for high values of the accumulated deficit. Such deviation is also observed as a larger value of critical area is considered, i.e. $a_c = .7$ (Figs. 6.8(b) and 6.9(b)). Nonetheless, such deviations occurr only for very low or very high quantiles and therefore for practical purposes they can be conseidered acceptable. The causes of such discrepancies can be ascribed in part to the sample variability of quantiles computed from generated sample, as well as to the small differences between the moments of accumulated deficit derived analytically and computed from the samples (see Tables 6.V, 6.VI). Also, the deviations may indicate that the gamma distribution that has been assumed for accumulated deficit is probably not adequate to model very extreme droughts. However, it should be pointed out that attempts to use different distributions (log Normal, Weibull, Gumbel), here not shown, led to a general poorer fit than the gamma case. Therefore, the latter distribution has been preferred, since it seems to be reasonably adequate for an ample range of accumulated deficit values.

In order to verify the applicability of the derived pdf to compute return period of critical regional droughts, a comparison between return periods computed by means of Eq. (6.53) and corresponding sample estimates computed as the average interarrival time between drought events E identified on the generated and historical samples has been carried out. First, return period of drought events E with regional accumulated deficit greater than D_0 have been considered. In particular theoretical return period of critical drought events $E = \{D > D_0\}$ has been computed by combining Eqs. (6.50)and (6.53) as:

$$T = \frac{1}{p_d(1-p_d)\int_{D_0}^{\infty} \frac{1}{\beta\Gamma(r_D)} e^{-\frac{d}{\beta_D}} \left(-\frac{d}{\beta_D}\right)^{r_D-1}} \qquad (6.56)$$

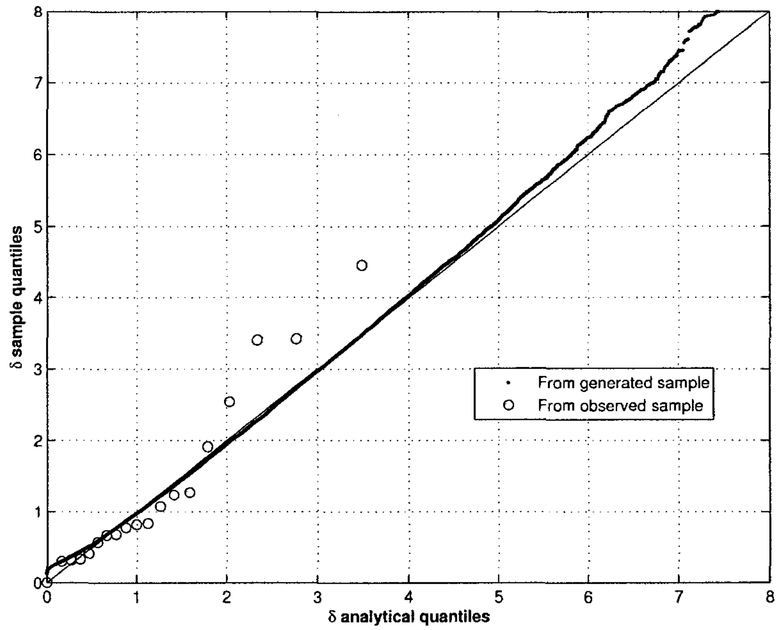


(a) $a_c = .5$

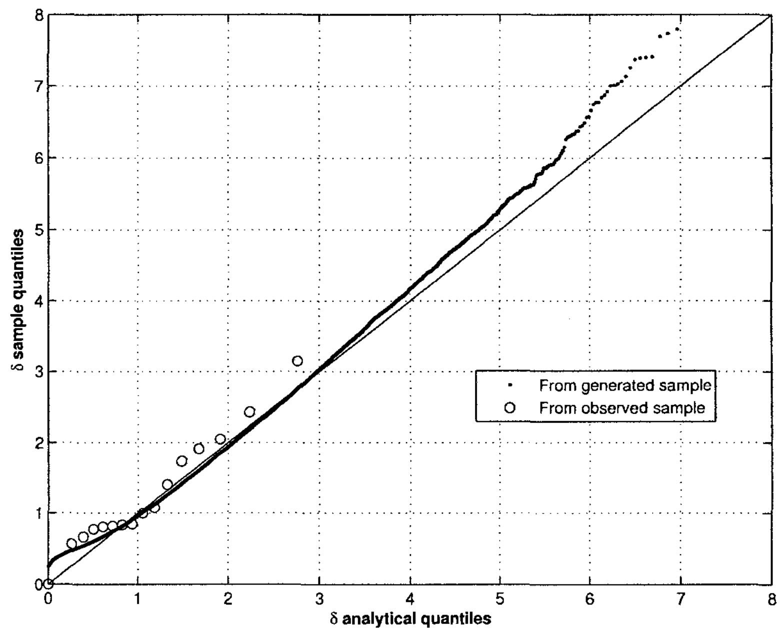


(b) $a_c = .7$

Figure 6.8: q-q plot between quantiles of regional accumulated deficit in Sicily computed analytically from Eq. (6.50) and the ones computed from the generated and observed samples for critical areas $a_c = .5$ and $a_c = .7$



(a) $a_c = .5$

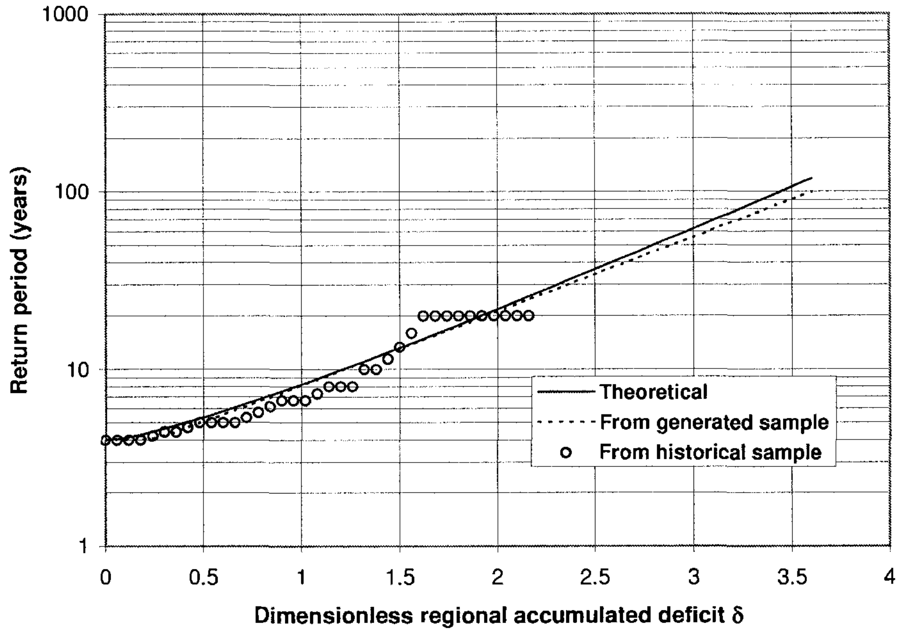


(b) $a_c = .7$

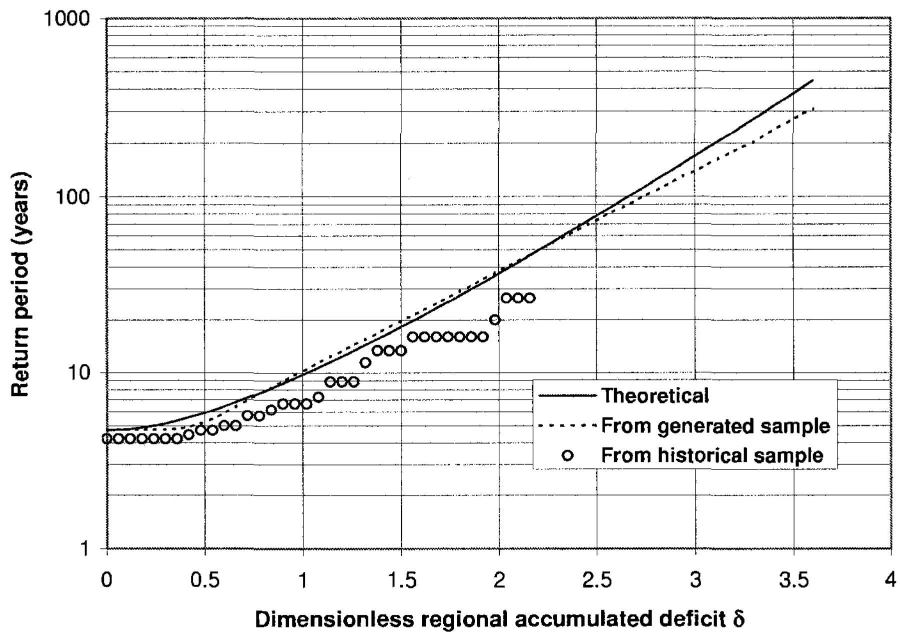
Figure 6.9: q-q plot between quantiles of regional accumulated deficit in Colorado computed analytically from Eq. (6.50) and the ones computed from the generated and observed samples for critical areas $a_c = .5$ and $a_c = .7$

In Figure 6.10 the comparison between return periods computed analytically and from generated and historical sample is shown. In order to somewhat increase the reliability of the estimation, return periods of critical droughts on generated and historical samples have been computed by considering, besides interarrival times between critical droughts, also the first arrival time (beginning from the first year of the series), as well as the time from the last critical drought to the last year of the series. In all cases the general pattern is similar, i.e. there is an excellent agreement between the return periods T estimated from (6.56) and those estimated from the generated and historical samples. This is especially so for the first few values of drought lengths, thereafter the generated sample estimates appear to drift below the analytical estimates. Also, the agreement is generally better for critical area $a_c = .5$ than for $a_c = .7$. This is not surprising since, as as already shown in 6.8, the fit of the analytical pdf of accumulated deficit (Eq. (6.50) to the corresponding sample pdf computed from the generated series is better for $a_c = .5$. Also Figure 6.10 indicates that for $a_c = .7$ return periods estimated from the historical sample are generally lower than those computed analytically or from the generated sample. This can be explained because as the accumulated deficit increases there are smaller numbers of drought episodes that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time become unreliable and underestimated. In fact, beyond a certain drought length no droughts may be observable from the historical sample and no return period can be determined. Furthermore, the sample mean interarrarival time will be obviously biased downward, since it has an upper bound equal roughly to the length of the series. This issue will be addressed with more details in what follows.

Figure 6.12 shows the return periods of droughts defined jointly in terms of accumulated deficit and length as in case (1) above, i.e. $E = \{ \ DC > D_0$ and $L = l_0$ (

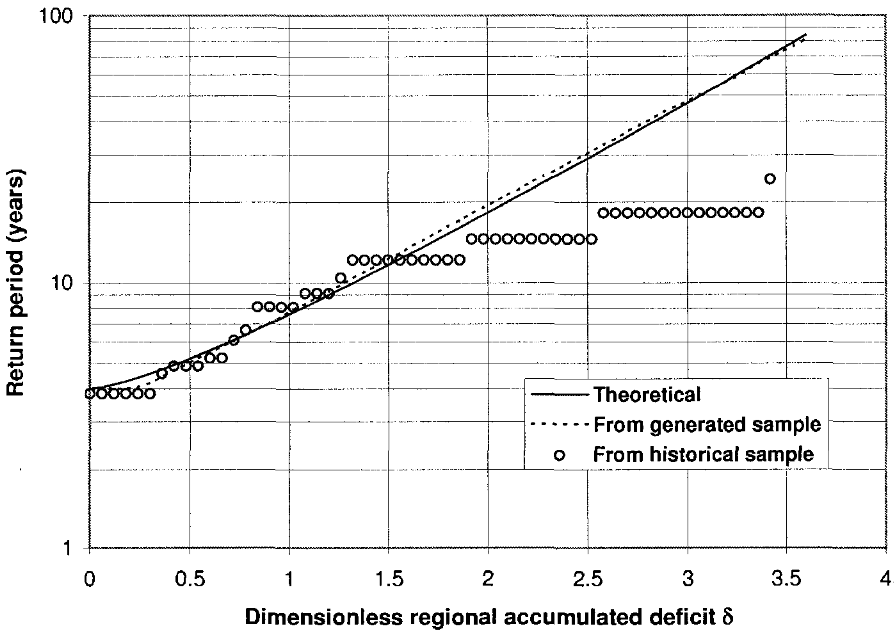


(a) $a_c = .5$

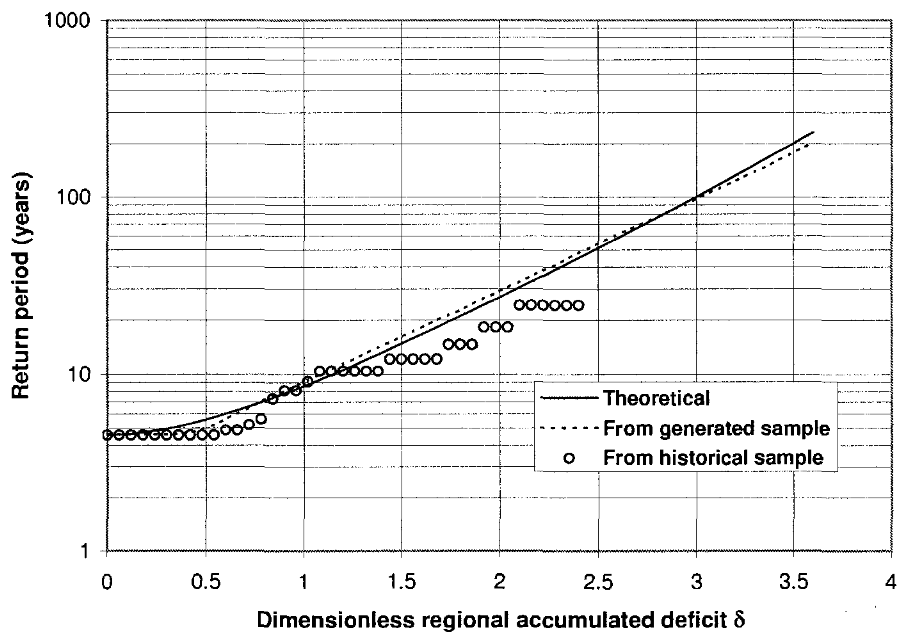


(b) $a_c = .7$

Figure 6.10: Return period of regional drought events in Sicily defined by $\{DC > D_0\}$ obtained from Eq.(6.56) and from the generated and historical sample for various values of the critical area a_c



(a) $a_c = .5$



(b) $a_c = .7$

Figure 6.11: Return period of regional drought events in Eastern Colorado defined by $\{DC > D_0\}$ obtained from Eq.(6.56) and from the generated and historical sample for various values of the critical area a_c

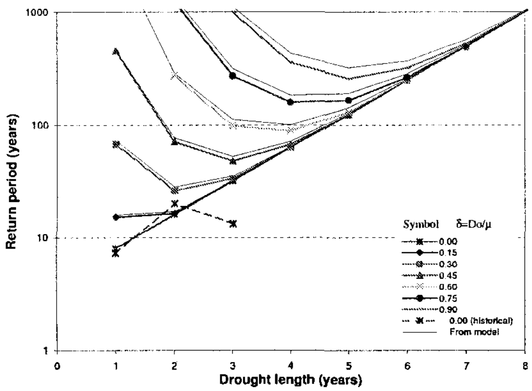
$l_0 = 1, 2, ...)\}$, which were obtained from the generated samples and from Eq.6.53 for various values of the deficit coefficient δ. From the figure it can be inferred that, for a given drought duration l_o, (the return period) increases rapidly with increasing δ, which means that estimating the return period T for large values of δ may require a very long sample. On the other hand, Fig. 6.12 indicates that using Eq.(6.53) one can always find the results for any δ and any l_o.

Figure 6.12 also shows that results obtained from the generated samples and from Eq.(6.53) are generally close. Furthermore, for all values of the deficit coefficient δ, as the drought duration l_0 increases, the return period T converges to the return period curve for $\delta = 0$. The comparison between the results obtained for the two critical areas considered reveals also some features. In particular, the curve related to $\delta = 0$, which represents critical droughts of fixed length, and of any accumulated deficit, is less steep in the case of $a_c = .5$ than in the other case. This is consistent with the fact that as the critical area increases, the occurrence probability of longer droughts will decrease (see Figure 6.6), and therefore longer return period of longer droughts are expected.

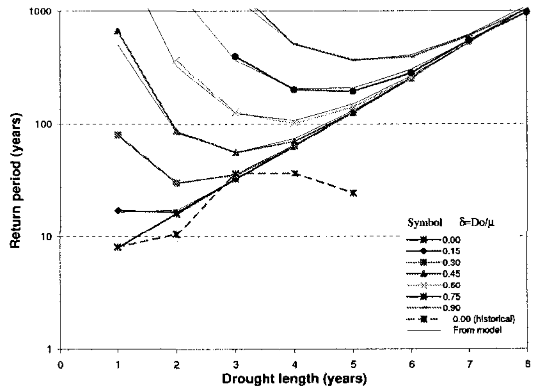
Fig.6.12 shows also the return periods obtained from the historical sample for δ=0 and for some l_0. In most cases, the return periods computed from the observed samples are lower than the generated or analytical counterpart. The can be explained by the fact that not many drought episodes can be observed from the historical records, and therefore the return periods estimated from the samples have a high variability. Just as an example, the return period of droughts with lengths equal to 3 years in Sicily, assuming $a_c = .7$ has been computed by averaging the interarrivals times between only two droughts. Furthermore, the return periods computed from the sample as average interarrival times has an obvious upper bound given roughly by the length of the sample. Thus, estimates of return period from the sample must be considered affected by a significant sampling variability

To illustrate the above point, the sampling variability of return period estimated as the mean interarrival time has been assessed from the generated sample. In particular, the 200000 years generated sample has been divided into 2500 80-years long subsamples, namely the same length of the observed sample. Then assuming $a_c = .7$, $\delta = 0$ and for each drought length $l_0 = 1, 2, ...8$, the averaged interarrival time has been estimated from each subsample. In figure 6.13 the 5% and 95% percentiles of these averaged interarrival times are shown by dotted lines, along with the return period estimated from the observed sample, from the generated sample (200000 years) and from Eq.(6.53). The latter three curves obviously coincide with the corresponding ones shown in 6.12c. Also note that the two percentiles curves represents approximate 90% confidence intervals for sample return period. From Figure 6.13 it can be inferred that the return periods estimated from the observed sample are always within the 90% confidence bounds. Furthermore, the return periods estimated from the sample roughly follows the patterns of the confidence limits, that tend to align horizontally as the drought length increases. Similar results (here not shown) are obtained for $a_c = .5$ and for the Estern Colorado region. This confirms the unreliability of return periods estimated as mean interarrival times from observed samples. On the contrary, the derived analytical expressions allow to estimate return period for any l_o and δ, which confirms their validity.

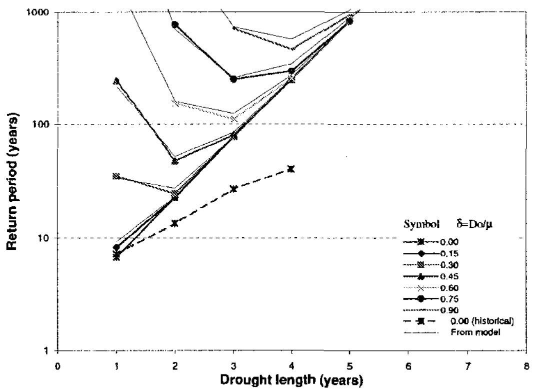
The return periods of droughts defined as in case (2) above i.e. $E = \{ DC > D_0$ and $L \geq l_0$ ($l_0 = 1, 2, ...)\}$, which were obtained from the generated samples and from Eq.(6.53) for various values of δ are shown in Figure 6.14. The figure shows that as in case (1) above, as l_0 increases all return period curves converge to a single curve that is independent of δ. Furthermore the same figure shows the close results that are obtained from the generated samples and from Eq.(6.53).



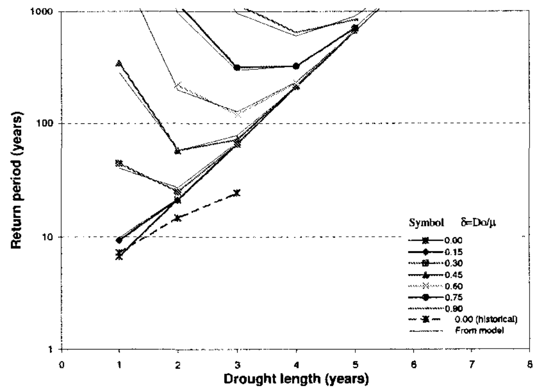
(a) Sicily $a_c = .5$



(b) Colorado $a_c = .5$



(c) Sicily $a_c = .7$



(d) Colorado $a_c = .7$

Figure 6.12: Return period of regional drought events in Sicily and Eastern Colorado defined by $\{\ DC > D_0$ and $L = l_0,\ l_0 = 1, 2, ...\}$ obtained from Eq.(6.53) and from the generated sample for various values of the deficit coefficient δ ($D_0 = \delta\overline{\mu}$) and for critical area $a_c = .5$ and $a_c = .7$. The return period curves for δ=0 obtained from the historical samples are also shown

.

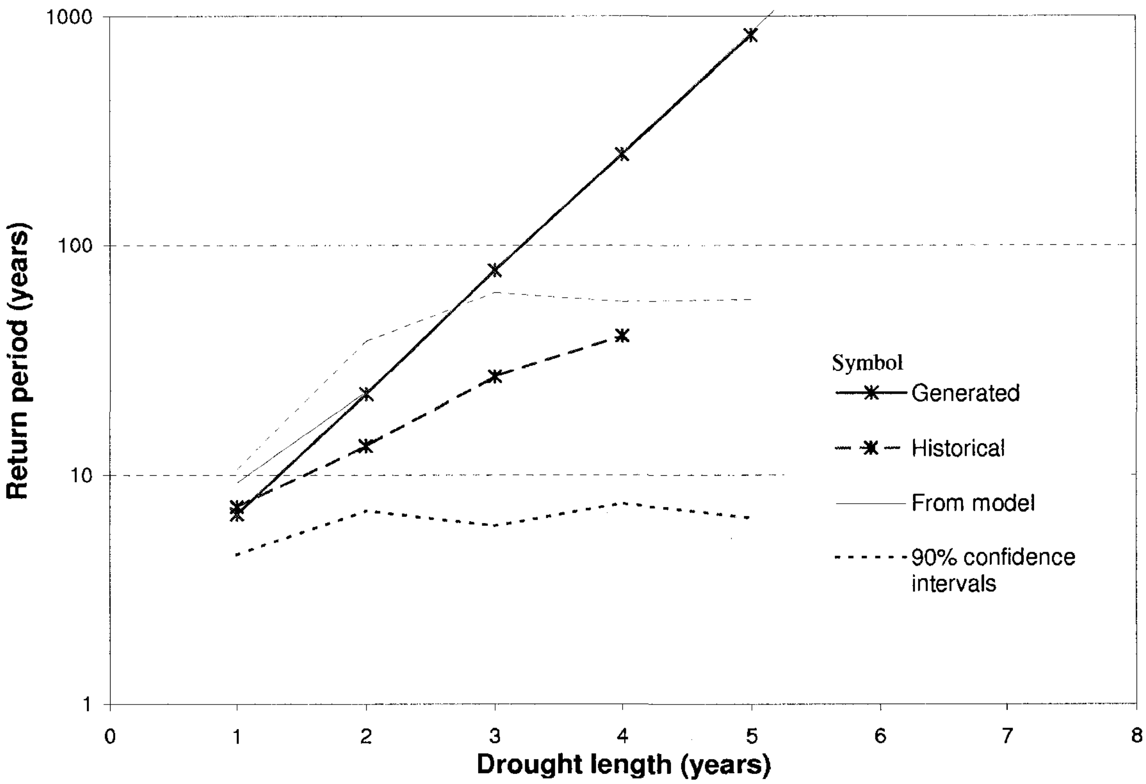
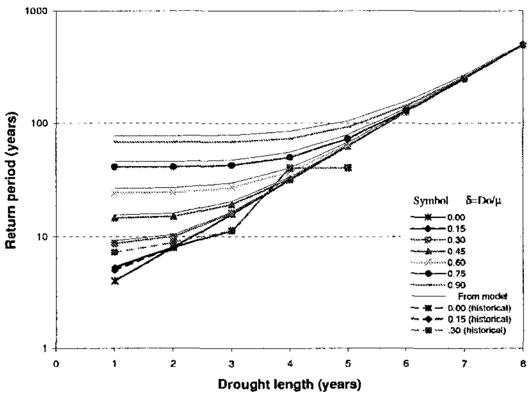


Figure 6.13: Return period of regional drought events in Sicily { $DC > 0$ and $L = l_0$, $l_0 = 1, 2, ...$} obtained from Eq.(6.53) and from the generated sample for critical area $a_c = .7$. The return period curves for δ=0 obtained from the historical samples are also shown along with the corresponding 90% approximated confidence intervals

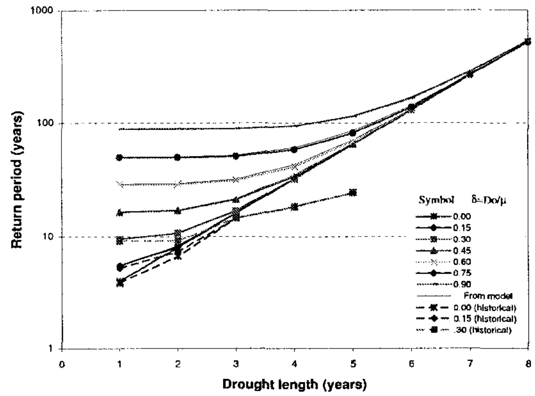
In addition, Fig. 6.14 also shows the return periods based on the historical record for some values of l_0 and δ. Unlike the case (1) droughts where only a few estimates of return periods can be made for very small values of l_0 and δ (Fig. 6.12), for case (2) droughts, estimates of return periods based on the historical records can be made for several values of l_0 and δ as shown in Fig. 6.14. This allows comparing the return periods obtained from the generated sample (or from Eq.(6.53) versus those obtained from the historical records so that further evaluation of the simulation model and the analytical formulation be made. However, also in this case the same considerations already made for case (1) regarding the sampling variability of return period estimated from the the observed sample can be made, and results similar to those shown in Figure 6.13 can be found. Thus, the results obtained from the historical sample must be used with caution because as l_0 increases and as δ increases the return periods estimated from the historical sample becomes unreliable. For instance, in Fig. 6.14 the results are not reliable for $l_0 > 4$ or $\delta > 1$.

Figure 6.15 shows the return periods for droughts defined as in case (3) above, i.e. $E = \{ I > I_0 \text{ and } L = l_0(l_0 = 1, 2, ...)\}$, which were obtained from the generated sample and from Eq. (6.53) for various values of the drought intensity coefficient ζ. In this case, the return period curves are increasing functions of both l_0 and ζ. The figure also shows results obtained from the historical sample for a few values of l_0 and $\psi = 0$. Again in this case the results are not reliable for $l_0 > 2$ for the same reasons already discussed above.

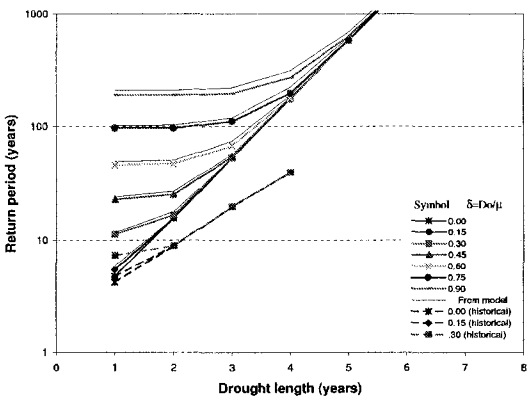
In addition, Fig. 6.15 shows that, as expected, as the critical area a_c increases, so does the steepness of the curves, which is somewhat in agreement with case (1) above. Furthermore, the results obtained from the generated samples and from Eq.(6.53) are quite comparable, with some exceptions, generally related to very high return



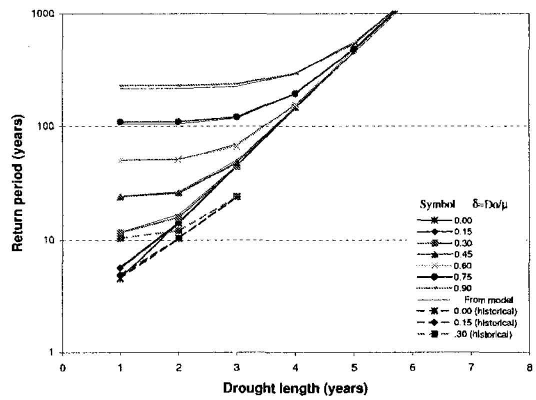
(a) Sicily $a_c = .5$



(b) Colorado $a_c = .5$



(c) Sicily $a_c = .7$



(d) Colorado $a_c = .7$

Figure 6.14: Return period of regional drought events in Sicily and Eastern Colorado defined by $\{\ DC > D_0$ and $L \geq l_0,\ l_0 = 1, 2, ...\}$ obtained from Eq.(6.53) and from the generated sample for various values of the deficit coefficient δ ($D_0 = \delta\overline{\mu}$) and for critical area $a_c = .5$ and $a_c = .7$. The return period curves for some δ values obtained from the historical samples are also shown

.

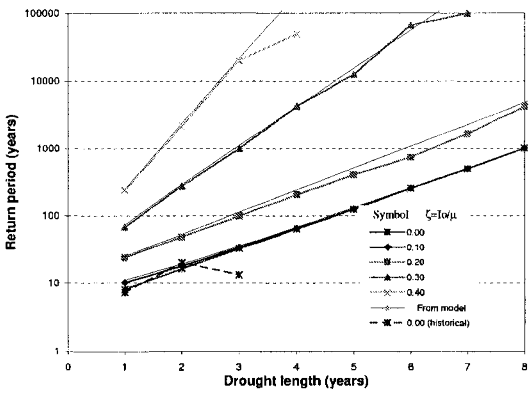
periods (>1000) and very severe droughts, and therefore they not be significant from a practical standpoint.

Also the results obtained from the observed samples and from Eq.(6.53) are similar for $a_c = .5$, whereas they drift below the theoretical ones for $a_c = .7$, in agreement with case (1) above and with Figure 6.10 b).

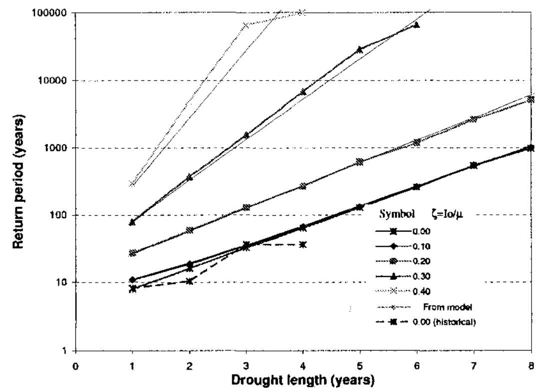
The return periods for drought events defined as in case (4) above, i.e. $E = \{ I > I_0 \text{ and } L \geq l_0\}$ are shown in Fig. 6.16 for various values of l_0 and ζ. The return period curves are increasing functions of both l_0 and ζ. From Fig. 6.16 a similar behaviour as in cases (1) and (3) above can be inferred, namely that as the critical area a_c increases, so does the steepness of the curves. The figure also shows that return periods obtained from the generated samples and from Eq.(6.53) are quite comparable for low values of ζ, while the differences generally increase with L_0 and ζ. Likewise, Fig. 6.16 shows that the return periods obtained from the historical record for ζ=0 and $a_c = .5$ correspond quite well with those obtained from the generated samples and from Eq.(6.53) whereas a downward drift is observed for $a_c = .7$, which is compatible with the sampling variability highlighted before.

6.6 Final remarks

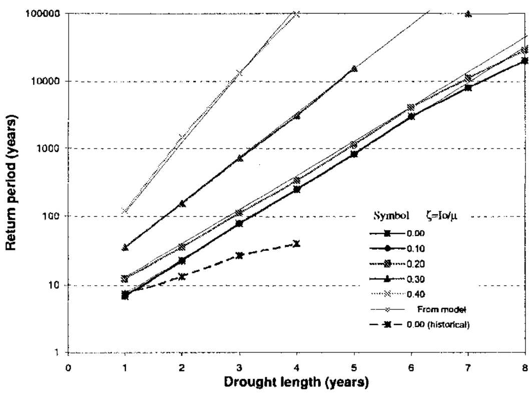
Probabilistic characterization of regional droughts is an important step toward a rational water planning and management. The knowledge of the regional features of drought over an area, such as its areal extension, duration and regional deficit or intensity, can provide useful information in order to trigger appropriate mitigation actions. This is particularly true when rainfed agriculture is predominant in the area, and therefore precipitation becomes the primary source of water. Also drought assessment on a regional scale can also be useful to evaluate the appropriateness of interbasin water transfers as a drought mitigation measure.



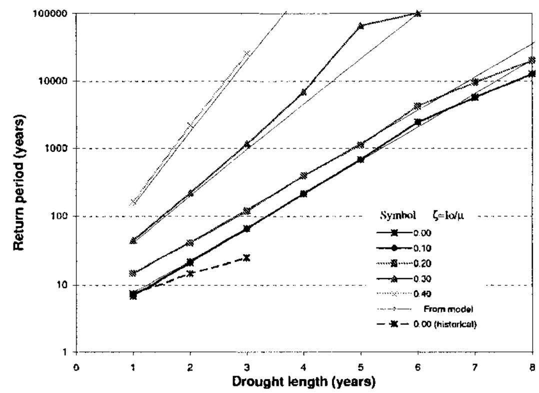
(a) Sicily $a_c = .5$



(b) Colorado $a_c = .5$



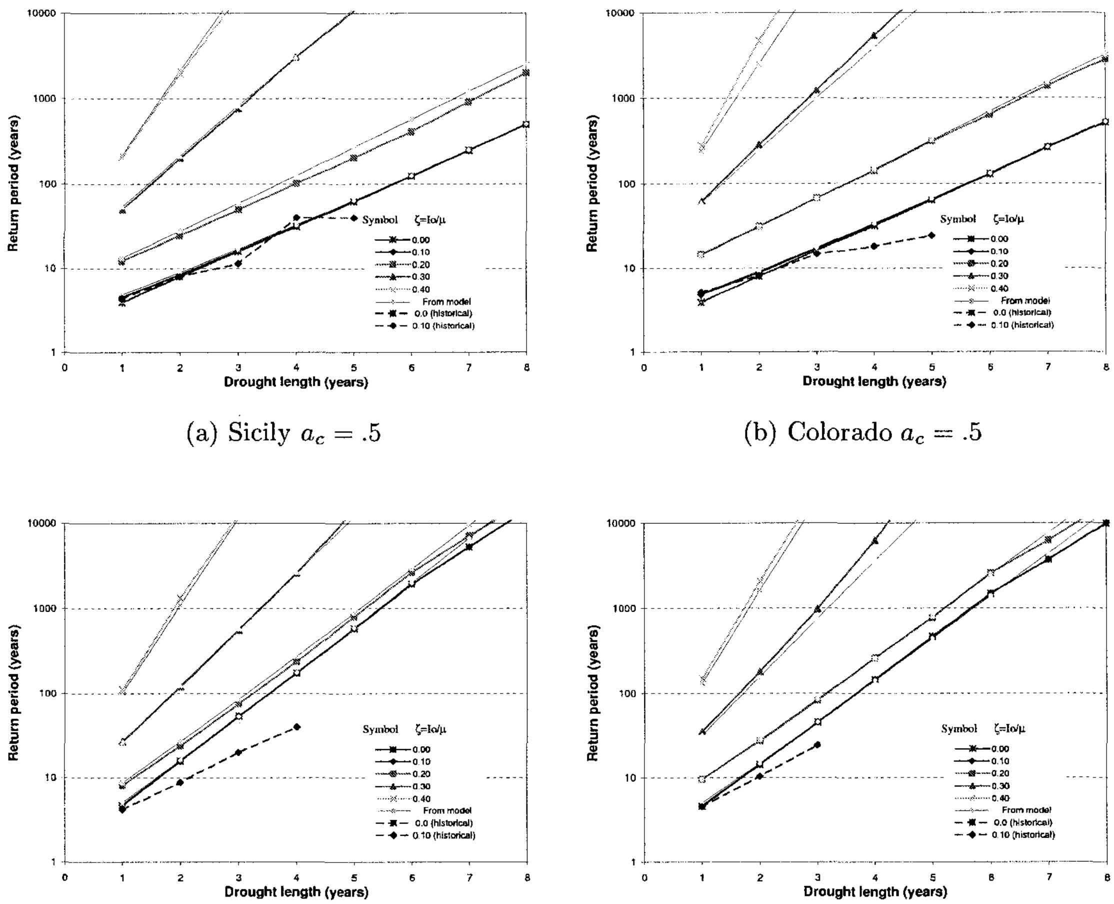
(c) Sicily $a_c = .7$



(d) Colorado $a_c = .7$

Figure 6.15: Return period of regional drought events in Sicily and Eastern Colorado defined by $\{ I > D_0$ and $L = l_0,\ l_0 = 1, 2, ...\}$ obtained from Eq.(6.53) and from the generated sample for various values of the intensity coefficient ζ ($D_0 = \zeta\overline{\mu}$) and for critical area $a_c = .5$ and $a_c = .7$. The return period curves for ζ=0 obtained from the historical samples are also shown

.



(a) Sicily $a_c = .5$

(b) Colorado $a_c = .5$

(c) Sicily $a_c = .7$

(d) Colorado $a_c = .7$

Figure 6.16: Return period of regional drought events in Sicily and Eastern Colorado defined by $\{\ I > D_0$ and $L \geq l_0,\ l_0 = 1, 2, ...\}$ obtained from Eq.(6.53) and from the generated sample for various values of the intensity coefficient ζ ($D_0 = \zeta\overline{\mu}$) and for critical area $a_c = .5$ and $a_c = .7$. The return period curves for some ζ values obtained from the historical samples are also shown

One of the main advantages of run method to characterize droughts is the possibility to extend it in order to identify and characterize regional droughts . Here run method has been adopted to derive analytically the distributional properties of regional drought characteristics, on the basis of the stochastic structure of the underlying precipitation series. The derived distributions enble to estimate return period of various types of critical droughts, also taking into account jointly different characteristics. The approach allows to overcome the difficulties associated with estimating probabilities and return periods of droughts identified on an historical sample, due to the limited number of droughts that are generally observed. The analytical pdf's of drought characteristics are expressed as a function of the basic statistics of the underlying series (mean, cross-covariance) as well as of the demand thresholds x_0 and of the areal threshold a_c. Since the basic statistics are computed from the whole available sample, their estimation can be considered reliable even with relatively short observation records.

Application of the developed methodology to two case studies has revealed that it yields return period estimates that can generally be considered more reliable than those obtained by averaging the interarrival times between observed droughts in an historical sample. Also the comparison with results obtained from long sinthetically generated samples indicates that the analytical expression yield very similar return periods for a wide range of drought severities, which is indicative of the adequacy of the proposed methodology.

Lastly, it should be noted that although here the methodology has been developed with explicit reference to precipitation series, its extension to other hydrological variables such as streamflows is straightforward. For instance, one may be interested in analyzing droughts occurring at several streamflow sites, whose water supply a given water system. The only required modification to the methodology is in the definition of the weights a_c assigned to each series, since in this case they may for example represent weighting factors assigned to each streamflow source, as a function of the senior/junior water rights structure.

CHAPTER VII

CONCLUSIONS

Characterizing periods of deficit and drought has been an important aspect of planning and management of water resources systems for many decades. For example, the design water supply capacity to satisfy the municipal consumption of a given city may be based on meeting water demands during a critical drought that may occur in a specified planning horizon. The implementation of water use restrictions in a given water district as an evolving drought reaches a certain degree of severity is just another example.

An extreme drought is a complex phenomenon that evolves through time and space in a random fashion. It may be characterized by its initiation, duration, severity (magnitude or intensity), termination, and spatial extent. These characteristics may be determined objectively by comparing a water supply time series versus a water demand series and defining droughts by applying concepts and theory of runs. Such an objective and simple drought definition may enable one characterizing droughts using probability and stochastic approaches. Another subject of much interest when analyzing extreme drought events is the determination of the associated return period. The return period can be defined in different ways for different applications such as, the average elapsed time between the occurrences of specified drought event (i.e., droughts exceeding a given critical drought) and the average time (e.g. years) to the first occurrence of a specified event.

The analytical derivation of probability distribution of drought characteristics has received much attention in several studies. Such analytical derivations are particularly useful in cases of short hydrologic records for which the applicability of an inferential approach to find the probability distributions of drought characteristics is either impractical or not feasible. In addition, they may be useful for checking approximations or results obtained for more complex cases.

The review of literature reveals that the major emphasis in analyzing drought properties using analytical methods has been for cases where the underlying processes are stationary and/or with a mild time dependence structure. Because annual hydrological processes, such as annual precipitation and annual streamflow, have been often assumed to be stationary and with a weak dependence, most developments and applications on drought characteristics using analytical methods have been centered about annual processes. Also, the derivation of probabilities and return period of regional features of droughts such as areal extension or areal deficit has not received much attention in literature.

The research described in this dissertation has addressed the need for stochastic modelling of drought characteristics, in cases when the underlying series is either periodic or with a non-negligible autocorrelation. In addition, extension of the derived models to the regional case has also been explored. Moments of drought characteristics, probability distribution functions and return period have been developed in all cases, thus providing tools to characterize droughts in a broad sets of situations that may occur in practice.

Although previous studies exist where droughts are analyzed in a bivariate fashion, however they have usually neglected the time-dependence in the hydrological series under investigation. Indeed, when time dependence is considered, even computing the first two moments of drought characteristics becomes difficult, since in this case the series of the deficits resulting after the hydrological variable is truncated by means of a water demand threshold is not stationary. Furthermore, when a time dependent process, such as an AR(1) is clipped by a constant threshold level, the resulting binary process follows only approximately a lag 1 Markov chain, and therefore the widely used geometric distribution to model drought length may become not appropriate.

Here a truncated multivariate normal has been applied to model the deficits, that enables to compute the first two moments of accumulated deficit or intensity, for droughts of fixed length. Such moments have then been used to define the distribution of the above characteristics, assumed beta. Also in order to model drought length, it has been assumed that the clipped binary process follows a DARMA(1,1) model, for which several theoretical and practical results are available. The result is a bivariate formulation for drought characteristics length and accumulated deficit or intensity pdf, that allows to compute the probability of occurrence of several types of critical droughts. The derived pdf has been applied to compute the return period of critical droughts. Also, empirical approximations of the moments of accumulated deficit, function of the lag-1 autocorrelation of the underlying hydrological variables and of the threshold level are proposed , that allow to derive the pdf of drought characteristics without resorting to numerical algorithms to solve the multivariate integrals. The empirical approximations are such that they can be applied also to skewed data.

Validation of the proposed methodology has been carried out by comparing the return periods of various critical drought obtained by means of the proposed empirical approximations with those computed from generated samples from an AR(1) model. The results indicate that the procedure can be considered sufficiently reliable to estimate return period of critical droughts, especially when the severity of the analyzed droughts remains within the range of those generally considered in engineering applications. The procedure has also been applied to four streamflow annual series, namely the annual flows of the Poudre River at Mouth of the Canyon, of the Colorado River at Lee Ferry, of the Nile River at Aswan, and the St. Lawrence River at Cornwall. The four series exhibit different level of persistence thus enabling to test the methodology when different hydrological regimes are considered. After an AR(1) model has been fitted to each series and synthetic series have been generated, the comparison between return periods obtained by means of the derived analytical expressions and those computed on the observed and generated samples has been carried out. The results of such comparison confirm the applicability of the proposed methodology, since it is able to capture reasonably well the stochastic variability of droughts in streamflow series with different degrees of persistence.

In order to model drought occurrences in periodic stochastic series, the probability mass function (pmf) of drought length, its associated low order moments, and the return period of drought events have been derived assuming a periodic two-state simple Markov. The derived pmf's allow estimating the occurrence probabilities of droughts of a given length where either the drought begins in a given season or regardless of the initial season. The applicability of the drought formulations has been illustrated using a variety of water supply series, and using both constant and periodic water demand thresholds.

The comparison of the empirical probability mass function (pmf) of drought length obtained from the historical sample and from the derived analytical expression generally showed a very good agreement between them.

Furthermore, the return period of droughts lengths for the data sets investigated were estimated from the derived analytical expressions and from the historical sample.

In all cases the results showed an excellent agreement between the two estimates for the first few values of drought lengths, thereafter the sample estimates appear to drift below the analytical estimates. This can be explained because as the drought length increases there are smaller numbers of droughts that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time become unreliable and generally underestimated.

The comparison between the return periods obtained assuming independence and simple Markov dependence illustrate clearly the effect of dependence in the behavior of the return period versus drought length, i.e. for a given drought length the return period is bigger for an independent process.

Probability distributions of drought characteristics in periodic series have been investigated both marginally and jointly. In particular, explicit expressions of probability density functions of drought characteristics, their associated low order moments, and the return period of drought events have been derived. The derived pdf's allow to estimate the occurrence probabilities and return period of droughts where either the drought begins in a given season or regardless of the initial season. The applicability of the drought formulations has been illustrated using six precipitation series assumed as supply series, and using periodic water demand thresholds. The comparison of the empirical probability mass function (pdf) of accumulated deficit obtained from the historical sample and from the derived expressions generally showed a very good agreement between them. The model pdf appears to represent quite well the empirical pdf for all series. Also the comparison between accumulated deficit moments for droughts starting at a given season computed by the model and estimated both on observed and generated sample indicates the general adequacy of the developed expressions.

Furthermore, the return period of droughts with accumulated deficit greater than a fixed value d for the six data sets investigated were estimated from the derived expressions and from the historical sample. In all cases the results showed an excellent agreement between the two estimates (analytical and historical sample) for smaller values of accumulated deficit, thereafter the sample estimates appear to drift above or below the analytical estimates. This can be explained because as the drought accumulated deficit increases there are smaller numbers of droughts that can be observed from the historical sample and consequently the estimates of the mean inter-arrival time become unreliable. Also the derived expressions for return period of droughts with accumulated deficit or intensity greater than a given value and length equal to a given duration have been applied to the 3-months precipitation series of Caltanissetta for droughts starting at a given season. The results indicate the validity of the proposed expressions, with a good agreement between return periods computed from analytical expressions and those estimated from generated samples.

Run method has been adopted to derive analytically the distributional properties of regional drought characteristics, on the basis of the stochastic structure of the underlying precipitation series. The derived distributions allow to estimate return period of various types of critical droughts, also taking into account jointly different characteristics. The approach allows to overcome the difficulties associated with estimating probabilities and return periods of droughts identified on an historical sample, due to the limited number of droughts that are generally observed. The analytical pdf's of drought characteristics are expressed as a function of the basic statistics of the underlying series (mean, cross-covariance) as well as of the demand thresholds x_0 and of the areal threshold a_c. Since the basic statistics are computed from the whole available sample, their estimation can be considered reliable even with relatively short observation records.

Application of the developed methodology to two case studies has revealed that it yields return period estimates that can generally be considered more reliable than those obtained by averaging the interarrival times between observed droughts in an historical sample.

The overall conclusion of the research is that exact and/or approximate analytical expressions of probability distributions of drought characteristics derived from the statistics of the underlying hydrological series enables one a more reliable probabilistic characterization than employing the inferential approach. In addition, such analytical derivations may be useful for checking approximations or results obtained for more complex cases. The examples using a variety of water supply series and climatologic and hydrological drought indices illustrate and confirm the applicability of the analytical derivations obtained for drought characteristics and associated return periods.

Also the comparison with results obtained from long sinthetically generated samples indicates that the analytical expression yield very similar return periods for a wide range of drought severities, although the results somehow differ when droughts characterized by a very high intensity (short and with elevated accumulated deficit) are considered.

APPENDIX A

ESTIMATION OF TRANSITION PROBABILITY MATRICES

A.1 Non parametric approach

Application of the previous results to observed hydrological series requires the estimation of the transition probabilities p_{ij_τ} $\quad i = 0, 1;\ j = 0, 1$ and of the marginal probabilities p_{i_τ} $i = 0, 1$ for each season $\tau = 1, \ldots, \omega$. Estimation of transition probabilities for the stationary case has been discussed by Fernandez and Salas (1999). Extension to the periodic case can be carried out following a non-parametric approach based on maximum likelihood. More specifically, with reference to an observed periodic series $\{x_{\nu,\tau}, \nu = 1, 2, \ldots,\ n\tau = 1, 2, \ldots, \omega\}$, let's assume that periodic thresholds have been fixed $\{x_{0\tau}, \tau = 1, 2, \ldots, \omega\}$. As already mentioned, we can associate to each observation $x_{\nu,\tau}$ the Bernoulli variable $y_{\nu,\tau}$ which takes value 1 if $x_{\nu,\tau} \geq x_{0\tau}$ and 0 if $x_{\nu,\tau} < x_{0\tau}$. If the $y_{\nu,\tau}$ follows a periodic lag 1 Markov chain, its likelihood (i.e. the joint probability of observing a given sequence of $y_{\nu,\tau}$) can be written as:

$$L = \prod_{\tau=1}^{\omega} {p_{0_\tau}}^{y_{1,1}} {p_{1_\tau}}^{1-y_{1,1}} {p_{00_\tau}}^{n_{00\tau}} {p_{01_\tau}}^{n_{01\tau}} {p_{10_\tau}}^{n_{10\tau}} {p_{11_\tau}}^{n_{11\tau}} \tag{A.1}$$

where n_{ij_τ}, $i = 0, 1$; $j = 0, 1$ is the number of observations $y_{\nu,\tau} = j$ for which $y_{\nu,\tau-1} = i$.

By letting $p_{01_\tau} = 1 - p_{00_\tau}$, $p_{10_\tau} = 1 - p_{11_\tau}$, substituting in Eq. (A.1), setting derivatives with respect to p_{00_τ} and p_{11_τ} equal to zero one gets:

$$\hat{p_{00_\tau}} = \frac{n_{00_\tau}}{n_{00_\tau} + n_{01_\tau}} \tag{A.2}$$

Similarly, the maximum likelihood estimators for the other transition probabilities can be derived. In particular:

$$\hat{p_{11_\tau}} = \frac{n_{11_\tau}}{n_{11_\tau} + n_{10_\tau}} \tag{A.3}$$

The remaining transition probabilities $\hat{p_{01_\tau}}$ and $\hat{p_{10_\tau}}$ can be easily derived through $p_{01_\tau} = 1 - p_{00_\tau}$, $p_{10_\tau} = 1 - p_{11_\tau}$. Finally, in order to estimate the marginal probabilities $\hat{p_{1_\tau}}$ the recursions (4.7) can be applied and $\hat{p_{0_\tau}}$ can be estimated as $\hat{p_{0_\tau}} = 1 - \hat{p_{1_\tau}}$.

A.2 Parametric approach

Transition probability matrices can also be estimated through a parametric approach, following the definition given in Eq. (4.4) (Fernandez and Salas, 1999; Salas et al., 2001). For example the conditional probability p_{00_τ} can be expressed in terms of joint and marginal probabilities as:

$$p_{00_\tau} = P\left[X_{\nu,\tau} \leq x_{0\tau} | X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \frac{P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right]}{P\left[X_{\nu,\tau-1} \leq x_{0\tau-1}\right]} \tag{A.4}$$

Now, assume the bivariate probability density function of $\{X_{\nu,\tau}, X_{\nu,\tau-1}\}$, $f_{X_{\nu,\tau},X_{\nu,\tau-1}}(x, y)$ is known. Then joint and marginal probabilities can be estimated by integration. In particular, the joint probability can be computed as:

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \int_{-\infty}^{x_{0\tau}} \int_{-\infty}^{x_{0\tau-1}} f_{X_{\nu,\tau},X_{\nu,\tau-1}}(x,y)dxdy \tag{A.5}$$

In what follows, the bivariate normal and lognormal case for $\{X_{\nu,\tau}, X_{\nu,\tau-1}\}$ will be developed.

Normal case In this case, the bivariate density takes the well known form:

$$f_{X_{\nu,\tau},X_{\nu,\tau-1}}(x,y) = \frac{1}{2\pi(1-{\rho_\tau}^2)^{1/2}} e^{-\frac{1}{2(1-{\rho_\tau}^2)}\left[\frac{(x-\mu_\tau)^2}{{\sigma_\tau}^2} + \frac{(y-\mu_{\tau-1})^2}{{\sigma_{\tau-1}}^2} - 2\rho_\tau \frac{(x-\mu_\tau)(y-\mu_{\tau-1})}{\sigma_\tau \sigma_{\tau-1}}\right]} \tag{A.6}$$

where μ_τ and σ_τ are the mean and standard deviation of $X_{\nu,\tau}$ respectively and ρ_τ is the correlation coefficient between $X_{\nu,\tau}$ and $X_{\nu,\tau-1}$. Now let:

$$\phi(x,y;\rho) = \frac{1}{2\pi(1-{\rho_\tau}^2)^{1/2}} \exp\left[-\frac{x^2+y^2-2\rho_\tau xy}{2(1-{\rho_\tau}^2)}\right] \tag{A.7}$$

the bivariate normal pdf of zero mean, unit variance random variables with correlation coefficient ρ_τ. The integral A.5 can be written as:

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \int_{-\infty}^{a} \int_{-\infty}^{b} \phi(x,y,\rho_\tau)dxdy \tag{A.8}$$

where $a = \frac{x_{0\tau}-\mu_\tau}{\sigma_\tau}$, $b = \frac{x_{0\tau-1}-\mu_{\tau-1}}{\sigma_{\tau-1}}$.

It can be shown that the double integration reduces to (Cramer and Leadbetter, 1967):

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \Phi(a)\Phi(b) + \int_{0}^{\rho_\tau} \phi(a,b;z)dz \tag{A.9}$$

Assume now that the periodic threshold takes the following general form:

$$x_{0\tau} = \mu_{X\tau} - \alpha\sigma_{X\tau} = \mu_{X\tau}(1 - \alpha Cv_\tau) \tag{A.10}$$

where Cv_τ is the coefficient of variation of $X_{\nu,\tau}$. Then, a and b become:

$$a = b = -\alpha \tag{A.11}$$

and therefore, Eq. (A.9) reduces to:

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \Phi(-\alpha)^2 + \int_0^{\rho_\tau} \phi(-\alpha, -\alpha; z)dz \tag{A.12}$$

which, substituting the expression of $\phi()$ becomes:

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \Phi(-\alpha)^2 + \int_0^{\rho_\tau} \frac{1}{2\pi(1-z^2)^{1/2}} e^{\left(\frac{-\alpha^2}{1+z}\right)} dz \tag{A.13}$$

The marginal probability $P\left[X_{\nu,\tau-1} \leq x_{0\tau-1}\right]$ in Eq. (A.4) can be easily computed as:

$$P\left[X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \Phi(b) \tag{A.14}$$

where $\Phi(.)$ is the standard normal (zero mean and unit standard deviation) cdf. Again, assuming a periodic threshold as in Eq. (A.10), such marginal probability becomes simply $\Phi(-\alpha)$.

Combining Equations (A.4), (A.12) and (A.14), the transition probability $p_{00\tau}$ can thus be estimated as a function of α as:

$$p_{00\tau} = \Phi(-\alpha) + \frac{1}{\Phi(-\alpha)} \int_0^{\rho_\tau} \frac{1}{2\pi(1-z^2)^{1/2}} e^{\left(\frac{-\alpha^2}{1+z}\right)} dz \tag{A.15}$$

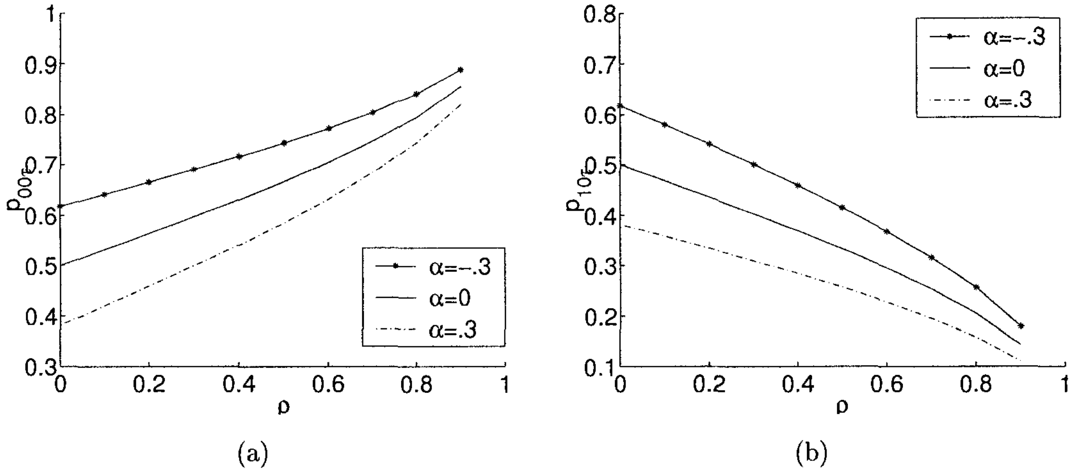


Figure A.1: Transition probabilities p_{00_τ} (a) and p_{10_τ} (b) vs. ρ_τ for different α levels

The other transition probabilities can also be estimated in a similar fashion, rearranging appropriately the integration bounds of (A.4) and (A.9). For example, p_{10_τ} can be estimated as:

$$p_{10_\tau} = \Phi(-\alpha) + \frac{1}{1-\Phi(-\alpha)} \int_0^{\rho_\tau} \frac{1}{2\pi(1-z^2)^{1/2}} e^{\left(\frac{-\alpha^2}{1+z}\right)} dz \tag{A.16}$$

Once p_{00_τ} and p_{11_τ} have been estimated, the remaining transition probabilities p_{01_τ} and p_{10_τ} can be found by means of $p_{01_\tau} = 1 - p_{00_\tau}$, $p_{10_\tau} = 1 - p_{11_\tau}$.

Figure A.1 depicts p_{00_τ} and p_{10_τ} as a function of ρ_τ for values of $\alpha = -.3, 0, .3$, computed by numerical integration of Equations (A.15) and (A.16).

Lognormal case If $\{X_{\nu,\tau}, X_{\nu,\tau-1}\}$ follow a bivariate lognormal distribution, the bivariate density takes the form:

$$f_{X_{\nu,\tau},X_{\nu,\tau-1}}(x,y) = \frac{1}{2\pi(1-\rho_\tau'^2)^{1/2}} e^{-\frac{1}{2(1-\rho_\tau'^2)}\left[\frac{(\ln x-\mu_\tau')^2}{\sigma_\tau'^2} + \frac{(\ln y-\mu_{\tau-1}')^2}{\sigma_{\tau-1}'^2} - 2\rho_\tau' \frac{(\ln x-\mu_\tau')(\ln y-\mu_{\tau-1}')}{\sigma_\tau'\sigma_{\tau-1}'}\right]} \tag{A.17}$$

where μ'_τ and σ'_τ are the mean and standard deviation of $\ln(X_{\nu,\tau})$ respectively and ρ'_τ is the correlation coefficient between $\ln(X_{\nu,\tau})$ and $\ln(X_{\nu,\tau-1})$.

The integral A.5 can be written in terms of the bivariate standard normal ϕ (Eq. A.7) as:

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \int_{-\infty}^{a'} \int_{-\infty}^{b'} \phi(x, y, \rho'_\tau) dx dy \tag{A.18}$$

where $a' = \frac{\ln x_{0\tau} - \mu'_\tau}{\sigma'_\tau}$, $b = \frac{\ln x_{0\tau-1} - \mu'_{\tau-1}}{\sigma'_{\tau-1}}$.

Again, the double integration reduces to:

$$P\left[X_{\nu,\tau} \leq x_{0\tau}, X_{\nu,\tau-1} \leq x_{0\tau-1}\right] = \Phi(a')\Phi(b') + \int_0^{\rho'_\tau} \phi(a', b'; z) dz \tag{A.19}$$

Recalling that for the log-normal distribution, the following relationships hold (Matalas, 1967):

$$\mu_{X_\tau} = e^{\frac{1}{2}{\sigma'_{X_\tau}}^2 + \mu'_{X_\tau}} \tag{A.20}$$

$$\sigma_{X_\tau} = \mu_{X_\tau}\sqrt{e^{{\sigma'_{X_\tau}}^2} - 1} \tag{A.21}$$

$${\sigma'_{X_\tau}}^2 = \ln(Cv_\tau + 1) \tag{A.22}$$

$$\rho'_\tau = \frac{\ln(1 + \rho_\tau\sqrt{(e^{\sigma'^2_{X_\tau}} - 1)(e^{\sigma'^2_{X_{\tau-1}}} - 1)}}{\sigma'_{X_\tau}\sigma'_{X_{\tau-1}}} \tag{A.23}$$

a' and b' become:

$$a' = \frac{\frac{1}{2}\ln\left(Cv_\tau^2 + 1\right) + \ln\left(1 - \alpha Cv_\tau\right)}{\sqrt{\ln(Cv_\tau^2 + 1)}} \tag{A.24}$$

$$b' = \frac{\frac{1}{2}\ln\left(Cv_{\tau-1}^2 + 1\right) + \ln\left(1 - \alpha Cv_{\tau-1}\right)}{\sqrt{\ln(Cv_{\tau-1}^2 + 1)}} \tag{A.25}$$

Combining Equations (A.22), (A.23), (A.24) and (A.25) and substituting into Eq. (A.19), the joint probability can be estimated in terms of Cv_τ, $Cv_{\tau-1}$, ρ_τ and α.

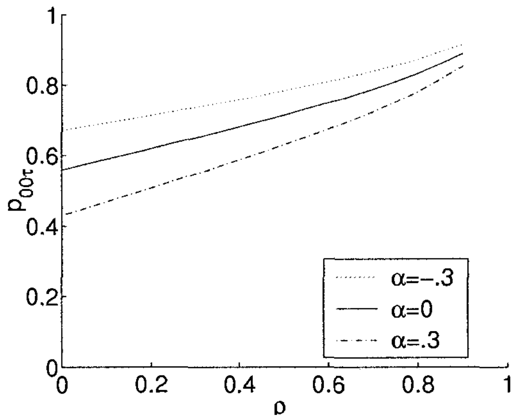
Similarly, the marginal probability $P\left[X_{\nu,\tau-1} \leq x_{0\tau-1}\right]$ in Eq. (A.4) can be computed as a function of the above parameters as $\Phi(b')$. Then, the transition probability p_{00_τ} becomes:

$$p_{00_\tau} = \Phi(a') + \frac{1}{\Phi(b')} \int_0^{\rho'_\tau} \phi(a', b'; z) dz \tag{A.26}$$

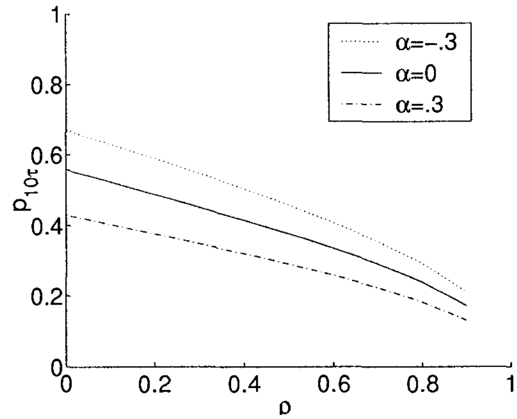
As in the normal case, the other transition probabilities can also be estimated in a similar fashion, rearranging appropriately the integration bounds of (A.4) and (A.19). For example, p_{10_τ} can be estimated as:

$$p_{10_\tau} = \Phi(a') + \frac{1}{1 - \Phi(b')} \int_0^{\rho'_\tau} \phi(a', b'; z) dz \tag{A.27}$$

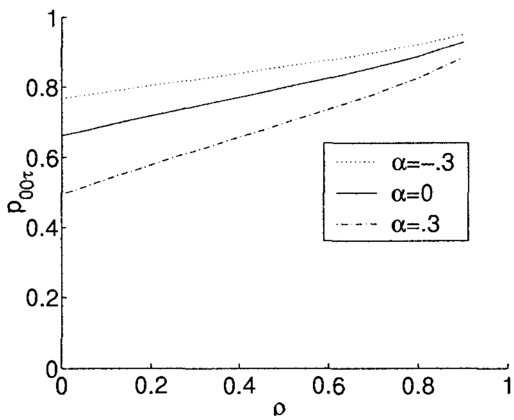
Figure A.2 depicts p_{00_τ} and p_{10_τ} as a function of ρ_τ for values of $\alpha = -.3, 0, .3$, and for different Cv_τ, $Cv_{\tau-1}$ combinations, computed by numerical integration of Equations (A.15) and (A.16).



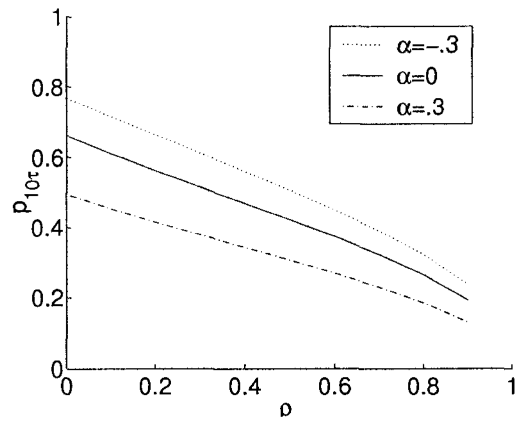
(a) $Cv_\tau = .3$, $Cv_{\tau-1} = .2$



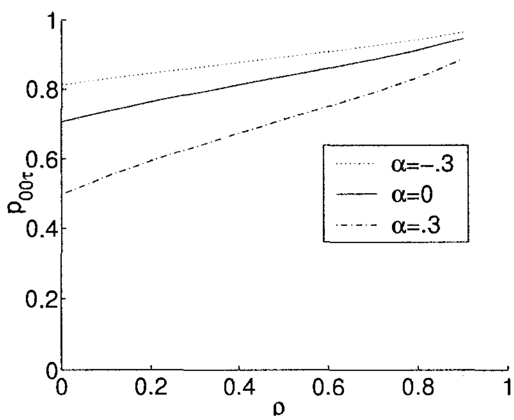
(b) $Cv_\tau = .3$, $Cv_{\tau-1} = .2$



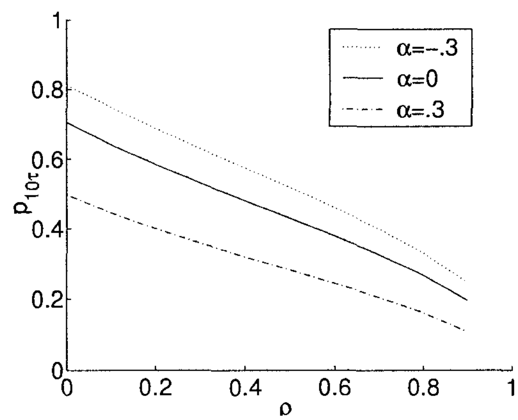
(c) $Cv_\tau = 1$, $Cv_{\tau-1} = .8$



(d) $Cv_\tau = 1$, $Cv_{\tau-1} = .8$



(e) $Cv_\tau = 1.5$, $Cv_{\tau-1} = 1.2$



(f) $Cv_\tau = 1.5$, $Cv_{\tau-1} = 1.2$

Figure A.2: Transition probabilities p_{00_τ} and p_{10_τ} vs. ρ_τ for different α and Cv_τ's combinations

APPENDIX B

DARMA(1,1) MODEL

B.1 General properties

The binary DARMA(1,1) model Y_t takes only the values 0 and 1 and is defined as (Jacobs and Lewis, 1977):

$$Y_t = \begin{cases} Z_t & \text{with probability } \beta; \\ U_{t-1} & \text{with probability } (1-\beta); \end{cases} \tag{B.1}$$

and

$$U_t = \begin{cases} U_{t-1} & \text{with probability } \lambda; \\ Z_t & \text{with probability } (1-\lambda); \end{cases} \tag{B.2}$$

where Z_t is a sequence of independent Bernoulli variables that take the value 0 with probability π_0 and 1 with probability $\pi_1 = 1-\pi_0$ and the variables are stationary and have the same marginal probability distribution π_0 and π_1. The autocorrelation function of Y_t is $\rho_k(Y) = c\lambda^{k-1}$, $k \geq 1$, where $c = (1-\beta)(\lambda+\beta-2\lambda\beta)$. The DARMA(1,1) process has a slower decaying *correlogram* (longer memory) than the DAR(1) process (as is the case for continuous valued processes).

Jacobs and Lewis (1978) showed that, even though the DARMA(1,1) process Y_t is not a Markov chain, the pair $\{Y_t, U_t\}$ forms a bivariate first-order Markov chain with transition probabilities $H_k(u,v) = P(Y_{t+1} = k), U_{t+1} = v|Y_t = m, U_t = u) = P(Y_{t+1} = k, U_{t+1} = v|U_t = u)$, since (Y_{t+1}, U_{t+1}) is indipendent of Y_t and u, v, k, m are $0, 1$ values. Then

$$H_0 = \begin{bmatrix} \lambda\,(1-\beta) + [1-\lambda\,(1-\beta)]\,\pi_0 & (1-\beta)\,(1-\lambda)\,\pi_1 \\ \beta\,(1-\lambda)\,\pi_0 & \beta\,\lambda\,\pi_0 \end{bmatrix} \tag{B.3}$$

$$H_1 = \begin{bmatrix} \beta\,\lambda\,\pi_1 & \beta\,(1-\lambda)\,\pi_1 \\ (1-\beta)\,(1-\lambda)\,\pi_0 & \lambda\,(1-\beta) + [1-\lambda\,(1-\beta)]\,\pi_1 \end{bmatrix} \tag{B.4}$$

The two-state bivariate Markov chain $\{Y_t, U_t\}$ is converted into a four -state simple Markov chain W_t by defining a "label variable" $W_t = 2Y_t + U_t$ as in (Lloyd and Saleem, 1979) such that

Variable	Values			
Y_t	0	0	1	1
U_t	0	1	0	1
W_t	0	1	2	3

Hence, the four-state Markov chain, W_t, takes on values $\{0, 1, 2, 3\}$ with transition probabilities such as

$$p_w\,(0,2) = P\,(W_{t+1} = 2|W_t = 0)$$
$$= P[Y_{t+1} = 1, U_{t+1} = 0|Y_t = 0, U_t = 0] = H_1\,(0,0)$$
$$p_w\,(1,0) = P\,(W_{t+1} = 0|W_t = 1)$$
$$= P[Y_{t+1} = 0, U_{t+1} = 0|Y_t = 0, U_t = 1] = H_0\,(1,0)$$

Thus the transition probability matrix Q of the univariate Markov chain, $\{W_t\}$, becomes

$$Q = \begin{bmatrix} H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1) \\ H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1) \\ H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1) \\ H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1) \end{bmatrix} \tag{B.5}$$

and its marginal distribution is

$$P[W_t = 0] = H_0(0,0)\pi_0 + H_0(1,0)\pi_1;$$
$$P[W_t = 2] = H_1(0,0)\pi_0 + H_1(1,0)\pi_1$$
$$P[W_t = 1] = H_0(0,1)\pi_0 + H_0(1,1)\pi_1;$$
$$P[W_t = 3] = H_1(0,1)\pi_0 + H_1(1,1)\pi_1$$

B.2 Estimation of DARMA(1,1) models

The DARMA(1,1) process only takes values from $\{0,1\}$ so that both of its probability distribution π_0 and π_1 sum to one, that is, $\pi_0 + \pi_1 = 1$. Estimates of π_0 and π_1 may be obtained from the run length property. Let $\overline{T}_0$ be the mean run length of state 0 and $\overline{T}_1$ be the mean run length of state 1, which are estimated from the sample 0-1 series. Then π_0 and π_1 are given by (Buishand, 1978)

$$\pi_0 = \frac{\overline{T}_0}{\overline{T}_0 + \overline{T}_1}$$

and

$$\pi_1 = 1 - \pi_0$$

respectively. Chang et al. (1984a,b) noted that this estimator based on the occurrence probabilities of runs is conceptually reasonable since that the 0 and 1 spells govern the persistence characteristics of the series through the run lengths of 0 and 1.

Three parameters, π_0 (orπ_1), λ and β, need to be estimated for fitting a DARMA(1,1) model. For a DARMA(1,1) model, the method of moments may be used to obtain the estimate λ by matching the sample and model autocorrelation functions. For the DARMA(1,1) model, based on the autocorrelation function $\rho_k(Y) = c\lambda^{k-1}$, $k \geq 1$, where $c = (1-\beta)(\lambda + \beta - 2\lambda\beta)$ is the lag-1 autocorrelation coefficient, an estimator of λ may be the ratio of the second to the first sample autocorrelations, i.e.,

$$\hat{\lambda} = \frac{r(2)}{r(1)} \tag{B.6}$$

And, since the estimation of c is the lag-1 autocorrelation coefficient, parameter β may be estimated as

$$\hat{\beta} = \frac{(3\hat{\lambda}-1) - \sqrt{(3\hat{\lambda}-1)^2 - 4(2\hat{\lambda}-1)(\hat{\lambda}-\hat{c})}}{2(2\hat{\lambda}-1)} \tag{B.7}$$

B.3 Run length properties of DARMA(1,1) models

Here their findings are summarized. The probability distributions of run lengths of states $i = 0, 1$ denoted by T_i can be derived from the conditional probability

$$P[T_i = n] = P[Y_0 = 1-i, Y_1 = i, \ldots, Y_n = i, Y_{n+1} = 1-i | Y_0 = 1-i, Y_1 = i]; \quad n = 1, 2, \ldots$$

for DARMA(1,1) process, the probability distributions are (Chang et al., 1984b)

:

$$P[T_i = n] = P[W_0 = 2(1-i)][H_i^n(0) - H_i^{n+1}(0)]$$
$$+P[W_0 = 2(1-i)+1][H_i^n(1) - H_i^{n+1}(1)]/P[Y_0 = 1-i, Y_1 = i], i = 0, 1$$

where $H_i^n(j) = H_i^n(j,0) + H_i^n(j,1), i = 0, 1$, and $j = 0, 1$, in which $H_i^n(j,k) = j, k$ element of the n-step transition probability matrix, and

$$P(Y_0 = 1-i, Y_1 = i) = \sum_{k=0}^{1}\sum_{j=0}^{1} H_i(j,k)\left[\pi_j - \sum_{l=0}^{1} H_i(l,j)\pi_l\right]; \quad i = 0, 1$$

The expected values of the length of a run of 0 or 1 are given respectively by:

$$E[T_0] = \frac{\pi(0)[1 - \beta\lambda + \beta(1-\lambda-\beta+2\beta\lambda)\{1-\pi(0)\}]}{\{1-\pi(0)\}[1-\lambda(1-\beta)\{1-\beta\pi(0)\} - \beta\pi(0)\{1-\beta(1-\lambda)\}]}$$

$$E[T_1] = \frac{\pi(1)[1 - \beta\lambda + \beta(1-\lambda-\beta+2\beta\lambda)\{1-\pi(1)\}]}{\{1-\pi(1)\}[1-\lambda(1-\beta)\{1-\beta\pi(1)\} - \beta\pi(1)\{1-\beta(1-\lambda)\}]}$$

APPENDIX C

VALIDATION OF THE PROPOSED PROCEDURE TO ESTIMATE RETURN PERIOD IN STATIONARY SERIES

Validation of the proposed expressions to estimate return period in stationary serially dependent series illustrated in Chapter 3 has been carried out by simulating several AR(1) processes of length 3,000,000 years with mean $\mu_X = 700$, standard deviations $\sigma_X = 200$ and lag-1 autocorrelation coefficients $r_1 = 0, .1, .2, .4, .6, .8$. Furthermore, in order to verify the procedure for skewed data, log normal transformation has also been applied, yielding X_t series with skewness $\gamma = .5$. The threshold has been parameterized as $x_0 = \mu_x - \alpha\sigma_x$, with $\alpha = 0, .2, .5$. For easy of reference the drought analysis has been performed by using the following notation. The deficit threshold D_0 is defined as a fraction δ of the threshold, i.e.

$$D_0 = \delta x_o$$

where $\delta > 0$ is called the deficit coefficient.

Return period of different critical droughts has been computed by means of eq.(3.67) where $P[E]$ has been computed by means of the joint cdf (3.52), estimating the parameters through the empirical relationships previously found. Also, the expected value of drought and non-drought length has been computed assuming a DARMA(1,1) model for the the clipped AR(1) process, using the expressions given in Appendix B. This allows one to quickly compute return period T corresponding to critical accumulated deficit D_o and length l_o only on the basis of the variance and autocorrelation coefficient of the series and of the threshold level x_0.

In Tables C.I-C.VI, the normalized difference between the return period computed by means of the empirical approximations (eq. 3.67) and the one obtained by simulation is reported for drought events defined by $\{ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$ and for different values of the lag-1 autocorrelation coefficient r_1, of the threshold coefficient $\alpha = .0, .2, .5$ and of the skewness $\gamma = 0, .5, 1$. In order to limit sampling errors when computing return period from the generated sample, the latter has been computed only when at least 50 interarrival times between critical droughts where available. When this is not the case, a "-" has been inserted in the tables. Also, the symbol ∞ is related to drought events for which the probability $P[E]$ was equal to 0, using double precision, and consequently the return period could not be computed from eq. (3.67). Finally the errors in the tables are shown in different fonts, according to the underlying value of the return period computed from the generated sample. More specifically, bold font indicates return periods less than 200 years, a roman font indicates return periods between 200 and 1000 years, and an italic font indicates return periods greater than 1000.

From Tables C.I-C.VI it can be inferred that the differences between theoretical and simulated return periods are generally lower than 10% for a for a wide range of drought lengths, deficit coefficient δ and autocorrelation coefficient r_1. In particular the errors are generally very low for return periods below 200 years, as indicated by bold font. Occasional very high values of the errors are noted, but for return periods presumably much greater than 1000 years. This probably is due to the combination of two factors namely the errors in the empirical expressions yielding the moments, especially for long droughts, and some inadequacy of the beta distribution to model appropriately the tail of the distribution of accumulated deficit. The Tables C.I-C.VI shows also that errors tend to increase with increasing r_1 values, although only for $r_1 = .8$ the errors become mostly above 10%.

In Tables C.VII-C.XII the results related to the case of drought events defined by $\{ D > D_0$ and $L \geq l_0,\ l_0 = 1, 2, ...\}$ are shown. Again it can be inferred that for return periods less than 200 years the errors are generally lower than 10%.It may also be worthwhile to note how the errors do not change with L for high values of the δ coefficient. This is consistent with the fact that for high δ values, i.e. drought with very high accumulated deficit, the return period does not depend on the drought length. Similar results are obtained when critical drought events defined by $\{ I > I_0$ and $L = l_0,\ l_0 = 1, 2, ...\}$ or $\{ I > I_0$ and $L \geq l_0,\ l_0 = 1, 2, ...\}$ are considered (see Tables C.XIII-C.XVIII and C.XIX-C.XXIV respectively).

Table C.I: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ D > D_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

		δ						
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**-0.1**	**2.8**	*-34.0*	-	∞	∞	∞
	2	**0.1**	**2.8**	-2.5	*-35.4*	-	-	∞
	3	**-0.3**	**-0.7**	**2.1**	-10.7	*-36.8*	-	-
	4	**-0.2**	**-0.9**	**2.1**	-0.5	*-14.7*	-	-
$r_1 = 0.0$	5	**1.1**	**0.5**	**1.4**	2.6	*-2.9*	*-22.7*	-
	6	1.7	1.6	1.7	4.2	*4.4*	*4.6*	-
	7	-1.4	-1.4	-1.1	-0.9	*0.4*	*-3.0*	-
	8	*0.1*	*0.1*	*-0.5*	*-1.4*	*-0.2*	*6.0*	-
	9	*9.1*	*9.1*	*8.8*	*7.7*	*8.9*	*12.7*	-
	10	*0.5*	*0.5*	*0.6*	*1.7*	*-0.4*	*-0.7*	-
	1	**0.1**	**-4.7**	*-44.5*	-	∞	∞	∞
	2	**-0.1**	**5.4**	12.5	*8.7*	-	-	∞
	3	**0.1**	**0.9**	**11.5**	16.2	*11.3*	-	-
	4	**-0.4**	**-1.2**	**2.5**	9.9	*18.0*	*26.2*	-
$r_1 = 0.2$	5	**-0.2**	**-0.7**	**-0.6**	4.3	9.7	*16.4*	-
	6	**0.1**	**-0.0**	**-1.0**	-0.2	1.9	*-0.4*	-
	7	1.0	1.0	0.2	-2.0	-2.6	*-3.0*	-
	8	-0.6	-0.6	-1.1	-2.6	-6.0	*-7.2*	-
	9	0.4	0.4	0.4	*-0.6*	*-5.2*	*-11.3*	*-23.4*
	10	*2.2*	*2.2*	*2.2*	*1.8*	*-0.1*	*-6.0*	*-40.2*
	1	**-0.0**	-16.8	-	-	∞	∞	∞
	2	**0.9**	**5.3**	12.0	*4.0*	-	-	∞
	3	**-0.7**	**1.7**	**13.0**	*29.9*	*59.8*	-	-
	4	**-0.9**	**-0.7**	**4.5**	15.7	*35.0*	*47.0*	-
$r_1 = 0.4$	5	**-0.7**	**-0.8**	**0.5**	5.7	8.5	*26.4*	-
	6	**-1.3**	**-1.3**	**-2.0**	-0.5	2.0	*8.2*	-
	7	1.3	1.3	1.0	-1.7	-2.7	-1.2	*7.7*
	8	1.8	1.8	1.7	0.1	-4.1	-9.8	*-30.7*
	9	0.4	0.4	0.4	-0.5	-2.9	-6.6	*-27.2*
	10	1.6	1.6	1.5	1.1	-1.9	-5.6	*-33.0*
	1	**0.1**	*-30.8*	-	-	∞	∞	∞
	2	**5.7**	**7.8**	*-9.5*	-	-	-	∞
	3	**-1.0**	**2.9**	10.6	*16.4*	-	-	-
	4	**-5.3**	**-3.7**	**3.4**	11.6	*26.8*	*46.8*	-
$r_1 = 0.6$	5	**-5.0**	**-4.5**	**-1.0**	6.0	17.0	*39.7*	-
	6	**-3.8**	**-3.6**	**-3.0**	-1.0	1.2	*9.4*	-
	7	**-3.1**	**-3.1**	**-3.2**	-3.1	-4.2	-5.2	*11.0*
	8	-2.5	-2.5	-2.4	-3.4	-4.8	-8.0	*-7.0*
	9	-0.4	-0.4	-0.4	-0.8	-2.5	-5.9	*-17.8*
	10	0.9	0.9	0.9	0.7	-0.5	-3.3	*-21.3*
	1	**-0.1**	-	-	∞	∞	∞	∞
	2	**25.8**	*41.2*	-	-	-	-	∞
	3	**5.8**	9.1	*39.3*	-	-	-	∞
	4	**-6.8**	**-5.5**	6.8	*40.9*	-	-	-
$r_1 = 0.8$	5	**-12.8**	**-13.0**	-9.9	*3.4*	*30.2*	-	-
	6	**-15.9**	**-16.2**	-15.2	-9.0	*16.0*	*58.0*	-
	7	-14.9	-15.0	-16.4	-14.7	-7.1	*6.3*	-
	8	-15.7	-15.7	-16.3	-17.2	-14.9	-6.2	*63.4*
	9	-13.0	-13.0	-13.2	-15.3	-16.4	-15.4	*24.8*
	10	-12.6	-12.6	-12.7	-14.2	-16.5	-19.4	*-2.1*

Table C.II: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ D > D_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.1**	**2.5**	*-35.5*	-	∞	∞	∞
	2	**0.3**	**3.0**	-3.4	*-27.6*	-	-	∞
	3	**-0.7**	**-0.6**	**0.5**	*-12.7*	*-31.9*	-	-
	4	**-0.1**	**-0.6**	3.0	4.2	*-4.5*	-	-
$r_1 = 0.0$	5	-2.2	-2.7	-2.2	0.7	*-2.9*	*-6.6*	-
	6	-0.9	-1.2	-1.7	*-0.5*	*-2.1*	*-6.5*	-
	7	*-0.2*	*-0.1*	*0.8*	*3.4*	*1.4*	*-3.1*	-
	8	*-1.8*	*-2.0*	*-2.8*	*-1.3*	*0.2*	*2.8*	-
	9	*-8.1*	*-8.1*	*-8.0*	*-10.6*	*-11.9*	*-7.3*	-
	10	*-9.2*	*-9.2*	*-9.1*	*-12.8*	*-13.4*	-	-
	1	**0.4**	**-4.8**	*-27.6*	-	∞	∞	∞
	2	**-0.9**	**4.0**	13.7	*8.6*	-	-	∞
	3	**-0.2**	**0.6**	**11.5**	*22.4*	*38.7*	-	-
	4	**-0.2**	**-0.9**	**4.1**	14.0	*19.4*	*42.6*	-
$r_1 = 0.2$	5	**-0.7**	**-1.4**	-1.1	2.6	*7.6*	*10.0*	-
	6	0.3	0.1	-0.5	-0.6	*-0.7*	*7.7*	-
	7	6.0	6.0	5.2	2.5	*2.7*	*3.4*	-
	8	*2.8*	*2.8*	*2.5*	*0.9*	*0.2*	*5.2*	-
	9	*3.1*	*3.1*	*2.6*	*1.8*	*2.4*	*-1.6*	-
	10	*6.4*	*6.4*	*6.2*	*5.6*	*0.1*	*-7.8*	-
	1	**0.7**	-16.1	-	-	∞	∞	∞
	2	**-0.1**	**4.7**	15.6	*34.2*	-	-	∞
	3	**-2.1**	**0.0**	**12.7**	*30.4*	*72.3*	-	-
	4	**-1.5**	**-1.4**	**4.2**	14.9	*36.4*	*59.3*	-
$r_1 = 0.4$	5	**0.3**	**-0.0**	**1.2**	7.4	15.7	*26.6*	-
	6	0.4	0.3	0.6	0.8	5.3	*10.3*	-
	7	1.3	1.3	0.9	0.2	-0.4	*-1.3*	-
	8	1.2	1.1	1.0	-0.1	-4.4	*-12.2*	*-19.9*
	9	*3.0*	*3.0*	*2.9*	*1.0*	*-3.1*	*-5.8*	*-22.8*
	10	*2.0*	*2.0*	*2.0*	*1.5*	*-0.3*	*-6.5*	*-33.5*
	1	**0.4**	*-28.7*	-	-	∞	∞	∞
	2	**3.9**	**6.6**	*-1.7*	-	-	-	∞
	3	**-1.7**	**2.9**	11.3	*21.0*	*49.5*	-	-
	4	**-5.2**	**-3.3**	**3.6**	9.8	*26.2*	*50.9*	-
$r_1 = 0.6$	5	**-3.9**	**-3.4**	**0.0**	6.2	*18.7*	*37.4*	-
	6	**-3.6**	**-3.4**	**-2.0**	0.7	5.4	*10.3*	-
	7	-1.0	-0.9	-0.6	0.0	-0.8	1.0	*22.3*
	8	-1.2	-1.2	-1.1	-1.6	-4.1	-7.3	*-10.2*
	9	-0.4	-0.4	-0.5	-1.2	-3.2	-5.6	*-22.4*
	10	1.3	1.3	1.3	1.2	-0.9	-3.7	*-26.6*
	1	**-0.3**	-	-	∞	∞	∞	∞
	2	**21.6**	32.7	-	-	-	-	∞
	3	**2.7**	8.7	*23.7*	-	-	-	∞
	4	**-8.6**	**-6.6**	5.1	*27.9*	*98.7*	-	-
$r_1 = 0.8$	5	**-13.5**	**-12.9**	-8.4	*9.8*	*36.3*	*99.6*	-
	6	**-13.0**	**-13.0**	-11.6	-4.8	*8.8*	*40.0*	-
	7	-15.1	-15.1	-14.7	-12.8	-5.3	*8.3*	-
	8	-12.5	-12.5	-12.8	-13.3	-10.0	*-3.7*	*68.3*
	9	-11.8	-11.8	-11.9	-13.7	-15.3	-14.3	*19.4*
	10	-7.2	-7.2	-7.3	-8.3	-11.5	-14.3	*-3.7*

Table C.III: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ D > D_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**0.5**	*-22.0*	-	∞	∞	∞
	2	**0.0**	**2.2**	-5.2	*-18.0*	-	-	∞
	3	**-1.0**	**-0.5**	1.9	*-14.7*	-	-	-
	4	0.7	0.3	5.0	*-1.3*	*-30.5*	-	-
$r_1 = 0.0$	5	-0.8	-1.9	*-1.1*	*-5.8*	*10.9*	-	-
	6	*9.2*	*8.9*	*11.4*	*16.2*	*11.4*	-	-
	7	*3.6*	*3.8*	*3.2*	*8.3*	*13.6*	-	-
	8	*11.9*	*11.9*	*12.0*	*16.6*	-	-	-
	9	-	-	-	-	-	-	-
	10	-	-	-	-	-	-	-
	1	**0.8**	**-4.3**	*-34.9*	-	∞	∞	∞
	2	**-2.6**	**1.7**	11.2	*31.1*	-	-	∞
	3	**-0.1**	**0.9**	13.9	*35.5*	*60.8*	-	-
	4	**1.1**	**0.3**	6.8	*19.2*	*39.9*	-	-
$r_1 = 0.2$	5	2.2	1.9	3.2	*10.2*	*17.2*	*32.8*	-
	6	7.1	6.7	*5.7*	*6.0*	*12.2*	*27.8*	-
	7	*6.0*	*5.9*	*6.2*	*6.2*	*8.3*	*22.9*	-
	8	*6.8*	*6.8*	*5.8*	*4.0*	*5.3*	*11.8*	-
	9	*8.1*	*8.1*	*7.6*	*2.4*	*4.9*	*3.6*	-
	10	*25.1*	*25.1*	*25.1*	*21.2*	*24.9*	-	-
	1	**1.6**	-12.4	-	-	∞	∞	∞
	2	**-3.1**	**2.2**	17.3	*44.9*	-	-	∞
	3	**-2.5**	**0.2**	16.2	*43.0*	*88.1*	-	-
	4	**-0.4**	**-0.2**	7.1	20.2	*38.1*	*92.3*	-
$r_1 = 0.4$	5	1.4	1.1	3.5	11.5	*22.7*	*41.6*	-
	6	5.0	4.9	4.8	6.4	*10.7*	*24.9*	-
	7	8.0	7.9	7.5	*6.4*	*4.7*	*14.5*	-
	8	*7.7*	*7.7*	*7.5*	*5.3*	*2.2*	*0.5*	-
	9	*16.0*	*16.0*	*16.0*	*14.5*	*7.7*	*7.2*	-
	10	*7.3*	*7.3*	*7.3*	*5.8*	*5.0*	*4.7*	-
	1	**0.7**	*-28.2*	-	-	∞	∞	∞
	2	**1.6**	**5.0**	*-1.3*	-	-	-	∞
	3	**-2.8**	**1.6**	10.9	*29.9*	-	-	-
	4	**-3.7**	**-1.4**	6.2	21.7	*40.4*	*89.8*	-
$r_1 = 0.6$	5	**-3.5**	**-2.9**	2.8	11.3	*24.9*	*52.2*	-
	6	-1.8	-1.7	-0.2	3.6	10.3	*20.5*	-
	7	0.8	0.7	1.7	2.8	4.6	*11.8*	*38.5*
	8	-2.0	-2.0	-2.1	-3.0	*-3.4*	*-2.5*	*1.3*
	9	*8.2*	*8.2*	*8.2*	*7.7*	*5.8*	*3.4*	*2.5*
	10	*13.7*	*13.7*	*13.8*	*14.1*	*11.8*	*7.0*	*-20.1*
	1	**-0.3**	-	-	∞	∞	∞	∞
	2	**17.3**	31.0	-	-	-	-	∞
	3	**-0.7**	5.0	*17.3*	-	-	-	-
	4	**-9.3**	**-6.3**	5.6	*31.9*	*55.2*	-	-
$r_1 = 0.8$	5	**-13.7**	**-12.9**	-7.9	*5.1*	*36.4*	*70.6*	-
	6	-13.4	-13.1	-10.2	-4.3	*15.4*	*48.4*	-
	7	-12.0	-11.9	-11.3	-9.6	-0.2	*21.4*	-
	8	-9.2	-9.3	-9.3	-8.4	-5.3	*4.0*	*88.6*
	9	-6.8	-6.7	-6.8	-6.8	-7.0	*-4.5*	*21.4*
	10	-3.2	-3.2	-3.2	-3.7	-5.1	*-6.8*	*18.2*

Table C.IV: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

		δ						
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.1**	**1.5**	-	-	∞	∞	∞
	2	**0.3**	**6.0**	-17.5	-	-	-	∞
	3	**-0.1**	**-0.4**	**3.7**	*-31.4*	-	-	-
	4	**-0.5**	**-1.9**	**1.8**	-5.7	*-47.4*	-	-
$r_1 = 0.0$	5	**-0.2**	**-0.9**	**0.1**	4.1	*-14.0*	-	-
	6	**0.4**	**0.1**	0.1	5.6	*0.9*	*-25.7*	-
	7	-1.9	-2.0	-2.6	-0.3	*-0.4*	*-19.7*	-
	8	-1.1	-1.1	-2.1	-2.5	*-2.5*	*-8.6*	-
	9	*-3.1*	*-3.1*	*-3.1*	*-2.3*	*-1.5*	*-2.6*	-
	10	*-3.3*	*-3.3*	*-3.2*	*-3.7*	*-2.2*	*-0.2*	-
	1	**-0.3**	-8.4	-	-	∞	∞	∞
	2	**0.5**	**7.3**	0.3	-	-	-	∞
	3	**0.1**	**0.0**	**13.5**	*2.2*	-	-	-
	4	**-0.1**	**-1.6**	**3.7**	10.3	*-12.8*	-	-
$r_1 = 0.2$	5	**1.0**	**0.3**	**-0.2**	2.9	*1.6*	*-25.7*	-
	6	**-0.9**	**-1.1**	**-2.9**	-3.2	-5.4	*-19.0*	-
	7	-0.4	-0.4	-1.7	-4.5	-6.9	*-15.5*	-
	8	1.1	1.1	0.4	-2.8	-8.5	*-15.2*	-
	9	-1.6	-1.6	-1.8	-4.1	-7.8	*-14.6*	-
	10	*-7.1*	*-7.1*	*-7.2*	*-8.0*	*-10.6*	*-18.6*	-
	1	**-0.4**	-23.1	-	-	∞	∞	∞
	2	**2.0**	**7.8**	*5.1*	-	-	-	∞
	3	**-1.3**	**-0.4**	14.6	*12.7*	-	-	-
	4	**-0.7**	**-1.8**	**4.2**	14.5	*2.7*	-	-
$r_1 = 0.4$	5	**-1.6**	**-2.2**	**-2.2**	1.9	*6.8*	*-8.7*	-
	6	**0.0**	**-0.2**	**-2.1**	-3.0	-4.4	*-9.7*	-
	7	**0.8**	**0.7**	**-0.7**	-4.2	-9.6	*-15.5*	-
	8	0.8	0.8	0.2	-3.2	-9.2	-17.1	-
	9	0.7	0.7	0.5	-1.3	-7.4	-19.2	*-56.5*
	10	-1.6	-1.6	-1.6	-2.5	-6.7	-16.4	*-53.8*
	1	**-0.2**	*-43.1*	-	-	∞	∞	∞
	2	**6.5**	11.4	*-25.6*	-	-	-	∞
	3	**-0.7**	**1.6**	10.5	*-9.2*	-	-	-
	4	**-4.3**	**-4.6**	3.2	*10.6*	*-2.4*	-	-
$r_1 = 0.6$	5	**-6.3**	**-6.6**	**-5.3**	1.0	*6.7*	*2.1*	-
	6	**-4.6**	**-4.7**	**-5.8**	-5.6	-4.9	*-2.5*	-
	7	**-3.5**	**-3.5**	**-4.2**	-6.9	-9.1	-13.0	-
	8	-1.0	-1.0	-1.5	-4.0	-8.3	-13.5	*-46.1*
	9	-1.2	-1.2	-1.6	-3.4	-8.5	-16.8	*-45.8*
	10	0.6	0.6	0.4	-0.3	-3.8	-11.2	*-48.4*
	1	**0.3**	-	-	∞	∞	∞	∞
	2	**27.9**	*85.2*	-	-	-	∞	∞
	3	**6.7**	17.6	*73.2*	-	-	-	∞
	4	**-4.7**	**-4.7**	*19.2*	*73.6*	-	-	-
$r_1 = 0.8$	5	**-12.3**	**-13.3**	-7.7	*19.9*	*62.6*	-	-
	6	**-18.0**	**-18.6**	-18.8	-7.5	*15.6*	*43.7*	-
	7	-17.4	-17.7	-20.1	-19.6	*-8.4*	*7.1*	-
	8	-16.6	-16.7	-18.3	-21.3	-19.5	*-13.3*	-
	9	-14.5	-14.5	-15.1	-18.0	-21.3	-18.0	*21.3*
	10	-14.4	-14.4	-14.6	-16.7	-21.1	-23.0	*-13.3*

Table C.V: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L = l_0$, $l_0 = 1, 2, \ldots\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.1**	**-0.8**	-	-	∞	∞	∞
	2	**0.1**	**5.6**	-25.6	-	-	-	∞
	3	**-0.9**	**-0.1**	0.3	*-44.8*	-	-	-
	4	**0.6**	**-0.7**	4.7	*-10.5*	-	-	-
$r_1 = 0.0$	5	**-1.6**	**-2.3**	-0.5	*2.9*	*-18.9*	-	-
	6	2.7	2.3	2.5	*7.8*	*-1.2*	-	-
	7	-2.2	-2.3	-3.0	*1.2*	*2.7*	-	-
	8	*0.2*	*0.1*	*-0.1*	*-0.1*	*3.2*	*-2.9*	-
	9	*-1.1*	*-1.1*	*-1.3*	*-0.8*	*5.8*	*-3.3*	-
	10	*2.9*	*2.9*	*3.1*	*1.1*	*-1.2*	*-2.4*	-
	1	**0.3**	-10.6	-	-	∞	∞	∞
	2	**-0.5**	**7.2**	-5.2	-	-	-	∞
	3	**-0.3**	**0.2**	14.0	*-1.0*	-	-	-
	4	**0.2**	**-1.3**	**6.3**	*10.0*	*-0.7*	-	-
$r_1 = 0.2$	5	**-0.7**	**-1.4**	**0.3**	5.9	*-2.2*	-	-
	6	2.8	2.5	0.7	0.4	*-2.5*	*-24.7*	-
	7	-2.1	-2.1	-3.9	-5.6	*-7.6*	*-22.2*	-
	8	1.1	1.1	0.4	*-3.0*	*-5.3*	*-10.9*	-
	9	*-0.0*	*-0.0*	*-0.9*	*-3.1*	*-6.7*	*-10.1*	-
	10	*10.2*	*10.2*	*9.7*	*7.9*	*5.3*	*-4.7*	-
	1	**0.7**	-26.5	-	-	∞	∞	∞
	2	**-0.0**	**7.2**	*-6.8*	-	-	-	∞
	3	**-1.6**	**-0.2**	14.9	*13.6*	-	-	-
	4	**-1.7**	**-2.9**	**6.1**	16.6	*9.8*	-	-
$r_1 = 0.4$	5	**-0.7**	**-1.5**	**-1.7**	7.0	*10.2*	*2.6*	-
	6	**0.3**	**0.1**	**-1.8**	-3.1	-3.3	*-5.3*	-
	7	1.2	1.1	-0.7	-3.5	-9.5	*-18.6*	-
	8	4.0	4.0	3.3	-1.0	-7.5	*-14.5*	-
	9	3.2	3.2	3.1	0.3	-6.2	*-15.9*	-
	10	*-0.0*	*-0.0*	*-0.1*	*-1.1*	-5.9	*-18.3*	*-55.0*
	1	**0.0**	*-52.6*	-	-	∞	∞	∞
	2	**4.6**	10.4	*-22.1*	-	-	-	∞
	3	**-1.7**	**1.8**	11.2	*-13.7*	-	-	-
	4	**-4.3**	**-3.7**	4.4	*14.1*	*8.0*	-	-
$r_1 = 0.6$	5	**-3.0**	**-3.1**	**-2.0**	4.7	*8.7*	*-7.7*	-
	6	**-4.6**	**-4.7**	**-5.1**	-2.3	-4.6	*-4.6*	-
	7	-4.0	-4.1	-4.8	-7.4	-9.7	*-11.6*	-
	8	0.3	0.2	-0.2	-3.7	-8.7	-13.6	-
	9	2.3	2.3	2.1	0.6	-4.3	-13.1	*-45.4*
	10	1.9	1.9	1.9	1.2	-4.0	-14.4	*-45.0*
	1	**-0.3**	-	-	∞	∞	∞	∞
	2	**25.2**	*73.2*	-	-	-	∞	∞
	3	**2.2**	11.6	*48.4*	-	-	-	∞
	4	**-8.4**	**-7.3**	*15.3*	*38.1*	-	-	-
$r_1 = 0.8$	5	**-11.8**	**-12.6**	-3.4	*25.5*	*42.2*	-	-
	6	**-14.9**	**-15.4**	-15.2	-5.9	*15.4*	*24.9*	-
	7	-15.6	-15.8	-17.4	-15.7	*-5.4*	*16.4*	-
	8	-13.3	-13.4	-14.6	-16.3	-15.8	*-3.2*	-
	9	-13.2	-13.2	-14.1	-16.5	-19.1	-18.7	*-5.1*
	10	-11.7	-11.7	-11.8	-14.3	-18.1	-21.2	*-20.8*

Table C.VI: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\, D > D_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**-0.1**	**-7.3**	-	-	∞	∞	∞
	2	**0.5**	**7.4**	*-37.7*	-	-	-	∞
	3	**-0.0**	**1.0**	-5.2	-	-	-	-
	4	**-0.5**	-0.9	2.0	*-20.1*	-	-	-
$r_1 = 0.0$	5	-1.2	-2.8	*-0.5*	*-10.8*	-	-	-
	6	*-1.3*	*-1.9*	*-3.5*	*-6.8*	-	-	-
	7	*-2.3*	*-2.9*	*-3.3*	*4.3*	-	-	-
	8	*-10.1*	*-10.1*	*-14.0*	*-17.4*	-	-	-
	9	-	-	-	-	-	-	-
	10	-	-	-	-	-	-	-
	1	**0.7**	-15.8	-	-	∞	∞	∞
	2	**-2.1**	**7.3**	*-8.7*	-	-	-	∞
	3	**0.2**	**2.8**	16.5	*-9.1*	-	-	-
	4	**0.8**	**-0.4**	9.5	*7.1*	-	-	-
$r_1 = 0.2$	5	1.4	0.7	4.8	*12.0*	*16.1*	-	-
	6	4.0	3.4	3.0	*11.7*	*9.5*	-	-
	7	*1.9*	*1.7*	*0.8*	*-8.3*	*-7.2*	-	-
	8	*5.6*	*5.4*	*2.2*	*-2.4*	*-3.4*	-	-
	9	*10.5*	*10.5*	*10.6*	*5.9*	*-1.0*	-	-
	10	*-2.3*	*-2.3*	*-2.2*	*-8.8*	-	-	-
	1	**1.4**	-30.7	-	-	∞	∞	∞
	2	**-2.3**	**7.0**	*-11.3*	-	-	-	∞
	3	**-2.5**	**0.6**	19.6	*21.1*	-	-	-
	4	**-2.2**	**-2.6**	9.4	*20.2*	*26.5*	-	-
$r_1 = 0.4$	5	**2.9**	**1.9**	4.4	*13.5*	*28.3*	-	-
	6	3.1	2.6	-0.4	0.6	*3.8*	*0.3*	-
	7	3.8	3.7	1.7	*-0.1*	*-0.2*	*-1.3*	-
	8	*10.2*	*10.2*	*9.3*	*3.8*	*-4.8*	*-14.2*	-
	9	*11.3*	*11.3*	*11.3*	*5.2*	*-6.0*	*-15.1*	-
	10	*7.4*	*7.4*	*7.5*	*5.1*	*-1.1*	*-14.6*	-
	1	**0.9**	*-55.6*	-	-	∞	∞	∞
	2	**1.6**	9.8	*-43.6*	-	-	-	∞
	3	**-2.9**	**2.8**	12.9	*-12.6*	-	-	-
	4	**-4.4**	**-3.0**	6.3	*16.7*	*8.6*	-	-
$r_1 = 0.6$	5	**-4.3**	**-4.4**	0.9	9.1	*14.5*	*17.3*	-
	6	-2.1	-2.2	-2.0	1.5	*6.0*	*7.9*	-
	7	-0.7	-0.7	-1.3	-3.4	*-3.6*	*-4.4*	-
	8	2.5	2.5	2.1	-0.6	*-4.8*	*-6.4*	-
	9	11.0	11.0	10.7	*8.5*	*-0.1*	*-6.8*	-
	10	*6.1*	*6.1*	*5.7*	*4.5*	*-0.2*	*-9.0*	-
	1	**0.0**	-	-	∞	∞	∞	∞
	2	**18.1**	*55.6*	-	-	-	∞	∞
	3	**-0.7**	10.4	*36.0*	-	-	-	∞
	4	**-9.2**	**-6.3**	*19.5*	*56.4*	-	-	-
$r_1 = 0.8$	5	**-13.0**	**-12.7**	-2.8	*25.1*	*57.1*	-	-
	6	-15.0	-15.4	-12.7	*0.9*	*29.5*	*70.2*	-
	7	-13.9	-14.0	-14.2	-12.2	*0.4*	*19.6*	-
	8	-12.9	-12.9	-13.8	-14.4	*-11.8*	*-1.8*	-
	9	-5.7	-5.7	-6.3	-7.6	-10.0	*-7.0*	-
	10	-6.0	-6.0	-6.0	-8.9	-13.2	*-15.9*	*-13.6*

Table C.VII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, \ldots\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

		δ						
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**1.1**	**0.6**	**-0.5**	-1.1	-0.1	*17.6*
	2	**0.1**	**0.8**	**1.2**	**-0.5**	-1.1	-0.1	*17.6*
	3	**0.1**	**-0.3**	**1.7**	**0.3**	-1.0	-0.0	*17.6*
	4	**0.5**	**-0.0**	**1.5**	**1.5**	-0.1	0.1	*17.6*
$r_1 = 0.0$	5	**1.1**	**0.8**	**1.2**	**2.1**	1.2	0.7	*17.6*
	6	**1.1**	**1.0**	**1.0**	**1.9**	2.2	2.5	*17.6*
	7	0.5	0.5	0.4	0.4	1.3	*2.2*	*18.1*
	8	2.3	2.3	1.9	1.4	1.9	*3.9*	*19.1*
	9	*4.6*	*4.6*	*4.4*	*4.0*	*3.3*	*3.0*	*23.7*
	10	*0.2*	*0.2*	*0.2*	*0.5*	*-1.0*	*-2.4*	*21.1*
	1	**0.0**	**0.9**	**3.3**	**3.0**	**0.7**	-3.8	*-26.1*
	2	**-0.1**	**1.4**	**3.4**	**3.0**	**0.7**	-3.8	*-26.1*
	3	**-0.0**	**-0.0**	**2.7**	**2.9**	**0.7**	-3.8	*-26.1*
	4	**-0.1**	**-0.6**	**0.5**	**2.0**	**0.6**	-3.8	*-26.1*
$r_1 = 0.2$	5	**0.1**	**-0.2**	**-0.5**	**0.4**	**-0.5**	-4.2	*-26.1*
	6	**0.3**	**0.3**	**-0.4**	**-1.0**	**-2.2**	-5.4	*-26.1*
	7	**0.5**	**0.5**	**0.0**	**-1.4**	-3.5	-6.1	*-26.0*
	8	0.1	0.1	-0.1	-1.0	-3.9	-7.0	*-26.5*
	9	0.7	0.7	0.7	0.1	-2.7	-6.8	*-26.0*
	10	1.0	1.0	1.0	0.7	-0.9	-4.5	*-26.1*
	1	**0.0**	**0.5**	**3.1**	**3.5**	**1.1**	**-1.8**	-22.1
	2	**0.0**	**1.2**	**3.1**	**3.5**	**1.1**	**-1.8**	-22.1
	3	**-0.5**	**0.3**	**2.8**	**3.5**	**1.1**	**-1.8**	-22.1
	4	**-0.3**	**-0.3**	**1.2**	**2.6**	**0.8**	**-1.9**	-22.1
$r_1 = 0.4$	5	**-0.0**	**-0.1**	**0.2**	**0.9**	**-0.4**	**-2.2**	-22.1
	6	**0.3**	**0.3**	**0.0**	**-0.4**	**-1.5**	**-3.4**	-22.2
	7	**1.2**	**1.2**	**1.1**	**-0.4**	**-2.4**	**-4.8**	-22.4
	8	**1.2**	**1.2**	**1.1**	**0.3**	**-2.2**	**-5.5**	-23.0
	9	**0.8**	**0.8**	**0.8**	**0.4**	-1.4	-4.1	-22.5
	10	1.1	1.1	1.1	0.8	-0.6	-3.1	-21.9
	1	**0.0**	**-0.6**	**0.7**	**1.8**	**1.6**	**0.3**	**-10.3**
	2	**-0.1**	**-0.3**	**0.7**	**1.8**	**1.6**	**0.3**	**-10.3**
	3	**-2.3**	**-1.1**	**0.8**	**1.8**	**1.6**	**0.3**	**-10.3**
	4	**-2.7**	**-2.1**	**0.1**	**1.6**	**1.6**	**0.4**	**-10.3**
$r_1 = 0.6$	5	**-1.8**	**-1.6**	**-0.4**	**1.0**	**1.2**	**0.2**	**-10.3**
	6	**-0.6**	**-0.5**	**-0.3**	**0.3**	**0.2**	**-0.5**	**-10.3**
	7	**0.6**	**0.6**	**0.6**	**0.6**	**0.1**	**-1.1**	**-10.4**
	8	**2.0**	**2.0**	**2.0**	**1.7**	**1.1**	**-0.6**	**-10.7**
	9	**3.6**	**3.6**	**3.6**	**3.5**	**2.8**	**1.0**	**-10.8**
	10	**5.1**	**5.1**	**5.1**	**5.0**	**4.5**	**2.8**	-10.2
	1	**0.0**	**-5.9**	**-6.6**	**-5.2**	**-3.0**	**-1.7**	**0.9**
	2	**0.0**	**-5.9**	**-6.6**	**-5.2**	**-3.0**	**-1.7**	**0.9**
	3	**-5.8**	**-6.7**	**-6.6**	**-5.2**	**-3.0**	**-1.7**	**0.9**
	4	**-8.2**	**-8.1**	**-7.0**	**-5.2**	**-3.0**	**-1.7**	**0.9**
$r_1 = 0.8$	5	**-8.5**	**-8.5**	**-7.6**	**-5.5**	**-3.1**	**-1.7**	**0.9**
	6	**-7.7**	**-7.7**	**-7.3**	**-5.8**	**-3.3**	**-1.7**	**0.9**
	7	**-5.9**	**-5.9**	**-6.1**	**-5.5**	**-3.9**	**-2.2**	**0.8**
	8	**-4.0**	**-4.0**	**-4.1**	**-4.3**	**-3.6**	**-2.4**	**0.7**
	9	**-1.6**	**-1.6**	**-1.6**	**-2.0**	**-2.3**	**-2.1**	**0.5**
	10	**0.8**	**0.8**	**0.8**	**0.6**	**-0.0**	**-0.7**	**0.2**

Table C.VIII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

		δ						
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**1.1**	**-1.1**	**-2.0**	-4.6	*-6.8*	-
	2	**-0.1**	**0.7**	**-0.3**	**-2.0**	-4.6	*-6.8*	-
	3	**-0.7**	**-1.0**	**0.3**	-1.0	-4.4	*-6.8*	-
	4	**-0.8**	**-1.3**	**0.2**	0.9	-3.3	*-6.6*	-
$r_1 = 0.0$	5	**-1.9**	**-2.3**	**-2.0**	-0.3	-3.1	*-5.2*	-
	6	-1.5	-1.6	-1.8	-0.9	*-3.0*	*-4.9*	-
	7	-2.2	-2.2	-1.9	-1.2	*-3.4*	*-4.2*	-
	8	*-4.8*	*-4.9*	*-5.4*	*-5.9*	*-6.6*	*-4.4*	-
	9	*-8.7*	*-8.7*	*-8.7*	*-11.1*	*-12.4*	*-8.1*	-
	10	*-9.1*	*-9.1*	*-9.0*	*-11.3*	*-12.2*	*-7.9*	-
	1	**0.0**	**0.7**	**5.0**	**5.6**	3.6	2.0	*-29.7*
	2	**-0.5**	**1.4**	**5.1**	**5.6**	3.6	2.0	*-29.7*
	3	**0.0**	**0.1**	**4.2**	**5.5**	3.6	2.0	*-29.7*
	4	**0.3**	**-0.3**	**1.8**	**3.9**	3.0	1.8	*-29.7*
$r_1 = 0.2$	5	**0.7**	**0.3**	**0.4**	**1.3**	1.5	1.0	*-29.7*
	6	**2.2**	**2.1**	**1.5**	0.6	0.2	0.2	*-29.7*
	7	4.1	4.1	3.5	1.5	0.6	-1.2	*-29.8*
	8	2.2	2.2	1.9	0.6	-0.6	*-2.7*	*-28.9*
	9	*1.6*	*1.6*	*1.3*	*0.4*	*-1.1*	*-6.7*	*-30.2*
	10	*0.2*	*0.2*	*0.1*	*-0.8*	*-4.0*	*-9.9*	*-30.4*
	1	**0.0**	**0.2**	**4.5**	**5.8**	**5.0**	0.0	*-20.3*
	2	**-0.6**	**1.1**	**4.5**	**5.8**	**5.0**	0.0	*-20.3*
	3	**-0.9**	**0.0**	**4.0**	**5.7**	**5.0**	0.0	*-20.3*
	4	**-0.0**	**-0.0**	**2.2**	**4.5**	**4.5**	-0.1	*-20.3*
$r_1 = 0.4$	5	**1.0**	**0.9**	**1.4**	**2.7**	**2.8**	-0.7	*-20.3*
	6	**1.6**	**1.5**	**1.5**	**1.2**	**0.8**	-2.3	*-20.4*
	7	**2.4**	**2.4**	**2.1**	**1.4**	-0.5	-4.2	*-20.6*
	8	3.1	3.1	3.0	2.1	-0.6	-4.9	*-20.9*
	9	4.6	4.6	4.5	3.5	1.6	-1.9	*-20.9*
	10	5.7	5.7	5.7	5.4	4.6	0.2	*-20.5*
	1	**0.0**	-0.1	**2.0**	**3.4**	**3.8**	**2.4**	-10.4
	2	**-0.2**	**0.3**	**2.0**	**3.4**	**3.8**	**2.4**	-10.4
	3	**-2.2**	**-0.5**	**2.1**	**3.4**	**3.8**	**2.4**	-10.4
	4	**-2.4**	**-1.6**	**1.2**	**3.1**	**3.7**	**2.4**	-10.4
$r_1 = 0.6$	5	**-1.1**	**-0.9**	**0.7**	**2.6**	**3.2**	**2.2**	-10.4
	6	**0.2**	**0.3**	**0.9**	**1.9**	**2.0**	**1.3**	-10.5
	7	**1.9**	**2.0**	**2.1**	**2.3**	**1.4**	**0.6**	-10.7
	8	**3.3**	**3.3**	**3.3**	**3.1**	**2.1**	**0.5**	-11.2
	9	**5.3**	**5.3**	**5.3**	**5.1**	**4.2**	**2.6**	-11.3
	10	8.0	8.0	8.0	8.0	7.3	5.3	-10.1
	1	**0.0**	**-4.7**	**-4.7**	**-2.4**	**-0.3**	**1.4**	**5.6**
	2	**0.2**	**-4.7**	**-4.7**	**-2.4**	**-0.3**	**1.4**	**5.6**
	3	**-5.6**	**-5.6**	**-4.7**	**-2.4**	**-0.3**	**1.4**	**5.6**
	4	**-7.8**	**-7.2**	**-5.0**	**-2.5**	**-0.3**	**1.4**	**5.6**
$r_1 = 0.8$	5	**-7.5**	**-7.3**	**-5.6**	**-2.7**	**-0.4**	**1.4**	**5.6**
	6	**-6.0**	**-6.0**	**-5.2**	**-3.3**	**-0.7**	**1.3**	**5.6**
	7	**-4.2**	**-4.2**	**-4.0**	**-3.2**	**-1.1**	**0.8**	**5.6**
	8	**-1.5**	**-1.5**	**-1.5**	**-1.5**	**-0.7**	**0.5**	**5.5**
	9	**1.4**	**1.4**	**1.3**	**0.9**	**0.6**	**0.9**	**5.2**
	10	**4.7**	**4.7**	**4.7**	**4.4**	**3.8**	**3.0**	**4.9**

Table C.IX: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**0.9**	**-0.2**	-3.3	*-5.2*	*-2.5*	-
	2	**-0.1**	**1.1**	**0.9**	-3.4	*-5.2*	*-2.5*	-
	3	**-0.4**	**-0.1**	**3.1**	-2.2	*-4.7*	*-2.5*	-
	4	**1.1**	**0.6**	4.0	*1.7*	*-1.2*	*-2.4*	-
$r_1 = 0.0$	5	1.9	1.1	2.8	*3.7*	*7.2*	*-0.3*	-
	6	*7.9*	*7.8*	*9.2*	*13.2*	*5.7*	-	-
	7	*5.3*	*5.4*	*5.2*	*9.7*	*2.1*	-	-
	8	*9.9*	*9.9*	*10.2*	*13.1*	-	-	-
	9	-	-	-	-	-	-	-
	10	-	-	-	-	-	-	-
	1	**0.0**	**0.5**	**8.4**	15.0	*17.0*	*23.8*	-
	2	**-1.2**	**1.5**	**8.9**	15.0	*17.0*	*23.8*	-
	3	**0.8**	**1.3**	**8.4**	14.6	*17.2*	*23.8*	-
	4	**2.3**	**1.8**	**5.6**	11.2	*15.7*	*22.2*	-
$r_1 = 0.2$	5	4.1	3.9	4.6	7.9	*12.0*	*20.9*	-
	6	7.3	7.0	6.4	6.2	*10.2*	*19.4*	-
	7	*7.5*	*7.5*	*7.4*	*6.5*	*9.2*	*17.1*	-
	8	*10.2*	*10.2*	*9.5*	*7.1*	*10.3*	*14.4*	-
	9	*16.0*	*16.0*	*15.7*	*12.0*	*16.3*	*17.4*	-
	10	*29.7*	*29.7*	*29.8*	*27.6*	*32.5*	*32.9*	-
	1	**0.0**	**0.4**	**9.0**	**13.7**	13.3	15.4	*1.1*
	2	**-1.8**	**1.5**	**9.1**	**13.7**	13.3	15.4	*1.1*
	3	**-0.4**	**1.2**	**8.5**	**13.5**	13.4	15.4	*1.1*
	4	**1.9**	**2.0**	**6.1**	**11.1**	12.4	15.2	*1.1*
$r_1 = 0.4$	5	**4.2**	**4.1**	**5.5**	8.7	10.1	13.6	*1.1*
	6	7.2	7.1	7.1	7.3	7.2	11.0	*1.3*
	7	9.5	9.5	9.3	8.1	5.7	*7.9*	*-0.6*
	8	11.2	11.2	11.1	*9.5*	*6.3*	*5.3*	*-1.8*
	9	*15.0*	*15.0*	*15.0*	*13.7*	*9.8*	*8.3*	*-3.2*
	10	*14.2*	*14.2*	*14.2*	*13.1*	*12.0*	*9.4*	*-8.4*
	1	**0.0**	**0.3**	**4.5**	**8.5**	**9.9**	10.9	*-2.8*
	2	**-0.6**	**1.0**	**4.5**	**8.5**	**9.9**	10.9	*-2.8*
	3	**-2.0**	**0.2**	**4.6**	**8.5**	**9.9**	10.9	*-2.8*
	4	**-1.5**	**-0.4**	**3.8**	**7.9**	**9.9**	11.0	*-2.8*
$r_1 = 0.6$	5	**-0.1**	**0.2**	**3.0**	**6.3**	**8.9**	10.5	*-2.8*
	6	**2.1**	**2.2**	**3.1**	**5.0**	**7.0**	8.7	*-2.8*
	7	**4.6**	**4.6**	**5.0**	**5.5**	6.2	7.3	*-3.5*
	8	7.1	7.1	7.1	7.0	6.8	6.2	*-4.5*
	9	12.9	12.9	12.9	12.9	11.7	9.4	*-4.9*
	10	15.9	15.9	16.0	16.1	15.1	12.2	*-6.0*
	1	**0.0**	**-3.5**	**-2.5**	**1.0**	**5.2**	**8.6**	14.2
	2	**0.2**	**-3.6**	**-2.5**	**1.0**	**5.2**	**8.6**	14.2
	3	**-5.7**	**-4.8**	**-2.5**	**1.0**	**5.2**	**8.6**	14.2
	4	**-7.3**	**-6.3**	**-2.9**	**0.9**	**5.2**	**8.6**	14.2
$r_1 = 0.8$	5	**-6.7**	**-6.3**	**-3.6**	**0.4**	**5.1**	**8.6**	14.2
	6	**-4.3**	**-4.2**	**-2.8**	**0.1**	**4.6**	**8.4**	14.2
	7	**-1.3**	**-1.3**	**-0.9**	**0.7**	**3.9**	**7.6**	14.1
	8	**2.2**	**2.2**	**2.3**	**3.1**	**4.5**	**6.8**	13.7
	9	**6.0**	**6.0**	**6.0**	**6.3**	**6.5**	**7.1**	12.9
	10	**10.3**	**10.3**	**10.3**	**10.3**	**9.9**	9.3	12.7

Table C.X: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**1.1**	**-0.8**	**-1.2**	-4.3	-7.6	*-0.1*
	2	**-0.0**	**1.1**	**-0.5**	**-1.2**	-4.3	-7.6	*-0.1*
	3	**-0.3**	**-1.0**	**0.7**	**-0.8**	-4.3	-7.6	*-0.1*
	4	**-0.5**	**-1.4**	**0.0**	**0.6**	-3.8	-7.6	*-0.1*
$r_1 = 0.0$	5	**-0.5**	**-1.0**	**-0.8**	**1.5**	-2.5	-7.3	*-0.1*
	6	**-0.9**	**-1.0**	**-1.4**	**0.8**	-1.3	-6.3	*-0.1*
	7	**-2.0**	**-2.0**	**-2.5**	-1.6	-1.7	-4.9	*-0.1*
	8	-2.0	-2.0	-2.4	-2.5	-2.2	*-2.6*	*0.1*
	9	-2.8	-2.8	-2.7	-2.4	-2.0	*-0.9*	*0.9*
	10	*-2.4*	*-2.4*	*-2.3*	*-2.5*	*-2.3*	*-0.1*	*-0.3*
	1	**0.0**	**0.7**	**1.9**	**-0.9**	**-6.0**	-12.9	*-27.1*
	2	**0.2**	**1.2**	**1.9**	**-0.9**	**-6.0**	-12.9	*-27.1*
	3	**0.0**	**-0.5**	**2.0**	**-0.8**	**-6.0**	-12.9	*-27.1*
	4	**-0.1**	**-0.9**	**0.0**	**-0.9**	**-5.9**	-12.9	*-27.1*
$r_1 = 0.2$	5	**-0.0**	**-0.4**	**-1.3**	**-2.1**	**-5.7**	-12.8	*-27.1*
	6	**-0.7**	**-0.8**	**-1.9**	**-3.5**	**-6.4**	-12.6	*-27.1*
	7	**-0.6**	**-0.6**	**-1.3**	**-3.6**	**-6.5**	-12.1	*-27.1*
	8	**-0.7**	**-0.7**	**-1.0**	**-3.1**	-6.4	-11.5	*-27.0*
	9	-1.9	-1.9	-2.0	-3.2	-5.4	-10.5	*-27.2*
	10	-2.1	-2.1	-2.2	-2.6	-4.1	-8.9	*-26.1*
	1	**0.0**	**0.0**	**1.4**	**-0.4**	**-5.0**	**-12.1**	-36.9
	2	**0.2**	**0.5**	**1.4**	**-0.4**	**-5.0**	**-12.1**	-36.9
	3	**-0.7**	**-0.8**	**1.4**	**-0.4**	**-5.0**	**-12.1**	-36.9
	4	**-0.4**	**-0.9**	**0.0**	**-0.5**	**-5.0**	**-12.1**	-36.9
$r_1 = 0.4$	5	**-0.2**	**-0.4**	**-1.0**	**-1.6**	**-5.1**	**-12.1**	-36.9
	6	**0.5**	**0.4**	**-0.6**	**-2.3**	**-5.8**	**-12.1**	-36.9
	7	**0.7**	**0.7**	**0.1**	**-2.1**	**-6.1**	**-12.2**	-36.9
	8	**0.7**	**0.7**	**0.5**	**-1.2**	**-5.1**	**-11.8**	-36.9
	9	**0.7**	**0.7**	**0.6**	**-0.3**	**-3.6**	**-10.5**	-36.7
	10	0.7	0.7	0.7	0.2	-1.9	-7.6	-35.6
	1	**0.0**	**-1.2**	**-0.9**	**-1.0**	**-2.5**	**-5.6**	**-22.5**
	2	**0.1**	**-1.0**	**-0.9**	**-1.0**	**-2.5**	**-5.6**	**-22.5**
	3	**-2.1**	**-1.9**	**-0.8**	**-1.0**	**-2.5**	**-5.6**	**-22.5**
	4	**-2.6**	**-2.7**	**-1.2**	**-0.9**	**-2.5**	**-5.6**	**-22.5**
$r_1 = 0.6$	5	**-2.1**	**-2.1**	**-1.8**	**-1.3**	**-2.5**	**-5.6**	**-22.5**
	6	**-0.6**	**-0.7**	**-1.0**	**-1.5**	**-2.8**	**-5.6**	**-22.5**
	7	**0.7**	**0.7**	**0.4**	**-0.7**	**-2.6**	**-5.7**	**-22.5**
	8	**2.1**	**2.1**	**1.9**	**0.9**	**-1.5**	**-5.1**	**-22.4**
	9	**3.1**	**3.1**	**3.0**	**2.4**	**0.2**	**-3.8**	**-22.1**
	10	**4.6**	**4.6**	**4.6**	**4.3**	**2.7**	**-0.9**	**-21.1**
	1	**0.0**	**-6.3**	**-7.5**	**-6.3**	**-5.1**	**-4.3**	**-5.5**
	2	**-0.1**	**-6.3**	**-7.5**	**-6.3**	**-5.1**	**-4.3**	**-5.5**
	3	**-5.9**	**-7.1**	**-7.5**	**-6.3**	**-5.1**	**-4.3**	**-5.5**
	4	**-8.4**	**-8.7**	**-7.8**	**-6.3**	**-5.1**	**-4.3**	**-5.5**
$r_1 = 0.8$	5	**-9.1**	**-9.3**	**-8.5**	**-6.5**	**-5.1**	**-4.3**	**-5.5**
	6	**-8.5**	**-8.6**	**-8.6**	**-6.9**	**-5.3**	**-4.3**	**-5.5**
	7	**-6.7**	**-6.7**	**-7.2**	**-6.9**	**-5.6**	**-4.4**	**-5.5**
	8	**-4.6**	**-4.6**	**-4.9**	**-5.5**	**-5.4**	**-4.5**	**-5.5**
	9	**-2.3**	**-2.3**	**-2.4**	**-3.1**	**-4.1**	**-4.2**	**-5.5**
	10	**0.1**	**0.1**	**0.0**	**-0.5**	**-1.8**	**-3.0**	**-5.6**

Table C.XI: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ D > D_0 \text{ and } L \geq l_0, l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**1.3**	**-1.2**	-2.1	*-2.7*	*-11.4*	-
	2	**-0.1**	**1.7**	**-0.8**	-2.1	*-2.7*	*-11.4*	-
	3	**-0.4**	**-0.4**	**1.5**	-1.5	*-2.7*	*-11.4*	-
	4	**0.2**	**-0.8**	**1.8**	1.1	*-2.2*	*-11.4*	-
$r_1 = 0.0$	5	**-0.4**	**-0.9**	**0.1**	3.3	*-0.4*	*-10.9*	-
	6	1.1	0.9	0.7	3.4	*2.0*	*-9.0*	-
	7	-0.8	-0.8	-1.2	0.9	*2.9*	*-6.8*	-
	8	*1.1*	*1.0*	*0.9*	*0.7*	*3.1*	*-1.5*	-
	9	*2.2*	*2.2*	*2.1*	*1.5*	*3.3*	*-0.7*	-
	10	*6.5*	*6.5*	*6.6*	*4.3*	*1.7*	*1.2*	-
	1	**0.0**	**0.9**	**3.4**	**0.8**	-4.6	-13.2	*-29.7*
	2	**-0.3**	**1.7**	**3.5**	**0.8**	-4.6	-13.2	*-29.7*
	3	**-0.0**	**-0.3**	**3.9**	**1.0**	-4.6	-13.2	*-29.7*
	4	**0.3**	**-0.7**	**1.8**	**1.0**	-4.5	-13.2	*-29.7*
$r_1 = 0.2$	5	**0.3**	**-0.1**	**-0.2**	**-0.2**	-4.5	-13.2	*-29.7*
	6	**1.2**	**1.1**	**-0.5**	**-2.1**	-4.8	-13.1	*-29.7*
	7	-0.2	-0.2	-1.4	-3.3	-5.3	-12.2	*-29.6*
	8	1.5	1.5	0.8	-1.7	-4.3	*-10.3*	*-29.4*
	9	1.8	1.8	1.2	-0.6	-3.7	*-10.0*	*-28.4*
	10	*3.5*	*3.5*	*3.3*	*1.6*	*-1.5*	*-9.9*	*-26.9*
	1	**0.0**	**0.1**	**2.5**	**1.6**	**-4.0**	-12.0	*-44.0*
	2	**-0.5**	**0.7**	**2.5**	**1.6**	**-4.0**	-12.0	*-44.0*
	3	**-0.8**	**-0.6**	**2.6**	**1.6**	**-4.0**	-12.0	*-44.0*
	4	**-0.2**	**-0.9**	**1.2**	**1.4**	**-4.0**	-12.0	*-44.0*
$r_1 = 0.4$	5	**0.7**	**0.3**	**-0.3**	**0.1**	**-4.2**	-12.0	*-44.0*
	6	**1.6**	**1.5**	**0.3**	**-1.5**	**-5.2**	-12.2	*-44.0*
	7	**2.5**	**2.4**	**1.6**	**-0.9**	**-5.6**	-12.6	*-44.1*
	8	**3.3**	**3.3**	**3.0**	**0.5**	-4.3	-11.7	*-44.0*
	9	2.8	2.8	2.8	1.3	-2.8	-10.9	*-43.4*
	10	2.6	2.6	2.6	2.0	-0.9	-8.8	*-42.4*
	1	**0.0**	**-0.5**	**0.2**	**0.3**	**-2.5**	**-6.5**	-25.7
	2	**-0.0**	**-0.2**	**0.2**	**0.3**	**-2.5**	**-6.5**	-25.7
	3	**-2.0**	**-1.2**	**0.3**	**0.3**	**-2.5**	**-6.5**	-25.7
	4	**-2.2**	**-2.0**	**-0.3**	**0.3**	**-2.5**	**-6.5**	-25.7
$r_1 = 0.6$	5	**-1.3**	**-1.3**	**-1.0**	**-0.2**	**-2.5**	**-6.5**	-25.7
	6	**-0.5**	**-0.6**	**-0.7**	**-0.8**	**-2.9**	**-6.5**	-25.7
	7	**1.1**	**1.1**	**0.8**	**-0.4**	**-2.7**	**-6.5**	-25.7
	8	**3.3**	**3.3**	**3.1**	**1.8**	**-1.3**	**-6.1**	-25.7
	9	**4.6**	**4.6**	**4.5**	**3.9**	**0.8**	**-4.7**	-25.5
	10	**5.6**	**5.6**	**5.5**	**5.3**	**2.7**	-2.5	-24.6
	1	**0.0**	**-5.6**	**-5.6**	**-4.1**	**-2.7**	**-1.7**	**-3.3**
	2	**0.2**	**-5.6**	**-5.6**	**-4.1**	**-2.7**	**-1.7**	**-3.3**
	3	**-6.1**	**-6.5**	**-5.7**	**-4.1**	**-2.7**	**-1.7**	**-3.3**
	4	**-8.0**	**-8.0**	**-5.8**	**-4.1**	**-2.7**	**-1.7**	**-3.3**
$r_1 = 0.8$	5	**-7.9**	**-8.1**	**-6.5**	**-4.2**	**-2.7**	**-1.7**	**-3.3**
	6	**-7.0**	**-7.1**	**-6.8**	**-4.8**	**-2.8**	**-1.7**	**-3.3**
	7	**-5.1**	**-5.2**	**-5.5**	**-4.8**	**-3.1**	**-1.8**	**-3.3**
	8	**-2.7**	**-2.7**	**-3.0**	**-3.3**	**-3.0**	**-2.0**	**-3.3**
	9	**-0.2**	**-0.2**	**-0.3**	**-1.0**	**-1.6**	**-2.0**	**-3.3**
	10	**2.9**	**2.9**	**2.9**	**2.3**	**1.1**	**-0.4**	**-3.3**

Table C.XII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ D > D_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, δ is the deficit coefficient ($D_0 = \delta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

					δ			
	L	0.00	0.25	0.50	0.75	1.00	1.25	2.00
	1	**0.0**	**1.2**	-7.4	*-16.9*	*-34.3*	-	-
	2	**0.2**	**3.2**	-6.6	*-16.9*	*-34.3*	-	-
	3	**-0.3**	**-0.1**	-2.0	*-15.8*	*-34.3*	-	-
	4	**-0.9**	**-1.6**	-0.5	*-10.8*	*-33.4*	-	-
$r_1 = 0.0$	5	-1.7	-2.9	-2.5	*-7.9*	*-31.9*	-	-
	6	*-2.5*	*-3.1*	*-5.0*	*-6.1*	*-28.2*	-	-
	7	*-4.9*	*-5.3*	*-7.0*	*-5.1*	*-20.0*	-	-
	8	*-9.8*	*-9.8*	*-13.0*	*-15.3*	-	-	-
	9	*-7.5*	*-7.5*	*-9.2*	-	-	-	-
	10	-	-	-	-	-	-	-
	1	**0.0**	**1.8**	**7.1**	4.6	*1.9*	*-8.5*	-
	2	**-0.9**	**3.6**	**7.2**	4.6	*1.9*	*-8.5*	-
	3	**0.7**	**1.7**	**8.3**	4.9	*1.9*	*-8.5*	-
	4	**1.5**	**0.6**	**6.0**	5.6	*1.8*	*-8.5*	-
$r_1 = 0.2$	5	**2.5**	**1.9**	3.8	5.4	*1.2*	*-8.1*	-
	6	4.0	3.6	2.8	2.6	*-0.7*	*-8.4*	-
	7	*4.0*	*3.8*	*2.7*	*-3.6*	*-3.6*	*-10.3*	-
	8	*7.3*	*7.1*	*5.3*	*1.1*	*-1.4*	*-11.0*	-
	9	*9.9*	*9.9*	*10.0*	*5.6*	*0.6*	*-3.2*	-
	10	*9.7*	*9.7*	*9.8*	*5.9*	*3.2*	*-9.1*	-
	1	**0.0**	**1.0**	**6.9**	6.9	2.7	*-8.7*	*-37.4*
	2	**-1.5**	**2.1**	**6.9**	6.9	2.7	*-8.7*	*-37.4*
	3	**-0.7**	**0.7**	**7.3**	6.9	2.7	*-8.7*	*-37.4*
	4	**1.0**	**0.7**	**5.4**	6.6	2.7	*-8.7*	*-37.4*
$r_1 = 0.4$	5	**4.1**	**3.5**	**3.7**	5.1	2.3	*-8.6*	*-37.4*
	6	**5.4**	**5.1**	3.3	2.7	-0.0	*-8.9*	*-37.4*
	7	7.6	7.5	6.4	3.7	-0.9	*-9.5*	*-37.4*
	8	11.3	11.3	10.9	6.4	*-1.1*	*-11.0*	*-37.1*
	9	*12.4*	*12.4*	*12.4*	*8.7*	*1.2*	*-9.7*	*-36.8*
	10	*13.6*	*13.6*	*13.6*	*12.2*	*6.8*	*-6.7*	*-34.8*
	1	**0.0**	**0.6**	**3.1**	**4.4**	2.5	-1.1	*-29.9*
	2	**-0.7**	**0.9**	**3.1**	**4.4**	2.5	-1.1	*-29.9*
	3	**-2.1**	**-0.1**	**3.3**	**4.4**	2.5	-1.1	*-29.9*
	4	**-1.7**	**-1.1**	**2.7**	**4.5**	2.5	-1.1	*-29.9*
$r_1 = 0.6$	5	**-0.1**	**-0.2**	**2.0**	**3.9**	2.5	-1.1	*-29.9*
	6	**2.3**	**2.3**	**2.4**	**3.1**	2.0	-1.2	*-29.9*
	7	**4.9**	**4.9**	**4.6**	**3.6**	1.5	-1.5	*-29.9*
	8	8.3	8.3	8.0	6.6	2.8	-1.2	*-29.7*
	9	11.7	11.7	11.5	10.3	5.7	-0.0	*-29.6*
	10	12.1	12.1	12.0	11.4	8.4	2.3	*-28.8*
	1	**0.0**	**-4.3**	**-2.5**	**0.7**	**3.3**	**5.3**	2.5
	2	**-0.0**	**-4.3**	**-2.5**	**0.7**	**3.3**	**5.3**	2.5
	3	**-5.9**	**-5.1**	**-2.5**	**0.7**	**3.3**	**5.3**	2.5
	4	**-7.5**	**-6.7**	**-2.7**	**0.7**	**3.3**	**5.3**	2.5
$r_1 = 0.8$	5	**-7.0**	**-6.8**	**-3.6**	**0.5**	**3.3**	**5.3**	2.5
	6	**-5.1**	**-5.2**	**-3.7**	**-0.2**	**3.1**	**5.3**	2.5
	7	**-2.0**	**-2.0**	**-1.8**	**-0.3**	**2.6**	**5.0**	2.5
	8	**1.7**	**1.7**	1.6	**1.7**	**2.7**	**4.7**	2.5
	9	**6.2**	**6.2**	**6.1**	**5.4**	**4.7**	**5.1**	2.5
	10	**9.9**	**9.9**	**9.9**	**9.0**	**7.7**	**6.6**	2.4

Table C.XIII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L = l_0,\ l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
$r_1 = 0.0$	1	**-0.1**	**1.5**	**5.1**	**-2.5**	-19.0
	2	**0.1**	**1.2**	**2.4**	-13.2	*-43.1*
	3	**-0.3**	**0.3**	-0.8	*-22.8*	-
	4	**-0.2**	**0.3**	-1.7	*-26.8*	-
	5	**1.1**	**1.4**	*-2.9*	-	-
	6	1.7	2.3	*2.4*	-	-
	7	-1.4	-1.1	*0.2*	-	-
	8	*0.1*	*-2.1*	*-18.4*	-	-
	9	*9.1*	*7.4*	-	-	-
	10	*0.5*	*-0.4*	-	-	-
$r_1 = 0.2$	1	**0.1**	**-1.8**	**-1.1**	-12.1	*-26.7*
	2	**-0.1**	**2.6**	**10.9**	*12.1*	*5.4*
	3	**0.1**	**2.4**	15.4	*19.8*	-
	4	**-0.4**	**0.1**	12.0	*22.4*	-
	5	**-0.2**	**-0.6**	9.7	*16.7*	-
	6	**0.1**	-1.3	*0.1*	-	-
	7	1.0	-1.1	*-1.1*	-	-
	8	-0.6	-2.5	*-6.9*	-	-
	9	0.4	*-3.2*	*-21.9*	-	-
	10	*2.2*	*-0.1*	*-40.2*	-	-
$r_1 = 0.4$	1	**-0.0**	**-4.4**	**-8.2**	-26.7	*-43.0*
	2	**0.9**	**3.3**	9.7	*11.2*	*18.3*
	3	**-0.7**	**3.3**	18.8	*49.0*	-
	4	**-0.9**	**1.8**	18.4	*50.2*	-
	5	**-0.7**	**0.5**	8.5	*34.7*	-
	6	**-1.3**	**-1.9**	6.6	*24.1*	-
	7	1.3	-0.6	*-2.5*	-	-
	8	1.8	-0.8	*-15.0*	-	-
	9	0.4	-2.2	*-19.1*	-	-
	10	1.6	-1.9	*-33.0*	-	-
$r_1 = 0.6$	1	**0.1**	**-4.1**	-17.7	*-51.6*	-
	2	**5.7**	**7.2**	1.4	*-22.7*	-
	3	**-1.0**	**5.2**	16.3	*24.5*	-
	4	**-5.3**	**-0.2**	14.9	*38.5*	-
	5	**-5.0**	**-1.0**	17.0	*51.6*	-
	6	**-3.8**	**-2.4**	*7.4*	*60.6*	-
	7	**-3.1**	-3.3	*-4.1*	*6.2*	-
	8	-2.5	-3.4	*-9.5*	*2.6*	-
	9	-0.4	-1.3	*-15.8*	-	-
	10	0.9	-0.5	*-21.3*	-	-
$r_1 = 0.8$	1	**-0.1**	25.4	*38.2*	-	-
	2	**25.8**	35.9	*37.2*	-	-
	3	**5.8**	12.6	*45.9*	-	-
	4	**-6.8**	-0.2	*44.0*	-	-
	5	**-12.8**	-9.9	*30.2*	-	-
	6	**-15.9**	-14.2	*46.7*	-	-
	7	-14.9	-15.8	*28.4*	-	-
	8	-15.7	-17.2	*17.1*	-	-
	9	-13.0	-16.3	*6.3*	-	-
	10	-12.6	-16.5	*-2.1*	-	-

Table C.XIV: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ I > I_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.1**	**1.5**	**4.3**	**-2.7**	-18.1
	2	**0.3**	**1.5**	**1.5**	-10.9	*-28.1*
	3	**-0.7**	**0.3**	-2.7	*-27.0*	-
	4	**-0.1**	**0.6**	*3.5*	-	-
$r_1 = 0.0$	5	-2.2	-2.2	*-2.9*	-	-
	6	-0.9	-0.5	*-9.7*	-	-
	7	*-0.2*	*2.6*	-	-	-
	8	*-1.8*	*-2.7*	-	-	-
	9	*-8.1*	*-9.9*	-	-	-
	10	*-9.2*	*-13.4*	-	-	-
	1	**0.4**	**-1.5**	**-0.9**	-8.9	*-21.2*
	2	**-0.9**	**1.6**	**10.9**	*14.1*	*13.5*
	3	**-0.2**	**2.3**	15.9	*23.4*	-
	4	**-0.2**	**1.3**	16.5	*36.3*	-
$r_1 = 0.2$	5	**-0.7**	-1.1	*7.6*	-	-
	6	0.3	-0.9	*6.6*	-	-
	7	6.0	3.0	*-6.9*	-	-
	8	*2.8*	*-0.1*	*-6.6*	-	-
	9	*3.1*	*1.7*	*-16.9*	-	-
	10	*6.4*	*0.1*	-	-	-
	1	**0.7**	**-3.9**	**-8.4**	-24.7	*-45.4*
	2	**-0.1**	**2.3**	10.7	*22.4*	*40.5*
	3	**-2.1**	**2.3**	18.4	*51.0*	-
	4	**-1.5**	**2.1**	18.7	*63.7*	-
$r_1 = 0.4$	5	**0.3**	**1.2**	15.7	*38.9*	-
	6	0.4	0.4	*10.1*	-	-
	7	1.3	0.3	*0.4*	-	-
	8	1.2	-0.1	*-19.9*	-	-
	9	*3.0*	*-1.1*	*-20.8*	-	-
	10	*2.0*	*-0.3*	*-33.5*	-	-
	1	**0.4**	**-4.4**	-15.8	*-43.6*	-
	2	**3.9**	**6.3**	0.2	*-12.8*	-
	3	**-1.7**	**4.1**	13.7	*46.0*	-
	4	**-5.2**	**0.4**	11.3	*32.7*	-
$r_1 = 0.6$	5	**-3.9**	**0.0**	*18.7*	*89.8*	-
	6	**-3.6**	-0.9	*10.0*	*60.9*	-
	7	-1.0	0.2	*6.6*	*48.7*	-
	8	-1.2	-2.1	*-10.4*	-	-
	9	-0.4	-1.8	*-16.4*	-	-
	10	1.3	-0.9	*-26.6*	-	-
	1	**-0.3**	24.0	*42.7*	-	-
	2	**21.6**	31.6	*37.0*	-	-
	3	**2.7**	12.3	*40.4*	-	-
	4	**-8.6**	-0.4	*35.8*	-	-
$r_1 = 0.8$	5	**-13.5**	-8.4	*36.3*	-	-
	6	**-13.0**	-9.7	*32.3*	-	-
	7	-15.1	-13.4	*19.5*	-	-
	8	-12.5	-13.0	*13.0*	-	-
	9	-11.8	-14.4	*6.1*	-	-
	10	-7.2	-11.5	*-3.7*	-	-

Table C.XV: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ I > I_0$ and $L = l_0$, $l_0 = 1, 2, \ldots\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
$r_1 = 0.0$	1	**0.0**	**1.5**	**2.8**	**-3.6**	-15.4
	2	**0.0**	**1.2**	**0.1**	*-14.2*	*-32.8*
	3	**-1.0**	**0.0**	-0.7	*-27.3*	-
	4	0.7	2.8	*-6.4*	-	-
	5	-0.8	*-1.1*	*10.9*	-	-
	6	*9.2*	*13.8*	-	-	-
	7	*3.6*	*7.3*	-	-	-
	8	*11.9*	*6.3*	-	-	-
	9	-	-	-	-	-
	10	-	-	-	-	-
$r_1 = 0.2$	1	**0.8**	**-2.1**	**-2.2**	-9.4	*-21.5*
	2	**-2.6**	**-0.4**	9.2	*15.1*	*39.7*
	3	**-0.1**	**2.7**	22.1	*48.9*	-
	4	**1.1**	3.1	*23.8*	-	-
	5	2.2	3.2	*17.2*	-	-
	6	7.1	*5.4*	*25.1*	-	-
	7	*6.0*	*5.0*	*35.7*	-	-
	8	*6.8*	*3.1*	-	-	-
	9	*8.1*	*3.0*	-	-	-
	10	*25.1*	*24.9*	-	-	-
$r_1 = 0.4$	1	**1.6**	**-4.0**	**-8.5**	-19.7	*-33.5*
	2	**-3.1**	**-0.4**	9.5	*26.1*	*50.1*
	3	**-2.5**	**2.2**	25.8	*64.2*	-
	4	**-0.4**	**3.8**	*25.9*	*72.7*	-
	5	1.4	3.5	*22.7*	*123.6*	-
	6	5.0	5.4	*20.1*	-	-
	7	8.0	*6.9*	*14.6*	-	-
	8	*7.7*	*5.5*	*-10.8*	-	-
	9	*16.0*	*10.4*	*5.8*	-	-
	10	*7.3*	*5.0*	-	-	-
$r_1 = 0.6$	1	**0.7**	**-4.4**	-15.5	*-44.6*	-
	2	**1.6**	**4.8**	1.0	*-4.2*	-
	3	**-2.8**	**3.2**	16.8	*31.8*	-
	4	**-3.7**	**3.2**	*24.3*	*76.5*	-
	5	**-3.5**	2.8	*24.9*	*83.3*	-
	6	-1.8	0.7	*20.1*	*57.6*	-
	7	0.8	2.3	*18.9*	-	-
	8	-2.0	-2.9	*-0.8*	-	-
	9	*8.2*	*6.7*	*-5.5*	-	-
	10	*13.7*	*11.8*	*-20.1*	-	-
$r_1 = 0.8$	1	**-0.3**	19.5	*17.5*	-	-
	2	**17.3**	26.5	*20.0*	-	-
	3	**-0.7**	7.1	*32.6*	-	-
	4	**-9.3**	0.7	*38.1*	-	-
	5	**-13.7**	-7.9	*36.4*	-	-
	6	-13.4	-8.6	*46.7*	-	-
	7	-12.0	-10.3	*43.2*	-	-
	8	-9.2	-8.4	*28.1*	-	-
	9	-6.8	-7.0	*10.2*	-	-
	10	-3.2	-5.1	*18.2*	-	-

Table C.XVI: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L = l_0$, $l_0 = 1, 2, \ldots\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.1**	**2.9**	**10.5**	-17.4	*-70.2*
	2	**0.3**	**2.9**	**2.0**	*-43.7*	-
	3	**-0.1**	**1.5**	-5.6	-	-
	4	**-0.5**	**-0.0**	*-10.8*	-	-
$r_1 = 0.0$	5	**-0.2**	**0.1**	*-14.0*	-	-
	6	**0.4**	1.8	*-18.8*	-	-
	7	-1.9	-1.3	*-34.6*	-	-
	8	-1.1	-2.7	-	-	-
	9	*-3.1*	*-3.0*	-	-	-
	10	*-3.3*	*-2.2*	-	-	-
	1	**-0.3**	**-1.5**	**3.5**	-29.8	*-79.6*
	2	**0.5**	**3.2**	11.4	*-25.3*	-
	3	**0.1**	**2.3**	16.4	*-23.1*	-
	4	**-0.1**	**0.0**	8.4	-	-
$r_1 = 0.2$	5	**1.0**	**-0.2**	*1.6*	-	-
	6	**-0.9**	**-3.5**	*-12.4*	-	-
	7	-0.4	-4.3	*-22.8*	-	-
	8	1.1	-3.4	*-37.3*	-	-
	9	-1.6	-6.2	*-33.5*	-	-
	10	*-7.1*	*-10.6*	-	-	-
	1	**-0.4**	**-1.6**	-5.5	*-45.0*	-
	2	**2.0**	**4.9**	13.9	*-14.8*	-
	3	**-1.3**	**1.9**	16.7	*-4.7*	-
	4	**-0.7**	**0.6**	14.6	-	-
$r_1 = 0.4$	5	**-1.6**	**-2.2**	*6.8*	-	-
	6	**0.0**	**-2.9**	*-9.2*	-	-
	7	**0.8**	-3.7	*-21.9*	-	-
	8	0.8	-4.4	*-36.6*	-	-
	9	0.7	-4.0	*-47.6*	-	-
	10	-1.6	-6.7	*-53.8*	-	-
	1	**-0.2**	**3.6**	-14.1	-	-
	2	**6.5**	**10.8**	*0.5*	-	-
	3	**-0.7**	**3.8**	*10.5*	-	-
	4	**-4.3**	**-1.4**	*8.5*	-	-
$r_1 = 0.6$	5	**-6.3**	**-5.3**	*6.7*	-	-
	6	**-4.6**	**-5.6**	*-3.5*	-	-
	7	**-3.5**	-6.1	*-14.7*	-	-
	8	-1.0	-4.6	*-19.6*	-	-
	9	-1.2	-5.4	*-37.6*	-	-
	10	0.6	-3.8	*-48.4*	-	-
	1	**0.3**	68.1	-	-	-
	2	**27.9**	63.8	-	-	-
	3	**6.7**	26.6	*109.6*	-	-
	4	**-4.7**	5.3	*62.3*	-	-
$r_1 = 0.8$	5	**-12.3**	-7.7	*62.6*	-	-
	6	**-18.0**	-15.9	*44.2*	-	-
	7	-17.4	-20.5	*26.2*	-	-
	8	-16.6	-21.8	*-2.4*	-	-
	9	-14.5	-20.1	*7.2*	-	-
	10	-14.4	-21.1	*-13.3*	-	-

Table C.XVII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L = l_0$, $l_0 = 1, 2, \ldots\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.1**	**4.3**	**9.4**	-22.2	*-72.4*
	2	**0.1**	**3.5**	-4.7	*-54.3*	-
	3	**-0.9**	**1.7**	-12.3	-	-
	4	**0.6**	**2.2**	*-14.0*	-	-
$r_1 = 0.0$	5	**-1.6**	-0.5	*-18.9*	-	-
	6	2.7	3.3	-	-	-
	7	-2.2	*-0.9*	-	-	-
	8	*0.2*	*-1.1*	-	-	-
	9	*-1.1*	*2.2*	-	-	-
	10	*2.9*	*-1.2*	-	-	-
	1	**0.3**	**0.4**	**2.4**	-31.2	*-76.5*
	2	**-0.5**	**3.1**	10.1	*-29.6*	-
	3	**-0.3**	**3.1**	13.0	-	-
	4	**0.2**	**1.6**	*8.7*	-	-
$r_1 = 0.2$	5	**-0.7**	**0.3**	*-2.2*	-	-
	6	2.8	-0.8	*-18.3*	-	-
	7	-2.1	-5.0	*-28.6*	-	-
	8	1.1	*-3.5*	-	-	-
	9	*-0.0*	*-5.2*	-	-	-
	10	*10.2*	*5.3*	-	-	-
	1	**0.7**	**-2.1**	-8.0	*-50.9*	-
	2	**-0.0**	**4.2**	8.9	*-35.7*	-
	3	**-1.6**	**2.7**	19.5	-	-
	4	**-1.7**	**0.7**	*16.2*	-	-
$r_1 = 0.4$	5	**-0.7**	**-1.7**	*10.2*	-	-
	6	**0.3**	-2.2	*-6.6*	-	-
	7	1.2	-2.6	*-22.3*	-	-
	8	4.0	-2.6	*-25.8*	-	-
	9	3.2	-3.4	*-45.4*	-	-
	10	*-0.0*	*-5.9*	*-55.0*	-	-
	1	**0.0**	**2.1**	-18.6	-	-
	2	**4.6**	**9.5**	*2.8*	-	-
	3	**-1.7**	**4.1**	*6.8*	-	-
	4	**-4.3**	**0.6**	*13.6*	-	-
$r_1 = 0.6$	5	**-3.0**	**-2.0**	*8.7*	-	-
	6	**-4.6**	-4.9	*-5.0*	-	-
	7	-4.0	-7.0	*-16.0*	-	-
	8	0.3	-4.7	*-24.6*	-	-
	9	2.3	-2.6	*-39.7*	-	-
	10	1.9	-4.0	*-45.0*	-	-
	1	**-0.3**	57.3	-	-	-
	2	**25.2**	53.9	-	-	-
	3	**2.2**	16.2	-	-	-
	4	**-8.4**	1.6	*26.4*	-	-
$r_1 = 0.8$	5	**-11.8**	-3.4	*42.2*	-	-
	6	**-14.9**	-13.0	*32.9*	-	-
	7	-15.6	-16.2	*22.4*	-	-
	8	-13.3	-16.8	*23.9*	-	-
	9	-13.2	-18.1	*-4.7*	-	-
	10	-11.7	-18.1	*-20.8*	-	-

Table C.XVIII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L = l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**-0.1**	**5.3**	**6.6**	-31.5	*-79.0*
	2	**0.5**	**5.3**	-10.3	*-64.8*	-
	3	**-0.0**	**2.7**	*-17.5*	-	-
	4	**-0.5**	2.3	*-28.9*	-	-
$r_1 = 0.0$	5	-1.2	*-0.5*	-	-	-
	6	*-1.3*	*-2.4*	-	-	-
	7	*-2.3*	*-1.1*	-	-	-
	8	*-10.1*	-	-	-	-
	9	-	-	-	-	-
	10	-	-	-	-	∞
	1	**0.7**	**0.5**	**-0.8**	-38.5	*-78.4*
	2	**-2.1**	**2.9**	6.4	*-27.9*	-
	3	**0.2**	**5.6**	*12.4*	-	-
	4	**0.8**	4.6	*9.5*	-	-
$r_1 = 0.2$	5	1.4	4.8	*16.1*	-	-
	6	4.0	*3.5*	-	-	-
	7	*1.9*	*-7.3*	-	-	-
	8	*5.6*	*-5.6*	-	-	-
	9	*10.5*	*6.6*	-	-	-
	10	*-2.3*	-	-	-	-
	1	**1.4**	**-0.3**	-10.1	*-57.5*	-
	2	**-2.3**	**3.1**	9.3	*-39.5*	-
	3	**-2.5**	**3.9**	*28.7*	-	-
	4	**-2.2**	**2.7**	*22.5*	-	-
$r_1 = 0.4$	5	**2.9**	4.4	*28.3*	-	-
	6	3.1	-0.5	*3.0*	-	-
	7	3.8	*-0.4*	*-2.3*	-	-
	8	*10.2*	*2.8*	-	-	-
	9	*11.3*	*-2.3*	-	-	-
	10	*7.4*	*-1.1*	-	-	-
	1	**0.9**	**1.9**	-21.3	-	-
	2	**1.6**	**7.9**	*-0.5*	-	-
	3	**-2.9**	**6.3**	*11.8*	-	-
	4	**-4.4**	**2.6**	*19.0*	-	-
$r_1 = 0.6$	5	**-4.3**	0.9	*14.5*	-	-
	6	-2.1	-1.1	*2.8*	-	-
	7	-0.7	-3.1	*0.2*	-	-
	8	2.5	-2.3	*-6.7*	-	-
	9	11.0	*3.8*	*-23.9*	-	-
	10	*6.1*	*-0.2*	-	-	-
	1	**0.0**	48.1	*70.0*	-	-
	2	**18.1**	42.9	-	-	-
	3	**-0.7**	17.2	-	-	-
	4	**-9.2**	5.5	*65.5*	-	-
$r_1 = 0.8$	5	**-13.0**	-2.8	*57.1*	-	-
	6	-15.0	-8.3	*65.4*	-	-
	7	-13.9	-12.3	*22.0*	-	-
	8	-12.9	-13.9	*21.4*	-	-
	9	-5.7	-9.4	*12.1*	-	-
	10	-6.0	-13.2	*-13.6*	-	-

Table C.XIX: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
$r_1 = 0.0$	1	**0.0**	**1.2**	**3.6**	**-4.6**	-20.2
	2	**0.1**	**0.9**	**1.0**	-15.1	*-42.5*
	3	**0.1**	**0.5**	**-1.2**	*-23.9*	-
	4	**0.5**	**0.7**	-1.8	*-28.3*	-
	5	**1.1**	**1.2**	*-1.9*	-	-
	6	**1.1**	**1.0**	*-0.2*	-	-
	7	0.5	-0.3	*-3.7*	-	-
	8	2.3	0.6	*-8.7*	-	-
	9	*4.6*	*3.2*	*8.0*	-	-
	10	*0.2*	*-0.9*	-	-	-
$r_1 = 0.2$	1	**0.0**	**0.2**	**4.7**	**-5.7**	*-24.0*
	2	**-0.1**	**1.3**	**10.4**	14.0	*6.8*
	3	**-0.0**	**0.5**	**10.0**	*18.5*	-
	4	**-0.1**	**-0.7**	5.4	*16.2*	-
	5	**0.1**	**-1.2**	0.0	*5.4*	-
	6	**0.3**	**-1.6**	-7.7	-	-
	7	**0.5**	**-1.9**	*-13.7*	-	-
	8	0.1	-2.5	*-23.4*	-	-
	9	0.7	-2.4	*-35.8*	-	-
	10	1.0	-1.8	*-45.9*	-	-
$r_1 = 0.4$	1	**0.0**	**0.0**	**1.3**	-8.0	*-30.9*
	2	**0.0**	**1.6**	**6.3**	*23.3*	*45.8*
	3	**-0.5**	**1.0**	**4.8**	*35.3*	-
	4	**-0.3**	**0.0**	**-0.6**	*24.7*	-
	5	**-0.0**	**-0.8**	**-8.1**	*7.2*	-
	6	**0.3**	**-1.4**	-14.6	*-9.5*	-
	7	**1.2**	**-1.2**	-23.0	*-28.2*	-
	8	**1.2**	**-1.5**	-31.1	-	-
	9	**0.8**	**-1.8**	-37.7	-	-
	10	1.1	-1.6	*-45.3*	-	-
$r_1 = 0.6$	1	**0.0**	**0.6**	**-8.5**	*-25.4*	-
	2	**-0.1**	**1.6**	**-6.8**	*-14.1*	-
	3	**-2.3**	**0.5**	**-7.9**	*-12.3*	-
	4	**-2.7**	**-0.7**	**-11.4**	*-18.1*	-
	5	**-1.8**	**-0.8**	**-15.5**	*-26.3*	-
	6	**-0.6**	**-0.7**	**-21.0**	*-37.2*	-
	7	**0.6**	**-0.2**	**-26.3**	*-51.0*	-
	8	**2.0**	**0.8**	-30.7	*-59.1*	-
	9	**3.6**	**2.3**	-35.1	*-67.9*	-
	10	**5.1**	**3.5**	-39.5	*-74.6*	-
$r_1 = 0.8$	1	**0.0**	**-2.1**	**-17.3**	*-62.1*	-
	2	**0.0**	**-3.8**	**-17.8**	*-62.0*	-
	3	**-5.8**	**-6.1**	**-18.2**	*-62.0*	-
	4	**-8.2**	**-7.5**	**-18.9**	*-62.0*	-
	5	**-8.5**	**-8.3**	**-19.9**	*-62.1*	-
	6	**-7.7**	**-8.1**	**-21.1**	*-63.0*	-
	7	**-5.9**	**-7.2**	**-23.2**	*-64.4*	-
	8	**-4.0**	**-5.8**	**-25.3**	*-66.4*	-
	9	**-1.6**	**-3.9**	**-27.4**	*-68.2*	-
	10	**0.8**	**-1.6**	-29.4	*-70.6*	-

Table C.XX: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L \geq l_0,\ l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.0**	**1.2**	**3.1**	**-4.0**	-18.5
	2	**-0.1**	**0.8**	**0.6**	-13.1	*-28.4*
	3	**-0.7**	**0.0**	-1.5	*-26.7*	-
	4	**-0.8**	**-0.4**	1.0	-	-
$r_1 = 0.0$	5	**-1.9**	**-1.6**	*-4.3*	-	-
	6	-1.5	-0.8	*-6.8*	-	-
	7	-2.2	-1.1	*0.8*	-	-
	8	*-4.8*	*-6.0*	-	-	-
	9	*-8.7*	*-10.0*	-	-	-
	10	*-9.1*	*-9.6*	-	-	-
	1	**0.0**	**0.2**	**4.5**	**-3.9**	*-18.9*
	2	**-0.5**	**1.4**	**11.9**	16.4	*15.2*
	3	**0.0**	**1.2**	**13.2**	*24.1*	-
	4	**0.3**	**0.3**	10.1	*27.0*	-
$r_1 = 0.2$	5	**0.7**	**-0.5**	2.8	-	-
	6	**2.2**	**-0.1**	*-2.4*	-	-
	7	4.1	0.7	*-12.2*	-	-
	8	2.2	-1.3	*-17.4*	-	-
	9	*1.6*	*-2.5*	*-28.0*	-	-
	10	*0.2*	*-6.4*	-	-	-
	1	**0.0**	**-0.0**	**2.9**	-6.9	*-33.6*
	2	**-0.6**	**1.9**	**11.0**	*31.8*	*58.5*
	3	**-0.9**	**1.6**	**11.2**	*44.9*	-
	4	**-0.0**	**1.2**	7.2	*38.5*	-
$r_1 = 0.4$	5	**1.0**	**0.8**	0.7	*14.2*	-
	6	**1.6**	**0.5**	-7.6	*-6.9*	-
	7	**2.4**	**0.6**	-17.5	-	-
	8	3.1	0.8	*-27.4*	-	-
	9	4.6	1.5	*-31.5*	-	-
	10	5.7	3.3	*-37.4*	-	-
	1	**0.0**	**0.9**	**-4.4**	*-13.4*	-
	2	**-0.2**	**2.4**	**-1.2**	*7.7*	-
	3	**-2.2**	**1.4**	**-1.4**	*14.6*	-
	4	**-2.4**	**0.5**	**-4.6**	*6.5*	-
$r_1 = 0.6$	5	**-1.1**	**0.6**	-8.1	*0.4*	-
	6	**0.2**	**0.8**	-14.4	*-19.7*	-
	7	**1.9**	**1.5**	-20.6	*-37.5*	-
	8	**3.3**	**2.0**	-27.8	*-56.3*	-
	9	**5.3**	**3.7**	-32.6	*-65.4*	-
	10	8.0	6.1	-37.4	-	-
	1	**0.0**	**0.9**	-8.9	*-34.6*	-
	2	**0.2**	**-1.1**	-9.8	*-34.4*	-
	3	**-5.6**	**-3.6**	-10.5	*-34.4*	-
	4	**-7.8**	**-5.2**	-11.5	*-34.2*	-
$r_1 = 0.8$	5	**-7.5**	**-5.9**	-12.9	*-35.0*	-
	6	**-6.0**	**-5.5**	-14.8	*-36.6*	-
	7	**-4.2**	**-4.7**	-17.1	*-38.9*	-
	8	**-1.5**	**-3.0**	-19.3	*-41.5*	-
	9	**1.4**	**-0.9**	-21.7	*-45.4*	-
	10	**4.7**	**2.2**	-24.1	*-50.3*	-

Table C.XXI: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Normal model. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.0**	**1.4**	**2.2**	**-4.6**	-15.9
	2	**-0.1**	**1.1**	**-0.2**	*-15.5*	*-32.2*
	3	**-0.4**	**0.9**	-1.3	*-27.8*	-
	4	**1.1**	2.8	*-3.3*	-	-
$r_1 = 0.0$	5	1.9	2.7	*8.8*	-	-
	6	*7.9*	*10.9*	-	-	-
	7	*5.3*	*5.2*	-	-	-
	8	*9.9*	*2.0*	-	-	-
	9	-	-	-	-	-
	10	-	-	-	-	-
	1	**0.0**	**-0.6**	**2.7**	-5.2	*-18.5*
	2	**-1.2**	**1.1**	**13.4**	*20.4*	*48.1*
	3	**0.8**	**3.1**	22.0	*48.5*	-
	4	**2.3**	**3.5**	*22.0*	*49.4*	-
$r_1 = 0.2$	5	4.1	4.2	*18.9*	-	-
	6	7.3	5.7	*22.4*	-	-
	7	*7.5*	*6.1*	*18.9*	-	-
	8	*10.2*	*7.9*	-	-	-
	9	*16.0*	*15.6*	-	-	-
	10	*29.7*	*35.9*	-	-	-
	1	**0.0**	**-0.7**	**3.2**	-5.3	*-25.2*
	2	**-1.8**	**1.7**	**16.4**	*40.1*	*64.8*
	3	**-0.4**	**3.4**	23.2	*70.7*	-
	4	**1.9**	**4.4**	20.8	*83.4*	-
$r_1 = 0.4$	5	**4.2**	**5.0**	*16.2*	*102.7*	-
	6	7.2	6.3	*10.5*	-	-
	7	9.5	7.2	*2.1*	-	-
	8	11.2	*7.6*	*-8.8*	-	-
	9	*15.0*	*9.7*	*-6.2*	-	-
	10	*14.2*	*9.3*	*-15.7*	-	-
	1	**0.0**	**1.3**	**2.2**	*-14.0*	-
	2	**-0.6**	**3.6**	**10.6**	*23.3*	-
	3	**-2.0**	**3.1**	13.9	*40.5*	-
	4	**-1.5**	**3.1**	12.9	*45.3*	-
$r_1 = 0.6$	5	**-0.1**	**3.1**	8.7	*32.0*	-
	6	**2.1**	**3.2**	2.7	*11.4*	-
	7	**4.6**	**4.7**	-4.2	*-6.0*	-
	8	7.1	6.1	*-13.6*	-	-
	9	12.9	11.4	*-18.9*	-	-
	10	15.9	14.2	*-24.7*	-	-
	1	**0.0**	**3.5**	10.8	*12.9*	-
	2	**0.2**	**1.5**	10.5	*15.6*	-
	3	**-5.7**	**-1.4**	10.2	*16.0*	-
	4	**-7.3**	**-2.6**	9.3	*15.5*	-
$r_1 = 0.8$	5	**-6.7**	**-3.2**	7.6	*14.7*	-
	6	**-4.3**	**-2.2**	5.6	*12.0*	-
	7	**-1.3**	**-0.7**	2.0	*7.3*	-
	8	**2.2**	**1.9**	-2.2	*-3.3*	-
	9	**6.0**	**4.8**	-5.8	*-6.4*	-
	10	**10.3**	**8.3**	-8.0	*-14.1*	-

Table C.XXII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ I > I_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.0\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.0**	**2.2**	**6.2**	**-20.9**	*-70.7*
	2	**-0.0**	**1.6**	**-1.7**	*-47.4*	-
	3	**-0.3**	**0.6**	-8.5	-	-
	4	**-0.5**	**-0.1**	-13.5	-	-
$r_1 = 0.0$	5	**-0.5**	**-0.1**	*-18.2*	-	-
	6	**-0.9**	**-0.2**	*-25.0*	-	-
	7	**-2.0**	-1.9	*-34.9*	-	-
	8	-2.0	-2.4	-	-	-
	9	-2.8	-2.2	-	-	-
	10	*-2.4*	*-1.4*	-	-	-
	1	**0.0**	**0.1**	**5.9**	-29.2	*-80.3*
	2	**0.2**	**0.9**	**8.6**	*-26.2*	-
	3	**0.0**	**-0.3**	5.9	*-28.8*	-
	4	**-0.1**	**-1.7**	-3.4	-	-
$r_1 = 0.2$	5	**-0.0**	**-2.8**	-13.1	-	-
	6	**-0.7**	**-4.3**	*-24.8*	-	-
	7	**-0.6**	**-4.8**	*-34.4*	-	-
	8	**-0.7**	**-5.0**	*-42.9*	-	-
	9	-1.9	-6.1	*-46.8*	-	-
	10	-2.1	-5.9	*-55.8*	-	-
	1	**0.0**	**-0.1**	**-3.1**	*-37.4*	-
	2	**0.2**	**0.4**	**-1.9**	*-17.8*	-
	3	**-0.7**	**-1.1**	**-7.9**	*-21.2*	-
	4	**-0.4**	**-2.1**	-16.6	*-36.8*	-
$r_1 = 0.4$	5	**-0.2**	**-3.2**	-27.2	-	-
	6	**0.5**	**-3.7**	-38.6	-	-
	7	**0.7**	**-4.0**	-48.4	-	-
	8	**0.7**	**-4.2**	-57.3	-	-
	9	**0.7**	**-4.1**	-64.3	-	-
	10	0.7	-4.1	*-70.0*	-	-
	1	**0.0**	**0.3**	**-26.5**	*-73.0*	-
	2	**0.1**	**-0.2**	**-28.3**	*-69.6*	-
	3	**-2.1**	**-2.1**	**-31.0**	*-73.0*	-
	4	**-2.6**	**-3.3**	**-35.1**	-	-
$r_1 = 0.6$	5	**-2.1**	**-3.8**	**-39.8**	-	-
	6	**-0.6**	**-3.4**	-45.4	-	-
	7	**0.7**	**-2.8**	-50.8	-	-
	8	**2.1**	**-1.8**	-56.0	-	-
	9	**3.1**	**-0.9**	-61.6	-	-
	10	**4.6**	**0.5**	-65.6	-	-
	1	**0.0**	**-2.4**	-47.5	-	-
	2	**-0.1**	**-5.2**	-47.9	-	-
	3	**-5.9**	**-7.8**	-48.1	-	-
	4	**-8.4**	**-9.7**	-48.7	-	-
$r_1 = 0.8$	5	**-9.1**	**-10.9**	-49.3	-	-
	6	**-8.5**	**-11.2**	-50.3	-	-
	7	**-6.7**	**-10.6**	-51.6	-	-
	8	**-4.6**	**-9.3**	-53.0	-	-
	9	**-2.3**	**-7.4**	-54.3	-	-
	10	**0.1**	**-5.4**	-56.3	-	-

Table C.XXIII: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{\ I > I_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.2\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
$r_1 = 0.0$	1	**0.0**	**3.5**	**5.1**	-24.8	*-72.6*
	2	**-0.1**	**2.6**	**-7.1**	*-56.4*	-
	3	**-0.4**	**1.5**	-13.4	-	-
	4	**0.2**	**1.4**	*-16.4*	-	-
	5	**-0.4**	**0.4**	*-22.3*	-	-
	6	1.1	1.5	*-29.6*	-	-
	7	-0.8	-0.6	-	-	-
	8	*1.1*	*-0.1*	-	-	-
	9	*2.2*	*1.2*	-	-	-
	10	*6.5*	*0.2*	-	-	-
$r_1 = 0.2$	1	**0.0**	**1.4**	**4.8**	-31.1	*-76.7*
	2	**-0.3**	**2.1**	**8.8**	*-30.8*	-
	3	**-0.0**	**1.3**	6.8	*-36.2*	-
	4	**0.3**	**-0.1**	-1.4	-	-
	5	**0.3**	**-1.4**	*-13.9*	-	-
	6	**1.2**	**-2.8**	*-27.9*	-	-
	7	-0.2	-4.5	*-38.6*	-	-
	8	1.5	-4.1	-	-	-
	9	1.8	-4.5	-	-	-
	10	*3.5*	*-3.8*	-	-	-
$r_1 = 0.4$	1	**0.0**	**0.3**	**-0.5**	*-47.0*	-
	2	**-0.5**	**1.3**	**5.6**	*-32.0*	-
	3	**-0.8**	**-0.0**	3.6	*-23.4*	-
	4	**-0.2**	**-1.3**	-5.1	-	-
	5	**0.7**	**-2.4**	-16.2	-	-
	6	**1.6**	**-2.7**	*-29.6*	-	-
	7	**2.5**	**-3.0**	*-40.9*	-	-
	8	**3.3**	-3.2	*-50.0*	-	-
	9	2.8	-3.6	*-61.8*	-	-
	10	2.6	-3.6	*-69.5*	-	-
$r_1 = 0.6$	1	**0.0**	**1.1**	**-18.7**	*-68.8*	-
	2	**-0.0**	**0.8**	-18.7	*-57.4*	-
	3	**-2.0**	**-1.1**	-22.2	-	-
	4	**-2.2**	**-2.4**	-26.7	-	-
	5	**-1.3**	**-3.3**	-33.4	-	-
	6	**-0.5**	**-3.7**	-40.8	-	-
	7	**1.1**	**-3.2**	-47.5	-	-
	8	**3.3**	**-1.9**	-53.7	-	-
	9	**4.6**	**-0.8**	-59.9	-	-
	10	**5.6**	**-0.1**	-64.4	-	-
$r_1 = 0.8$	1	**0.0**	**-0.1**	-43.5	-	-
	2	**0.2**	**-3.2**	-44.1	-	-
	3	**-6.1**	**-6.0**	-44.7	-	-
	4	**-8.0**	**-7.5**	-45.1	-	-
	5	**-7.9**	**-8.4**	-45.9	-	-
	6	**-7.0**	**-9.1**	-47.2	-	-
	7	**-5.1**	**-8.5**	-48.9	-	-
	8	**-2.7**	**-7.2**	-50.8	-	-
	9	**-0.2**	**-5.5**	-53.5	-	-
	10	**2.9**	**-3.0**	-55.7	-	-

Table C.XXIV: Difference (%) between return period computed analytically by means of eq. (3.67) and by simulation for drought events defined by $\{ I > I_0$ and $L \geq l_0$, $l_0 = 1, 2, ...\}$. L is the drought length, ζ is the intensity coefficient ($I_0 = \zeta x_0$) and r_1 is the lag-1 autocorrelation coefficient of the underlying process. Threshold $x_0 = \mu_x - 0.5\sigma_X$. Log normal model, skewness=.5. Bold values $Tr < 200$, roman values $200 \leq Tr < 1000$, italic values $Tr \geq 1000$.

				ζ		
	L	0.00	0.10	0.20	0.30	0.40
	1	**0.0**	**4.9**	**3.8**	-32.8	*-79.2*
	2	**0.2**	**4.2**	-11.9	*-66.3*	-
	3	**-0.3**	**2.1**	*-20.2*	-	-
	4	**-0.9**	1.1	*-33.1*	-	-
$r_1 = 0.0$	5	-1.7	-1.3	-	-	-
	6	*-2.5*	*-2.8*	-	-	-
	7	*-4.9*	*-3.3*	-	-	-
	8	*-9.8*	*-6.6*	-	-	-
	9	*-7.5*	-	-	-	-
	10	-	-	-	-	∞
	1	**0.0**	**2.0**	**1.3**	-37.8	*-78.7*
	2	**-0.9**	**3.8**	7.9	*-28.4*	-
	3	**0.7**	**4.8**	*11.7*	-	-
	4	**1.5**	**3.7**	*10.4*	-	-
$r_1 = 0.2$	5	**2.5**	2.6	*13.0*	-	-
	6	4.0	-0.3	-	-	-
	7	*4.0*	*-5.2*	-	-	-
	8	*7.3*	*-2.2*	-	-	-
	9	*9.9*	*2.9*	-	-	-
	10	*9.7*	*-1.0*	-	-	-
	1	**0.0**	**1.7**	**-0.1**	*-53.9*	-
	2	**-1.5**	**3.0**	15.2	*-30.2*	-
	3	**-0.7**	**3.0**	22.0	-	-
	4	**1.0**	**2.3**	*15.2*	-	-
$r_1 = 0.4$	5	**4.1**	**1.9**	*8.2*	-	-
	6	**5.4**	-0.1	*-10.2*	-	-
	7	7.6	0.3	*-21.4*	-	-
	8	11.3	0.9	*-37.0*	-	-
	9	*12.4*	*-0.7*	-	-	-
	10	*13.6*	*0.8*	-	-	-
	1	**0.0**	**3.2**	-8.8	-	-
	2	**-0.7**	**3.7**	-2.0	-	-
	3	**-2.1**	**2.3**	-2.5	-	-
	4	**-1.7**	**0.9**	-6.8	-	-
$r_1 = 0.6$	5	**-0.1**	**0.2**	*-14.7*	-	-
	6	**2.3**	**-0.2**	*-23.8*	-	-
	7	**4.9**	**0.3**	*-32.1*	-	-
	8	8.3	2.0	*-42.7*	-	-
	9	11.7	4.3	*-54.8*	-	-
	10	12.1	4.6	*-65.4*	-	-
	1	**0.0**	**5.0**	-24.4	-	-
	2	**-0.0**	**1.5**	-26.2	-	-
	3	**-5.9**	**-1.5**	*-27.0*	-	-
	4	**-7.5**	**-3.4**	*-27.6*	-	-
$r_1 = 0.8$	5	**-7.0**	**-4.6**	*-29.7*	-	-
	6	**-5.1**	**-4.9**	*-32.4*	-	-
	7	**-2.0**	**-4.2**	*-36.3*	-	-
	8	**1.7**	**-2.4**	*-39.1*	-	-
	9	**6.2**	**0.3**	*-42.8*	-	-
	10	**9.9**	**2.7**	*-46.8*	-	-

APPENDIX D

RELATIONSHIP BETWEEN T-YEAR EVENTS AND EVENTS WITH T-YEARS RETURN PERIOD

The rarity of an event is univocally determined once the probability distribution of the underlying process is defined. However in engineering practice such probability is seldom applied "as is", but instead, it is expressed in terms of abstract concepts such as return period T or T-year event. Assuming the event in question can be measured in terms of a random variable Y, the return period of an event Y_o is defined as the mean inter-arrival time between two events $Y > Y_o$, whereas the risk in T years is defined as the probability of observing at least one event $Y > Y_o$ in T years. Alternatively, the so called T-year event is often used, whereas Y_o is defined as the expected value of the maximum Y that can occur in T-years.

Note that in general the event Y_o with return period T will not coincide with the T-year event. Put in another way, the return period of a T-year event is not T. In order to show the above point, we will assume first that the event Y in question occurs once every year. Furthermore, for the purpose of illustration and in order to obtain an analytically tractable solution, we will assume Y distributed according to a standard exponential with cdf:

$$F_Y(y) = 1 - e^{-y} \tag{D.1}$$

The event Y_o with return period T can be derived from the well known equation linking the return period to the cdf:

$$T = \frac{1}{1 - F_Y(Y_o)} \tag{D.2}$$

Solving Eq. (D.2) for Y_o one gets:

$$Y_o(T) = \ln T \tag{D.3}$$

By definition, the T-year event $Y_o^*(T)$ is the expected value of the largest among T events Y , i.e. the expected value of the largest order statistic, which for a standard exponential is given by (Johnson et al., 1994):

$$Y_o^*(T) = \sum_{j=1}^{T} \frac{1}{j} \tag{D.4}$$

Note how the two critical events differ. Also it may be shown that for $T > 1$, $Y_o^*(T) > Y_o(T)$.

Combining Eqs. (D.2) with (D.4), the return period T of a T^*-year event can be computed as:

$$T = \frac{1}{\exp(-\sum_{j=1}^{T^*} \frac{1}{j})} \tag{D.5}$$

It may be of some use to derive an asymptotic expression for the ratio between the return period of the T^*-year event and T^*, i.e. $\kappa = \lim_{T^* \to \infty} \frac{T}{T^*}$. To this extent, we will make use of the following identity for the harmonic sum in D.4 (Abramowitz and Stegun, 1965):

$$\sum_{j=1}^{T^*} \frac{1}{j} = \gamma + \psi(T^* + 1) \tag{D.6}$$

where $\gamma = 0.577215665\ldots$ is the Euler's constant and $\psi(T + 1)$ is the digamma function. Thus:

$$\kappa = \lim_{T^* \to \infty} \frac{e^{\gamma + \psi(T^*+1)}}{T^*} \tag{D.7}$$

In order to solve the above limit, we will make use of the following series expansion for the digamma function (Abramowitz and Stegun, 1965):

$$\psi(z) = \ln z - \frac{1}{2z} - \frac{1}{12z^2} + \frac{1}{120z^4} - \ldots \tag{D.8}$$

Neglecting all the terms after the first, since they will vanish quickly as $z \to \infty$, substituting into Eq. (D.7) one get:

$$\kappa = \lim_{T^* \to \infty} \frac{e^{\gamma + \ln(T^*+1)}}{T^*} = \lim_{T^* \to \infty} e^{\gamma} \frac{(T^* + 1)}{T^*} = e^{\gamma} = 1.781\ldots \tag{D.9}$$

Thus, for large T^*, $T \approx 1.781T^*$, which confirms that the return period of the T^*-year event is greater than T^*.

The above analysis refers to a simple case, i.e. when the hydrological variable is distributed according to a standard exponential. Nonetheless, it should be expected that similar results should hold for a broad spectrum of distributions. For instance, if Y has a standard logistic distribution (Arnold et al., 1992):

$$F_Y(y) = \frac{1}{1 + e^{-y}} \tag{D.10}$$

it can be shown that also in this case $\kappa = e^{\gamma}$. Indeed $Y_o^*(T^*) = \psi(T^*) + \gamma$ (Arnold et al., 1992), and therefore the relationship between the return period T and $T*$ becomes:

$$T = \frac{1}{1 - \frac{1}{1+e^{\psi(T^*)+\gamma}}} \tag{D.11}$$

Dividing by T^*, making use again of the approximation $\psi(z) \approx \ln z$ and taking the limit $T^* \rightarrow \infty$, the result $\kappa = e^{\gamma}$ is obtained.

On the other hand, for other heavy tail distributions, the above result may not hold, not even as an approximation. As an example, we will repeat the above analysis assuming that the Y is distributed according to a heavy tailed Pareto with cdf:

$$F_Y(y) = 1 - \left(\frac{k}{y}\right)^a \tag{D.12}$$

where $a > 0$ and $k > 0$ are parameters and $y \geq k$. The expected value of the largest T^* order statistic (T^*-year event) is given by (Johnson et al., 1994):

$$Y_o^*(T^*) = \frac{\Gamma(T^*+1)\Gamma(1-a^{-1})}{\Gamma(T^*+1-a^{-1})}k \tag{D.13}$$

From equation D.12, the return period T of the T^*-year event is:

$$T = \frac{1}{1 - F_Y(Y_o^*(T^*))} = \left[\frac{Y_o^*(T^*)}{k}\right]^a = \left[\frac{\Gamma(T^*+1)\Gamma(1-a^{-1})}{\Gamma(T^*+1-a^{-1})}\right]^a \tag{D.14}$$

Once again we are interested to the limit $\kappa_P = \lim_{T^* \to \infty} \frac{T}{T^*}$, where the subscript in κ stands for Pareto. After letting $\alpha = 1 - a^{-1}$, making use of the identity $\Gamma(x+1) = x\Gamma(x)$ one gets:

$$\frac{T}{T^*} = \frac{1}{T^*}\left[\frac{T^*\Gamma(T^*)\Gamma(\alpha)}{\Gamma(T^*+\alpha)}\right]^a \tag{D.15}$$

For $1 < \alpha < 0$ the following result holds:

$$\lim_{x \to \infty} \frac{1}{x^{\alpha}} \frac{\Gamma(x+\alpha)}{\Gamma(x)} = 1 \tag{D.16}$$

Thus, combining the above result, with Eq. (D.15), one finally gets:

$$\kappa_P = \Gamma^a \left(1 - a^{-1}\right) \tag{D.17}$$

In Figure D.1 the plot of κ_P vs. a is shown, together with the value $\kappa = e^{\gamma}$ previously derived for the exponential. It can be inferred how $\kappa_P > \kappa$, as expected due to the heavy tail of the Pareto. Also $\kappa_P \to \infty$ as $a \to 1$, which means that T can be much larger than T^*. Interestingly enough, as $a \to \infty$, κ_P approaches κ, which seems to indicate that for large values of the parameter a, the Pareto behaves like an exponential.

In order to verify the above analysis, a simulation experiment has been set up using six different distributions, namely normal, log-normal, exponential, standard logistic, gamma, and Pareto. The parameters of the normal, log-normal, exponential, and Gamma have been fitted by method of moments assuming a mean equal to 700 mm and a standard deviation equal to 200 mm, which are typical values for annual precipitation in Mediterranean climates. For the Pareto, $k = 1$ and several values of $a = 1.5, 2, 5, 10$ have been considered. Then, for fixed T^*, the T^*-event has been estimated by averaging the maximum values extracted from 1000 non-overlapping T^* years generated series, and the return period T of such a value has then been computed making use of the known cdf.

In Figure D.2, return period T of T^*-year events computed by simulation using the aforementioned distributions is plotted vs. T^*. In the same figure, the asymptotic

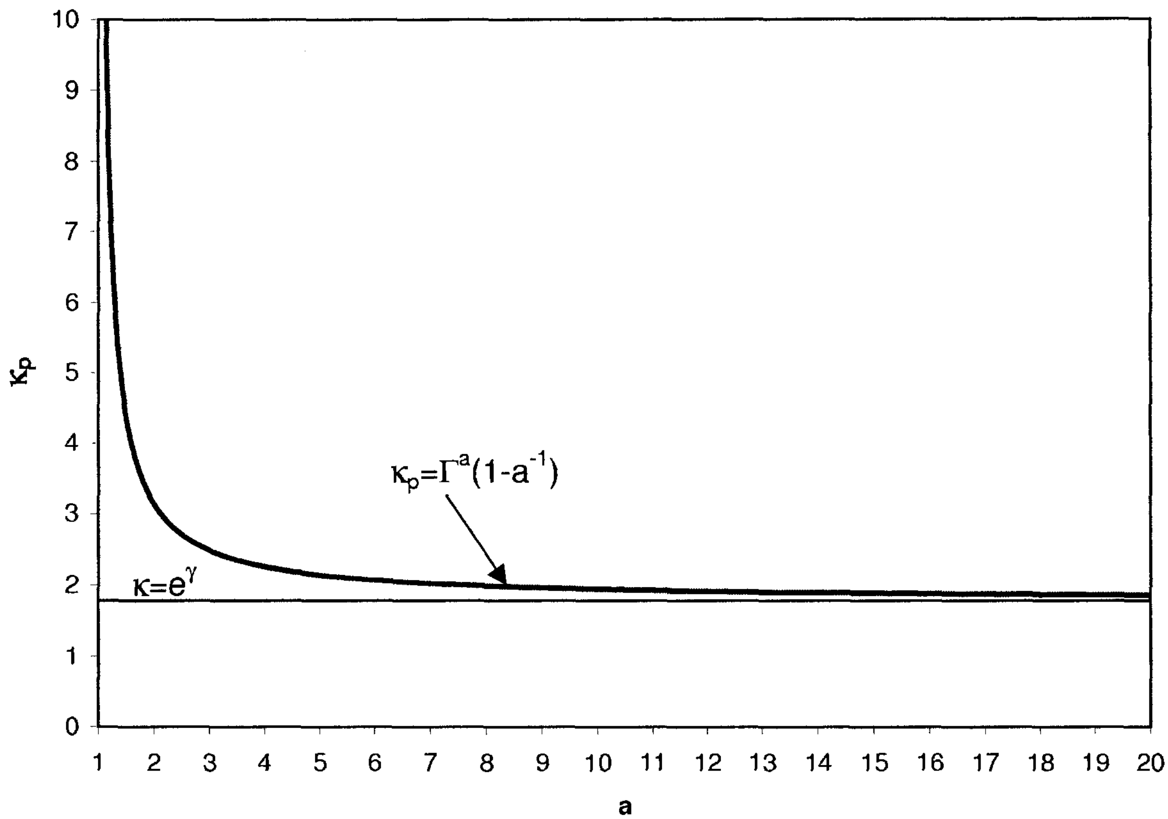


Figure D.1: Asymptotic ratio $\kappa_P = \frac{T}{T^*}$ for a Pareto distribution as a function of the parameter a. Note the vertical asymptote as $a \to 1$, and how $\kappa_P \to \kappa$ as $a \to \infty$

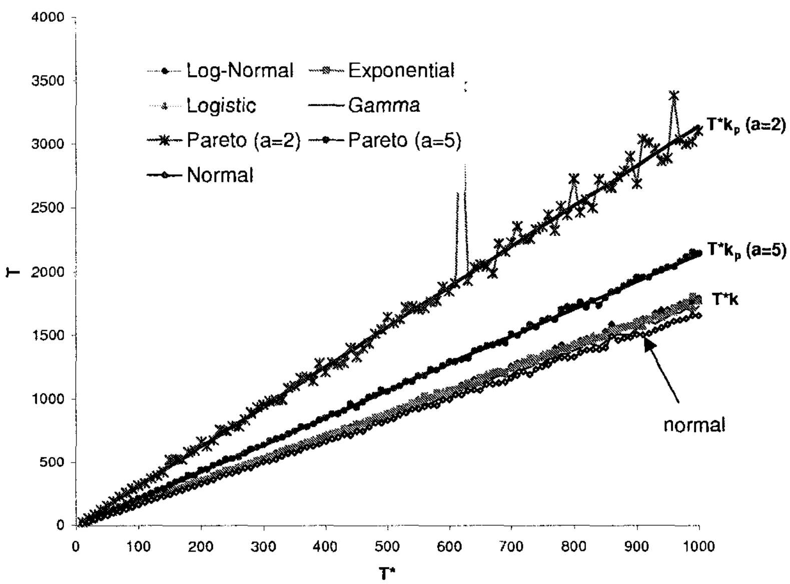


Figure D.2: Return period T of T^*-year events for different distributions, computed by simulation. The asymptotic linear relationships are also plotted

linear relationship $T = \kappa T^*$ and $T = \kappa_P T^*$, derived for the exponential or logistic and for the Pareto respectively are also plotted. It can be inferred that the curves, apart from the sampling variability, tend to align around the corrisponding asymptotic linear relationships. In particular the log-normal and gamma distributions, tend to align around the line $T = \kappa T^*$, which seems to indicate that the asymptotic ratio κ may be valid for other distributions besides the exponential and logistics. The normal distribution on the other hand, appears to yield values of the ratio $\frac{T}{T^*}$ slightly lower than κ, although the latter can still be assumed as a very good approximation.

When complex events are considered, i.e. events with duration that spans several years such as droughts, similar conclusions can be reached using similar arguments. More specifically, with reference to an hydrological series X_t and to a demand threshold level x_o, let $F_L(l)$ and $F_D(d)$ be the cdf's of drought length and accumulated deficit respectively. For the sake of illustration, we will assume that D can be modelled by a standard exponential distribution.

When multi year events are considered, computation of the T^*-year critical drought is somewhat more involving then in the single year events case, since the number of droughts $N(T^*)$ in a T^* years is a random variable. More specifically, let's start by observing that, for a fixed number of droughts n the cdf of the maximum of D is given by:

$$F_{Y|n}(y) = F_D^n(y) = e^{-yn} \tag{D.18}$$

where $f_{N(T)}(n)$ is the pdf of the number of drought in T-years $N(T)$. Such pdf has been derived by Shiau and Shen (2001) for the case when the sequence of deficits and surplus can be modeled by a Markov chain:

$$\begin{aligned} f_{N(T)}(n) \quad &= \sum_{i=2n-1}^{T} {p_{01}}^{n-1}{p_{10}}^{n} \sum_{j=0}^{i-2n+1} \binom{n+j-2}{j} \binom{i-n-j}{i-2n-j+1} {p_{00}}^{j}{p_{11}}^{i-2n-j+1} - \\ &- \sum_{i=2n+1}^{T} {p_{01}}^{n}{p_{10}}^{n+1} \sum_{j=0}^{i-2n-1} \binom{n+j-1}{j} \binom{i-n-j-1}{n} {p_{00}}^{j}{p_{11}}^{i-2n-j-1} \end{aligned} \tag{D.19}$$

where $p_{00} = 1 - p_{01}$ and $p_{11} = 1 - p_{10}$.

In order to avoid the analytical complications involved in using Eq. (D.19), following Sen (1977), we will assume that the number n of droughts in a T^* year period can be approximately modelled by a Poisson distribution, with parameter $\lambda = \frac{T^* p_{01} p_{10}}{p_{01}+p_{10}} = T^*\beta$, i.e.:

$$f_{N(T^*)}(n) = \frac{e^{-T^*\beta}(T^*\beta)^n}{n!} \tag{D.20}$$

Substituting Eq. (D.20) into (D.28), assuming $T^* \to \infty$ one gets:

$$Y_o^*(T^*) = \sum_{n=0}^{\infty} \frac{e^{-T^*\beta}(T^*\beta)^n}{n!}(\gamma + \ln n) \tag{D.21}$$

$$= \gamma + \sum_{n=0}^{\infty} \frac{e^{-T^*\beta}(T^*\beta)^n}{n!} \ln n \tag{D.22}$$

$$= \gamma + \sum_{n=0}^{\infty} \frac{e^{-T^*\beta}(T^*\beta)^n}{n!} \ln n \tag{D.23}$$

As shown previously, assuming that the sequence of surplus and deficits follows a lag-1 Markov chain, the return period of a drought $D > D_o$ can be expressed as:

$$T = \frac{p_{01} + p_{10}}{p_{01}p_{10}\left(1 - F_D(D_o)\right)} \tag{D.24}$$

where $p_{01} = \mathrm{P}[X_t \leq x_o | X_{t-1} > x_o]$ and $p_{10} = \mathrm{P}[X_t > x_o | X_{t-1} \leq x_o]$. From the above equation, solving for D_o one gets:

$$D_o(T) = F^{\leftarrow}\left[1 - \frac{p_{01} + p_{10}}{p_{01}p_{10}T}\right] \tag{D.25}$$

Assumind D distributed according to a standard exponential, the above equation simplifies as:

$$D_o(T) = \ln\left[1 - \frac{p_{01} + p_{10}}{p_{01}p_{10}T}\right] \tag{D.26}$$

If we fix the number of droughts $N(T) = n$, $Y_o^*(T)|_{N(T)=n}$ is given, similarly as before, by:

$$Y_o^*(T^*)|_{N(T^*)=n} = \sum_{j=0}^{n} \frac{1}{j} \approx \gamma + \ln n \tag{D.27}$$

where the approximation is valid for large n. The asymptotic unconditional value can be computed by:

$$Y_o^*(T^*) \approx \sum_{n=0}^{(T^*-1)/2} f_{N(T^*)}(n)(\gamma + \ln n) \tag{D.28}$$

Bibliography

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions.* Dover Pubblications, Inc., New York.

Arnold, B., Balakrishnan, N., and Nagaraja, H. (1992). *A first course in order statistics.* Wiley & Sons, New York.

Bayazit, M. and Onoz, B. (2005). "Probabilities and return periods of multisite droughts." *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, 50(4), 605.

Ben-Zvi, A. (1987). "Indices of hydrological drought in israel." *Journal of Hydrology*, 92, 179–191.

Bonaccorso, B., Cancelliere, A., and Rossi, G. (2003). "An analytical formulation of return period of drought severity." *Stochastic Environmental Research and Risk Assessment*, 17(1), 157–174.

Bras, R. L. (1990). *Hydrology: An introduction to hydrologic science.* Addison-Wesley, Reading, Mass.

Buishand, R. 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.

Cancelliere, A., Bonaccorso, B., Rossi, G., and Salas, J. (2003). "On the probabilistic characterization of drought events." 23rd AGU Hydrology Days, Fort Collins, CO, April 2003.

Cancelliere, A. and Rossi, G. (2003). "Droughts in Sicily and comparison of identified droughts in Mediterranean regions." *Tools for drought mitigation in Mediterranean regions*, G. Rossi, A. Cancelliere, T. Oweis, L. Pereira, M. Shatanawi, and A. Zairi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 103–122.

Cancelliere, A. and Salas, J. (2004). "Drought length properties in periodic-stochastic hydrologic data." *Water Resources Research*, 40(2), 13.

Chang, T., Kavvas, M., and Delleur, J. (1984a). "Daily precipitation modeling by discrete autoregressive moving average processes." *Water Resources Research*, 20(5), 565–580.

Chang, T., Kavvas, M., and Delleur, J. (1984b). "Modeling sequences of wet and dry days by binary discrete autoregressive moving average processes." *J. Climate and Appl. Meteorology*, 23, 1367–1378.

Cheebane, M., Salas, J., and Boes, D. (1995). "Product periodic autoregressive processes for modeling intermittent monthly streamflows." *Water Resources Research*, 31(6), 1513–1518.

Chung, C. (1999). *Probability distribution, risk and return period of dependent hydrological events.* Ph.D. Thesis, Colorado State University, Fort Collins, CO.

Chung, C. and Salas, J. (2000). "Drought occurrence probabilities and risk of dependent hydrologic processes." *Journal of Hydrological Engineering*, 5(3), 259–268.

Clausen, B. and Pearson, C. P. (1995). "Regional frequency analysis of annaul maximum streamflow drought." *Journal of Hydrology*, 173, 111–130.

Cramer, H. and Leadbetter, M. (1967). *Stationary and related stochastic processes.* John Wiley & Sons, New York.

de Moivre, A. (1967). *The doctrine of chances, or, A method of calculating the probabilities of events in play.* Cass, London (Facsimile of the original print by H. Woodfall, 1738).

Douglas, E. M., Vogel, R. M., and Kroll, C. N. (2002). "Impact of streamflow persistence on hydrologic design." *Journal of Hydrologic Engineering*, 7(3), 220–227.

Downer, R., Siddiqui, M., and Yevjevich, V. (1967). "Application of runs to hydrologic droughts." Proceedings of the International Hydrology Symposium, Fort Collins, CO, September 1967.

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

Feller, W. (1968). *An introduction to the probability theory and its application*, Vol. I. Wiley, New York.

Fernandez, B. and Salas, J. (1999). "Return period and risk of hydrologic events. I: mathematical formulation." *Journal of Hydrologic Engineering*, 4(4), 297–307.

Fernandez, B. and Vergara, A. (1998). "Risk of scarcity of monthly precipitation and streamflows in semiarid regions." *Hydrological Sciences Journal*, 43(5), 759–773.

Filliben, J. J. (1975). "The probability plot correlation coefficient test for normality." *Technometrics*, 17(1), 111–117.

Finney, D. (1962). "Cumulants of truncated multinormal distributions." *Journal of the Royal Statistical Society. Series B (Methodological)*, 24(2), 535–536.

Foley, J. (1957). "Drought in australia." *Bulletin 43*, Bureau of Meteorology, Malbourne, Australia.

Frank, M. (1979). "On the simultaneous associativity of f(x,y) and x+y-f(x,y)." *Aequationes Math.*, 19, 194–226.

Frick, D. M., Bode, D., and Salas, J. D. (1990). "Effect of drought on urban water supplies. i: Drought analysis." *Journal of Hydrologic Engineering*, 116(6), 733–753.

Gonzalez, J. and Valdes, J. (2003). "Bivariate drought recurrence analysis using tree ring reconstructions." *Journal of Hydrologic Engineering*, 8(5), 247–258.

Gonzalez, J. and Valdes, J. (2006). "New drought frequency index: Definition and comparative performance analysis." *Water Resources Research*, 42, 1–13.

Guven, O. (1983). "A simplified semiempirical approach to probabilities of extreme hydrologic droughts." *Water Resources Research*, 19(2), 441–453.

Herbst, P. H., Bredekamp, D. B., and Barker, H. M. G. (1966). "A technique for the evaluation of drought from rainfall data." *Journal of Hydrology*, 4(3), 264–272.

Jacobs, P. and Lewis, P. (1977). "A mixed autoregressive-moving average exponential sequence and point process." *Adv. Appl. Prob.*, 9, 87–104.

Jacobs, P. and Lewis, P. (1978). "Discrete time series generated by mixtures i: Correlational and and runs properties." *J. Roy. Stat. Soc.*, B40, 94–105.

Johnson, N., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions*, Vol. 1. Wiley & Sons, New York.

Katz, R. and Parlange, M. (1995). "Generalization of chain dependent processes: Application to hourly precipitation." *Water Resources Research*, 31(5), 1331–1341.

Kingery, R. J. (1992). *A stochastic analysis of spatial droughts in Colorado*. M.S. Thesis, Colorado State University, Fort Collins, CO.

Kotz, S., Balakrishnan, N., and Johnson, N. (2000). *Continuous Multivariate Distributions*, Vol. 1. Wiley & Sons, New York.

Leppard, P. and Tallis, G. (1989). "Algorithm as249: Evaluation of the mean and covariance of the truncated multinormal distribution." *Applied Statistics*, 38(3), 543–553.

Llamas, J. and Siddiqui, M. (1969). "Runs of precipitation series." *Hydrology paper 33*, Colorado State University, Fort Collins, Colorado.

Lloyd, E. and Saleem, S. (1979). "Waiting time to first achievement of specified levels in reservoirs subjects to seasonal markovian inflows." *Inputs for risk analysis in water systems*, E. Mc Bean, K. Hipel, and T. Unny, eds., Water resources publications, Littleton, CO.

Lloyd, E. H. (1970). "Return period in the presence of persistence." *Journal of Hydrology*, 10(3), 202–215.

Loaicigica, H. and Leipnik, R. B. (1996). "Stochastic renewal model of low-flow streamflow sequences." *Stoch. Hydrol. Hydraul.*, 10, 65–85.

Loaicigica, H. A. and Mariño, M. A. (1991). "Recurrence interval of geophysical events." *Journal of Water Resources Planning and Management*, 117(3), 367–382.

Matalas, N. (1967). "Mathematical assessment of synthetic hydrology." *Water Resources Resources*, 3(4), 937–945.

Mathier, L., Perreault, L., and Bobée, B. (1992). "The use of geometric and gamma related distributions for frequency analysis of water deficit." *Stochastic Hydrology and Hydraulics*, (6), 239–254.

McKee, T., Doesken, N., and Kleist, J. (1993). "The relationship of drought frequency and duration to time scales." 8th Conference on Applied Climatology, Anaheim, California. 179–184.

Millan, J. and Yevjevich, V. (1971). "Probabilities of observed droughts." *Hydrology paper 50*, Colorado State University, Fort Collins, Colorado.

Mood, A., Graybill, F., and Boes, D. (1974). *Introduction to the theory of statistics.* McGraw-Hill, New York.

Palmer, W. C. (1965). "Meteorological drought." *Research Paper 45*, U.S. Weather Bureau, Washington, D.C.

Rossi, G. (1988). "On the identification of regional droughts." *Excerpta of the Italian Contributions to the Field of Hydraulic Engineering*, 3, 145–166.

Rossi, G. (2003). "Analisi della siccita in una regione." *Piene e siccita. Atti del corso su eventi idrologici estremi*, E. Guggino, G. Rossi, and E. Todini, eds., CLUP, Milano, 361–387.

Rossi, G., Benedini, M., Tsakiris, G., and Giakoumakis, S. (1992). "On regional drought estimation and analysis." *Water Resources Management*, (6), 249–277.

Rossi, G. and Cancelliere, A. (2003). "At site and regional identification by REDIM model." *Tools for drought mitigation in Mediterranean regions*, G. Rossi, A. Cancelliere, T. Oweis, L. Pereira, M. Shatanawi, and A. Zairi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 37–54.

Russel, C. S., Arey, D. G., and Kates, R. W. (1971). *Drought and water supply: implications of the Massachusetts experiences for municipal planning.* John Hopkins Press, Baltimore, Maryland.

Salas, J. (1993). "Analysis and modeling of hydrologic time series (chap. 19)." *Handbook of Hydrology*, D. Maidment, ed., McGraw Hill Book Co., N. York.

Salas, J., Chung, C., and Cancelliere, A. (2005). "Correlations and crossing rates of periodic-stochastic hydrologic processes." *Journal of Hydrological Engineering*, 10(4), 280–287.

Salas, J., Chung, C., and Fernandez, B. (2001). "Relating autocorrelation and crossing rates of continuous- and discrete-valued hydrological processes." *Journal of Hydrologic Engineering*, 6(2), 109–118.

Salas, J., Fu, C., Cancelliere, A., Dustin, D., Bode, D., Pineda, A., and Vincent, E. (2005). "Characterizing the severity and risk of droughts of the poudre river, colorado." *Journal of Water Resources Planning and Managament*, 131(5), 383–393.

Saldarriaga, J. and Yevjevich, V. (1970). "Applications of run lengths to hydrologic series." *Hydrology paper 40*, Colorado State University, Fort Collins, Colorado.

Santos, M. (1983). "Regional droughts: a stochastic characterization." *Journal of Hydrology*, 66(1-4), 183–211.

Schwager, S. J. (1983). "Run probabilities in sequences of markov-dependent trials." *Journal of American Statistical Association*, 78, 168–175.

Sen, Z. (1976). "Wet and dry periods of annual flow series." *Journal of the Hydraulics Division*, 102(HY10), 1503–1514.

Sen, Z. (1977). "Run-sums of annual flow series." *Journal of Hydrology*, 35(3/4), 311–324.

Sen, Z. (1980a). "Regional drought and flood frequency analysis: Theoretical consideration." *Journal of the Hydrology*, 46(3/4), 265–279.

Sen, Z. (1980b). "Statistical analysis of hydrologic critical droughts." *Journal of the Hydraulics Division*, 106(HY1), 99–104.

Sharma, T. (1995). "Estimation of drought severity on independent and dependent hydrologic series." *Water Resources Management*, 11, 35–49.

Sheppard, W. (1899). "On the application of the theory of error to cases of normal distribution and normal correlation." *Philosophical Transactions of the Royal Society of London, Series A*, 192, 101–167.

Shervish, M. (1984). "Algorithm as195: Multivariate normal probability with error bound." *Applied Statistics*, 33(3), 81–94.

Shiau, J. and Shen, H. (2001). "Recurrence analysis of hydrologic droughts of differing severity." *Journal of Water Resources Planning and Management*, 127(1), 30–40.

Tallis, G. (1961). "The moment generating function of the truncated multi-normal distribution." *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(1), 223–229.

Tase, N. (1976). *Area-deficit-intensity characteristics of droughts.* Ph.D. Thesis, Colorado State University, Fort Collins, CO.

Taylor, H. and Karlin, S. (1998). *An Introduction to Stochastic Modeling.* Academic Press, San Diego.

Todorovich, P. (1970). "On some problems involving a random number of random variables." *Annals of mathematical statistics*, 41.

Vogel, R. M. (1987). "Reliability indices for water supply systems." *Journal of Water Resources Planning Management*, 113(4), 645–654.

Yevjevich, V. (1967). "An objective approach to definitions and investigations of continental hydrologic droughts.." *Hydrology paper 23*, Colorado State University, Fort Collins, Colorado.

Yevjevich, V. (1972). *Stochastic Processes in Hydrology.* Water Resources Publication, Fort Collins, Colorado.

Yevjevich, V., Da Cunha, L., and Vlachos, E. (1983). *Coping with Droughts.* Water Resources Publications, Littleton, Colorado.

Zelenhasic, E. and Salvai, A. (1987). "A method for streamflow drought analysis." *Water Resources Research*, 23(1).