Hydrological Drought Analysis-occurrence, severity, risks: the case of Wabi Shebele River Basin, Ethiopia

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
(Dr.-Ing.)

vom Fachbereich Bauingenieurwesen
der Universität Siegen

genehmigte

Dissertation

von

Adane Abebe Awass, M.Sc

Referent	Univ. Prof. Dr.-Ing Gerd Förch,	Universität Siegen
Korreferent	Univ. Prof. Dr.-Ing habil. Hans-B. Horlacher,	TU Dresden

Tag der mündlichen Prüfung 16. Januar 2009

Siegen 2009

Declaration of Authorship

I, ADANE ABEBE AWASS, declare that this thesis titled, 'HYDROLOGICAL DROUGHT ANALYSIS-occurrence, severity, risks: the case of Wabi Shebele River Basin, Ethiopia' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University
- Where I have consulted the published work of others, this is always clearly attributed
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
- I have acknowledged all main sources of help

Signed: _____

Date: _____

To

Victims of hydrological extremes in the region.

"When the well is dry, we know the worth of water,"

Benjamin Franklin (1706-1790)

Acknowledgements

This research work is financially supported mainly from Gessellschaft fur Technische Zusammenarbeit/German Technical Cooperation (GTZ) under the framework of the support program to the Water Technology Institute in Arba Minch University, Ethiopia. The financial support for the last year is obtained from Universitat Siegen through the ACP-EU water facility programme. I would like to express my great gratitude to all of them for making the fund availble. Thanks are also due to Deutscher Acadamischer Austauschdienst(DAAD) for the scholarship and supports I obtained.

I would like to thank my supervisor Prof. Dr.-Ing Gerd Förch for his invaluable advices and supports throughout the research period. I thank Prof. Dr.-Ing. Hans-B. Horlacher (TU Dresden) for his continuous very helpful comments. I acknowledge the administrative support I obtained from Dr. Stefan Thiemann, Dipl.-Ing. Ingrid Althoff and Dipl.-Geecol. Beate Böhme, all members of the Research Institute for Water and Environment, Universitat Siegen.

I am grateful to the dendrochronology group at Wageningen University, the Netherlands for hosting and supporting the 2-month laboratory work. I am thankful to the staffs at the center for ecosystems studies, Wageningen University for the great experience and social atmosphere. Particularly, I would like to extend my appreciation to Dr. Ute Sass-Klaassen for all the help. Thanks Ute! A multitude of experts with whom I had the opportunity to discuss my views during the course of the research helped me while I was writing this dissertation. I would like to express my deepest appreciation and heart felt thanks to all who gave their time so generously.

I am thankful to the hydrology section of Ministry of Water Resources and National Meteorological Services, Ethiopia for the provision of streamflow and Meteorological data respectively. I am thankful to Arba Minch University for providing the facilities during my fieldwork in Ethiopia.

I would like to thank Kassa Tadele for sharing all the fun and sorrow during my stay in Siegen. Last but by no means least I would like to convey a heartfelt gratitude to my family and friends for their unfailing support.

Adane Abebe Awass,
Universität Siegen, Germany, June 2008

Abstract

by Adane Abebe Awass

This study analyses hydrological drought with due emphasis to ungauged catchments. Identification of hydrological drought and methods of unveiling its intrinsic multi-variate characteristics are investigated. The severity of drought has been examined using a multitude of methodologies. Quite often, absence of recorded long time streamflow data hinders a reliable drought analysis and understanding of the phenomenon in the past. Signatures of water stress are imprinted on tree rings. In this study streamflow reconstruction is achieved by coalescing proxy data from riparian tree rings and climatic indices. The generated data are used for extracting the multivariate features of this extreme hydrological event. The case study is demonstrated in Wabi Shebele river basin in Ethiopia. The methods proposed here are applicable to other similar river basins.

Extracting hydrological drought entailed defining a possible threshold level. Threshold levels of low exceedance probability are found to be appropriate unlike the commonly used higher exceedance probabilities in temperate climate. It was corroborated that the frequency of some notable recent droughts as revealed with data generated from proxy records well matches that of instrumental data. Stochastic simulation of hydrological drought is done using SARIMA models from time series of instrumental monthly streamflow records in the study area. The nonlinear dependency between severity and duration of hydrological drought is studied using copula models. Different copula families and parameter estimation techniques are evaluated. Joint and conditional probabilities of severity and duration of drought in the area is specified from the derived relationship to provide empirical insight on the nature of the extreme events. It is demonstrated that droughts have distinct spatial patterns regarding temporal evolution and variability in the study area in which the geographic and climatic characteristics have a high degree of diversity. The relative strengths in association between the climatic, morphometric and geologic features of the catchment to the base flow estimates are weighted and a plausible relationship is produced. However, in view of the tremendous spatio-temporal heterogeneity of climatic and landscape properties extrapolation of response information or knowledge from gauged to ungauged basins remains fraught with considerable difficulties and uncertainties. Catchment characteristics can be related to low flows thus are used to delineate hydrologically homogeneous pools. Severity-area-frequency analysis of drought in the area using nonparametric kernels shows high variability of drought events within the pools formed. Associated relative risks of drought are also scrutinized using multicriteria analysis. No single variable is sufficient to portray the complexity of the vulnerability of an area to drought.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ACF	Autocorrelation Function
ADI	Aggregate Drought Index
AIC	Akaike Information Criterion
ARIDE	Assesment of the Regional Impacts of Drought in Europe
ARIMA	Auto Regressive Integrated Moving Average
AVHRR	Advanced Very High Resolution Radiometer
AWC	Accumulated Water-holding Capacity
BFI	Base Flow Index
BIC	Bayes Information Criterion
CMI	Crop Moisture Index
CWT	Continuous Wavelet Transforms
DCMM	Double Chain Markov Model
DFID	Department for International Development (UK)
ENSO	El Niño Southern Oscillation Index
EOF	Empirical Orthogonal Functions
HMM	Hidden Marcov Model
IFM	Inference of Margins
MAM	Mean Annual Minimum
MoWR	Ministry of Water Resources (Ethiopia)
MTD	Mixture Transition Distribution model
NDVI	Normalized Difference Vegetation Index
NMSA	National Meteorological Services Agency (Ethiopia)
NOAA	National Oceanic and Atmospheric Administration (USA)
PACF	Partial Autocorrelation Function
PCA	Principal Component Analysis

PDSI	Palmer Drought Severity Index
PHDI	Palmer Hydrologic Drought Index
RBF	Radial Basis Functions
SBC	Schwarz Bayesian Criterion
scPDSI	Self calibrating Palmer Drought Severity Index
scPHDI	Self calibrating Palmer Hydrologic Drought Index
SNNPR	Southern Nations Nationalities and Peoples Region
SOI	Southern Oscillation Index
SPI	Standardized Precipitation Index
SST	Sea Surface Temperature
SVM	Support Vector Machines
SWSI	Surface Water Supply Index
UNEP	United Nations Environment Programme
USAID	United States Agency for International Development
WAPCOS	Water and Power Consulting Services Ltd

Chapter 1

Introduction

This study addresses the issues associated with hydrological drought analysis with special focus at data sparse ungauged catchments. A general background, research objectives, scope of the study, conceptual framework of the research and thesis outlines are introduced in this chapter.

1.1 Background and Statement of the Problem

River systems are most stressed in low flow periods and thus an understanding of drought conditions in time and space is fundamental to a wide range of water management problems. Low flow estimates are vital for planning water supplies, hydropower, irrigation systems, water quality management, issuing and renewing waste disposal permits, and for assessing the impact of prolonged drought on aquatic ecosystems. The ever-increasing demand on water resources calls for better management of the water deficit situation, may it be unusual droughts, or yearly recurring low flows. The consequences of droughts are felt most keenly in areas which are in any case arid (Beran and Rodier, 1985).

Unlike aridity, which is restricted to low rainfall and high potential evapotranspiration regions and is a permanent feature of climate, drought is a temporary aberration. It occurs in virtually all climatic zones. It is not possible to avoid droughts. Proper definition of droughts and quantification of its characteristics is essential for drought preparedness and management of its impacts (Smakthin and Hughes, 2004). Drought characteristics vary significantly from one region to another. Defining drought is therefore difficult;

it depends on differences in regions, needs, and disciplinary perspectives. Such notions always involve a degree of subjectivity. There is no way to objectively choose a threshold of drought. Drought can be defined as the naturally occurring phenomenon when the natural water availability is significantly below the normally recorded level. The notion of drought should not be confused with aridity, where water is always in short supply. Nor should it be used to describe the normal year-to-year dry season, which is part of the climatic regime. The term is rarely applied to short periods, say of weeks or one or two months duration. Droughts include low flow periods, but a continuous seasonal low-flow event does not necessarily constitute a drought.

Time variability of streamflow, at all time scales, is more important from a water resources point of view than the irregularity of the rainfall. There is an increasing tendency to rely on streamflow, groundwater, etc than precipitation as sources of water supply. The irregularity of runoff and the low flows that are often experienced require careful analysis before attempts are made to utilize the water for development. An extensive review on low flow studies is provided in Smakthin (2001). In certain regions, where water supplies mainly depend upon river diversions, when dealing with a regional drought it may be necessary to consider not only precipitation but also stream-flow. Infact, streamflow is also a dependent variable, controlled by the current and antecedent precipitation. In many cases, where the river discharges are regulated by dams and other hydraulic structures, the drought definition may need to be more a reflection of the river management decisions than of the natural supply.

The interdependence between climatic, hydrologic, geologic, geomorphic, ecological and societal variables makes it very difficult to adopt a definition that fully describes the drought phenomena and the respective impacts. Moving the definition of drought from the conceptual level to the operational level typically requires quantification of “normal” or “expected” conditions within specified geographic boundaries and variations in societal conditions. For example in a fast responding tropical catchment a few days without rain might lead to a streamflow drought whereas in a semi-arid climate a stream could fall dry for several months and might still not be considered to be in a drought situation. Consequently, a method to derive streamflow drought characteristics developed in one region is not necessarily appropriate or even applicable in another region. In the semi-arid region for example, a good way of characterizing drought events might be the duration of zero-flow periods, in other regions however, streams never fall dry and one

might hastily conclude that these streams never experience a drought. Different methods of deriving streamflow drought characteristics are therefore needed in order to precisely describe the whole variety of streamflow droughts and for at-site drought studies one has to be careful selecting a method that suite the characteristics of the stream under study. Unlike floods the literature on hydrological drought is rather scant. The EU supported ARIDE project investigated the nature of European droughts; their duration, magnitude, extent and sensitivity to environmental change (Demuth and Stahl, 2001). This study, focused at the pan-European scale, investigated the temporal and spatial variability in drought behaviour.

Long-term data are required for any meaningful analysis of the flow regime in areas of high variability of flow. Quite often, absence of recorded long time streamflow data hinders a reliable drought analysis and understanding of the phenomenon in the past. Statistical techniques dealing with the duration aspects of drought are reasonably well developed, whereas techniques for severity aspects are less satisfactory and require considerable improvement and refinement (Panu and Sharma, 2002). Separate analysis of drought characteristics may not reveal the significant correlation relationships among drought characteristics. Recently Tallaksen and van Lanen (2004) provided extensive discussions on the processes and estimation methods for hydrological droughts and presented variety of techniques for assessment of low flows and hydrological droughts, including their prediction, forecasting and estimation at gauged/ungauged sites but it fall short on multivariate characterization of drought.

Regional estimate of drought severity is a good tool for water management purposes in a river basin. Developing a history of drought frequency, severity, duration and impacts for a region provides a greater understanding of the region's drought characteristics and the probability of drought recurrence at various levels of severity. As such, stochastic investigation is increasingly germane to analysis of hydrological drought. Although droughts are natural hazards, society can reduce its vulnerability and therefore lessen the risks associated with drought events.

Classification of Drought

Many classifications of drought from different perspectives exist (Yevjevich, 1967; Wilhite and M.H.Glantz, 1987; Tate and Gustard, 2000; Dracup et al., 1980). As drought

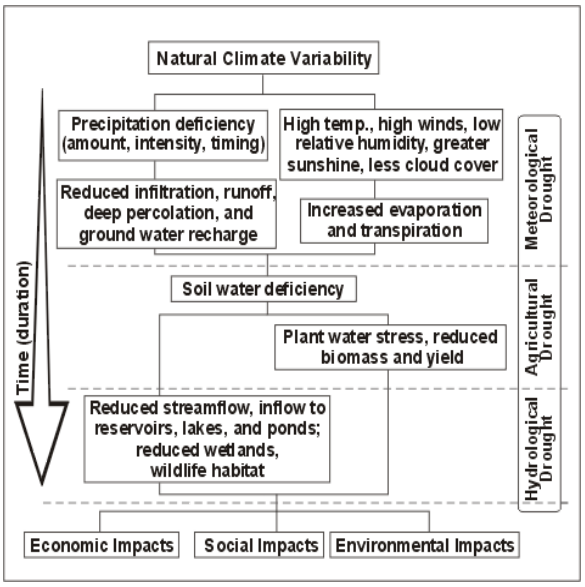


FIGURE 1.1: Propagation of drought through the hydrological cycle (NDMC, 2005)

propagates through the hydrological cycle, the different classes of drought are manifested (Figure 1.1).

1. **Meteorological drought:** Longer time intervals of no, or very little, rain are typically referred to as meteorological drought.

2. **Agricultural drought:** refers to low soil moisture and its effect on cultivated vegetation. It is a climatic excursion involving a shortage of precipitation sufficient to adversely affect crop production or range production. The term environmental drought is used to emphasize adverse consequences of water deficits on ecosystems. Nowadays extreme low flow events are diligently analyzed and given focus in the emerging field of ecohydrology.

3. **Hydrological drought:** Hydrological drought is a sustained occurrence of below expected natural water availability in rivers, lakes, groundwater level etc over large areas. It implies low flows and low levels of surface water (rivers, lakes) and of groundwater. In this research groundwater is not considered in detail for lack of time series of observations.

4. **Socioeconomic drought:** associates the supply and demand of some economic good with elements of meteorological, hydrological, and agricultural drought. Socioeconomic drought occurs when the demand for an economic good exceeds supply

as a result of a weather-related shortfall in water supply. A large body of theoretical and empirical literature has been developed that focuses on appropriate approaches for measuring direct economic impact of changes in water use levels and economic analysis of water resource developments to drought (Rogers et al., 1998).

1.2 Research Objective

The overall objective of this study is to develop appropriate tools for assessment of hydrological drought occurrence for water resources planning and management of especially ungauged catchments. It aims to stochastically model hydrological drought for prediction of its severity.

The specific objectives are to:

- develop approaches for assessment and prediction of hydrological drought in semi-arid regions.
- strengthen understanding of the spatial and temporal variability of hydrological drought in the study area.
- improve understanding of severity of hydrological drought at ungauged catchments using proxy data, paleohydrology.
- identify multivariate characteristics of hydrological drought.
- devise a mechanism to integrate or contribute to enhancing recent regionalization and prediction methodologies available using above. Develop a methodology to allocate ungauged basins within pooled clusters of hydrologically homogeneous drought regions.
- contribute to better identification of drought risk area and its prediction.

In the process, identification of appropriate tools to model the magnitude, duration and severity of drought in the study area will be outlined. The research will focus on investigating the magnitude, frequency, spatial extent, and seasonality of drought behavior, and associated risks and implications of drought incidents. It collates information on

the previous application of models on hydrological drought in the Wabi Shebele river basin. The study will seek to establish empirical relationships between the climatic, morphologic and geologic features of a catchment to its baseflow in an attempt to reveal low flow response in ungauged basins. A user-friendly interface will be proposed to allow users to easily acquire low-flow information in the study area. Investigating a history of drought frequency, severity and duration for a region provides a greater understanding of the region's drought characteristics and the probability of drought recurrence at various levels of severity. This type of information is indispensable in the development of drought response and mitigation strategies and preparedness plans.

1.3 Scope of the Study

The study will not address the phenomenal and causative aspects of drought. Seleshi (1998) has attempted to identify the climatic variables like ENSO and sea level changes that lead to changes in the precipitation magnitude and seasonality in North Eastern Africa. This study is bounded in scope to hydrological drought only. However, for a better understanding it will provide a cursory glance at the basics of meteorological drought and its link to hydrological drought. It will partly deal about human induced hydrological drought.

1.4 Conceptual Framework of the Research

The study will employ a plethora of techniques available in the subject-matter and critically evaluates in relation to the area of investigation. It will mainly rely on stochastic and data-driven models. Figure 1.2 presents the conceptual framework of the research.

The research mainly relies on meteorological, hydrological and proxy (like dendrochonologic) data sources. Characterising hydrological drought, analysing its temporal and spatial pattern from instrumental records and proxy data are considered in tandem. Multivariate characteristics of droughts are examined using copula modeling. The relative strengths of the risks of drought among the subbasins in the study area are investigated using multicriteria analysis. Each of the methodology employed as outlined in the components of the framework of the research and their justification are discussed

in detail under separate chapters. The associated results in relation to the envisaged objectives are also presented in subsequent sections.

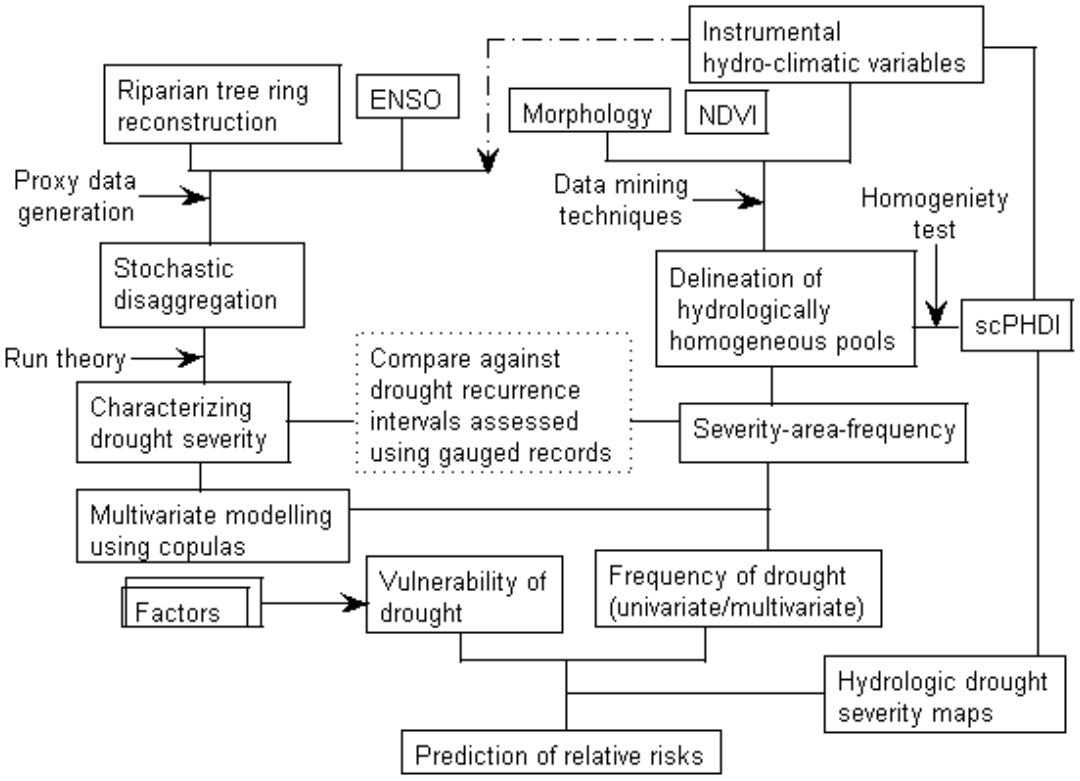


FIGURE 1.2: Conceptual framework of the research.

1.5 Organization of the Thesis

The thesis is structured into 10 chapters, a reference list and annexes.

In Chapter 1 the introductory background on the main theme, the research objectives and conceptual framework of the research are presented. The various chapters that follow systematically analyze different issues on the basis of this framework.

In Chapter 2 a description of the Wabi Shebele river basin, including its physiographic features, hydroclimatic condition and socio economic condition is presented.

Chapter 3 treats the different ways of characterizing properties of drought in general and hydrological drought in particular. It shows relevance of different drought indices to hydrological drought analysis.

Chapter 4 deals with hydrologically homogeneous drought pools. It presents methods of regionalization of low flows. The chapter illustrates catchment characteristics as predictors of baseflow.

Chapter 5 deals with spatial patterns of hydrological drought in the study area. This chapter presents the spatial coverage of recent hydrological drought.

Chapter 6 is dedicated to paleohydrology. Description and analysis of dendrochronologic records will be made. It is mainly a result of laboratory work. Reconstruction of streamflow from these proxy data is also dealt here.

Chapter 7 deals with temporal patterns of hydrological drought variables. It covers stochastic simulation of severity of drought in the area.

Chapter 8 deals with double chain markov modelling.

Chapter 9 treats prediction of risks of hydrological drought. It deals with simulating univariate frequency of drought using support vector machines and multivariate characteristics from proxy data using copula modelling.

Chapter 10 presents summaries and conclusions of the research. It also outlines recommendations for further research.

The reference list outlines the bibliography of the materials to which the respective citations allude.

The Appendix provides supplementary information to the materials presented above.

Chapter 2

Description of Study Area and Data

The area of investigation in this research is the Wabi Shebele river basin in Ethiopia. In this chapter main features of the study area are explained. It describes the location of the study area, its topographic characteristics, the geology, the land use/land cover and the hydroclimatic features. The socio-economics of the study area is also briefly discussed. Historical drought phenomenon in the basin are also highlighted. Metadata on hydroclimatic features are presented.

2.1 General Physiographic Condition

2.1.1 Location, Coverage

The Wabi Shebele river basin is one of the twelve major river basins in Ethiopia which is situated in the South Eastern part of the country. It lies between 4 ° 45' N to 9 ° 45'N latitude and 38 ° 45'E to 45 ° 30'E longitude, including the closed watershed of the Fafen and the Bio Ado (Figure 2.1). It springs from the Bale mountain ranges of the Galama and the Ahmar about 4000 m above mean sea level and drains into Indian ocean crossing Somalia. About 72% of the catchment (202, 220 Km^2) is lying in Ethiopia. The basin covers about 19% of the area of the country.

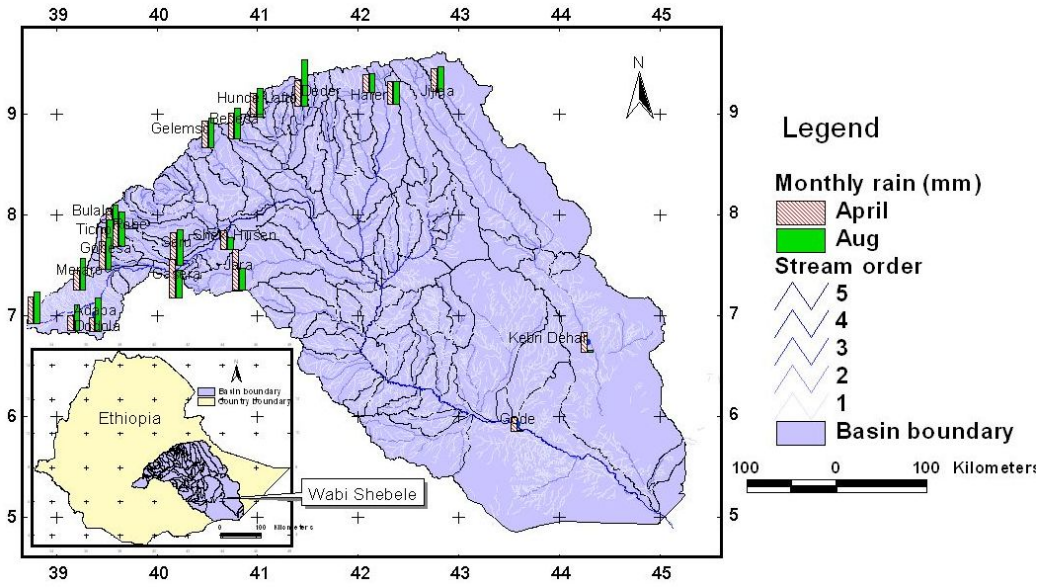


FIGURE 2.1: Location of Wabi Shebele river basin in Ethiopia.

It is bounded by Genale basin in the South West, rift valley in the West and North West, Awash basin in the North, Aysha Dewele in the North East, Ogaden in the East and Somalia in the South. The basin includes Oromiya, Somali, Harari and a small area at the source of the Wabi river in SNNP regional states. Out of these Oromiya and Somali regional states cover about 38% and 60% of the basin area respectively.

2.1.2 Topographic Features

The Wabi Shebele river emerges from the mountaineous areas of the North Western borders of the river basin near a place called Hebeno. About 11% of the total basin area is highland. The altitude varies from 200 m above mean sea level (msl) north of Mustahil in Somali region to about 4000 m msl on the highlands of Bale (Figure 2.2). From its source, the Wabi river flows eastward until it meets with another major component of the main river joining from Harari region where most of the left bank tributaries originates and then it changes its course to flow southwards. Downstream of Melka Wakena hydropower dam site, the river flows through a deep gorge up to north of Imi and emerges in its lower valley. The lower valley is a vast alluvial plain stretching up to Somalia border with a very gentle slope of 0.25 to 0.35 m/km. Most of the tributaries after this portion do not add a substantial flow to the main water course. The Fafen and Jerer watersheds are closed watersheds. Hypsometric curve shows the plot of the

proportion of the total basin elevation to the proportion of the total watershed area above. The area under the curve is the hypsometric integral. Approximately 75% of the area is below the mean elevation of the basin, which is approximately 2100 m above msl (Figure 2.3).

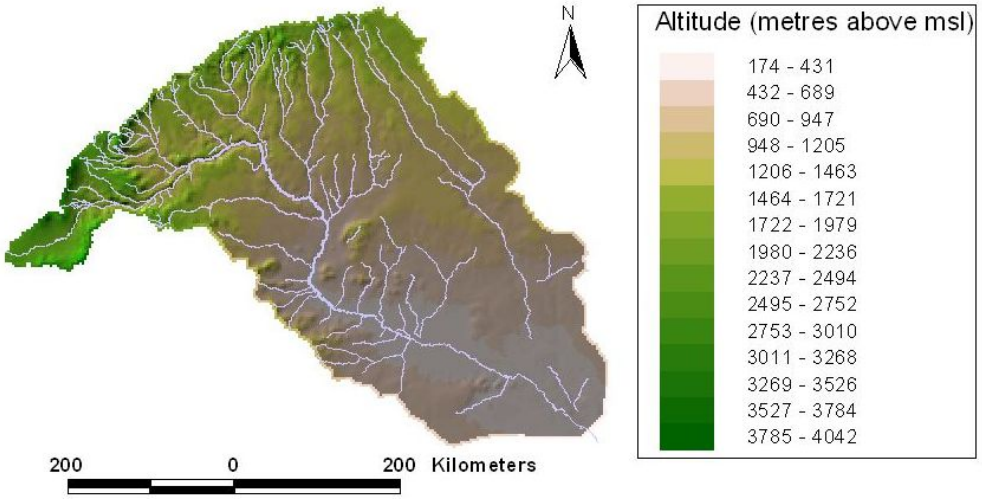


FIGURE 2.2: Digital elevation model of Wabi Shebele river basin.

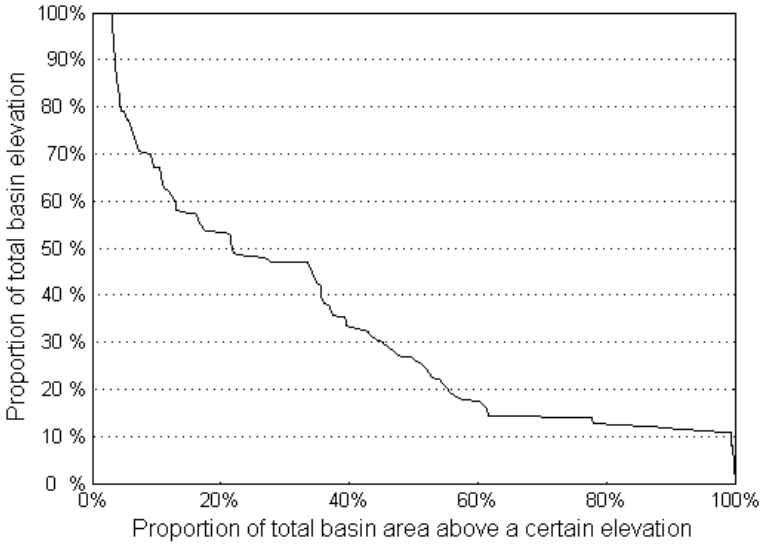


FIGURE 2.3: The hypsometric curve in Wabi Shebele river basin.

2.1.3 Geologic Features

The study area is dominated by Mesozoic sedimentary formations, to some extent there are also volcanic rocks at the North West of the basin and isolated ridges and hills within

the sedimentary basin. Metamorphic rocks outcrops in a small extent at the northern part of the study area. Alluvial deposits are also distributed linearly along the Wabi Shebele, Jerer, and Fafen rivers and fan deposits of seasonal floods and stream beds. The volcanic rocks of Arsi-Bale basalt bordering the rift valley are highly fractured. Numerous springs outcrops along faults and fractures in this area and form substantial part of the base flow of Wabi Shebele river. The Southern part of the basin is overlain by thick gypsum and limestone.

The water level monitoring for one hydrologic cycle on two wells at Gode showed that the water level is deep always lower than the river bed and the phreatic water level is practically the same during the hydrologic cycle and no interaction with Wabi Shebele river water indicating the permeability is very low. On the other hand at Kelafo there is interaction of the alluvial ground water with Wabi Shebele river. At Mustahil there is infiltration of the river floodwaters into the alluvial ground water (MoWR, 2003).

2.1.4 Land Use/Land Cover

A small dense forest is found at the North Western portion of the basin. Dense shrubland is the predominant land cover in the basin. The shrubland occurs mainly on the semi-arid parts and often consists of patches of shrubs interspersing grasslands with some scattered low trees. Patches of exposed rock or sand surface are found in parts of Bale and Hararghe lowlands in the southeast. Parts of central Arsi and northern Bale have afro-alpine and sub-afroalpine vegetation. These consist mostly of short shrub and heath vegetation used partly for sedentary grazing and browsing; and, where the terrain permits, for some cultivation of barley. Riparian woodland and bushland occur along the river banks and on floodplains and are important in the semi-arid and arid parts of the basin where they are used for grazing and browsing and scattered seasonal crop cultivation on some of the flood plains.

The landuse consists a large part of silvipastoral type. Areas of intensively cultivated land are found on the highlands of Arsi and parts of highland Harerge, and northern Bale. The major seasonal crops in the basin include maize, barley, wheat and sorghum while the perennial crops include coffee, chat and fruit trees.

Lakes/Reservoirs: Lake Alemaya and Adele, which are completely dried out recently, were the major natural lakes in the basin. Lake Alemaya was a principal source of the

water-supply system in Harar. Lake Adele is an adjoining lake to Alemaya. There had been a steady decline over the past decade in both the quality and quantity of water being delivered from this lake to the residents of Harar. Large deposits of silt have compounded the problem reducing the lake's water capacity greatly. The fresh-water lake that once provided recreation for swimmers, boaters and fishermen, slowly choked and diminished. Unregulated irrigation, with farmers pumping water to reach fields up to 12 Kms away, has depleted already strained resources even further. In 2004 it is completely exhausted, no longer able to supply the pipes with even ration levels. This is one of the recent painful lessons of inappropriate water management in the area.

At present there is one large man made reservoir by Melka Wakena dam that is commissioned in 1988 for hydropower generation. The river assumes a ravine character above the site and carries water through a vast trench suitable for creation of the reservoir. A rather large-capacity regulating reservoir with a volume of 763 million m^3 was created by the construction of a relatively low dam with a maximum height of 38 m. Below the cataracts, which have a drop of the order of 80 m, the river flows into a deep and narrow canyon having significant slopes. The magnitude of the head at the hydroelectric plant is 300 m. Its rated capacity is 153 Mw. The mean annual electric-power generation at the plant is 543 million Kwh (Cheryachukin and Sitnin, 2000).

The full reservoir level has never been reached since its commissioning except in the year 1998 when the spillway spilled for only two months in the rainy season. The plant is not generating up to its full capacity and requires immediate rehabilitations (MoWR, 2003). This may possibly be because the reservoir might be provided with huge capacity considering absence of binding downstream commitment of water as the river flows in deep gorges. A change in the rainfall pattern during the years might also have led the plant to the present critical level. The time series of flow used in the design of the reservoir may not have failure years due to exaggerated figures of inflows and the plant may have been designed with overcapacity which it is unable to generate in the present scenario.

2.1.5 Soils

About 50% of the soils in the basin are soils of calcarious or gypseous differentiation types. Vertisols comprise 12% of the basin covering significant areas at the middle

belt. The soils with calcarious differentiation are characterized by the redistribution of lime in the profile while the soils with gypseous differentiation are characterized by the redistribution of gypsum in the profile. The latter predominantly occurs in the southern arid part of the basin. In this study, the potential soil moisture storage capacity dataset is obtained from the Food and Agriculture Organization digital soil map of the world using same methodology as specified in Reynolds et al. (1999). The water-holding capacity of the soils is subdivided into nine classes, from wetlands (which are given a water-holding capacity of 1000 mm) to soils with a water-holding capacity of < 20 mm. The water-holding capacity in a gridbox is taken to be one of the most dominant soil types in that gridbox. The data is regridded from 5' resolution to 10' resolution by simple interpolation so that the grid of the water holding capacity of the soils coincides with the grids of the temperature and precipitation data. Figure 2.4 shows the spatial distribution of the average accumulated water holding capacity (in mm) in Wabi Shebele river basin.

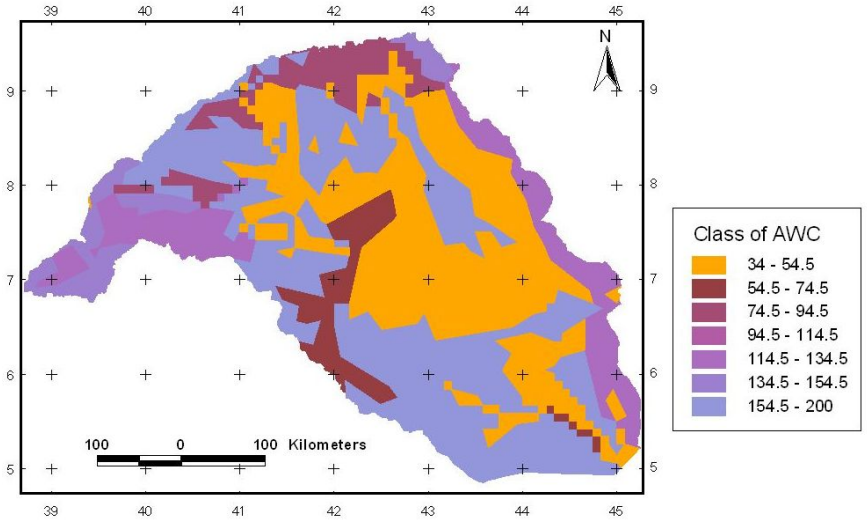


FIGURE 2.4: Spatial distribution of accumulated water holding capacity (in mm) in Wabi Shebele river basin.

2.2 Hydro Climatic Conditions

2.2.1 Climatic Situation

The climate is dependent on the altitude. The highlands are cool and densely populated while the lowlands are arid and sparsely populated. There are about 48 meteorological

stations in the basin. Most of them are clustered in or near urban centers of the upper portion of the basin. Analysis from long range rainfall records indicates that the highest mean annual rainfall recorded is 1487 mm in Seru woreda of Arsi zone, the lowest being 220 mm in Kelafo area in Somali region. The rainfall is bimodal taking place March -May and June - September. The maximum for each season is about 450 mm and 750 mm respectively. Nearly 60% of the basin receives less than 500 mm annually (Annex Figure A.2). During the main rainy season only 15% of the basin receives rainfall over 500 mm. The average annual potential evaporation in the basin is about 1500 mm. The low rainfall amount coupled with its erratic nature makes crop growth unreliable even during the main rains. Aridity index can serve to identify and delimit regions that suffer from a deficit of available water. A number of aridity indices have been proposed in literature. Here the one recently used by UNEP (1992) is employed, as the ratio between annual precipitation and annual potential evapotranspiration expressed in same units. In this classification a value below 0.2 is considered arid. Figure 2.5 shows the spatial distribution of aridity index in Wabi Shebele river basin.

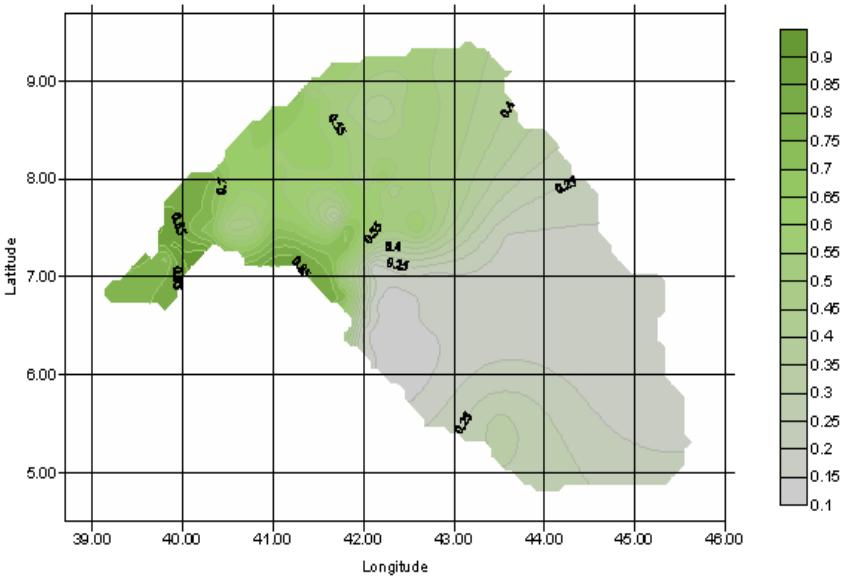


FIGURE 2.5: Spatial distribution of Aridity index in Wabi shebele river basin.

2.2.2 Hydrological Situation

The Wabi Shebele river basin is one of the water scarce basins in the country. While having the largest area coverage, its annual runoff and water availability are one of

TABLE 2.1: Estimated annual water yields of selected Ethiopian river basins (data drawn mainly from WAPCOS (1995))

Basin	Wabi Shebele	Genale Dawa	Awash	Abbay	Baro Akobo	Tekeze
Annual rainfall(mm)	425	528	557	1,420	1,419	1,300
Annual evaporation(mm)	1,500	1,450	1,800	1,300	1,800	1,400
Annual runoff($10^9 m^3$)	3.4	6.0	4.90	54.40	23.23	8.20
Basin area(Km^2)	202,220	172,259	110,000	199,912	75,912	82,350
Annual runoff coefficient(-)	0.04	0.07	0.08	0.19	0.22	0.08
Annual specific yield ($l/s/Km^2$)	0.53	1.10	1.41	8.63	9.70	3.16
Water availability(m^3 per capita)	578	1,176	445	3,823	10,507	1,737

the lowest among the major river basins (Table 2.1). This is one of the more reasons for the selection of the basin as the study region. The available hydrological data in the basin are very sparse. Some attributes of the hydrological gauging stations in the river basin are provided in Annex Table A.2. There are fewer than 50 stream gauging stations in the basin. Only less than 30 are operational at present, a significant portion of which have operated intermittently. Streamflow data at Dodolla (upper reach) and at Gode (lower reach) have relatively been continuously gauged since 1967/68. A good coverage of streamflow data over the basin is available only in the period 1967-1972. MoWR (2003) provides monthly flow data generated by rainfall-runoff relationships at 13 stations in the basin. The distribution of the meteorological and hydrological gauging stations is not well integrated reducing the plausibility of such data.

Inspection of most of the stations has shown that the stations are not reliable to capture the low flow situation in the basin. Some of them are installed on raised structures anchored to bridges, while others are offsetted in pockets of the main course of the stream. Low flow series often contain years with zero values. In some arid areas, zero flows are recorded more often than non-zero flows. Streamflows recorded as zero imply either that the stream was completely dry or that the actual streamflow was below a recording limit. Zero-values should not simply be ignored, nor do they necessarily reflect accurate measurements of minimum flow in a channel. Based on the hydraulic configuration of a gauge, and knowledge of the rating curve and recording policies, one

can generally determine the lowest discharge which can reliably be estimated and would not be recorded as zero.

Preliminary data checks

Before engaging in the development of a model on a given data set, it is always necessary to check the quality of the data. The goal of the preliminary checks is principally to look for possible inconsistency in the data which may be caused by changes in measurement device, changes in climate, natural disasters, hydraulic constructions, etc.

Measurement of low flows is quite sensitive to systematic errors. A modest systematic error of 10 mm in stage of Q_{95} at most of the stations in Wabi Shebele basin, changes the amount of discharge by more than 10% (Figure 2.6).

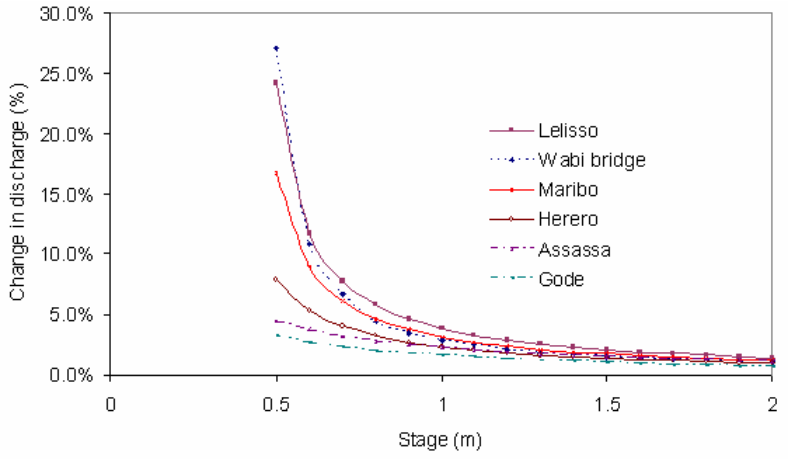


FIGURE 2.6: Percentage change in discharge for a systematic error of 10 mm at different stages in Wabi Shebele river basin.

The comparison of a precipitation series with the precipitation observed in the surrounding stations is useful for the analysis of the possible changes in stations location and equipment conditions. A candidate precipitation series can be considered homogeneous if the cumulative residuals are not biased. The bias hypothesis can be tested using an ellipse defining the confidence limits. For each candidate precipitation series, the axes of the ellipse are defined by (Allen et al., 1998):

$$\begin{cases} \alpha_i = N/2 \\ \beta_i = \frac{N}{\sqrt{N-1}} z_p \cdot S_{e,i} \end{cases}$$

where N-is the number of observations; z_p - stands for the standard normal variate for the desired probability p (confidence level); $S_{e,i}$ - is the standard deviation of the residuals of the precipitaion series of the suspected station i.

Thus, with uniformly varying angle θ[in radians] varying from 0 to 2π, the parametric equation of the ellipse for the suspected station i is:

$$\begin{cases} X_i = \alpha_i \cos(\theta) \\ Y_i = \beta_i \sin(\theta) \end{cases}$$

Plotting the cumulative residuals against time, using the time scale (interval) of the variable under analysis, the accumulated residual curve is obtained. If all the cumulative residuals lie inside the ellipse then the hypothesis of homogeneity is not rejected for the significance level considered. This test is then capable of locating the period (year) where a break is likely to occur. This type of inconsistencies can be detected using the accumulated residual method known as the ellipse test (Allen et al., 1998) and the double mass curve analysis. Heterogeneity detection using the accumulated residual method at Gode station for the years 1967-2000 is shown below (Figure 2.7). The zero mark indicates the year 1983/84. This plot shows that there is inconsistency in the years 1968-1977. The inconsistencies need to be removed first from the data by using

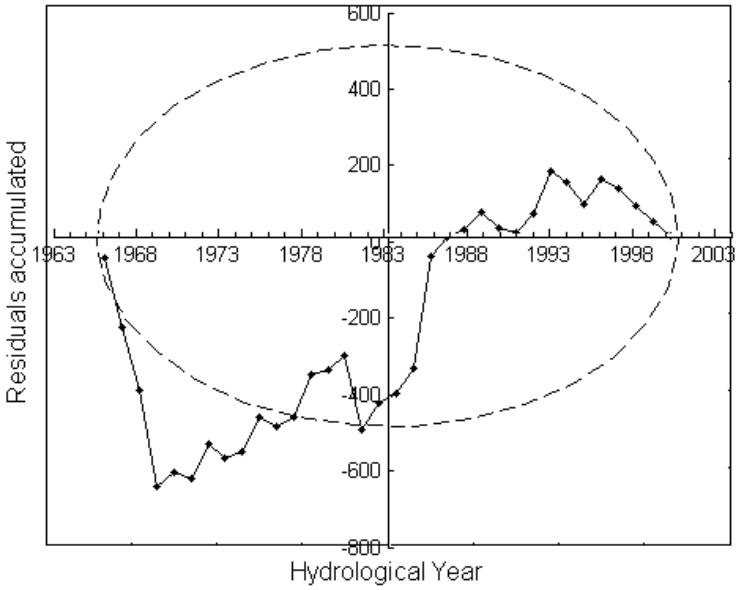


FIGURE 2.7: Heterogeneity detection using the accumulated residual method (ellipse test) for Gode station.

appropriate methods. The Wilcoxon rank-sum test is applied in this study (Helsel and Hirsch, 1992) along with the average of similar nearby gauging stations . Comparing with the t-test, the advantage of this test is that it does not require normal distribution of the data.

2.3 Socio-economics

There are about 76 administrative woredas that lie within the basin. About 7% of the population is urban. The major accesses to the basin are a gravel road along the highlands of the Bale mountains and the recently completed asphalt road to Harar. The density of feeder roads is quite low in the lowlands (See Annex Figure A.6).

2.3.1 Population and Settlement

As per the census of 1997, the total population of the basin is about 5.8 million. Out of this, 70.1% of the population belongs to parts of four zones of Oromiya region while about 27.4% are that of Somali region. The rest is that of Harari and SNNP regions. Population density is the highest in Arsi (78.5 people/Km^2) whereas the lowest is in Warder zone (4.2 people/Km^2). Large percentage of the population in the highlands depend on agriculture while the lowlanders in general are pastoralists. About two-third of the area is populated with less than 20 person/Km^2. Most of the less populated area lies in the arid to semi-arid lowland areas.

2.3.2 Water Consumption

Currently there are about 10 state farms in the basin involved in mechanized large scale rainfed agriculture. There are over 334 traditional irrigation schemes covering an estimated area of 12,000 ha. A total of 72 small and medium modern irrigation schemes with an estimated irrigable area of 7045 ha exist in the basin (MoWR, 2003). The only existing dam under operation on the Wabi Shebele river is the Melka Wakena hydropower dam. There is a large scale irrigation project at Gode but it is not functional. About 140 medium and small scale potential irrigation sites have been identified in MoWR (2003). Higher population density is observed in the upstream portion of the basin and large irrigation schemes are clustered in the downstream portion (Figure 2.8). The deficiency of

sufficient water for hydropower generation has inclined some consultants to contemplate interbasin water transfer for regulation of a proposed hydropower production scheme at Kuldash from Weyib river (WAPCOS, 1995). Sometimes at places where the num-

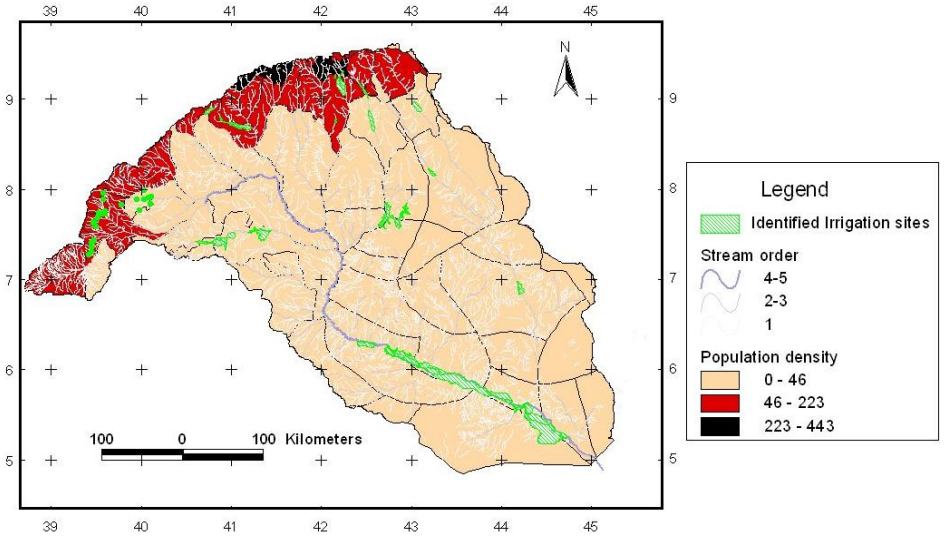


FIGURE 2.8: Population density and potential irrigation sites in Wabi shebele river basin.

ber of good quality flow records is insufficient for an analysis to proceed, stream flow naturalization procedures may be important to convert gauged flows to natural flows that would have occurred in the absence of water users and water management facilities. The procedure requires a systematic record of the influences, including times, rates and durations of the abstractions, discharges and compensation flows and adjusting the observed flow accordingly. The adjustments are governed largely by data availability. However, these data are not available for numerous smaller reservoirs.

2.3.3 Major Drought Episodes

Recurrent drought has ravaged economic activity in Ethiopia. Analysis of the chronological events of Ethiopias drought/famines have been divided into four parts and the analysis suggested some interesting features (NMSA, 1996). During the period from 253 B.C. to A.D., one drought /famine was reported in a seven years period. From the beginnings of A.D. to 1500 AD, there were some cruel famines which killed millions. In this period there were 177 droughts/famines, about one in nine years, reported in the country. From 1500 to 1950, the information is relatively based on recorded data and is more reliable. From 16^{th} to the first half of the 20^{th} century, 10,14,21,16 and

8 drought/famines were reported, respectively suggesting 69 events in a period of 450 years. This signifies that on average the occurrence of one drought in 7 years. The two notorious famines known as QUACHINE and KIFUKEN which devastated major areas of the country were reported in this period. The reports from 1950 onwards are well documented with scientific data. The analysis of the rainfall data in this period indicated 18 droughts/famines in 38 years, suggesting the occurrence of drought in every two years. The decadal analysis shows that the decade 1970-1979 was the worst having 7 disaster years in terms of frequency. The highest frequency of droughts/famines occurred in the 2^{nd} century AD followed by the first part of the 20^{th} century and an increasing trend in the 16^{th} century onwards. 20,000 people were thought to have lost their lives by the 1973/74 famine and 18% of the total population was affected by the 1984/85 famine in which over a million people are said to have lost their lives. The northern parts of the country appeared to be highly vulnerable to recurring droughts and famine (NMSA, 1996). Among the major drought episodes that of the 1984-85, 1991-92 and 1998 are

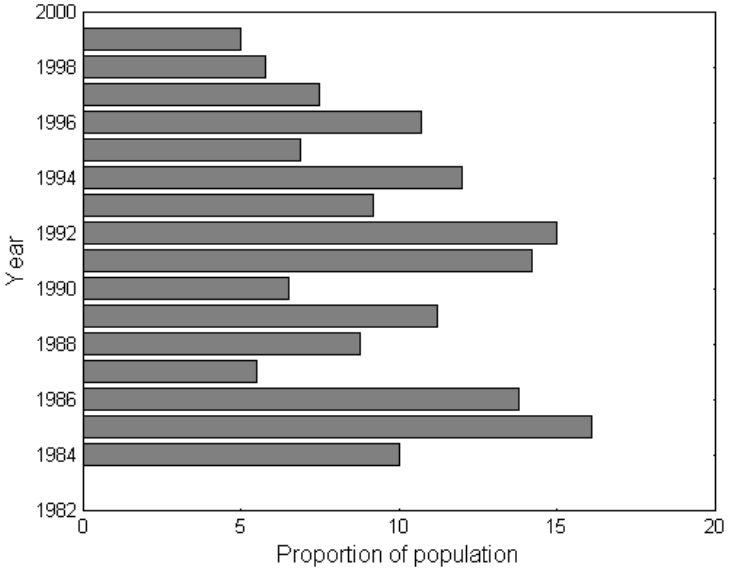


FIGURE 2.9: Proportion of population affected by drought in different years in Ethiopia.[1]

the very recent ones in which lots of impacts had been observed. The highest proportion of population affected by drought occurred in 1985 (more than 16%) and 1992 (15%) (Figure 2.9).

[1] As per DPPC, Report on assistance requirements for disaster prevention and preparedness activities, Addis Ababa, Ethiopia from various years

Chapter 3

Characterising Hydrological Drought

Chapter 3 deals with characterising hydrological drought through indices. Event definition of duration, severity, intensity and frequency of drought will be incorporated. Genesis of the hydrological drought process shall be briefed at this point.

3.1 Background

The interdependence between climatic, hydrologic, geologic, geomorphic, ecological and societal variables makes it very difficult to adopt a definition that fully describes the drought phenomena and the respective impacts. A single definition of drought applicable to all spheres is difficult to formulate since concept, observational parameters and measurement procedures are different for experts of different fields. Beside, the concept of drought varies among regions of differing climates (Dracup et al., 1980). Consequently, a method to derive drought characteristics developed in one region is not necessarily appropriate or even applicable in another region. For example, in Egypt before construction of the Aswan Dam, failure of the Nile River to flood constituted a drought regardless of rainfall (Hudson and Hazen, 1964). Conceptual definitions of drought lack to provide the specifics about the severity, the duration and extent of drought. Operational definitions of drought typically require quantification of "normal" or "expected" conditions within specified regions and variations in societal conditions. They are often formulated in terms of drought indices. A drought index value is typically a single number, far more

useful than raw data for decision making. There are various methods and indices for drought analysis and they measure different drought-causative and drought-responsive parameters, and identify and classify drought accordingly. Understanding what causes drought helps to predict it. However, since these parameters are not linearly correlated with each other, correlation among various kinds of drought is also difficult. It is important to investigate the consistency of results obtained by different drought indices.

3.2 Genesis of the Process

The major causes of drought are anomalies in the weather or climate that lead to less precipitation than normal for meeting water demands in agriculture, industry, households, hydropower generation, etc. Meteorological drought is the result of many causes, often synergistic in nature. Climate fluctuations is the underlying natural physical factor in determining the severity of a drought but human activity can also contribute to it. The location and persistence of high-pressure centres, or anticyclones, has a major influence on rainfall and temperature across a region. As the air subsides it is adiabatically warmed and dried. Because of this adiabatic warming, areas of high pressure are often free from cloud, any cloud that develops has little vertical extent and thus precipitation hardly ever forms.

A great deal of research has been conducted in recent years on the role of interacting systems, or teleconnections, in explaining regional and even global patterns of climatic variability. One such teleconnection is the El Niño Southern Oscillation (ENSO). El Niño Southern Oscillation (ENSO) is a coupled air and ocean phenomenon with global weather implications. ENSO stands for the coupling of El Niño (oceanic component) and the Southern Oscillation (atmospheric component). ENSO occurrences are global climate events that are linked to various climatic anomalies. Not all anomalies, even in ENSO years, are due to ENSO. In fact, statistical evidence shows that ENSO can account at most for about 50% of the interannual rainfall variance in eastern and southern Africa (Ogallo, 1994), but many of the more extreme anomalies, such as severe droughts, flooding, and hurricanes, have strong teleconnections to ENSO events. The immediate cause of drought is the predominant sinking motion of air (subsidence) that results in compressional warming or high pressure, which inhibits cloud formation and results in lower relative humidity and less precipitation. There is a remarkable correspondence

between annual rainfall in Ethiopia and ENSO events (Haile, 1988). A statistical analysis shows that ENSO episodes are negatively teleconnected with the flows of the Blue Nile and Atbara rivers that originate in Ethiopia and that knowledge of these events could be used to improve the predictability of its annual flow (Attia and Abulhoda, 1992; Eltahir, 1996). The lower Wabi Shebele received high rainfall during the El Niño year of 1997/98. Residents of the area, the Somali communities, refer to that year as the 'pouring Wednesday' [1](SC-UK et al., 2001).

Empirical data also indicate an association between ENSO events and droughts in Ethiopia. The principal cause of drought is asserted to be the fluctuation of the global atmospheric circulation, which is triggered by the Sea surface Temperature (SST) anomalies occurring during ENSO events coupled with Southern Oscillation Index (SOI). SOI is the difference in sea level pressure between Darwin and Tahiti. During an ENSO event, the SOI reaches a minimum. These phenomena have significant impact on the displacement and weakening of the rain-producing mechanisms in Ethiopia (Haile, 1988). Thus, an ENSO-based early warning system, used effectively by policymakers, could help to reduce the societal impacts of drought in Ethiopia. The time period from the arrival of precipitation until usable water is available differs considerably. Thus, the time scale over which precipitation deficits accumulate becomes extremely important and functionally separates different types of drought. It takes longer for precipitation deficiencies to show up in components of the hydrological system such as soil moisture, streamflow, and ground water and reservoir levels.

Low Flows

Low flow is a seasonal phenomenon, and an integral component of a flow regime of any river. Low flows are normally derived from groundwater discharge or surface discharge from lakes, marshes, or melting glaciers. Low flow studies focus among other things on the magnitudes and variability of annual low flows, duration of continuous low-flow events and relative contribution of low flows to the total streamflow hydrograph. The arbitrary upper bound to low-flow may be given by the Mean Annual Runoff (MAR), which is a mean value of the available flow time series of annual flow totals and one

[1]Somali communities, mark their traditional years by giving them names that correspond to the days of the week; years are known as Monday year, followed by Tuesday year, etc, and after the seventh year (i.e. Sunday), the cycle begins again with Monday. Years with the same name would be differentiated by a nickname related to a major event (droughts, floods, war, regime change, epidemics, etc), that took place during particular year; for example Arbaca Shuba (meaning the 'Pouring Wednesday') referred to the El Niño year of 1997/98, which was a Wednesday year.

TABLE 3.1: Mean annual minimum n-day flows at selected stations in the Wabi Shebele basin (m^3/s) [For location of the stations please refer to Annex Table A.2]

Station	MAM(1)	MAM(7)	MAM(30)
Assassa	0.4623	0.6926	0.7916
Erer	0.5143	0.5995	0.7132
Fruna	0.0746	0.0788	0.1086
Tebel	0.0247	0.0257	0.0284
Wabi at the bridge	1.3638	1.4038	1.5042

of the most fundamental hydrological characteristics. A middle value in a ranked flow time series, Median Flow (MF)- may represent a more conservative upper bound for low flows, because streamflow time-series are often positively skewed and therefore MF is frequently smaller than MAR. The MF corresponds to the discharge equaled or exceeded 50% of the time, Q_{50}. For the same record length, the positive skewness of the data normally increases as the time resolution of the streamflow data decreases from annual to daily and therefore the gap between higher mean flow value and lower median flow value increases. An extensive review on low flow studies is provided in Smakthin (2001). The ratio (Q_{50}/Q_{90}) may represent the variability of low flow discharges. Low flows can be described as annual lowflows, Q_{95}, annual minimum 7 day flow etc. The average of the annual series of minimum 7-day average flows is known as Mean Annual 7-day Minimum flow (MAM7) and is used in some countries e.g. the UK for abstraction licensing. The 7-day period covered by MAM7 eliminates the day-to-day variations in the artificial component of the river flow. Also, an analysis based on a time series of 7-day average flows is less sensitive to measurement errors. Mean annual minimum n-day flows at selected stations in the Wabi Shebele basin(m^3/s) is provided in Table 3.1.

The relative contribution of low flows to the total streamflow hydrograph is often specified using Base Flow Index. The Base Flow Index (BFI) is used as a measure of the base flow characteristics of catchments. It provides a systematic way of assessing the proportion of base flow in the total runoff of a catchment. It indicates the influence of soil and geology on river flows, and is important for low flow studies.

3.3 Drought Indices

Drought indices may be used for drought warning and lead time assessment (McKee et al., 1995; Lohani and Loganathan, 1997). Drought indices summarize different data on rainfall, snowpack, streamflow, and other water supply indicators into a comprehensible

TABLE 3.2: Different drought indices and their pros and cons

S.No	Index	Pros	Cons	Citation
1	PDSI/ PHDI	Non-dimensional, Widely accepted especially in USA	Arbitrary threshold, may lag emerging droughts by several months less well suited for mountainous or of frequent climatic extremes	Palmer 1965
2	SPI	Identifies emerging droughts months sooner than the PDSI, Limited data input	Arbitrary threshold,	McKee et al. 1995
3	CMI	Identifies potential agricultural droughts.,	not a good long-term drought monitoring tool	Palmer 1965
4	SWSI	Representative measure of water availability across a basin, region,	it is difficult to compare SWSI values between basins or regions -it is difficult to maintain a homogeneous time series of the index	Shafer and Dezman 1982 Doesken et al. 1991; Heddinghaus and Sabol 1991
5	ADI	Multivariate index that considers the bulk quantity of water across the meteorological, hydrological and agricultural regimes	More general	Keyantash and Dracup 2004

big picture. Drought indices are quantitative indicators, used to identify, characterize and analyze drought events with relative accuracy and objectivity. There are several indices that measure how much precipitation for a given period of time has deviated from historically established norms like PDSI/PHDI, SPI, CMI, SWSI, ADI. The various drought indices are usually discipline specific and have their own pros and cons, some indices are better suited than others for certain uses (Table 3.2). Most water supply planners find it useful to consult one or more indices before making a decision.

3.3.1 Standardized Precipitation Index

Standardized precipitation index (SPI) is based on an equi-probability transformation of aggregated monthly precipitation into a standard normal variable (Figure 3.1). In practice, computation of the index requires fitting a probability distribution to aggregated monthly precipitation series (e.g. k= 3, 6, 12, 24 months, etc), computing the non-exceedance probability related to such aggregated values and defining the corresponding standard normal quantile as the SPI. McKee et al. (1993) assumed an aggregated precipitation gamma distributed and used a maximum likelihood method to estimate the parameters of the distribution. Shorter or longer time scales may reflect different lags in the response to precipitation anomalies. SPI has advantages of statistical consistency, and the ability to describe both short-term and long-term drought impacts through the different time scales of precipitation anomalies. Its limitation is that it relies on one input. In general, different studies have indicated the usefulness of the SPI to quantify different drought types (Edwards and McKee, 1997; Hayes et al., 1999). The long time scales (over 6 months) are considered as hydrological drought indicators (river discharges or reservoir storages)(Hayes et al., 1999).

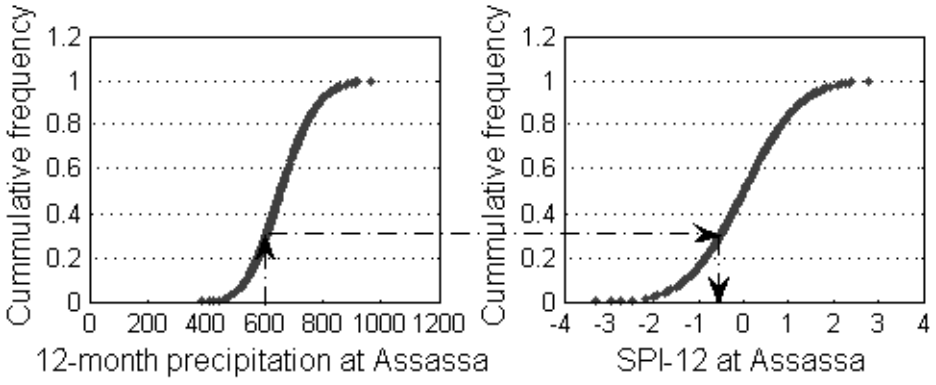


FIGURE 3.1: Equiprobability transformation from fitted gamma distribution of monthly precipitation aggregated at 12 month to the standard normal distribution at Assassa Station.

Using the threshold curve for each aggregation time scale, time series of SPI can be established for the different stations. In 1973/1974 moderate to extremely severe drought that lasted longer duration hit the area around Gedeb-Assassa. Similarly the drought that occurred in 1984/1985 is moderate to extreme nature in its severity and lasted longer duration (Figure 3.2).

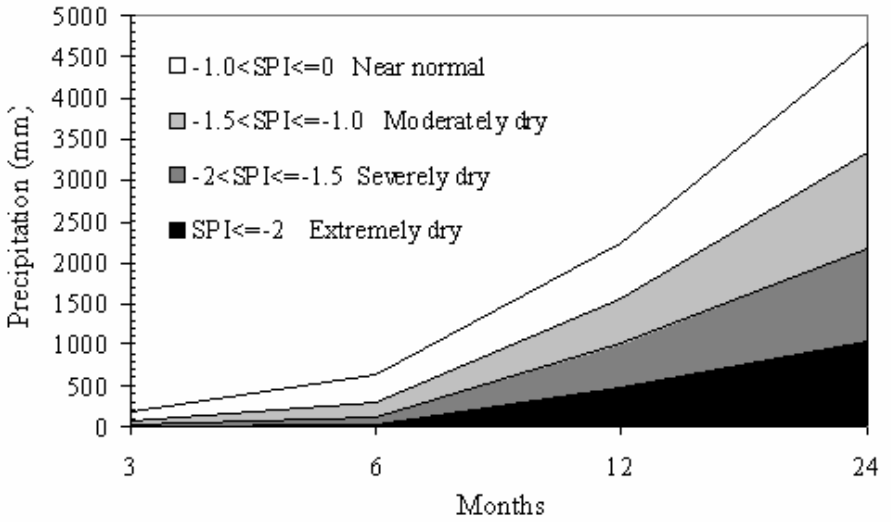


FIGURE 3.2: Threshold curves of SPI for different aggregation time scales at Assassa meteorological station.

Temporal variability of SPI at different time scales

Figure 3.3 shows the continuous evolution of SPI at different time scales. At shorter time scales (e.g., 3 months) the dry ($SPI < 0$) and moist ($SPI > 0$) periods show a high temporal frequency, whereas when the time scale increases the frequency of dry periods decreases. At the time scale of 24 months only two important dry periods are recognised: in the middle of 1970's and in the latter half of 1980's. The most recent dry period is building up from around 1997 onward. The average duration of the dry periods ($SPI < 0$) change noticeably as a function of the time scales. At the time scale of 12 months the average duration is 3.5 months and the longest mean duration is recorded at the time scale of 24 months with an average duration of 9.5 months.

To identify the main dry periods it is necessary to analyse the time scales larger than 6 months because the high frequency of SPI values at the shorter time scales hide the most important dry periods. Time scales shorter than 6 months show non-significant autocorrelations considering lags shorter than 4 months, whereas considering the SPI at time scales larger than 6 months the autocorrelations with lags of 4 of more months increase noticeably. Therefore, with time scales shorter than 6 months, it is difficult to identify periods of consecutive 4 months with dry conditions (Figure 3.4). Although the SPI is widely used, there are not many empirical studies that provide evidence about the usefulness of the different time scales for drought monitoring in surface water resources.

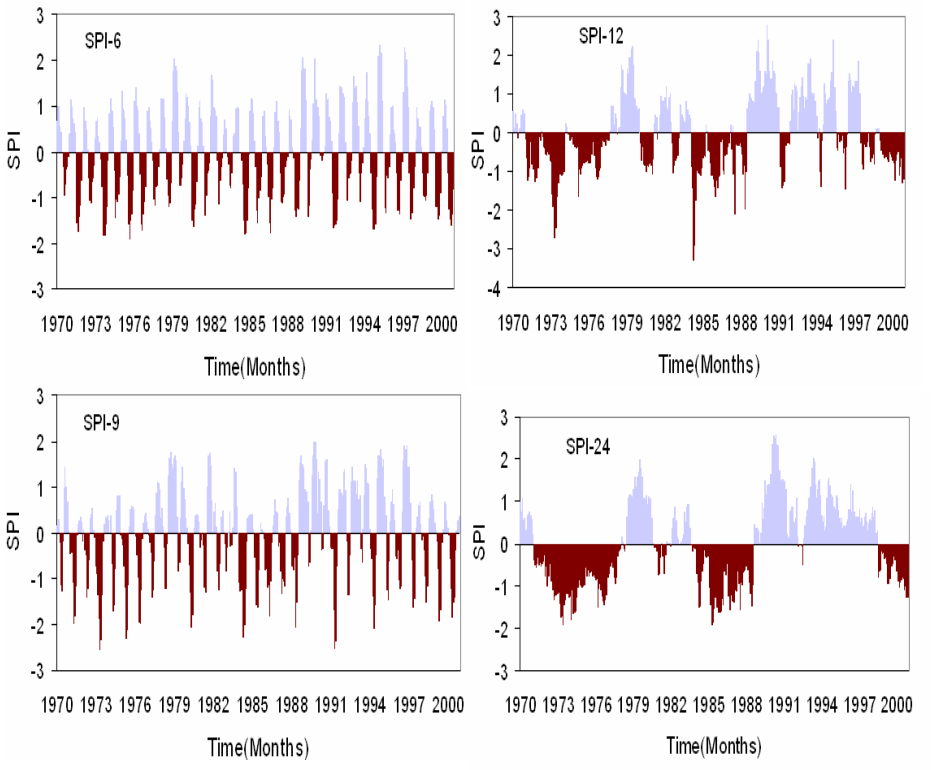


FIGURE 3.3: Evolution of the SPI for different time scales of 6 month to 24 month at Assassa meteorological station for the period between January 1970 to January 2000.

In Hungary Szalai et al. (2000) analysed the relationships between time scales of SPI, river discharges and reservoir storages, showing important spatial differences.

3.3.2 Normalized Difference Vegetation Index

Spatial and temporal variability of vegetation has been successfully observed at scales ranging from local to global using Normalized Difference Vegetation Index (NDVI), derived from satellite data (e.g., Holben and Justice 1981; Malo and Nicholson 1990). This index utilizes the reflectance spectra of healthy green vegetation, which is characteristically high in the near infrared, for example 0.73-1.10μm in the case of the NOAA Advanced Very High Resolution Radiometer (AVHRR), and low in red wavelengths (0.58-0.68μm for the AVHRR).

Computation of NDVI

Initially there is a need to geolocate the pixels from satellite data and remap to the chosen projection. Sensor calibration of individual channels must be implemented and digital numbers converted to spectral reflectances to enable NDVI calculation. Atmospheric

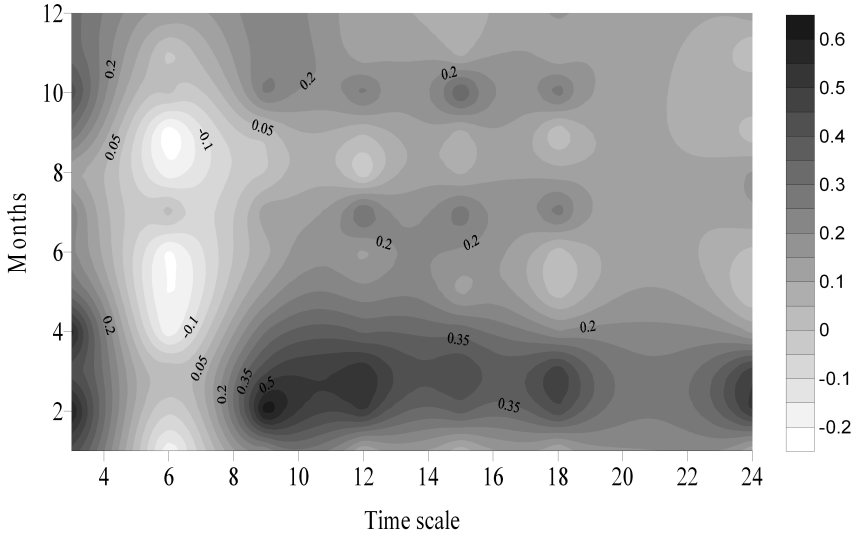


FIGURE 3.4: Correlation coefficient between time series of monthly streamflow at Adaba and times series of SPI for different time scales.

corrections must be applied and a method employed to screen for clouds, after which NDVI can be calculated (Rouse et al., 1974) as:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{3.1}$$

Where NIR is reflectance in the near infrared and RED is the red waveband reflectance. The differential reflectance in these bands provide a means of monitoring density and vigour of green vegetation growth using the spectral reflectivity of solar radiation. Green leaves commonly have larger reflectances in the near infrared than in the visible range. As the leaves come under water stress, become diseased or die back, they become more yellow and reflect significantly less in the near infrared range. Clouds, water, and snow have larger reflectances in the visible than in the near infrared while the difference is almost zero for rock and bare soil. Vegetation NDVI typically ranges from 0.1 up to 0.6, with higher values associated with greater density and greenness of the plant canopy. Since each vegetation type has a different seasonal NDVI signal, especially in cases where the cover classes are fairly broad, it should be possible to ascertain when the vegetation type over a particular geographical area has changed, given the seasonal NDVI signals of all the possible cover classes. Africa-wide NDVI images with 7.6 km resolution are available from NASA on a dekadal (10-day) basis throughout the year. Regions of high variability in NDVI depict regions which are either highly variable in precipitation regime, or extremely sensitive to slight changes in other climate variables,

soil properties, and/or human activity. These areas are typically found in semi-arid climatic regions that are characteristically variable in rainfall.

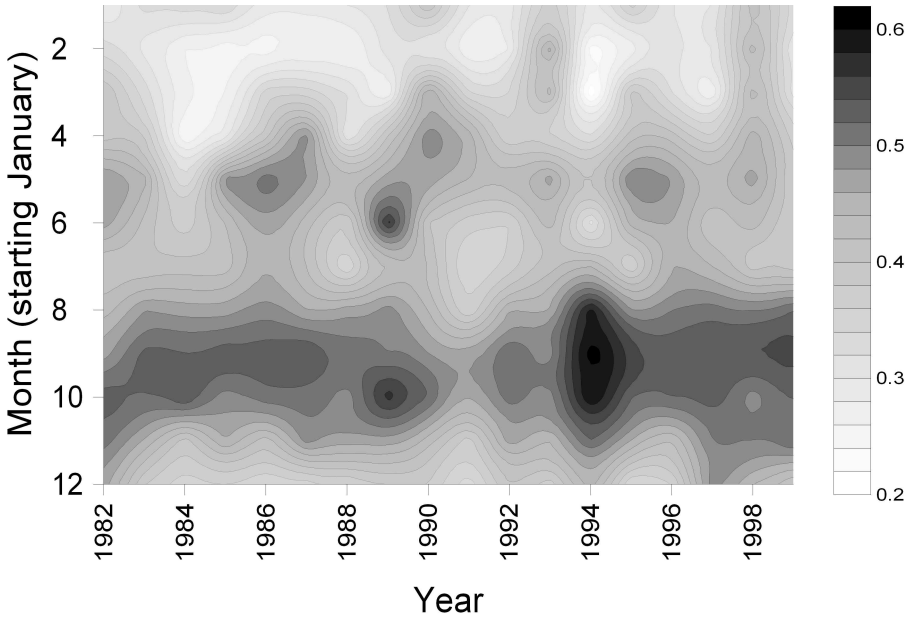


FIGURE 3.5: Temporal NDVI values in the North Western portion of Wabi Shebele at monthly scale.

Figure 3.5 shows the temporal variation of NDVI in north western portion of Wabi Shebele river basin. Although the main rainfall season (locally called Kiremt) starts in June the NDVI value starts to show this effect lagged because of storage of moisture. The relationship between NDVI and baseflow index is shown in Section 4.1. The notorious drought of the 1984/85 year has significantly affected the vegetation in the belg season (March, April, May) more than the main rainy season (June, July, August). It has intensified and elongated the dry season. But the 1991 drought has targeted predominantly the main rainfall season. In the year 1994 a significant fluctuation between dry and wet extremes is noticeable.

3.3.3 Palmer Drought Severity Index

The computation of the PDSI incorporates a water balance model using historic records of monthly precipitation , potential evapotranspiration and a simple 2-layer soil moisture reservoir. The upper layer is assumed to contain 1 inch (25.4 mm) of available moisture at field capacity. The underlying layer has an available capacity that depends on the soil characteristics of the site. Moisture cannot be removed from the lower layer until the

top layer is dry. Runoff (RO) is assumed to occur when both layers reach their combined moisture capacity (AWC). The AWC value should be representative of the soil of the area in general.

Four potential values are computed including: Potential evapotranspiration (PE, e.g. by Hargreaves equation or other), Potential recharge (PR) - the amount of moisture required to bring the soil to field capacity; Potential loss (PL) - the amount of moisture that could be lost from the soil to evapotranspiration provided precipitation during the period was zero; and Potential runoff (PRO) - the difference between the potential precipitation and the PR.

The climate coefficients are computed as a proportion between averages of actual versus potential values for each of 12 months. These climate coefficients are used to compute the amount of precipitation required for the Climatically Appropriate for Existing Conditions (CAFEC). The difference, d, between the actual (P) and CAFEC precipitation ($\hat{P}$) is an indicator of water deficiency for each month.

$$d = P - \hat{P} = P - (\alpha \cdot PE + \beta \cdot PR + \gamma \cdot PRO + \sigma \cdot PL) \tag{3.2}$$

where $\alpha = \bar{ET}/\bar{PE}$,$\beta = \bar{R}/\bar{PR}$, $\gamma = \bar{RO}/\bar{PRO}$ and $\sigma = \bar{L}/\bar{PL}$ for 12 months. E, R, RO and L are actual evapotranspiration, recharge, runoff and loss respectively. The value of d is regarded as a moisture departure from normal because the CAFEC precipitation is an adjusted normal precipitation. The difference between this precipitation and the actual precipitation is multiplied by a weighting factor, termed the "climatic characteristic," to produce a "moisture anomaly index." The purpose of the weighting is to adjust departures from normal precipitation such that they are comparable among different areas and different months (Alley, 1984). A Palmer Moisture Anomaly Index, Z, is then defined as

$$Z = K \cdot d$$

where K is a weighting factor. The value of K is determined from the climate record before the actual model calculation. Palmer suggested empirical relationships for K. By plotting Z versus duration for the worst drought episodes, generally a linear relationship of the following type can be obtained for drought severity:

$$PDSI_t = \varphi \cdot PDSI_{t-1} + \epsilon \cdot Z_t \tag{3.3}$$

where φ and ϵ are coefficients. The PDSI of the initial month in a dry or wet spell is equal to $\epsilon \cdot Z_t$. The Z index indicates how wet or dry it was during a single month without regard to past precipitation anomalies. Accordingly, Palmer(1965) obtained the following coefficients for the American mid west.

$$PDSI_t = 0.897 \cdot PDSI_{t-1} + 0.333 \cdot Z_t$$

These averaged coefficients have since become a fixed part of the PDSI computation (Wells et al., 2004). In Equation 3.3 the PDSI has built into its formulation a Markovian persistence term of φ from one time step to the next. Its combination with past PDSI means that the PDSI for a given month is a weighted function of current moisture conditions and an exponentially damped contribution of PDSI over previous months. The extreme drought lines at different stations in Wabi Shebele show variation with that of Palmer (1965) (Figure 3.6). Ntale and Gan (2003) reported, for stations in Kenya and Tanzania, a value of φ less than 0.897, which Palmer (1965) originally used. ϵ differs by four or five times from Palmer's original ϵ of 0.333. In Wabi Shebele rather higher values of φ are consistently noted (Table 3.3). Thus it is not appropriate to use the same φ and ϵ to compute PDSI at all the stations as suggested by Palmer, unless they all exhibit the same drought characteristics. The PDSI rapidly approaches a value of zero during the first month that weather begins to change from dry spell to a moist spell (or vice versa), regardless of whether soil moisture, ground water, or reservoir levels are still below normal. For operational purposes the Palmer hydrological drought index (PHDI) is used. The PHDI emulates conditions in hydrologic systems by lagging behind both the Z index and PDSI during the initial months of a drought and is considered a hydrologic drought index. In practice the PDSI is not a pure AR (1) process as it appears in equation 3.3, because the final value PDSI_t at times shifts abruptly between three pseudo-indices according to some predefined rules. Numerical differences in the values of the PDSI and the PHDI are small during the middle of a drought or a moist spell, but they are sizeable at the onset or endings of dry or wet periods, with the severity index moving toward normal conditions earlier than the PHDI.

The PDSI have been criticized for its complexity and empiricism (Karl, 1986; Alley, 1984; Heddinghaus and Sabol, 1991; Keyantash and Dracup, 2002), but it remains one of the most widely used drought indices in the world and is a fundamental part of the North American drought monitors (Lawrimore et al., 2002).

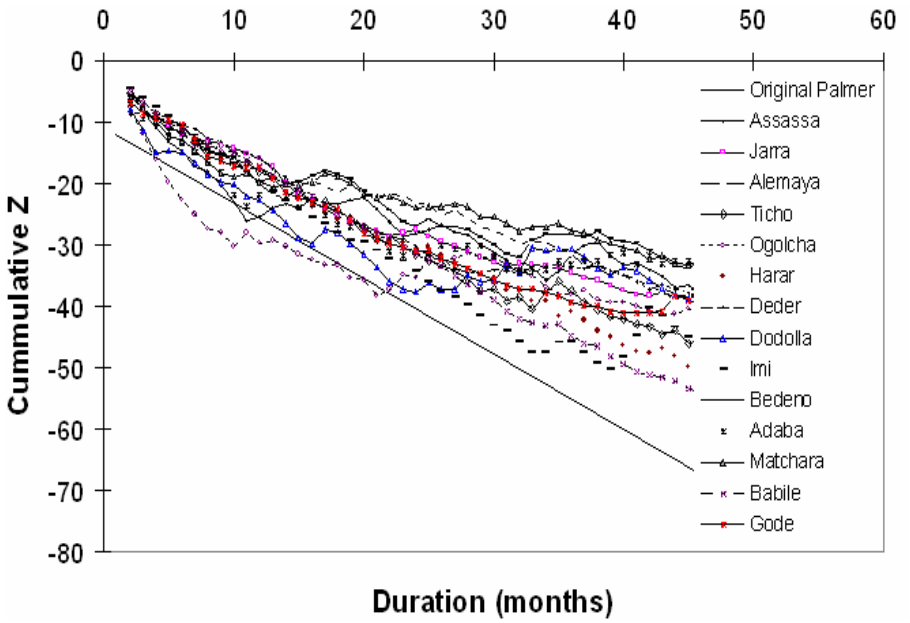


FIGURE 3.6: Plots of the most negative cumulative moisture index Z versus the duration (ranging from 2 to 48 months) and corresponding original PDSI estimate for selected stations in Wabi Shebele

TABLE 3.3: Estimated parameters of the original PDSI at different stations in Wabi Shebele and the goodness of fit. [For location of the stations please refer to Annex Table A.1]

Station	Intercept	Slope	φ	ϵ	R^2
Assassa	-13.901	-0.523	0.964	0.277	0.869
Jara	-9.086	-0.717	0.927	0.408	0.949
Alemaya	-8.488	-0.650	0.929	0.438	0.960
Ticho	-7.989	-0.877	0.901	0.451	0.968
Ogolcha	-21.233	-0.466	0.979	0.184	0.729
Harar	-6.717	-0.979	0.873	0.520	0.993
Deder	-9.399	-0.705	0.930	0.396	0.908
Dodolla	-16.946	-0.513	0.971	0.229	0.719
Imi	-8.822	-0.949	0.903	0.409	0.886
Bedeno	-10.702	-0.518	0.954	0.356	0.887
Adaba	-14.978	-0.475	0.969	0.259	0.742
Matchara	-11.296	-0.476	0.960	0.340	0.927
Babile	-4.755	-1.102	0.812	0.683	0.996
Gode	-9.790	-0.769	0.927	0.379	0.936
Original PDSI	-10.764	-1.236	0.897	0.333	

The spatial comparability of the PDSI across diverse climate regions has been questioned (e.g., Karl 1986) because Palmer (1965) derived coefficients and empirical weighing factors used in estimating PDSI from a very geographically limited region of the American Midwest. Another deficiency of the PDSI is that it produces values which are 4 or 4 up to 15% or more of the time (Wells et al., 2004): hardly corresponding to the classification "extreme." Wells et al. (2004) addressed this issue through the development of a

TABLE 3.4: Classification of drought and wet conditions as defined by Palmer (1965) for PDSI

PDSI/PHDI Value	Scale
Above +4.00	Extreme wet spell
3.00 to 3.99	Severe wet spell
2.00 to 2.99	Moderate wet spell
1.00 to 1.99	Mild wet spell
0.50 to 0.99	Incipient wet spell
-0.49 to +0.49	Near normal
-0.99 to -0.50	Incipient drought
-1.99 to -1.00	Mild drought
-2.99 to -2.00	Moderate drought
-3.99 to -3.00	Severe drought
Below -4.00	Extreme drought

self-calibrating PDSI (scPDSI) that locally adapts to the characteristics of the climate data being analyzed. This is achieved by weighting, for every location, the climatic characteristic and the duration factors with data from only that location. This scales the departure from normal precipitation and the sensitivity of the index with factors uniquely appropriate. The self-calibrating PHDI is scaled to have an extreme wet(dry) spell near two percent of the time. The scPHDI actually adjusts the value of K necessary to obtain the correct range of PDSI values (-4.0 to +4.0). Table 3.4 offers classification of drought and wet conditions as defined by Palmer (1965) for PDSI. A detailed description of how the PDSI is computed is given by Alley (1984); Karl (1986). A detailed description of the modifications to this algorithm to obtain the self calibrating PDSI is given by Wells et al. (2004) and by van der Schrier et al. (2006).

3.3.4 Surface Water Supply Index

Surface water supply index (SWSI) compares historic data of precipitation, reservoir storages and streamflow with current values and integrates them into a single index (Shafer and Dezman, 1982). It is designed to be an indicator of surface water conditions. Rouhani and Cargile (1989), using a kriging technique, have developed a drought lead time indicator based on reservoir storages. Obviously surface water drought indices are regionally unique and require long term records for their calibration.

It is difficult to compare SWSI values between basins or regions because the SWSI calculation is unique to each basin or region (Doesken et al., 1991). Several characteristics of the SWSI limit its application. Additional changes in the water management within a basin, such as flow diversions or new reservoirs, mean that the entire SWSI algorithm

for that basin needs to be redeveloped to account for changes in the weight of each component. Thus, it is difficult to maintain a homogeneous time series of the index (Heddinghaus and Sabol, 1991).

3.3.5 Multivariate Aggregate Drought Index[2]

3.3.5.1 Background

A clear understanding of an evolving drought in time and space is essential for a proper management of water resources. Drought Indices can quantitatively describe the characteristics of drought which otherwise is conceptually put. These indices may be used for drought warning and lead time assessment (McKee et al., 1995). Over the years, multitudes of drought indices are developed and are in use these days. However, most of the indices are well suited to describe a specific form of drought (meteorological, agricultural, hydrological drought). The cumulative effect of the phenomenon may better be illustrated using aggregate index. The aggregate drought index (ADI) employed in this study considers time series of monthly precipitation, temperature, normalized difference vegetation index (NDVI), and stream flow records to assess the combined effects of meteorological, agricultural and hydrological drought. Principal component analysis (PCA) is used as a numerical approach to distill the essential hydrologic information from the input data set, which leads to the construction of Aggregate drought index (ADI). Drought severity indicators are needed to manage the water resources and to issue warnings of impending droughts.

3.3.5.2 Data and Methodology

The data used for computation of the aggregate drought index (ADI) are monthly precipitation, temperature, normalized difference vegetation index (NDVI) processed from satellite images and monthly stream flow records. First the sub-basins are regionalized into different hydrologically homogeneous pools (see Section 4.4.1) and then computation of ADI is carried out for each pool. Similarly standardized precipitation index is computed for each pool (region) identified. In computation of the ADI monthly data of all

[2]A.A. Awass and G. Foerch (2006) Regional assessment of severity of drought using multivariate drought index, Proceedings of the III International Symposium on Transboundary Waters Management, Ciudad Real, Spain

the variables for a concurrent period of 1982 to 1999 is obtained and is areally weighted using a GIS environment. However, for the calculation of Standardized precipitation index a longer time series of monthly precipitation is employed (1967-1999).

3.3.5.3 Aggregate Drought Index

Principal component analysis (PCA) is used as a numerical approach to distill the essential hydrologic information from the input data set, which leads to the construction of Aggregate drought index (ADI). The principle behind the PCA applied in time series analysis is that it linearly transforms spatially correlated series from a region into two sets of orthogonal and thus uncorrelated functions. Principal components are a re-expression of the original p-variable data set in terms of uncorrelated components z_j $(1 < j < p)$. Specifically the p-mode PCA is used where the analysis describes temporal fluctuations of input variables at a region. Computation of the principal components requires constructing a square (p x p, where p is the number of variables), symmetric, correlation matrix C_x to describe the correlations between the original data.

$$C_x = E\{(x - \mu_x)(x - \mu_x)^T\}$$

where C_x is correlation matrix, x is vector of observational data , μ_x is the mean of x and T is the transpose.

Note that correlations were only computed among data representing the same month. Therefore there were twelve C_x per pool (Region) to describe the cross correlations between the variables considered. These correlation matrices underwent principal component analysis (PCA). Eigenvectors derived through PCA are unit vectors (i.e. magnitude 1) that establish the relationship between the principal components and the standardized data (Keyantash and Dracup, 2004):

$$Z = X \cdot E \tag{3.4}$$

where Z is the n x p matrix of principal components, X is the n x p matrix standardized observational data, and E is the p x p matrix of eigenvectors.

The aggregate drought index (ADI) is the first principal component (PC1), normalized by its standard deviation:

$$ADI_{i,k} = \frac{Z_{i,1,k}}{\sigma} \tag{3.5}$$

where $ADI_{i,k}$ is the ADI for month k in year i, $Z_{i,1,k}$ is the first principal component during year i for month k, and σ is the standard deviation of Z over all years i.

The ADI utilizes only the first principal component(PC) because it explains the largest fraction of the variance described by the full p-member, standardized data set. The first PC is deseasonalized to enable each month's ADI to represent a normalized expression of variability. Without standardization, months that routinely possess a higher degree of hydrologic variability cause a chronological plot of ADI values to predictably jump.

3.3.5.4 Results and Discussion

Hydrologically homogeneous regions (pools), those regions likely to have similar low-flow producing characteristics in the basin, were identified using multivariate statistical analysis (See Figure 4.6 for the pools). The ADI is computed for each month separately and a chronological time series of ADI is established for each region (Pool). The volume of deficiency below the 0 line threshold indicates the severity of drought. Time series plot of ADI in Pool A (Upper part of basin) shows that severe drought is observed in 1984 and 1991. The severest episode of drought in the period 1982-1999 has occurred in 1991 in this region (Figure 3.7).

In Pool B in addition to the years 1984 and 1991 there is an indication that severe drought happened in 1998. In this region the drought of 1984 is much more severe. This region usually receives the highest precipitation in the basin. In Pool E (around Gode) a wider range of anomaly is observed in the ADI series. Time series plot of ADI at Pool E indicates the recent drought of 1998 is severe in the period considered for analysis. The intensity of drought is more pronounced at Gode in 1998 than the well reported 1984 drought. Comparison of ADI with Standardized precipitation index is also undertaken in the basin. Standardized precipitation index (SPI) calculates the standardized deviation of the precipitation quantiles from the long term mean of the same station (precipitation anomalies). The index has the advantages of being easily calculated, having modest data requirements and being independent of the magnitude of mean rainfall and hence comparable over a range of climatic zones. It does, however,

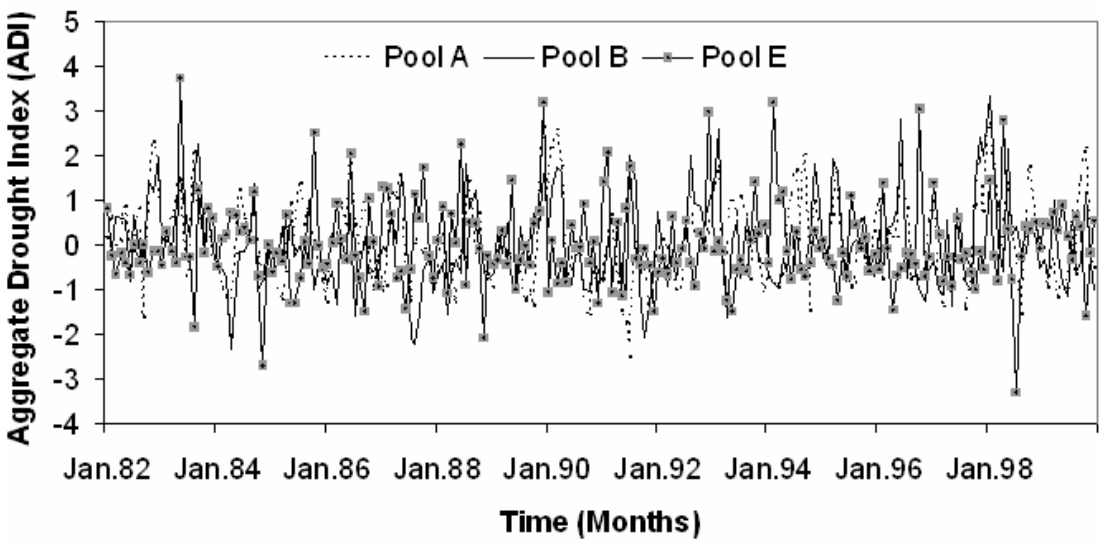


FIGURE 3.7: Aggregate Drought Index for Pool A near Dodolla(Upper valley), Pool B near Imi(Middle valley) and Pool E near Gode(Lower valley) in Wabi Shebele river basin.

assume the data are normally distributed, and this can introduce complications for shorter time periods. Guttman (1999) has examined the SPI properties and indicated that the Pearson III distribution, which has three parameters, is the best model for SPI calculation because it is more flexible and robust than Gamma. Parameters were calculated by means of the L-moments method. The extremes have a higher chance of occurrence in Pool A (Upper portion of basin) than Pool B and E. The occurrence of extremely and very dry condition in Pool E is almost negligible. This may be because the precipitation anomaly in this very arid area is very low. When one compares the severity of the aggregate drought in Pool E during the severest years of 1984 and 1998 with that of the other regions using ADI, it shows extreme condition. Even though, precipitation anomaly may not be observed during this period at Gode, stream flow deficiency is discernible. Thus the use of SPI for hydrological drought analysis in arid areas is questionable. The severest drought episodes of 1984 and 1998 are more clearly captured by the aggregate drought index in all the regions classified.

There should be greater emphasis on monitoring and drought management during normal years than during the actual drought periods. A gradual shift from rainfed agriculture to irrigated schemes in the upper portion of the basin entails all riparian countries to give due focus to analysis of drought in its whole domain.

3.4 Hydrologic Drought Event Definition

3.4.1 Categories of Hydrological Drought

Surface water Drought

Surface water drought results directly from reduced precipitation from lower surface runoff and indirectly from lower groundwater discharge to surface water (spring flow). The main indicators are reduced river flows and low water levels in lakes and reservoirs. Consequently, surface water droughts are the most visible and important in terms of human perception. However, this drought type is not necessarily a naturally induced event. It results from a complex combination of meteorological drought, water resource infrastructure and operational management decisions.

Groundwater Drought

Groundwater drought is caused by insufficient recharge through permeation and inflow from sub-catchments (van Lanen and Peters, 2000); as such it is a highly subjective phenomenon (Tate and Gustard, 2000). Ideally, it would be measured in terms of volumetric groundwater storage, but these data are not readily available and are considered to be an unsuitable indicator. Aquifer level is seen as a better indicator. Groundwater drought can also be assessed by looking at its secondary effects, such as the baseflow into rivers. Groundwater is an important source of water supply in areas where the surface water shows high temporal variability. Lack of time series of data on it hindered detail analysis of the groundwater drought in the area. Generalized cases of groundwater availability are used in assessing relative risks of drought in the region. A map showing groundwater availability during drought for Ethiopia, shown in Appendix Figure A.10, was developed by interpreting existing published information (Calow et al., 2000).

3.4.2 Drought Variables

Hydrological drought is a sustained occurrence of below expected natural water availability in rivers, lakes, groundwater level etc over large areas. Drought events can be objectively identified by the threshold level method (Yevjevich, 1967). If a hydrological drought variable Q (e.g. stream flow, ground water level, etc) is truncated at a threshold

level Q_0, then the event of surpluses ($Q > Q_0$) and deficits ($Q < Q_0$) would emerge along the time axis (Figure 3.8). Thus it is possible to define different hydrological drought variables like:

- drought severity
- drought duration
- drought intensity

and analyse their frequency.

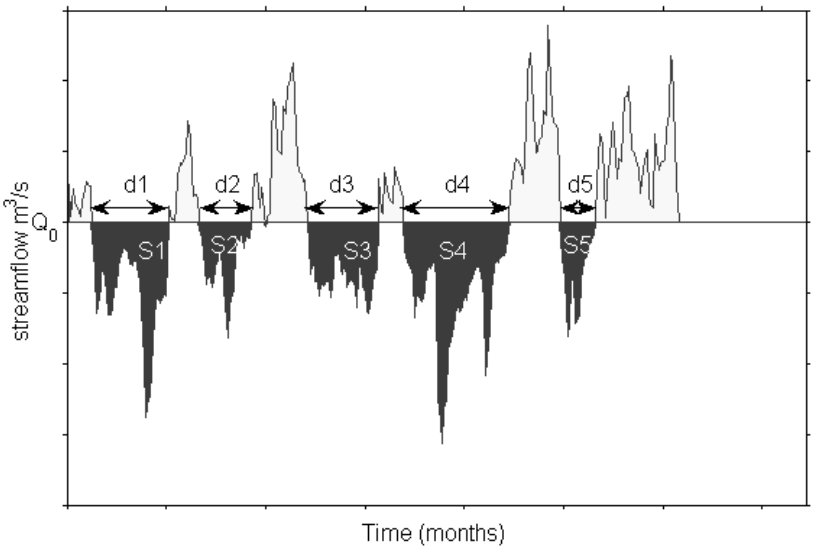


FIGURE 3.8: Definition sketch of hydrological drought events (Q_0=threshold level, S=drought severity, d= drought duration)

Drought Duration: When the streamflow, groundwater or lake level falls below the threshold value, a drought event starts and when the level rises above the threshold the drought event ends. Thus both the start and end of the drought can be defined. The run length of the drought deficit is referred as drought duration.

Drought Severity: Drought severities are normally expressed in terms of the cumulative shortages (deficit-sums) in sequences of drought variable below a desired demand level (Yevjevich, 1967). Any uninterrupted sequence of deficits can be regarded as a drought length (duration) designated by d_i(i = 1, 2, 3... j). Each drought duration is associated with the deficit sum S, i.e. the sum of the individual deficits in the successive epochs of the spell (Figure 3.8). This deficit sum S is termed as drought severity. Each drought

TABLE 3.5: Recent literatures on different classifications of frequency analysis of drought

Distribution	Bivariate (multivariate)	Univariate
Parametric	Gonzalez and Valdes 2003	Bonaccorso et al. 2003 Moy and Kapadia 1995 Salas et al. 2005
Non-parametric	Kim and Valdes 2002 2003	Sen 1998

spell has severity designated as S_i(i=1, 2, 3...). The severity of hydrological drought can be computed as:

$$S_i = \int_0^{d_i} (Q_0 - Q(t))dt \tag{3.6}$$

where S_i is severity of drought event i, d_i is the duration of drought event i, Q(t) is the stream flow at time t and Q_0 is the threshold level. No doubt the duration aspect is of importance in terms of frequency characterization and forecasting, yet the severity aspect is no less important. For instance, severity is the crucial parameter in sizing storage reservoirs towards combating droughts. There is also a need to standardize- to accept a unique definition of severity, which could be used in all drought scenarios universally. The same statement applies to drought intensity and magnitude (Sen 1980; Sharma 2000).

Drought Intensity: Drought intensity, sometimes referred as drought magnitude, is a drought deficit characteristic that represents the ratio between drought deficit volume and drought duration.

Drought frequency: Frequency distribution of drought variables can depict associated risks of droughts. There is a good deal of research on parametric or non-parametric and multi-variate or univariate frequency analysis on droughts (Table 3.5) Univariate parametric methods for frequency analysis may not reveal significant relationships among drought characteristics. It may provide two different probabilities for a single event (Frick et al., 1990). As opposed to parametric methods, nonparametric methods provide local estimates of the univariate and multivariate density function by using weighted moving averages of the data in a small neighborhood around the point of estimation (Kim et al., 2003). Separate analysis of drought characteristics cannot reveal the significant correlation relationships among drought characteristics (Shiau and Shen, 2001). On the other hand severity and duration are highly correlated.

3.4.3 Assessing Threshold Levels[3]

The threshold level approach is based on the theory of runs or crossing theory, which studies the statistical properties of runs above and below a given threshold level. The lowest values of flows for a certain period in a stream are not necessarily deficit amounts because of the possibility that the threshold value for a particular drought may be smaller than the lowest value of a given set. The truncation level specifies some statistic of the drought variable and serves to divide the time series of the variable in question into deficit and surplus sections. It may be constant or a function of time . When using a constant threshold level, the absence of trend in the streamflow needs to be checked.

Given a time series of streamflow, lake level or groundwater level, Q(t), the threshold function $Q_{T(c)}$, (modified after Peters et al. 2003), can be defined as follows:

$$\int_0^M (Q_{T(c)} - Q(t))_+ dt = c \int_0^M (Q_b - Q(t))_+ dt \tag{3.7}$$

where

$$Q_+ = \begin{cases} Q & if\ Q \geq 0; \\ 0 & if\ Q < 0 \end{cases}$$

$Q_b = min(average\ supply, average\ demand)$,

M is the length of the time series and c is the drought criterion. The drought criterion c is the ratio of the deficit below the threshold to the deficit below the average. It determines the height of the threshold level. If c = 1 the threshold level is equal to the average Q. If c = 0 the threshold level is equal to the minimum of Q. This definition of the threshold also ensures that the total drought deficit decreases with decreasing amplitude of Q(t). The last line of the equation is added with the argument that deficiency is a function of both availability of supply and demand put on it (Awass and Foerch, 2006a). Thus there is a need to substantiate the truncation level of drought from both supply and demand perspectives. In general, Q_T is an arbitrary value that depends on the objective of the drought study. Water scarcity is a relative concept and can occur at any level of supply or demand. Scarcity may be a social construct, a product of affluence, expectations and customary behaviour, or a change in supply pattern due to climate change. The truncation level which is used to objectively demarcate the onset and end of drought

[3]A.A Awass and G. Foerch (2006) Assessing the Threshold Levels used to detect Hydrological And Environmental Drought in semi- arid areas. Proceedings of Man and river systems II, Paris, France

is subjective. In fact, it is difficult to get time series data on actual water demands in Wabi Shebele river basin. Assuming the progressive change on demand of water on the basis of population change alone may not represent the actual demand because it is often noted that the increase in demand is higher than the increase of population.

There is a focus to use mean flow, median flow, a percentile of the daily flow duration curve like Q_{70} or more of the river as threshold levels for drought. The threshold level can either represent a certain demand, e.g. water supply or ecological minimum discharge, or it can represent a normal low flow condition of the stream. In the latter case a percentile of the daily flow duration curve can be applied as threshold level. The use of a statistical value (e.g. from the flow duration curve) allows the regional comparison. Tallaksen and Hisdal (1997) applied Q_{50}, Q_{70} and Q_{90} as a threshold value for a drought investigation in Norway. High thresholds would lead to the problems of multi-year droughts. In contrast, low thresholds especially in regional studies with a lack of long time series lead to fewer droughts, which results in interpretation difficulties. At Gode station, the flow exceeded 70% of the time, Q_{70}, approximately matches with an average truncation level set with c = 0.38. But the drought volume established using a threshold level of Q_{70} belittles the historical droughts observed in typical years of 1988 and 1992. In many stations in temperate regions thresholds of Q_{70} or Q_{90} were found to be suitable (Fleig et al., 2005). However, for the many ephemeral rivers both Q_{70} and Q_{90} are zero. Thus relatively higher threshold levels need to be considered.

The threshold level affects the drought analysis considerably. The time series of monthly stream flows at Dodolla and Gode are truncated with various threshold levels and the total number of crossings (upcrossing and downcrossing) are observed. The total number of crossings increases as the threshold per mean flow increases by up to 1-1.5 (Figure 3.9). Further increase in the threshold per mean flow decreases the total number of crossings at both stations. It shows that the total number of crossings is higher at a threshold level a bit more than the average flow. Specifying the threshold level enables one to determine the drought severity (run sum) and duration (run length).

The frequency analysis of the severity of drought evaluated for various threshold levels at the above stations is shown in Figure 3.10. It is discernible that the drought severity values at a higher truncation level (i.e. less flow exceedance) are higher than those at a low truncation level for any return period. This increase becomes dramatic at upstream station Dodolla for a truncation level below a flow exceedance of 50%. On

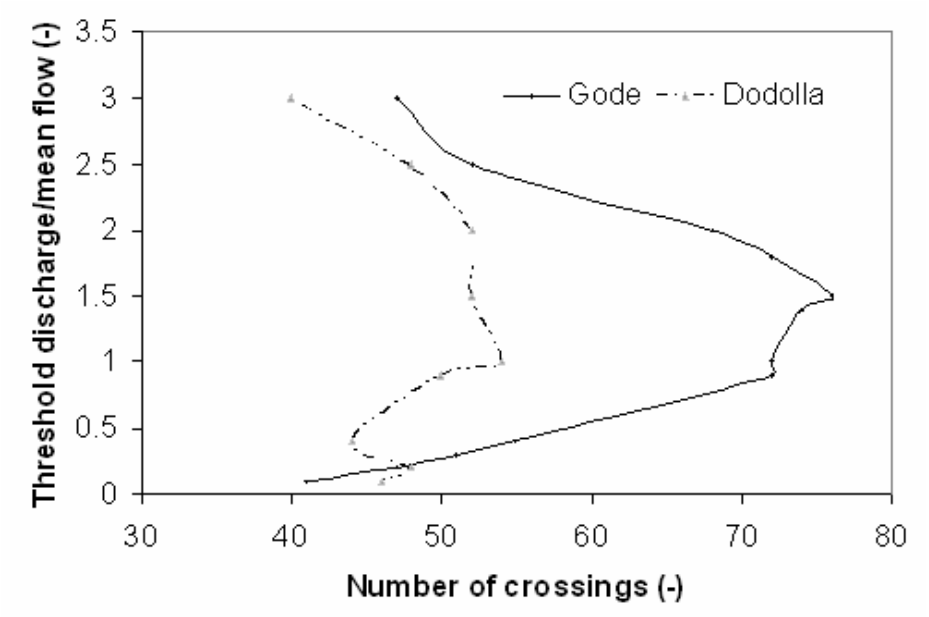


FIGURE 3.9: Threshold discharge per average flow vs mean number of crossings at Dodolla and Gode gauging stations

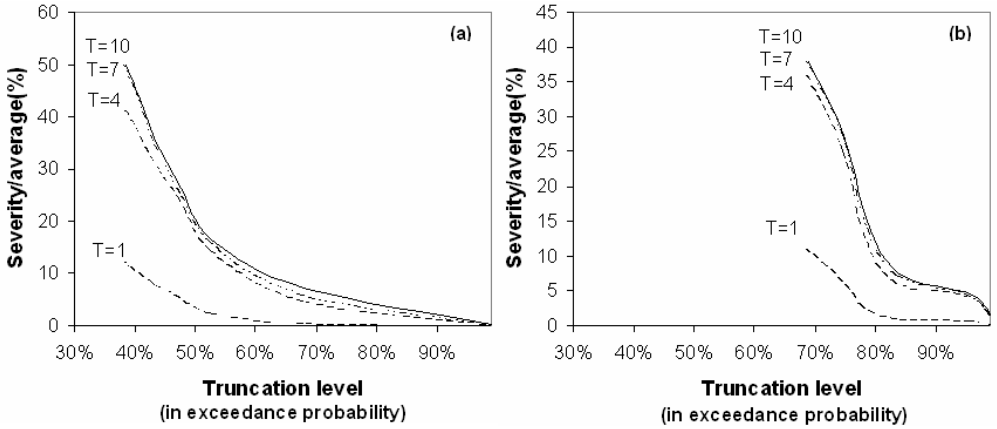


FIGURE 3.10: Frequency of severity of drought per average annual flow at different threshold levels in (a) Dodolla (b) Gode gauging stations.

the otherhand setting a threshold level at near 80% of the flow exceedance, barely any severity of drought can be noticed for smaller return periods. In case of Gode the increase in drought severity gets larger for a truncation level below a flow exceedance of 80%. For the analysis of short droughts, many literatures have used thresholds based on percentage points of the flow duration curve. This is advantageous in the sense that the value can be related to characteristics of the rivers and to abstraction levels for typical water uses, helping to provide a meaningful expression of drought.

Chapter 4

Hydrologically Homogeneous Drought Pools

Streamflow prediction in ungauged basins provides essential information for water resources planning and management and ecohydrological studies yet remains a fundamental challenge to the hydrological sciences. The most desperate situation exists when no flow records are available anywhere in the catchment. A difficulty associated with frequency analysis in hydrology is the lack of sufficient data. Large standard errors of estimate result because of small sample sizes. Records from nearby stations are often used to substitute spatial to temporal characteristics.

It is important to form groups of sites that approximately satisfy the homogeneity condition, i.e. the sites' frequency distributions are identical apart from site-specific scale factor. A homogeneous region may be viewed as a collection of catchments, which are similar in terms of catchment hydrological response, but not necessarily geographically contiguous. The use of the term region for both spatially contiguous and spatially non-contiguous regional classes is widespread among hydrologists. In general, a region denotes a contiguous zone identified by geographical or political boundaries. In applied hydrology, a region is often defined by hydrometric or operational boundaries. With respect to regional frequency analysis the term is satisfactory when data are pooled from fixed regions but less suitable for focused pooling schemes. In pooling schemes which group catchments according to hydrological similarity rather than geographical position, the scope for confusion is high. Such groups of sites are geographically dispersed and continued use of the word region can not be justified. The Flood estimation Handbook

proposes a change in terminology. Thus a scheme for choosing how to group sites for frequency analysis (formerly regionalization) is termed as pooling scheme, while a region for pooled frequency analysis is called a pooling group (IH, 1999). In this study those terminologies are adopted.

Similarities in characteristics of low flows from different gauging stations are used to identify hydrologically homogeneous pools and compute the frequency analysis in the area. Pooled estimates are helpful to reduce the uncertainty in a quantile estimate at gauged sites and more importantly to estimate the quantiles at ungauged sites. Classification of catchments into groups may be based on standardised flow characteristics estimated from the available observed or simulated streamflow records (Wiltshire, 1985). Alternatively, the regions are delineated using catchment physiographic and climatic parameters (Acreman and Wiltshire, 1989; Krokli, 1989) obtained from maps and hydrometeorological data. This last approach has more relevance to the estimation of flow characteristics at ungauged sites since an ungauged catchment for which, for example, low flow estimation is undertaken, should first be assigned to one of the identified groups/regions and this can only be done on the basis of catchment physiographic information.

Regardless of the way regions are formed, they have to satisfy homogeneity criteria. Each of the regions should have two basic properties, that is, dissimilarity from other regions and, homogeneity within the region. The importance of regional homogeneity has been demonstrated by Hosking et al. (1985), Hosking and Wallis (1993), Wiltshire (1986) and Lettenmaier et al. (1987). There are several tests available to examine regional homogeneity in terms of the hydrologic response of stations in a region (Zrinji and Burn, 1996). In order to ensure that the resulting regions are unique internally to a given level of tolerance, homogeneity and heterogeneity test need to be performed. Mostly the regional characteristics of severe seasonal droughts have been analysed by looking at the extreme value properties of annual maximum series (AMS) of drought characteristics using both the L-moment diagram and the EOF method (Tallaksen and Hisdal, 1997). It is also interesting to evaluate how the various catchment characteristics relate to the low flow response of a basin (Awass and Foerch, 2006b).

4.1 Catchment Characteristics as Predictors of Base Flow Index

4.1.1 Background

Understanding the relative importance that ground water discharge plays in maintaining stream flows is essential. River systems are often augmented by their base flows during lean seasons. The Base Flow Index (BFI) is used as a measure of the base flow characteristics of catchments. It provides a systematic way of assessing the proportion of base flow in the total runoff of a catchment. It indicates the influence of soil and geology on river flows, and is important for low flow studies. Nowadays extreme low flow events are more diligently analysed and given focus in the emerging field of ecohydrology. However, many of the catchments in developing countries are ungauged, thus, it is often difficult to get recorded data on base flows of rivers. This study seeks to establish a relationship between the climatic, morphologic and geologic features of a catchment to its base flow in the Wabi Shebele river basin, East Africa. It employs the variables catchment size, stream density, climate index, hypsometric integral, normalized difference vegetation index (NDVI) extracted from satellite images and geologic features to derive the base flow index of a catchment.

4.1.2 Data and Methodology

Unlike event based methods, continuous base flow separation techniques do not normally attempt to simulate the base flow conditions for a particular flood event, nor are they appropriate for the identification of the origin of base flow. These methods are rather aimed at the derivation of objective quantitative indices related to the long term base flow response of the catchment (e.g. base flow index -BFI) and at the estimation of continuous time series which specifically characterize the high frequency, low amplitude base flow regime (Hughes et al., 2003). Monthly base flow contribution to stream flow at gauged stations were estimated using automated analysis tool developed by USGS, HYSEP (Sloto and Crouse, 1996). In deriving the relation ship between catchment characteristics and base flow index, various variables are employed. These include the catchment size (A), stream density (D_d), climate index or aridity (humidity) index (AI), hypsometric integral (I), normalized difference vegetation index (NDVI) extracted from

satellite images and geologic features. A number of morphometric indices have been used to summarize landform geometry. The hypsometric integral I is the area beneath the curve which relates the percentage of total relief to cumulative percent of area. This provides a measure of the distribution of landmass volume remaining beneath or above a basal reference plane. Integration of the hypsometric curve gives the hypsometric integral I. It is proved mathematically that the elevation-relief ratio E which is defined as (Pike and Wilson, 1971):

$$E = \frac{(mean\,elevation - min\,elevation)}{(max\,elevation - min\,elevation)} \tag{4.1}$$

is identical to the hypsometric integral I but has the advantage that it is much more easy to obtain numerically. In this work, E is therefore used instead of I.

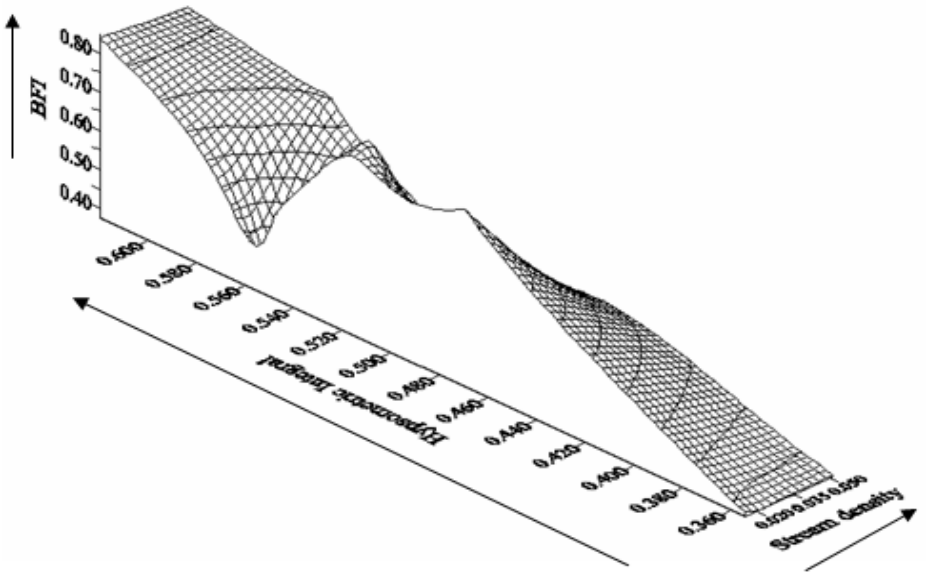


FIGURE 4.1: Base flow index versus hypsometric integral and stream density in the Wabi Shebele river basin.

The base flow index is higher for places with an average hypsometric integral (0.48 - 0.54) and lower stream density. An increase in stream density and hypsometric integral is matched by an increase in base flow index (Figure 4.1). The coefficient of variation (CV) of base flow index in the basin ranges from 14% to 79%. Climate index, often denoted as aridity/humidity index, is the ratio of mean annual precipitation to potential evapotranspiration. It affects the base flow of catchments. A positive correlation between climate index and normalized digital vegetation index as high as 0.824 is noticeable in Wabi Shebele basin. Spectral measurements made in the red (580 - 680 nm) and near

infrared (725 - 1100 nm) regions of the spectrum allow certain land and vegetation characteristics to be derived. Normalized difference vegetation index (NDVI) is the normalized differences between the reflectance of the red and near infrared spectrum.

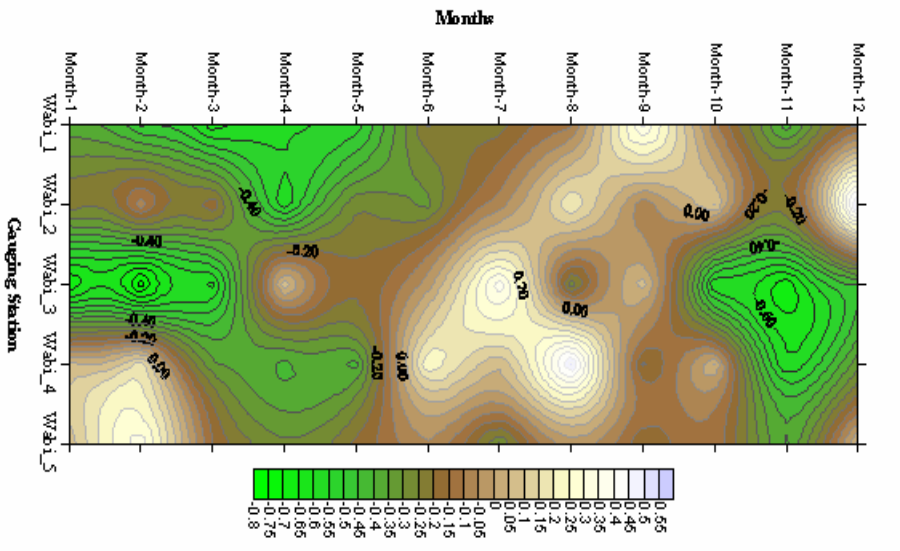


FIGURE 4.2: Spatial and temporal correlation of base flow index (BFI) and normalized difference vegetation index (NDVI) at selected stations in Wabi Shebele basin (Month 1 is January).

A high negative correlation between BFI and NDVI is observed in the months of November to March in the middle valley of Wabi Shebele basin where as a positive correlation that gradually narrows in time upstream from Gode through June to September is noticeable (Figure 4.2). The rainfall season is mainly in June-September in the upstream portion of the river basin where as in the downstream it is predominantly falling in March-May (See Figure 2.1). It can be observed that the relationship between BFI and NDVI becomes highly negatively correlated in the dry seasons. This is also observed from time series plot of the two at the middle valley near Tebel station (Figure 4.3). During the famous drought years of 1990, 1991 and 1998 the correlation is appreciably higher than at other years.

It was not easy to get data on aquifer properties. Pumping test results of most of the boreholes in the basin is missing. Thus it was not possible to include more rigorous parameters in the analysis like permeability that could better designate hydrogeological features of the basin. Thus average static water level (SWL) measured from ground surface is used. Values of base flow index determined for a network of stream flow gauges are matched to the composite morphometric and climatic data using spatial and

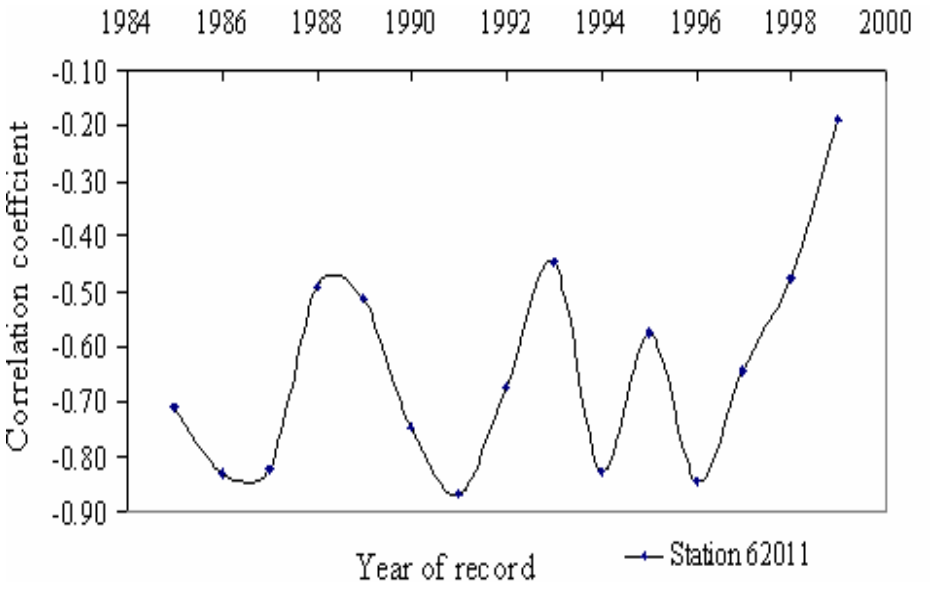


FIGURE 4.3: Time series plot of correlation between BFI and NDVI at the middle valley in Wabi Shebele basin.

regression analysis. To relate the BFI to a usable flow statistic, a relationship was derived between BFI and Q_{70}, the point on the flow duration curve at which flows are exceeded 70% of the time. Q_{70} was chosen because it is the critical point that has been often used in most previous works (Tallaksen and Hisdal, 1997).

4.1.3 Results and Discussion

Area of the catchment, average slope and drainage density were identified through factor analysis to affect base flows (Zecharias and Brutsaert, 1988). There is a strong negative correlation between catchment size and hypsometric integral, about -0.814, in Wabi Shebele basin. The average base flow index in the basin ranges from 0.375 to 0.856 where as the range of climate index is wider reaching from 0.134 to 0.980. The BFI has a strong relationship with climate and geology. If the storage is not large, catchments with high rainfall and low evapo-transpiration (high climate index) underlain with impervious formations tend to give high base flow. Among the topographical parameters tested, drainage density index has better relationship with BFI. Using least squares method different parameters were fitted against the base flow and it is found that the following relationship yielded the best result for the area under investigation:

$$BFI = 0.956D_d + 2.7951I - 0.164NDVI - 8.428SWL \quad r^2 = 0.97 \qquad (4.2)$$

The standard error of this fit is 0.15. When flows are expressed as a percentage of the long term mean flow (standardized), the dependencies on the climatic variability across the basin and on the scale effect of catchment area are minimised. The shape of the standardised flow duration curve indicates the characteristic response of the catchment to rainfall. The gradients of the log normal flow duration curves for a range of catchments with differing geology illustrate that impermeable catchments have high gradient curves reflecting a very variable flow regime; low storage of water in the catchment results in a quick response to rainfall and low flows in the absence of rainfall.

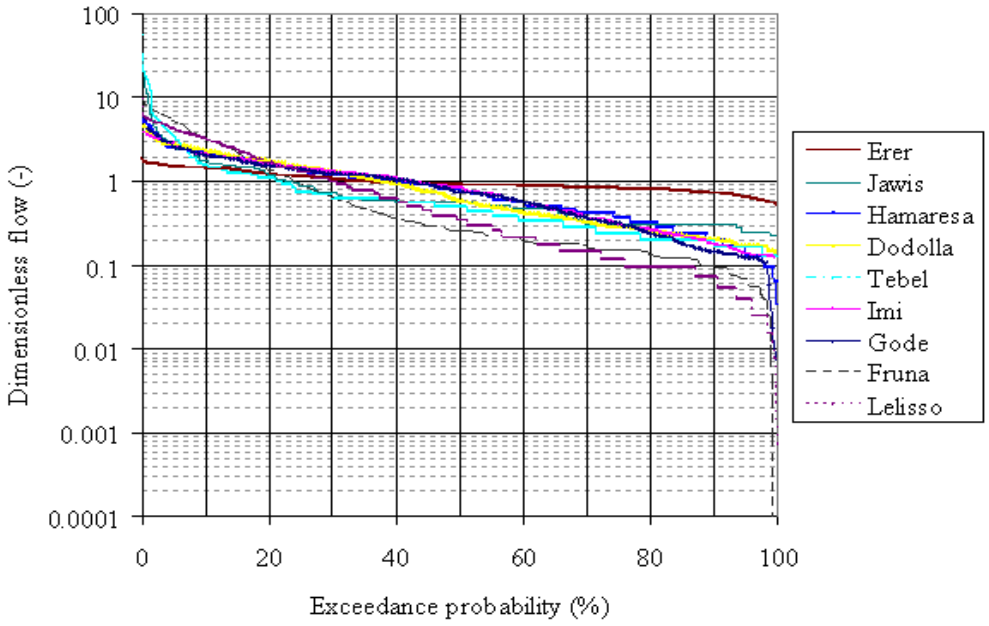


FIGURE 4.4: Semi-log plot of stream flow duration curves at different stations in Wabi Shebele river basin.

Catchments around Erer (station 62013) and Jawis (station 62015) experience low variability and average base flow index as high as 0.84 (Figure 4.4). In Australia a fit of flow duration curve with an equation of the form

$$y = \frac{1}{b} log(\frac{a}{x} - 1) \tag{4.3}$$

where y is log of normalised flow, x is exceedance probability (%), a and b are coefficients, yielded an interesting result (Post, 2005). The parameters a and b have somewhat physical connotation of the percentage at which the flow ceases to exist. The model is especially applicable for humid regions. In Zimbabwe, an exponential model of the following form can best describe the relationship between exceedance probability and

TABLE 4.1: Goodness of fit of the different models between dimensionless flow and exceedance probability.

Station code	Normalized flow						Log(Normalized flow)	
	$m \cdot log(x) + n$		$b_0 \cdot exp(-b_1 p)$		$a_0 \cdot exp(a_1 p) + a_2 \cdot exp(a_3 p)$		$\frac{1}{b} log\left(\frac{a}{x} - 1\right)$	
	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE
61001	0.959	0.274	0.993	0.118	0.994	0.106	0.541	0.443
61005	0.751	0.226	0.663	0.263	0.751	0.226	0.641	0.203
61004	0.987	0.125	0.946	0.255	0.997	0.055	0.953	0.237
62013	0.851	0.042	0.489	0.078	0.981	0.015	0.935	0.027
61016	0.909	0.507	0.991	0.158	0.997	0.091	0.991	0.158
62007	0.992	0.077	0.927	0.233	0.996	0.053	0.81	0.171
62015	0.528	1.472	0.781	1.004	0.983	0.281	0.513	0.238
62011	0.603	1.511	0.855	0.915	0.959	0.488	0.427	0.312
62018	0.981	0.113	0.975	0.128	0.978	0.128	0.779	0.187

stream flow (Mazvimavi et al., 2005):

$$q_p = b_0 exp(-b_1 p) \tag{4.4}$$

where q_p is the dimensionless stream flow, p is the exceedance probability, b_0 and b_1 are coefficients. The above equation fits better for the low flow estimates (high exceedance probability). Equations 4.3 and 4.4 were applied to rivers in Wabi Shebele basin. Their performance is compared against a different variant of the exponential model (Table 4.1).

It can be observed from Table 4.1 that the following equation has the highest coefficient of determination (r^2) and the lowest root mean square error (RMSE).

$$q_p = a_0 exp(a_1 p) + a_2 exp(a_3 p) \tag{4.5}$$

Where q_p is the flow exceeded p percent of the time, a_0, a_1, a_2, a_3 are coefficients.

It better explains the relationship between the dimensionless flow and exceedance probability in Wabi Shebele river basin. The coefficients of these flow duration curve (FDC) are regressed against the base flow index for different sub-basins. These coefficients are relatively better estimated for the upper and middle valleys of Wabi Shebele basin using

the following fits:

$$a_0 = 110.95BFI - 63.83 \quad r^2 = 0.90 \tag{4.6}$$

$$a_1 = -3.36BFI + 1.51 \quad r^2 = 0.71 \tag{4.7}$$

$$a_2 = -20.41ln(BFI) - 3.084 \quad r^2 = 0.68 \tag{4.8}$$

$$a_3 = -0.033BFI - 0.014 \quad r^2 = 0.68 \tag{4.9}$$

Station 63001 described well with the exponential model given in equation 4.4, but in the regional fit of the coefficients against the BFI it did not fit well with the above formulae. Thus it should be treated separately. The water level monitoring for one hydrologic cycle on two wells at Gode showed that the water level is deep always lower than the Wabi Shebele river bed and the phreatic water level is practically the same during the hydrologic cycle and no interaction with Wabi Shebele river water indicating the permeability is very low (MoWR, 2003). Validation of the exponential fit in the basin using chi-square test between measured and calculated Q_{70}, the point on the flow duration curve at which flows are exceeded 70% of the time, at four other stations including Gode, yielded 0.94. Thus it is fair to use the developed relationship for estimating the base flows. The availability of gauged stations with long years of record restricted the validation not to be carried out on more stations. Region of influence approach for estimating flow duration curves at ungauged sites is common. Delineating hydrogeologically similar catchments using soils data as surrogate of hydrogeological data yielded good results in UK (Holmes et al., 2002). Analysis at hydrogeologically homogeneous catchment level may improve the search for a better estimate of base flow index at ungauged catchments.

4.1.4 Conclusions

Estimation of continuous time series of base flow index which specifically characterize the high frequency, low amplitude base flow regime are done at gauged catchments in Wabi Shebele basin. The relative strengths in association between the climatic, morphometric and geologic features of the catchment to the base flow estimates are weighted and a plausible relationship is produced. There is a strong negative correlation between BFI and NDVI which is even more pronounced during the dry seasons in the basin. Values of base flow index determined for a network of stream flow gauges are matched to the composite morphometric and climatic data using spatial and regression analysis.

An exponential variant model fitted to the flow duration curves is used as an aid in deriving a relationship between BFI and Q_{70}. The developed relationship can be used for fairly estimating the base flows in the ungauged portion of the river basin considered. However, in view of the tremendous spatio-temporal heterogeneity of climatic and landscape properties extrapolation of response information or knowledge from gauged to ungauged basins remains fraught with considerable difficulties and uncertainties.

4.2 Pooled Estimates of Low Flow Quantiles[1]

4.2.1 Background

Low flow estimates are vital for planning water supplies, water quality management, issuing and renewing waste disposal permits, hydropower, irrigation systems and for assessing the impact of prolonged drought on aquatic ecosystems. Nowadays extreme low flow events are more diligently analysed and given focus in the emerging field of ecohydrology. Estimation of low flow is also important for small-scale irrigation projects that contribute significantly to poverty alleviation by means of increased cereal production and generation of rural employment.

The basic problem is to find the relation between the low flow magnitude and return period at a site for which no record of flows is available. Even a site which has only a few years of record has to be dealt with, in some aspects, as if it were ungauged catchment because the information contained in it has to be augmented. Short years of streamflow record leads to unreliable estimates of parameters and low-flow quantiles. Thus one seeks to employ spatial information to make up for the paucity of temporal information. In this section methods which are based on low flow frequency analysis are dealt with. Methods for ungauged catchments using rainfall-runoff models are also available but are not dealt here.

Two adjacent river catchments may have rather different topography, soils or other local anomalies. Geographical proximity of two catchments is no guarantee that they are similar from the hydrological response point of view (Cunnane, 1989). An alternative to geographical proximity as a measure of affinity offered by some clustering algorithms is

[1] A.A. Awass(2003) Pooled estimates of low flow quantiles in the Wabi shebele river basin. Water, Ethiopian Journal of Water Science and Technology. May 2003, AWTI.

Euclidean distance in the n-dimensional feature space that is defined by the n characteristics that have been adopted for site description. The final test of the homogeneity of the pools is independent of the method of identifying the pooling groups. Due to the complexity of factors that affect the generation of stream flows, there are different objective methods of pooling schemes. Index flood procedure is used here. Different authors identify hydrologically homogeneous pools based on either geographical considerations or hydrologic response characteristics or a combination of the two.

A possible disadvantage to methods of classification based upon Euclidean distance, including Kohonen networks, is the absolute certainty of the allocation to a particular class so artificial neural network is advised (Hall and Minns, 1999). Since the number of sites considered in this study is few, almost every possibility of cluster aggregation is attempted.

The study focuses on one of the water scarce basins in Ethiopia, the Wabi shebele river basin. There is no generally agreed upon methodology for estimating low flow quantiles and associated parameters in the country. Parameter estimation using L-moments gives estimates that are highly efficient, almost unbiased and not much influenced by outliers in the data (Hosking, 1986; Hosking and Wallis, 1997). However, extreme sample values are given little weight, thus the quantiles at the tails may not be efficiently estimated when L-moments are used (Gelder, 2000). The trends shown by low flows are expected to be better captured by the use of LL-moments.

The hydrogeology of the study area is dominated by Mesozoic sedimentary formations, to some extent there are also volcanic rocks at the north west of the basin and isolated ridges and hills within the sedimentary basin (see Section 2.1.3).

4.2.2 Data and Methodology

Hydrometeorological Data: the largest share of the Wabi Shebele basin is located in Semi-arid to arid zones (about 102,220 Km^2) but there are only few meteorological stations where as most of the meteorological stations are concentrated around the upper valley. Table A.1 and Table A.2 in the Appendix provide location of the hydrometeorological stations in and around the Wabi Shebele river basin.

Statistical Tests : It is supposed that the flows used in this study are natural discharge records unaffected by major abstractions or effluent discharge. Outliers were removed from all the data series. A non-parametric test of the lag-one serial correlation test was applied at a significance level of 10% for testing the independence of the data series.

Cluster Analysis : Statistically homogeneous regions, those regions likely to have similar low flow producing characteristics in the basin were identified using multivariate statistical analysis . In the cluster analysis each gauging site is associated with one or more catchment characteristics. Then a measure of similarity between two sites is measured by a scaled Euclidean distance between them in catchment characteristic space. A number of variables are considered in the cluster analysis. The procedure involved superimposing the location of each gauging station on the digital basin boundary map and identifying the mean area, the mean annual rainfall, mean elevation, hydrogeologic features and NDVI, obtained from satellite images. The Euclidean distance, D_{ij} between two sites in catchment characteristic space is given as:

$$D_{ij} = \sqrt{\sum_{k=1}^{p} \left(\frac{X_{ik} - X_{jk}}{stdev(X_k)} \right)^2} \tag{4.10}$$

where p is the number of attributes, X_{ik} and X_{jk} are values of the attribute k at subbasins i and j, $stdev(X_k)$ is the standard deviation of the attribute k.

In this analysis hierarchical clustering using Ward's method is used. This procedure attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics, using an algorithm that starts with each case (or variable) in a separate cluster and combines clusters until only one is left.

In the dendrogram (Figure 4.5) it can be noticed that if only two clusters are required all the subcatchments except one stand as one cluster and station no. 63001 (at Gode) stands out as a separate cluster. If three clusters are required the bigger cluster splits to two. Station no. 62011 (at Tebel) belongs hydrologically more to the category of Upper valley; Station No. 61009 (at Melka Wakena) and Station no. 61015 (Maribo at Kara Birole) belong to the middle valley category.

Discriminant Analysis: After homogeneous pools have been identified, a test was carried out to find out whether the proposed pool was reasonably homogeneous. This involved

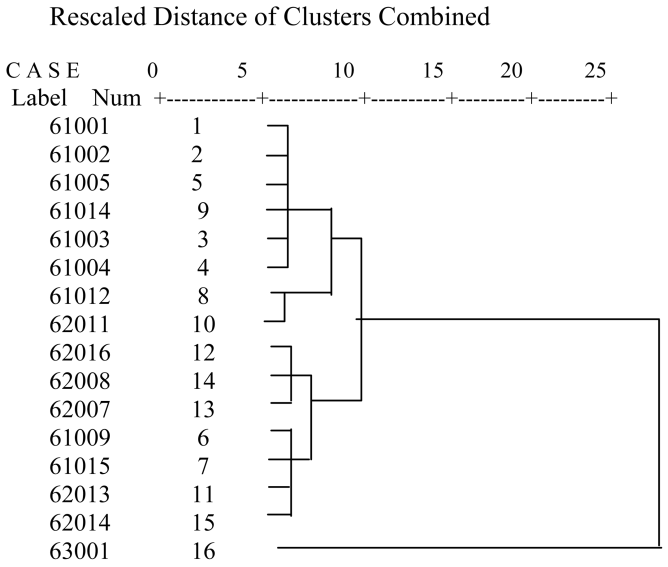


FIGURE 4.5: Dendrogram using Wards Method.

TABLE 4.2: Coefficient of variation of λ_2 for each cluster proposed.

Trial	Sub-basin (Station number)	Region	Cc
1	61001; 61002; 61003; 61004; 61005; 61006; 61009; 61015	Whole	0.694
2	61001; 61002; 61003; 61004; 61005; 61015; 61009;	Pool A	0.388
3	61001; 61002; 61003; 61004; 61005; 61015;	Pool B	0.168
4	61006; 61009; 61015	Pool C	0.26
5	61006; 61009;	Pool D	0.066

computing a statistic Cc, known as the coefficient of variation of the at-site second L-moment coefficients (λ_2). Excluding station no. 61006, which has large coefficient of variation, pool A yielded Cc of 0.388 which is higher than the reasonable limit. In the absence of station no. 61009 pool B yielded Cc of 0.168 (Table 4.2). The null hypothesis that pool B is hydrologically homogeneous cannot be rejected. Thus station no. 61015 cannot be statistically justified to be out of pool A as it's indicated in the cluster analysis. However it is more likely that station no. 61009 belongs to another cluster.

LL-moments : Analogous to LH-moments applied for upper tail of the distribution, LL-moments give more weight to the smaller observations and place an emphasis on fitting the lower tail of the selected probability distribution. Quantiles have lower biases when estimated by LL-moments. LL-moments can be defined using samples of size $r + m$

$(m = 1, 2, \ldots)$ and computing the expectations for the r smallest elements of the sample (Bayazit and Onoz, 2002):

$$\lambda_{Lr}^{m} = r^{-1} \sum (-1)^{k} \binom{r-1}{k} E[x_{r-k/r+m}] \quad r = 1, 2, \cdots . \tag{4.11}$$

where $m = 0$ corresponds to the L-moments. An increase of m increases the weight of the lower part of the data. The first two LL-moments are:

$$\lambda_{L1}^{m} = E[x_{1/1+m}] \qquad \lambda_{L2}^{m} = E[x_{2/2+m} - x_{1/2+m}] \tag{4.12}$$

Using equation 4.12 above, LL-moments can be expressed in terms of the distribution function F:

$$\lambda_{L1}^{m} = (1+m) \int_{0}^{1} x(F)(1-F)^{m}\, df \tag{4.13}$$

$$\lambda_{L2}^{m} = \frac{1}{2}[(1+m)(2+m) \int_{0}^{1} x(F)(1-F)^{m}\, df - (2+m) \int_{0}^{1} x(F)(1-F)^{1+m}\, df] \tag{4.14}$$

Expressions to estimate the LL-moments from a sample of size n arranged in ascending order can be derived following the approach of Wang (1996). Consider equation 4.12 for λ_{L1}^{m}. In a sample of size n, there are $i-1$ values less than (or equal to) x_i and $n-i$ values greater than (or equal to) x_i. For x_i to be the smallest of any combination of $1+m$ values from the sample, the other m must come from the $n-i$ larger values. There are $\binom{n-i}{m}$ such combinations. The total number of combinations is $\binom{n}{1+m}$. Therefore the estimate of λ_{L1}^{m} is :

$$\lambda_{L1}^{m} = \sum \binom{n-1}{m} x_i / \binom{n}{1+m} \tag{4.15}$$

Similarly for x_i to be the second smallest of any combination of $2+m$ values from the sample, m must come from the $n-i$ larger values ($\binom{n-i}{m}$ combinations) and one element must come from $i-1$ smaller values. For x_i to be the smallest of any combination of $2+m$ values from the sample, $1+m$ must come from the $n-i$ larger values ($\binom{n-i}{1+m}$ such combinations). The total number of combinations is $\binom{n}{2+m}$. Therefore, λ_{L2}^{m} can be estimated as:

$$\lambda_{L2}^{m} = \sum [\binom{n-i}{m}\binom{i-1}{1} - \binom{n-i}{1+m}] x_i / 2 \binom{n}{2+m} \tag{4.16}$$

Pooling schemes

Weibull distribution: Many literatures recommend Weibull distribution to analyze the frequency of low flows because it has a lower bound and also because of the appeal of the extreme value theory in statistics, in the context of annual minima. The probability density function of x (f(x)) and the distribution function of the Weibull distribution of x(F(x) is presented in different ways and with different notations in the literature. In this study the following form is used (Tallaksen and van Lanen, 2004):

$$f(x) = \frac{\kappa}{\alpha}\left(\frac{x-\xi}{\alpha}\right)^{\kappa-1} exp\left[-\left(\frac{x-\xi}{\alpha}\right)^{\kappa}\right] \tag{4.17}$$

$$F(x) = 1 - exp\left[-\left(\frac{x-\xi}{\alpha}\right)^{\kappa}\right] \tag{4.18}$$

Where κ=shape parameter,

α=scale parameter,

ξ=lower bound (displacement) parameter.

If the lower bound is equal to zero, i.e. if the probable minimum flow is zero or that there is the extreme possibility of the stream completely drying up, then the distribution takes the two parameter form. In fitting a three parameter distribution the following three types of failures occur:

(i) lower bound parameter is less than zero ($\xi < 0$),

(ii) lower bound parameter is larger than the lowest observed, discharge ($\xi > Q_{min}$)

(iii) procedure fails to produce results.

There are no failures in the two parameter procedures. Hence, unless clear evidence strongly suggest otherwise, for a particular set of data, it is always desirable to use two parameter procedures. The inverse of the two parameter Weibull probability distribution function is

$$x(F) = \alpha[-ln(1-F)]^{1/\kappa} \quad x \geq 0 \tag{4.19}$$

LL-moments can be computed using equation 4.13

$$\lambda^m_{L1} = (1+m)\int_0^1 \alpha[-ln(1-F)]^{\frac{1}{\kappa}}(1-F)^m\, df = \alpha\Gamma(1+\frac{1}{\kappa})/(1+m)^{\frac{1}{\kappa}} \tag{4.20}$$

$$\lambda^m_{L2} = \alpha\Gamma(1+\frac{1}{\kappa})(\frac{2+m}{2})[\frac{1}{(1+m)^{\frac{1}{\kappa}}} - \frac{1}{(2+m)^{\frac{1}{\kappa}}}] \tag{4.21}$$

where Γ is gamma function (see Appendix equation B.13). When $m = 0$, the above

TABLE 4.3: Low flow quantiles (m^3/s) estimated by Weibull and power distributions using LL and L- moments of the lower 80% of the data.

Return period	Weibull (LL moment)			Power (LL moment)			Weibull (L moment)	
	m=0	m=1	m=2	m=0	m=1	m=2	lower 80%	Full data
2	0.995	0.980	0.996	1.363	1.210	1.332	1.232	0.927
5	0.850	0.828	0.849	1.132	1.019	1.145	1.099	0.309
10	0.766	0.740	0.764	0.984	0.895	1.021	1.019	0.150
50	0.609	0.578	0.605	0.711	0.662	0.782	0.863	0.030
100	0.552	0.521	0.548	0.618	0.582	0.697	0.804	0.015
150	0.522	0.490	0.518	0.569	0.539	0.652	0.772	0.010

expressions would yield L-moment estimates of Weibull distribution as provided in Appendix B. Solving for the parameters α and κ:

$$\alpha = (1+m)^{\frac{1}{\kappa}}\lambda_{L1}^{m}/\Gamma(1+\frac{1}{\kappa}) \quad \kappa = ln(\frac{1+m}{2+m})/ln\left(1-\frac{2\cdot CV_L^m}{2+m}\right) \tag{4.22}$$

$$CV_L^m = \lambda_{L2}^m/\lambda_{L1}^m$$

where CV_L^m is the LL-moment coefficient of variation.

Power distribution: The power distribution has the upper bound x_0 and therefore it is a conditional distribution $F(x/X \leq x_0)$ (Bayazit and Onoz, 2002). Its inverse is given as:

$$x(F) = x_0 F^{\frac{1}{c}} \qquad 0 \leq x \leq x_0 \tag{4.23}$$

From equation 4.23 and equation 4.13, one obtains:

$$\lambda_{L1}^m = (1+m)x_0\Gamma(1+\frac{1}{c})\Gamma(1+m)/\Gamma(2+\frac{1}{c}+m) \tag{4.24}$$

$$\lambda_{L2}^m = (2+m)x_0[(1+m)\Gamma(2+\frac{1}{c})\Gamma(1+m) - \Gamma(1+\frac{1}{c})/\Gamma(2+m)]/2\Gamma(3+\frac{1}{c}+m) \tag{4.25}$$

Solving for the parameters c and x_0, one obtains:

$$c = \left[(1+\frac{m}{2})\frac{1}{CV_L^m} - 1\right]/(2+m) \quad x_0 = \frac{(1+c)(1+2c)\ldots[1+(1+m)c]\lambda_{L1}^m}{(1+m)!c^{1+m}} \tag{4.26}$$

There is a drastic decline in the estimated quantiles using L-moments with increase in return period when the full data is considered (Table 4.3).

4.2.3 Conclusion

Quantile estimation of the fitted Weibull and power distributions differ when L-moments or LL-moments of various m values are used. Differences are larger when T increases. Estimates of quantiles for the lower 80% of the data using L-moments show minimum differences as that of the ones computed using LL-moments. Parameter estimation using LL moments gives more weight to the smaller observations in the sample.

4.3 Pattern Recognition

As a common data mining technique, clustering is useful in finding suitable groupings and representatives for homogeneous groups, and in detecting unusual data objects. A cluster is therefore a collection of objects which are 'similar' among them and are 'dissimilar' to the data objects belonging to other clusters.

One can use Principal Component Analysis (PCA) to identify the total variance in the temporal or spatial variance among a time series (See Chapter 5). PCA is a statistical method aiming to express the observed data in terms of a linear combination of underlying latent variables. The latent variables are assumed to be non-Gaussian and mutually independent. The task is to identify both the latent variables and the mixing process.

Support vector machine (SVM), proposed by Vapnik (Vapnik, 1998), is one type of machine learning technique. Based on the structured risk minimization problems SVMs seek to minimize an upper bound of the generalization error instead of the empirical error in the other neural networks. SVMs have been developed to the problem of pattern recognition and with recent inclusion of ξ-insensitive loss function for non-linear regression estimation problems. SVMs are successfully used in time series forecasting problems (Cao, 2003). Application of SVM to hydrological drought prediction is provided in Section 9.1. It provides a new approach to the problem of pattern recognition (together with regression estimation) with clear connections to the underlying statistical learning theory. They differ radically from comparable approaches such as neural networks: SVM training always finds a global minimum, and their simple geometric interpretation provides fertile ground for further investigation. SVM is largely characterised by the choice of its kernel, and thus link the problems they are designed for with a large body of existing work on kernel based methods. Perhaps the biggest limitation

of the support vector approach lies in the choice of the kernel. SVMs were originally developed to perform binary classification. However, classification of data into more than two classes, called multiclass classification, is frequent in remote sensing applications. Details of multiclass classification with their merits and demerits can be found in (Watanachaturaporn and Arora, 2004).

4.4 Formation of Homogeneous Pools at Basin Scale

Various approaches have been proposed for grouping of sites that define hydrologic homogeneity. Cluster analysis technique has been used for analysis of low flow (Gustard and Gross, 1989). The method of empirical orthogonal functions (EOF) has been applied for delineation of hydrologically homogeneous regions with respect to annual drought duration series (Tallaksen and Hisdal, 1997). The EOF method is essentially a principal component analysis technique (see section on principal component analysis). Unlike cluster analysis the EOF method uses a hydrological time series for the definition of regions, thus the method does not directly support assignment of ungauged sites into regions. Assignment of an ungauged site to a pooling group requires the use of physical catchment attributes like land use, morphometry, soil type, hydrogeology and the like. Moreover, Climate conditions are also important to describe drought characteristics. Since the method is meant with a special emphasis for the case of ungauged sites, the hydrological attributes are not considered at the identification stage.

In the previous section pooled estimates of low flows were reported using some hydrological gauging stations. However, the density of the hydrometric stations in Wabi Shebele river basin is very low to substantiate the pools for the entire study area. In this section the relatively better networks of meteorological stations are used to produce time series of scPHDI as surrogates for hydrological drought. The incidence of mutual dependence of hydrological data as the consequence of nested basins should also be eliminated by the correct constitution of a sample of basins.

4.4.1 Delineation of the Hydrological Pools

An exhaustive coverage of the study area by basic spatial units (small basins) is an essential step for inclusion of ungauged sub-basins into the pooling group identified in

a sample of gauged basins and for the cartographic presentation of the delineation. The Wabi Shebele river basin is sub divided at stream order of 5 into 100 sub-basins (Figure A.5).

Data vectors containing various attributes like mean annual precipitation, average NDVI, hydrogeological features (average static water level), stream density are made for each sub-basin. As specified in the previous section, the hierarchical clustering using Wards method is used to group individuals (sub-basins) on the basis of the similarity between the data vectors. These procedures are used to identify groups of similar sites based on the Euclidean distance between each pair of rescaled data vectors. The clustering algorithms start with each site in separate clusters and combine clusters until only one is left. Although, the cutoff point is subjective, a balance has to be sought that compromises introduction of bias in quantile estimation as a consequence of large pools and diminished value of using pooling schemes as a result of small groups. As proposed in the previous section, 3 pools appear as an option at a higher clustering level (Upper, middle and lower valley). This seems mainly influenced by the dramatic variation in altitude in the basin. In order to avoid very large pools, a second higher level of 5 pooling groups is adopted (Figure 4.6). Then the identified plausible grouping of the sites has to be analyzed using the H-statistics.

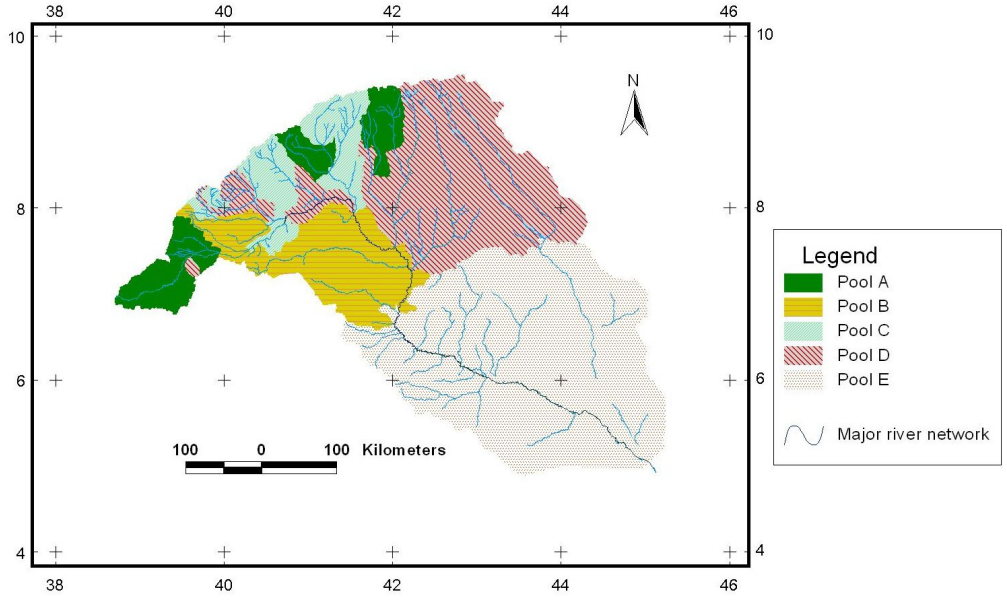


FIGURE 4.6: Hydrological pooling groups in Wabi Shebele river basin.

The time series of scPHDI averaged over the drought seasons of a year are used in

computation of L-moments at 42 stations. The hydrological pooling groups must satisfy with regard to hydrological response both the condition of intra-class similarity and that of inter-class dissimilarity.

4.4.2 Homogeneity of Pool Tests

For testing homogeneity of the pools, the discordancy measure test and heterogeneity measure test have been used.

Discordancy Measure

The discordancy measure is considered as a means of screening analysis, and the aim is to identify those sites that are grossly discordant with the group as a whole. In this test, L-moment ratio (L-coefficient of variation L-Cv, L-skewness, and L-kurtosis) of a site is used to describe that site in three-dimensional space. A group of homogeneous sites may form a cluster of such points. A point, which is far from the centre of the cluster, can be discarded. The discordancy measure, D_i , is used to screen out the data from unusual sites; and the test is applied by calculating the D-statistic, which is defined in terms of L-moments.

Let $u_i = [\tau_i, \tau_{3i}, \tau_{4i}]^T$ be a vector containing the L-moment ratios for site i (Hosking and Wallis, 1993). If the group average and the sample covariance matrix S are defined as:

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i \tag{4.27}$$

$$S = \frac{1}{(n-1)}\sum_{i=1}^{n}(u_i - \bar{u})(u_i - \bar{u})^T \tag{4.28}$$

then the discordancy measure for site i is:

$$D_i = \frac{1}{3}(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}) \tag{4.29}$$

where n is the total number of sites. When a single site does not appear to belong to the cloud of (τ, τ_3, τ_4) points on the L-moment diagram, the test of discordance shows large values of D_i. This indicate sites that are the most discordant from the group as a whole and are most worthy of investigation for the presence of data errors (Hosking and

Wallis, 1997). Two possibilities can be investigated: either there may be an error in the data, or the station may properly belong to another region or no region at all. However, it is advisable to examine the data for the sites with the largest D_i values, regardless of the magnitude of these values (Hosking and Wallis, 1993). Hosking and Wallis (1997) considered that the D_i largely depends on the number of sites in the group, and they identified critical values for it. Thus, initially, Hosking and Wallis (1993) suggested the criterion $D_i \geq 3$, but this is not satisfactory for small regions. In their recent studies, Hosking and Wallis (1997) defined the following algebraic bound for the critical D_i value: $Di < (N-1)/3$. Thus, the value of $D_i > 3$ only occur in regions having 11 or more sites.

Results of the discordancy anaylsis reveals that stations 39070253 (Hunte) and 38070013 (Koffele) are discordant from the rest of the stations in Pool A. Any possible outliers in these stations were not identified. Classifying them under another pool is even worse. So these stations are not involved in subsequent analysis of next section.

Heterogeneity Measure (H)

The purpose of this step is to form groups of stations that satisfy the homogeneity condition-that is, stations with frequency distributions that are identical apart from a station-specific scale factor. To compare the inter-site variation in sample L-moments for the group of sites with what would be expected of a homogeneous region, a test statistic H termed as heterogeneity measure is used (Hosking and Wallis, 1997). To determine what would be expected, repeated Monte Carlo simulations of a homogeneous region with sites having record lengths equal to those of the observed data are performed. A large positive value of the H-statistic indicates that the observed L-moments ratios are more dispersed than is consistent with the hypothesis of homogeneity. A number of, say 1000, data regions are generated based on the regional weighted statistics using a four parameter distribution, eg, Wakeby or Kappa distribution. This is to avoid committing to a particular two-or three parameter distribution (Hosking and Wallis, 1993). The four-parameter Kappa distribution includes, as special cases, the generalized logistic (GLO) when h = -1, the generalized extreme-value (GEV) when $h \rightarrow 0$, and the generalized Pareto (GPA) when h = +1 distributions (see Appendix B).

There are three measures of the H-statistic. The first, H_1, is the standard deviation, weighted according to record length, of the at-site L-CVs. The second measure, H_2, is the average distance from the site coordinates to the regional average on a plot of L-CV versus L-skewness. The third measure, H_3, is the average distance from the site coordinates to the regional average on a plot of L-skewness versus L-kurtosis.

Suppose that the proposed region has N sites, with site i having record length n_i and sample L-moment ratios $t^{(i)}$ (L-CV), $t_3^{(i)}$ (L-skewness) and $t_4^{(i)}$ (L-kurtosis) of the data. The test statistic is:

$$H_k = \frac{(V_k - \mu_{Vk})}{\sigma_{Vk}}; k = 1, 2, 3 \tag{4.30}$$

where μ_{Vk} and σ_{Vk} are the mean and standard deviation of the simulated values of the V respectively.

V_k is calculated from the regional data for each of the corresponding V-statistics (V_1, V_2, and V_3) as follows:

$$\begin{aligned} V_1 &= \sqrt{\frac{\sum_{i=1}^{n} n_i (t^i - \bar{t})^2}{\sum_{i=1}^{n} n_i}} \\ V_2 &= \frac{\sum_{i=1}^{n} n_i \sqrt{(t^i - \bar{t})^2 + (t_3^i - \bar{t_3})^2}}{\sum_{i=1}^{n} n_i} \\ V_3 &= \frac{\sum_{i=1}^{n} n_i \sqrt{(t_3^i - \bar{t_3})^2 + (t_4^i - \bar{t_4})^2}}{\sum_{i=1}^{n} n_i} \end{aligned} \tag{4.31}$$

The value of H (in equation 4.30) less than 1.0 indicates that the region is acceptably homogeneous. Similarly, the value of H between 1.0 to 2.0 indicates possibly homogeneous region. If, H is greater than 2.0, the region may be considered as heterogeneous.

Regional weighted L-moment ratios, parameter estimates of fitted Kappa distribution, summary statistics of simulated values and results of the heterogeniety statistics are provided in Table 4.5. The H_1 statistic indicates all the pools are acceptably homogeneous. Pool D falls marginally under possible homogeneous group using H_2 statistic, it is identified as heterogenous by H_3 statistic. The H_1 statistic has been shown to have much better power than H_2 and H_3 statistics (Hosking and Wallis, 1997). Removing the station with the largest discordancy measure in Pool D will turn the pool to acceptably homogeneous category but this achievement is at the expense of reducing the number

TABLE 4.4: Discordance analysis

Pool	Station code	Nearby town	t	t_3	t_4	D_i
Pool A	39070101	Adaba	0.2062	0.1806	0.1048	0.0977
	42090214	Alemaya	0.2347	0.1204	-0.0140	0.8693
	39070011	Assassa	0.2016	0.0483	-0.0498	0.6876
	40080244	Bedessa	0.2261	0.1911	0.0308	0.2212
	39060131	Dodolla Edo	0.2222	0.1786	0.1216	0.0877
	42080054	Fedis	0.2345	0.2444	0.1168	0.0488
	41090061	Girawa	0.2079	0.0977	0.0354	0.2938
	39070253	Hunte	0.1896	0.1296	0.3373	**3.1426**
	39070144	Indeto	0.2409	0.3928	0.2022	1.2527
	38070013	Koffele	0.3309	0.5102	0.3777	**3.3791**
	42080044	Medaga lole	0.1946	0.3034	0.2424	1.9296
	39070073	Meraro	0.1725	0.0441	0.0521	0.6780
	39070084	Sedika	0.2551	0.3624	0.1483	0.7047
	38070274	Sire	0.2254	0.1890	0.0587	0.0924
	39070031	Ticho	0.2038	0.1222	-0.0349	0.5149
Pool D	42090043	Babile	0.1876	0.0809	0.0411	0.0891
	43080053	Degehabur	0.2025	0.0194	-0.1093	1.1246
	42090164	Gursum	0.1564	-0.0375	0.0697	1.4034
	42090031	Harar	0.2222	0.4265	0.1673	1.5672
	42090011	Jijiga	0.1963	0.0340	0.0186	0.8390
	43090013	Kebri Beyah	0.2051	0.1265	-0.0176	0.2786
	40080304	Matchara	0.2074	0.0351	-0.0429	0.6980

TABLE 4.5: Heterogeneity measure

Item			Pool A	Pool D	Pool D (excluding station 42090031)
Regional weighted L-moment ratios	t		0.21616	0.20054	0.19546
	t_3		0.18177	0.12347	0.05235
	t_4		0.07273	0.02150	-0.01272
Parameter estimates of fitted Kappa distribution	ξ		1.47800	1.58900	1.68800
	α		0.92200	0.95400	0.98000
	κ		0.14990	0.27100	0.43630
	h		0.49998	0.48986	0.48978
Summary statistics of simulated values	Test 1	μ_{v_1}	0.44982	0.41429	0.36928
		σ_{v_1}	0.09008	0.07096	0.05462
	Test 2	μ_{v_2}	0.09152	0.07366	0.05813
		σ_{v_2}	0.05508	0.04100	0.03093
	Test 3	μ_{v_3}	0.06169	0.04998	0.29838
		σ_{v_3}	0.03727	0.02764	0.22157
Heterogeneity measure	H_1		-4.79664	-5.66473	-6.55965
	H_2		-0.16802	*1.07148*	-0.47841
	H_3		*1.39331*	**3.31977**	-1.03658

of stations. Considering only the H_1 statistic, thus all the pools are acceptably taken as homogeneous.

Chapter 5

Spatial Patterns of Hydrological Drought

In this chapter the spatial patterns of hydrological drought is addressed using the self calibrating palmer hydrologic severity index (scPHDI). Regional estimate of drought severity is a good tool for water management purposes in a river basin. It is specifically essential for reservoir operation and water quality management. Frequency distribution of drought variables can depict associated risks. The severity-areal extent-frequency curve of drought is illustrated in the study basin. Empirical orthogonal function is employed for dimensionality reduction of the data set. The spatial analysis is carried out using a geostatistic technique to estimate regional variables. Temporally evolving interactions in the drought time series is investigated using wavelet analysis. Latitudinal and Longitudinal variations in power wavelets are provided in Hovmoller plots. The spatial analysis shows that most of the north western portion of the basin faced mild to severe drought in 1973. The drought of 1984 is very severe in a small portion of north eastern part and mild to severe drought is observed in the eastern and southern portions of the basin.

5.1 Background

The Wabi Shebele river basin has different climatic regimes and altitudinal variation. Whereas the upper portion is mountainous the lower half of the basin is almost plain (Figure 2.2).

5.2 Data and Methodology

The data used is the scPHDI that is computed based on the net balance between water supply and demand. It is calculated using a rather complex water budget system using historical records of precipitation and temperature and the soil characteristics of the site being considered.

The scPHDI is scaled to have an extreme wet(dry) spell near two percent of the time. The scPHDI actually adjusts the value of K necessary to obtain the correct range of PDSI values (-4.0 to +4.0). It improved the performance of the PDSI by automating the calculations unlike Palmer method of deriving the empirical constants used in the PDSI algorithm. This is achieved by weighting, for every location, the climatic characteristic and the duration factors with data from only that location.

Time series of average scPHDI in Wabi Shebele river basin portrays that very severe events have occurred in the middle of the 1970's and the last half of the 1980's (Figure 5.1).

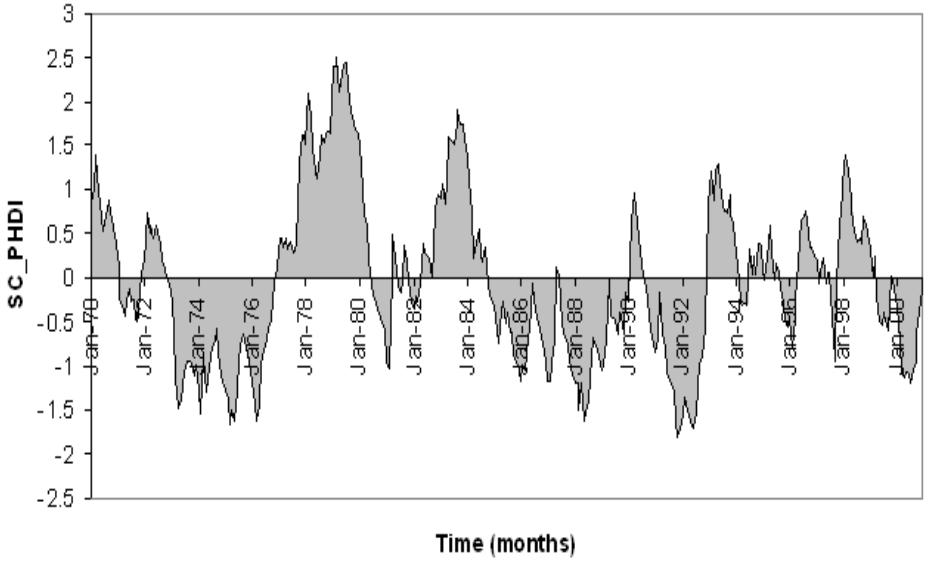


FIGURE 5.1: Time series of average self calibrating Palmer hydrologic drought index (scPHDI) in Wabi Shebele river basin.

In order to illustrate how the severity of drought has spatially affected the basin, a principal component analysis is done.

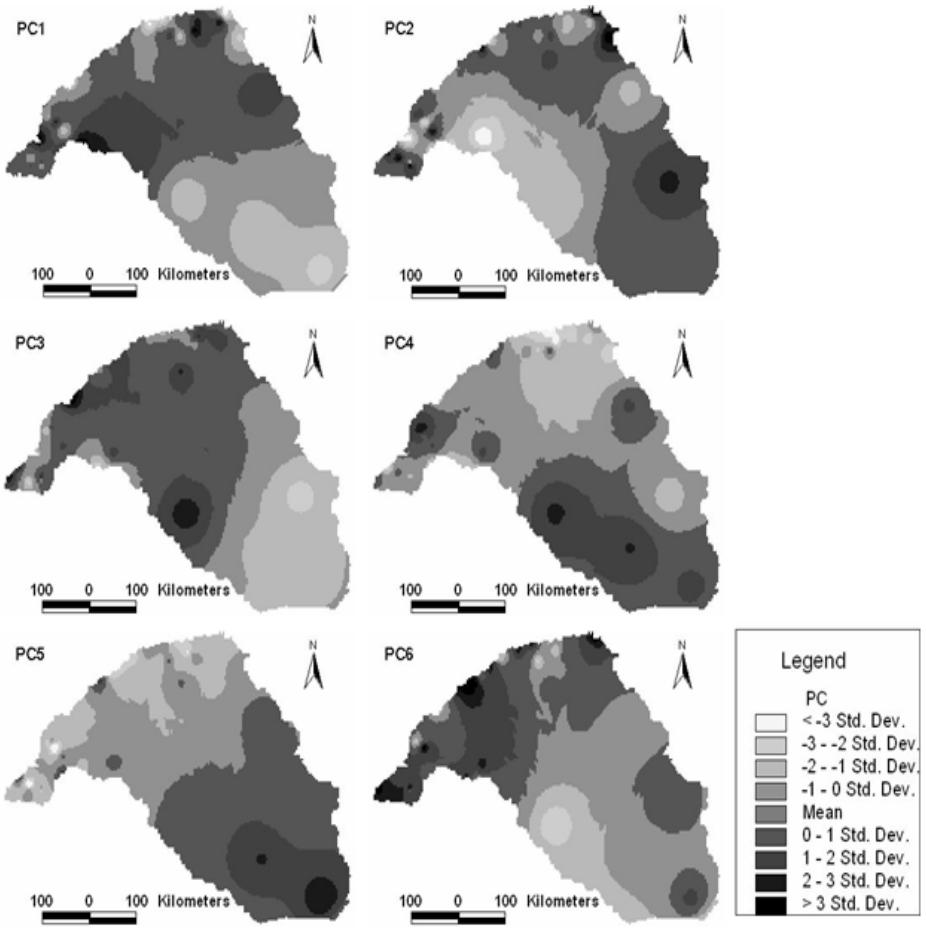


FIGURE 5.2: Spatial distribution of the first six principal components in Wabi Shebele river basin

5.2.1 Principal Component Analysis

Principal component Analysis (PCA) is a standard statistical method, often used in meteorological studies, to reduce the original intercorrelated variables in a small number of new linearly uncorrelated ones that explain most of the total variance. The new (uncorrelated) variables are called principal components (or PCs scores) and consist of linear combinations of the original variables. The coefficients of the linear combinations are called loadings' and they represent the weight of the original variables in the PCs.

In brief, this method consists in computing the eigenvalues and the eigenvectors of the covariance matrix, where the eigenvectors, properly normalized, are the loadings (i.e. the spatial patterns), while each eigenvalue tells about the fraction of the total variance explained by each loading.

The principal components of the scPHDI is computed for 56 meteorological stations in

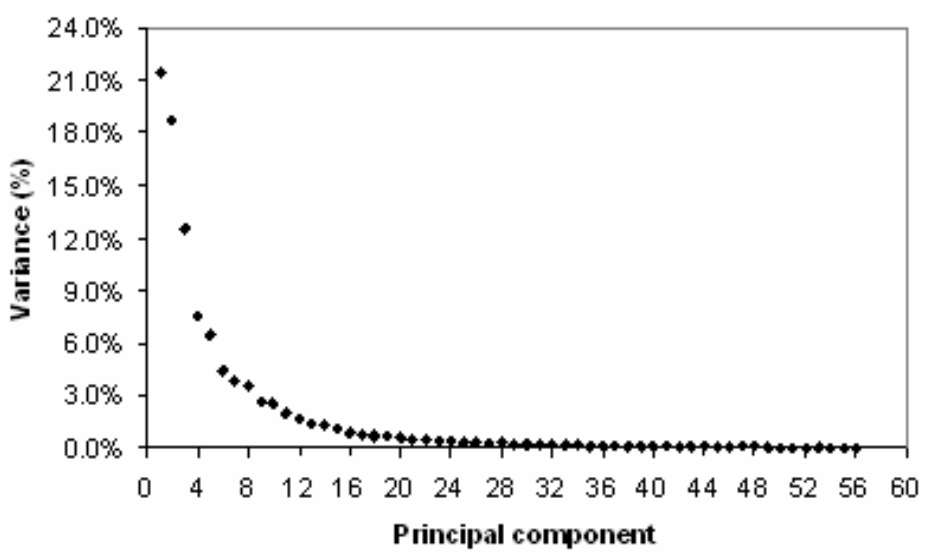


FIGURE 5.3: Percentage explained variance for the different principal components of the scPHDI series

and near the Wabi Shebele river basin. The most dominant components are spatially interpolated using kriging (Figure 5.2). The first four components make about 50% of the variability in the scPHDI series (Figure 5.3).

5.2.2 Wavelet Analysis

Customary spectral techniques applied to data do not lend themselves well to a proper understanding of the intermittent and non-stationary nature of interactions between hydrological variables. A shortcoming of standard Fourier transform is that it does not provide an accurate time-frequency localization, nor does it perform well on irregularly spaced events or non-stationary signals (Smith et al., 1998).

A major advantage of using the wavelet transform over the Fourier transform is that wavelet analysis is scale independent and hence there is no need for a predetermined scale (or response interval), which would limit the frequency range. Wavelet transformation techniques offer a better understanding of such temporally evolving interactions. Wavelet analysis was used to investigate the stability of the oscillations with time and to detect changes of oscillation frequency and strength during the investigation period. The wavelet transform is done by convolution of the time series with a scaled version of the chosen basis function (wavelet) as described in Torrence and Compo (1998).

Wavelet and cross wavelet spectral analysis offers a framework to analyze observed time series with respect to time and scale ($\tilde{1}$/frequency) and has therefore become popular in water resources research to characterize the temporal or spatial variability of hydrometeorological time series. By decomposing a time series into time-frequency space, one is able to determine both the dominant modes of variability and how those modes vary in time.

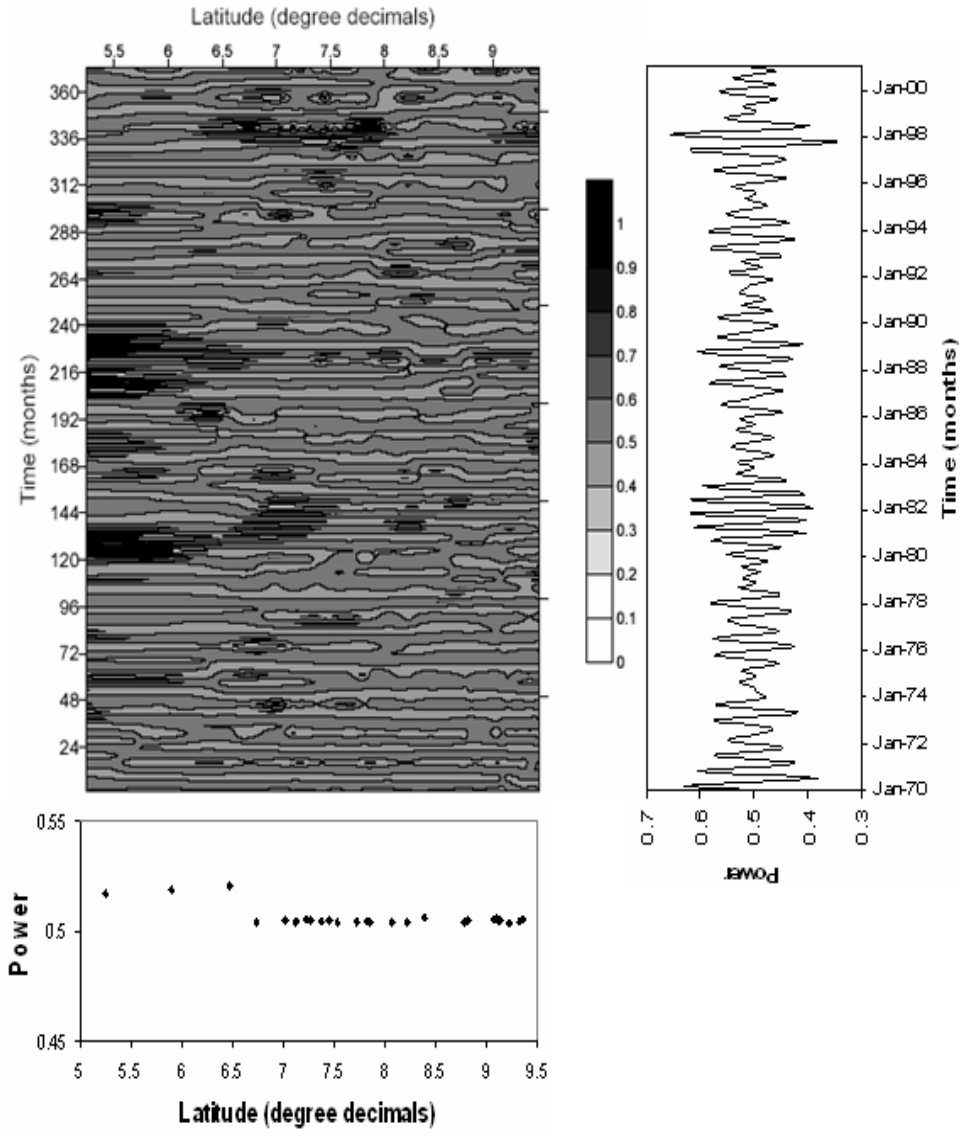


FIGURE 5.4: Latitudinal Hovmoller plot of power of wavelets of scPHDI for a time band of 6 months in Wabi Shebele

Additional insight into the spatial variability of the drought series can be gained through the examination of the latitudinal/longitudinal structure of the associated power of wavelets. The wavelet, chosen for this work was the Morlet wavelet, which is a (complex)

plane wave modulated with a Gaussian. The spatio-temporal frequency of the scPHDI latitudinally and longitudinally is further analyzed using Wavelets. First the values of the scPHDI is standardised. Then at selected time bands of 0.5 year, 1.5 year and 2.5 year, the wavelet analysis is performed on the scPHDI at 28 meterological stations in the basin.

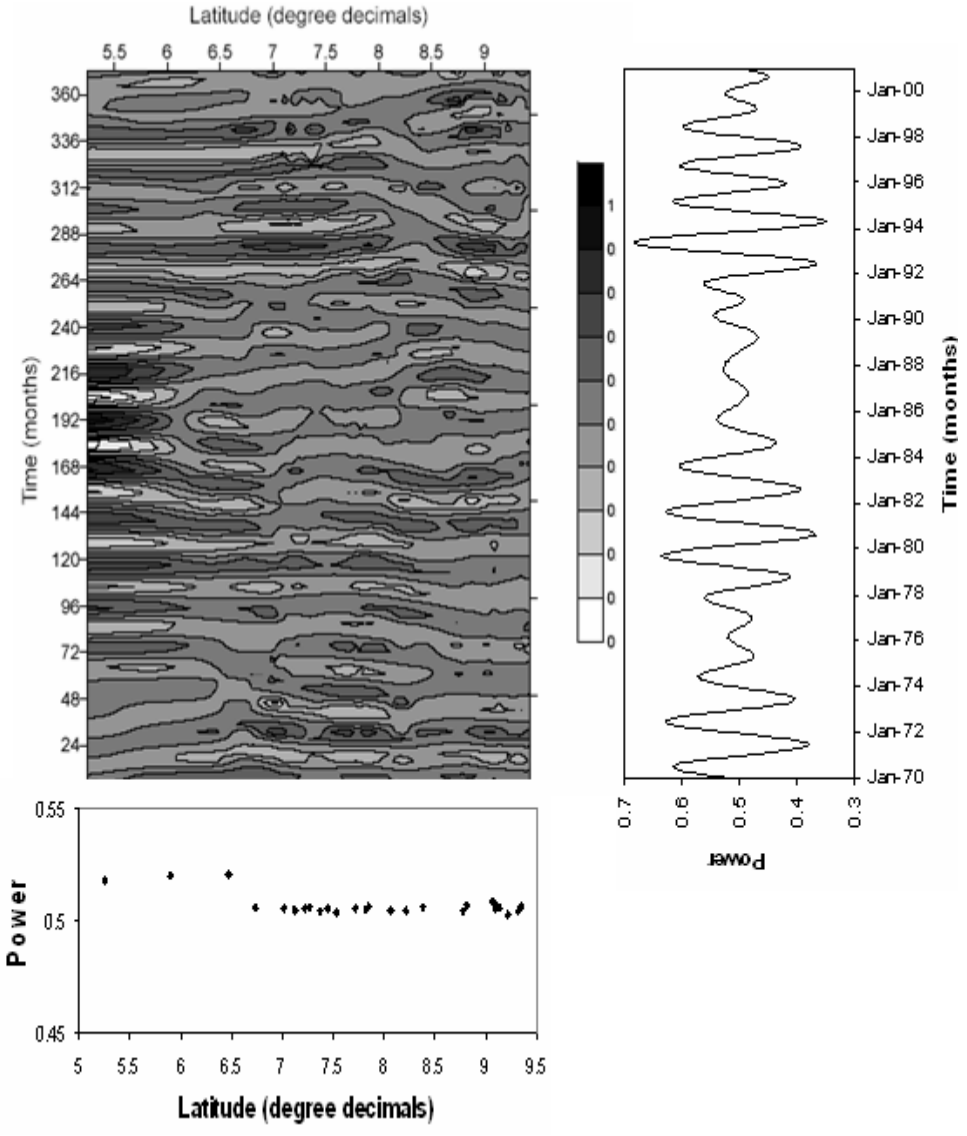


FIGURE 5.5: Latitudinal Hovmoller plot of power of wavelets of scPHDI for a time band of 18 months in Wabi Shebele

Figure 5.4- 5.6 show a power Hovmoller (time-latitude diagram) of the wavelet variance for the scPHDI at three different time bands at each latitude. Figure 5.7 - 5.9 show a power Hovmoller (time-longitude diagram) of the wavelet variance for the scPHDI at three different time bands at each longitude. Except for a moderate increase in the

middle of the decade the 1970's saw uniform fluctuation in the power of the variance in the 0.5 year band throughout the basin. That moderate increase is observed in the lower valley. At the 0.5 year band strong fluctuations are noticeable in 1980-1982 and 1987-1990 at latitudes below 6.5^0. More recently at the end of 1990's a slight increase in the strength for a short span can be seen widely around the middle valley.

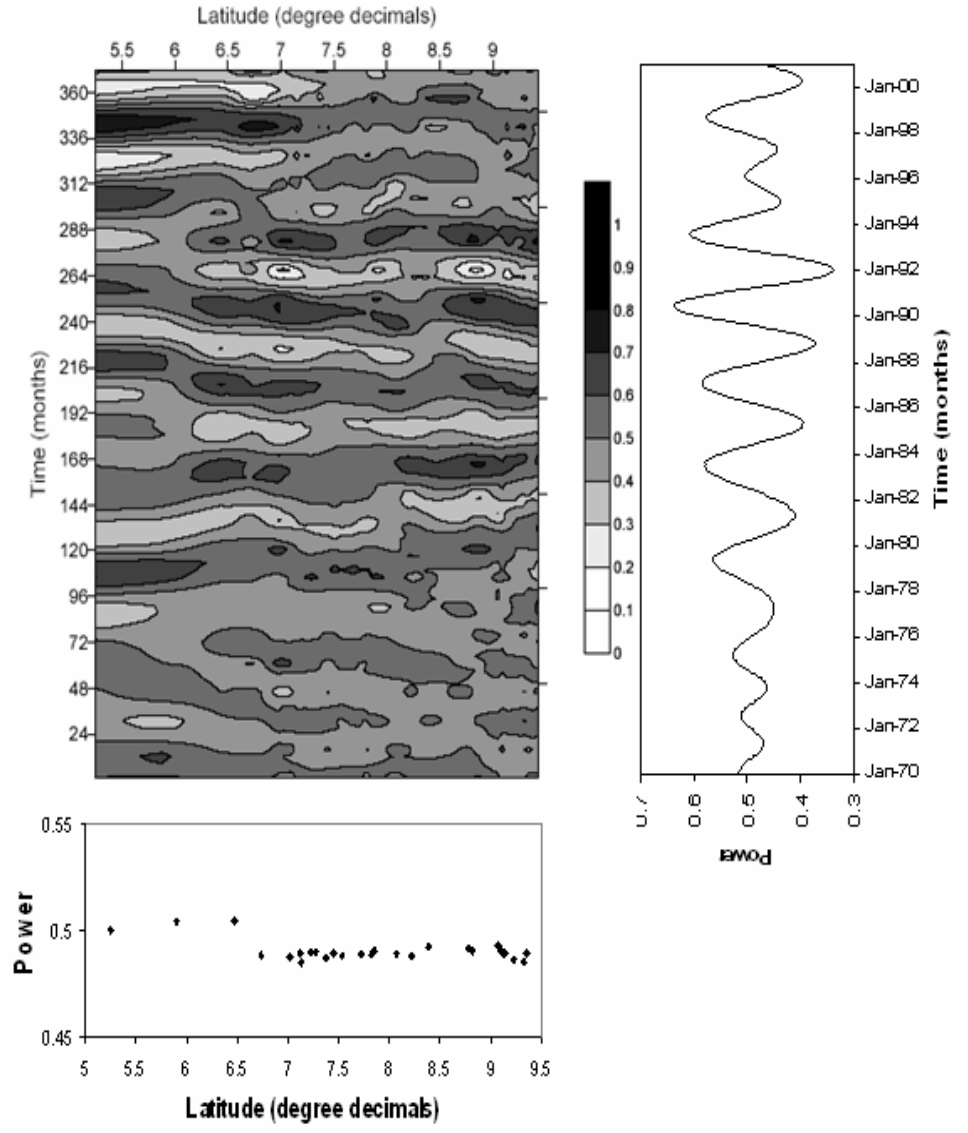


FIGURE 5.6: Latitudinal Hovmoller plot of power of wavelets of scPHDI for a time band of 30 months in Wabi Shebele

The time-averaged power as a function of latitude shows a rather abrupt change in the propagation pattern at around 6.5^0 North in all three bands. The time-latitude power hovmoller plot for all time bands show that as one traverses from North to South there is an abrupt increase in the frequency after Imi (6.28 °N). However there is no

significant variation in the strength laterally. This may suggest that the classification of hydrologically homogeneous regions can be refined by considering the space-time variability of the drought series.

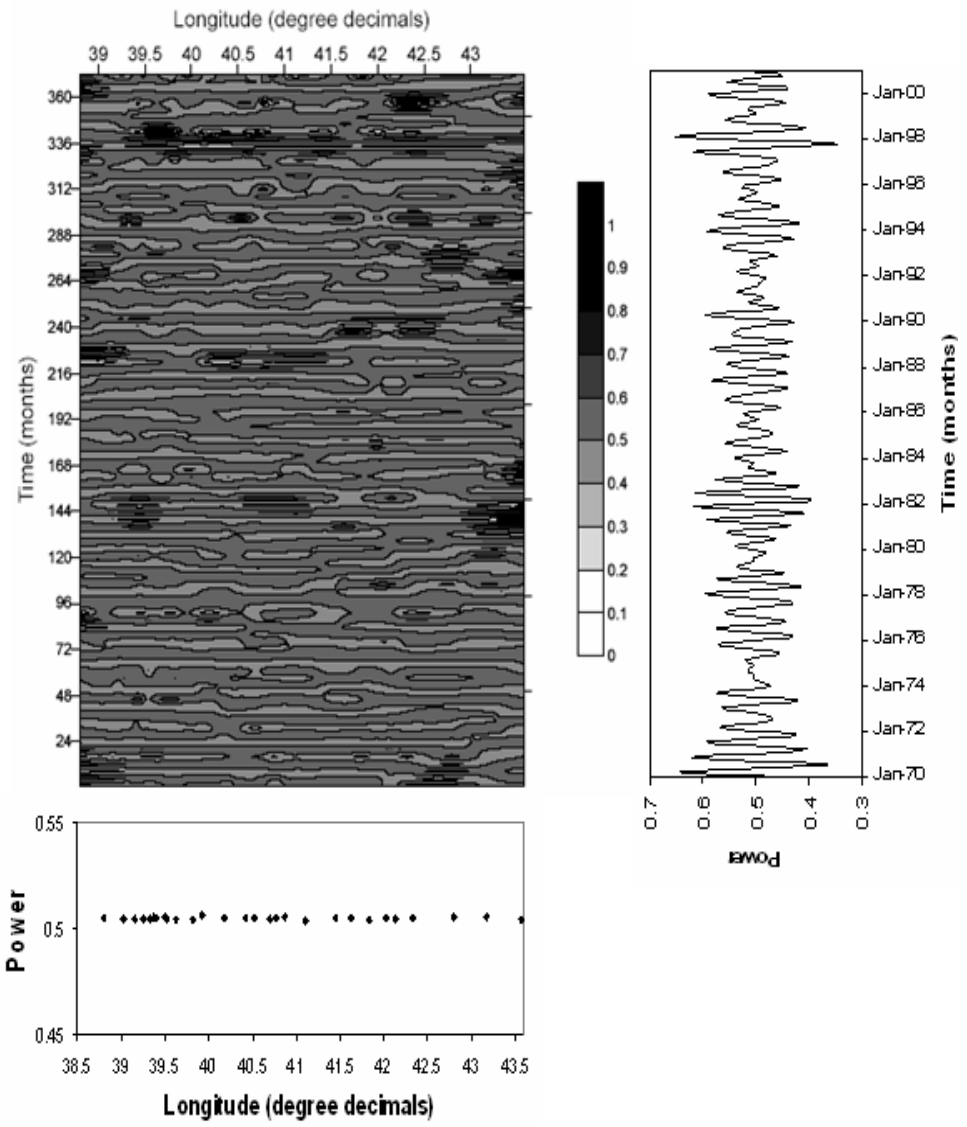


FIGURE 5.7: Longitudinal Hovmoller plot of power of wavelets of scPHDI for a time band of 6 months in Wabi Shebele

The wavelet analysis from the three time bands shows that at a time band of around 18 months there is high variability and at a time band of 30 months, it is smaller. Whereas at a time band of 6 months, its near to an average. These coincides with the expected time of the inter arrival of drought events in Wabi Shebele of 6.4 months. The identification of these periods being also of prime importance for good drought predictions,

this underlines that wavelet spectral analysis offers powerful tools for hydrological process understanding and modelling. At larger time bands the temporal frequency of the scPHDI smooths out.

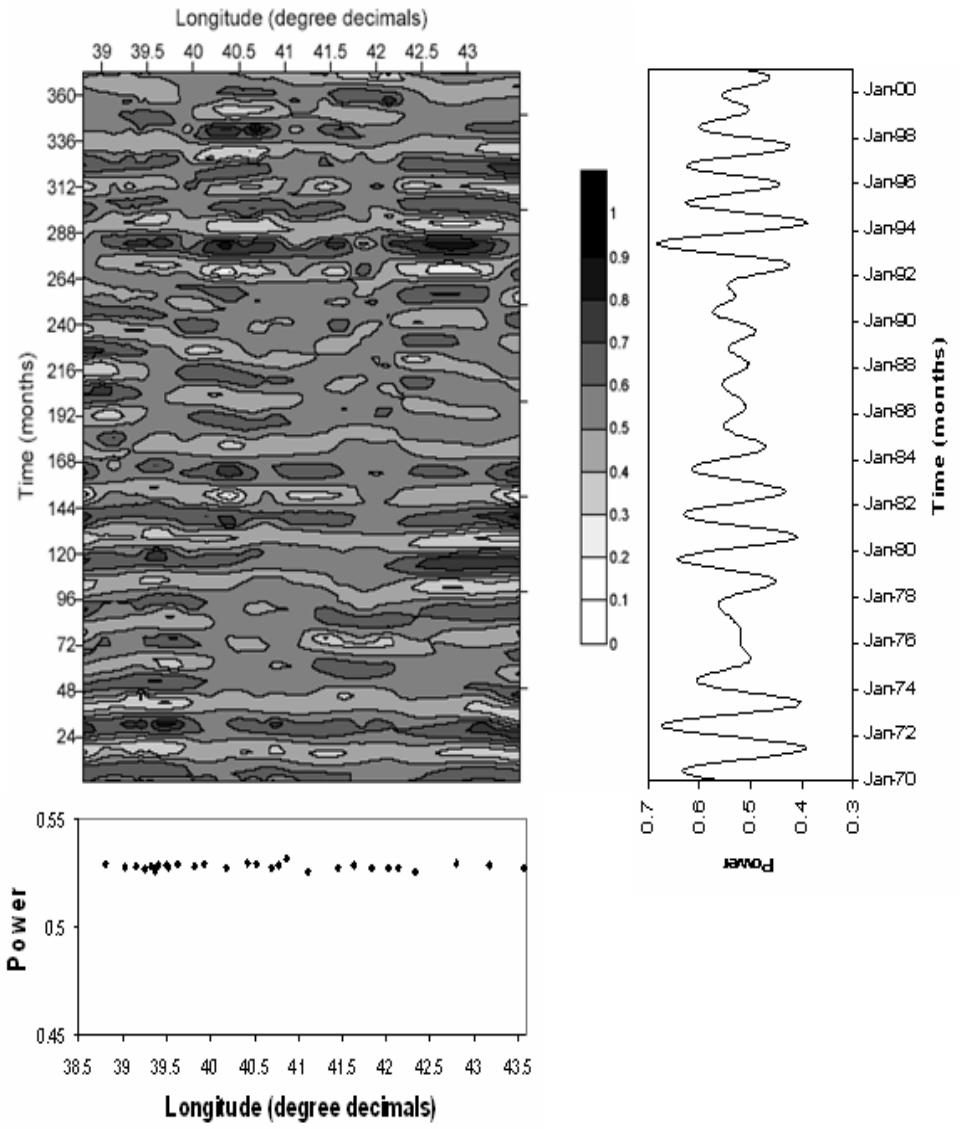


FIGURE 5.8: Longitudinal Hovmoller plot of power of wavelets of scPHDI for a time band of 18 months in Wabi Shebele

The generally low power observed in the 1.5 year band (Figures 5.5 and 5.8) between 1984 and 1991 mainly reflects a lack of power in the upstream portion and more significantly around Harerghe region,with the western areas experiencing relatively increased fluctuations at the end of the 1980s. The high power regions are observed clustered in time at the lower eastern portion of the basin where as the western area has relatively even occurrences at a time band of 1.5 years (Figure 5.8). The spatially averaged power

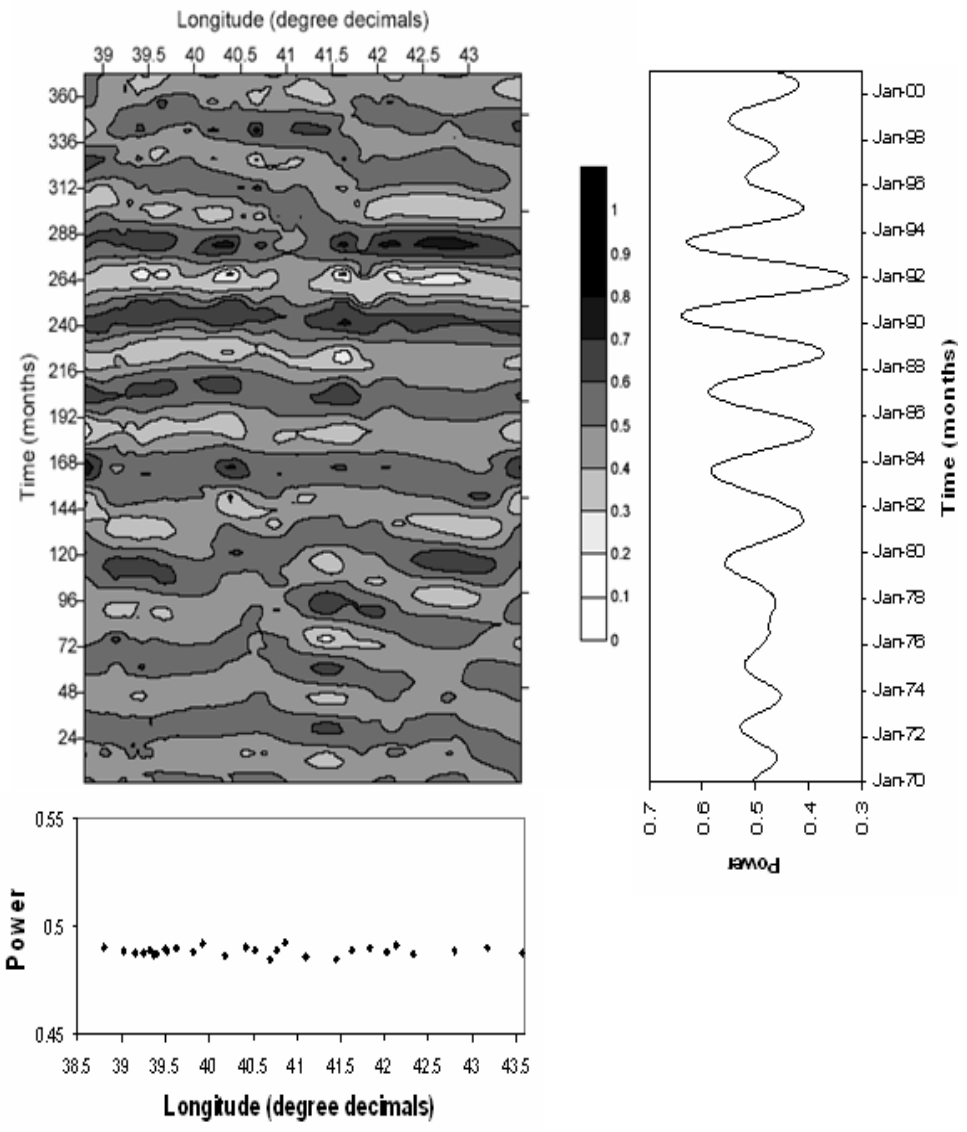


FIGURE 5.9: Longitudinal Hovmoller plot of power of wavelets of scPHDI for a time band of 30 months in Wabi Shebele

shows an increase in high frequency fluctuations at the beginning of the 1980's and with larger amplitude around end of the 1990's. The low-frequency fluctuations were stronger at the beginning of the 1990's. High amplitude power is seen around the middle of the 1990's in the 1.5 year band.

5.3 Area Coverage of Recent Worst Droughts

The area coverage of severe and extreme droughts is variable in different months. A close look at the areal distribution of these events from 1970-2000 shows how the events

progress spatially. Figure 5.10 shows the evolution of the percentage of the study area affected by droughts of varying severity.

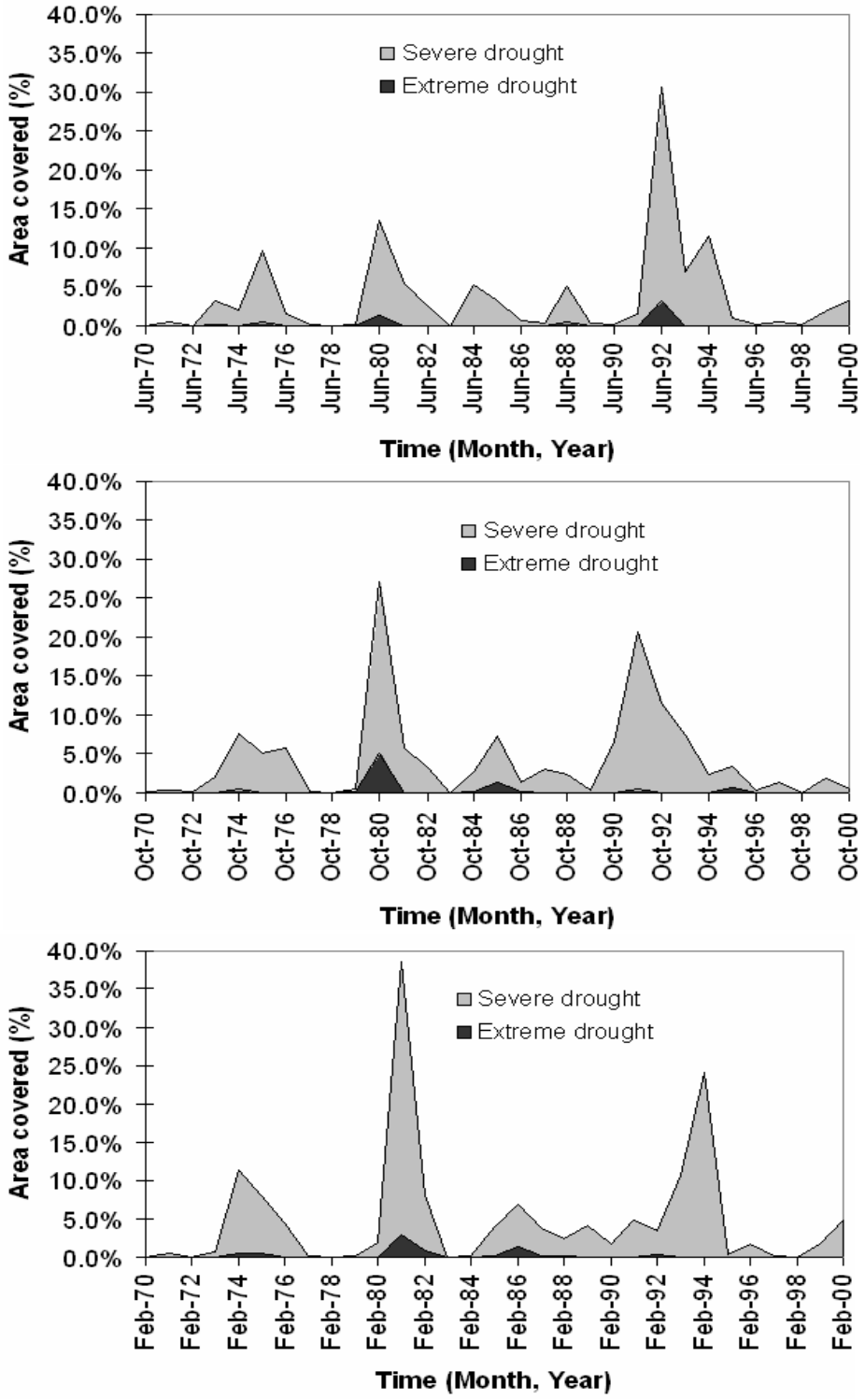


FIGURE 5.10: Area covered by severe and extreme droughts from 1970-2000

Severe drought events hit as wide as 40% of the total area of the basin during the

beiginning of the 1980‘s. The severe drought events persisted on relatively smaller area for quite long time between the middle of the 1980s to the middle of the 1990’s. Extreme drought events occurred longer at the beginning of the 1980’s covering as much as 5% of the total area.

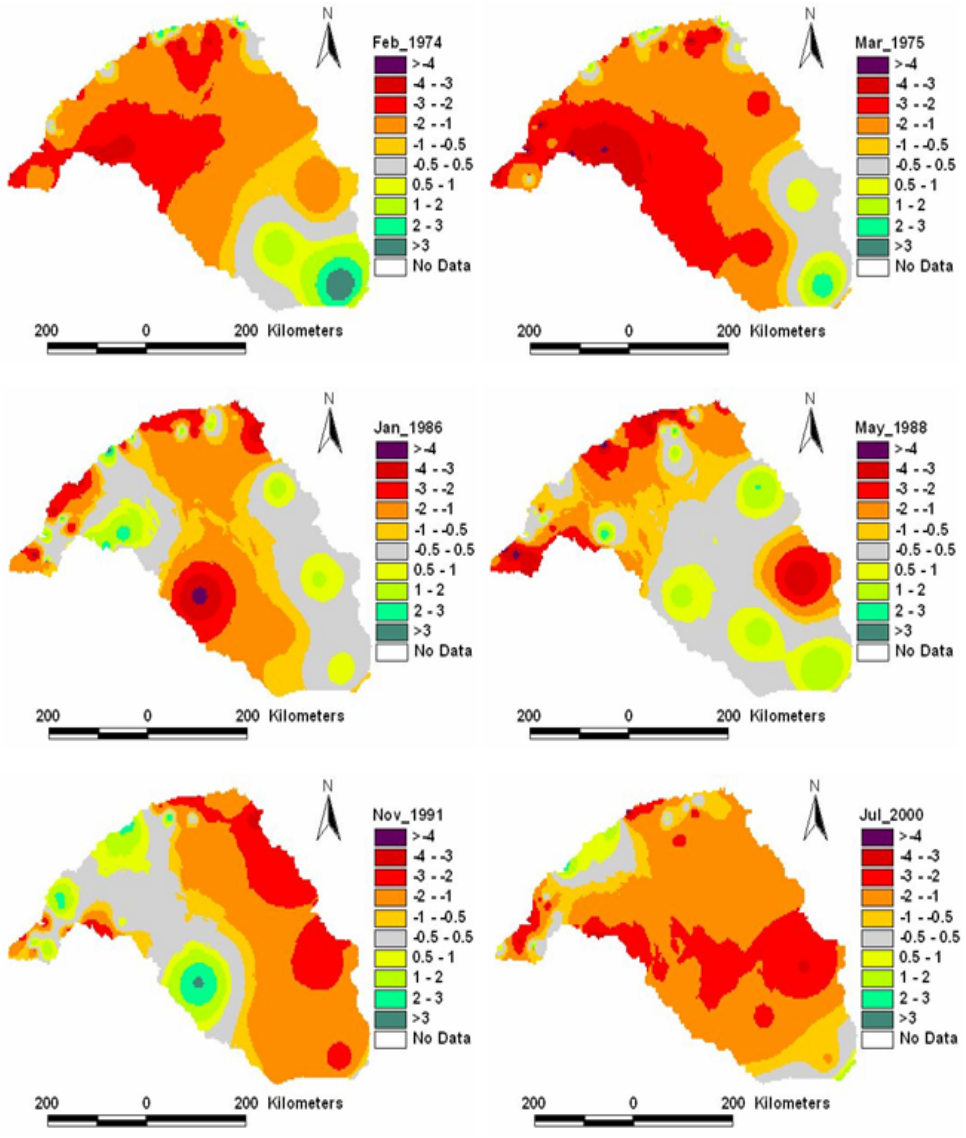


FIGURE 5.11: Spatial patterns of the worst droughts in the last four decades in Wabi Shebele

This observation indicates that drought severity can be highly diverse spatially and extreme drought can be recorded in some areas while in other places it may have a moderate or severe strength. It is essential to note that drought is spatially very complex because the changes in precipitation variability are more local than other elements such as temperature (Eder et al., 1987). Consequently, detailed spatial scales are required to

perform studies on drought. The notorious droughts which claimed many lives in the last 3 decades are spatially shown in Figure 5.11.

The drought that occurred recently in 2000 has affected an area more than 75% of the total basin area with an intensity above moderate scale. Very intense drought hit relatively larger area in 1985/1986. The devastating drought that erupted at the beginning of the 1970's and continued to the mid 1970's combines vast coverage, duration of drought and intensity. In the mid 1992 a very intense drought was widely observed througout the basin but it lasted for a shorter duration compared to the one that hit the area in the mid 1970's.

Analysis of the persistence of intensity of the drought events from 1970-2000 indicates that moderate and severe droughts are widely persistent throughout the river basin. Extreme drought events are very highly persistent at the mid uppermost portion of the river basin (Figure 5.12).

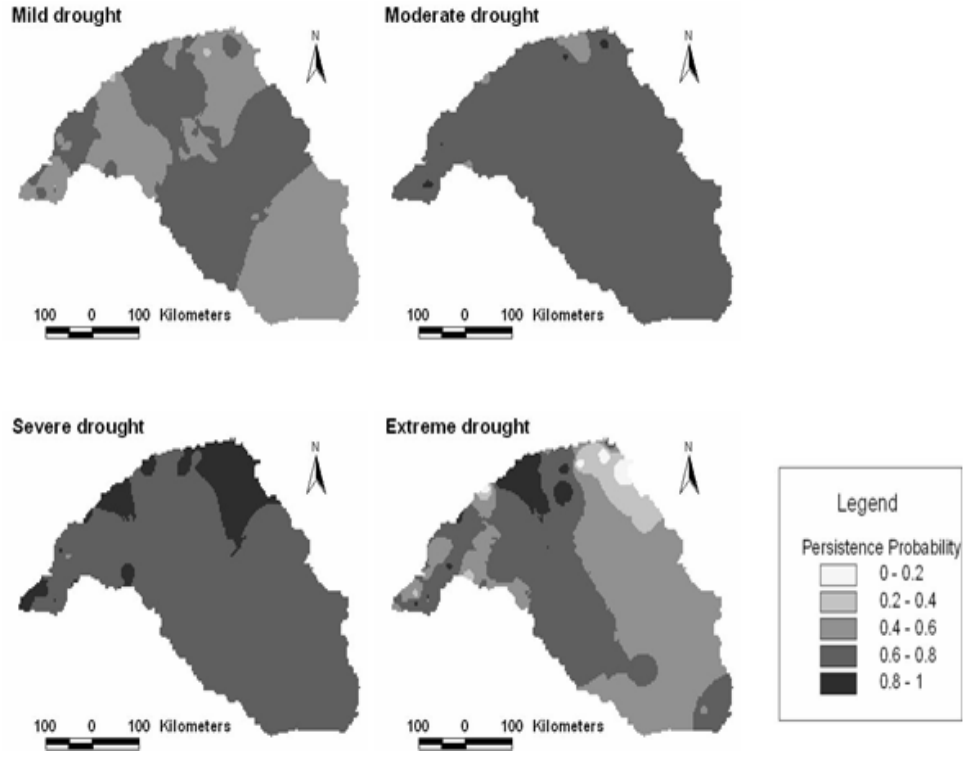


FIGURE 5.12: Persistence probability of mild, moderate, severe and extreme droughts

5.4 Severity-area-frequency

In this section the severity-area frequency curves for each of the hydrologically homogeneous pools identified in chapter 5 is presented. The frequency analysis methods do not predict the future with certainty, but they do offer good models for explaining and making efficient use of the extreme events that had occurred in the past. In the identification of the pools annual average scPHDI was used to enable use of established homogeneity tests using parametric techniques of L-moments. Here frequency analysis of seasonal scPHDI below the threshold of drought (i.e.scPHDI≤ -0.5) is performed at 56 meteorological stations in and near the basin.

In practice, the assumption of independence breaks down when the sampling frequency of hydro-meteorological data is quite high. In such a case, any particular observation is not independent of its previous observations thus alternate approaches to the classical frequency analysis should be used.

Khaliq et al. (2006) briefly review the techniques that have been presented to explicitly take into account the dependence in observations when fitting distributions to empirical observations. These techniques are reviewed under three categories: (i) the decorrelation approach, (ii) Lettenmaier's technique of pruning the data sample and (iii) the probability density estimation by wavelets and kernels. The decorrelation and Lettenmaier's approaches are useful for handling the problem of dependence in observations but their applicability is restrictive, e.g., the former is only suitable under the assumption of an AR(1) process and the latter involves a reduction of the sample and that may not always be justified. The time series of monthly scPHDI may not be an AR1 process (see Section 8.2.1). In this study nonparametric kernels are used.

5.4.1 Kernel Density Estimation

As opposed to conventional methods, nonparametric distribution fitting does not require the somewhat subjective choice of parent distribution. A nonparametric kernel estimation model, as Guo et al. (1996) reported, yields smaller bias and root-mean-square error in low flow quantile estimates, so is a viable alternative to the Weibull models in applications to real data. Kernel density estimation entails a weighted moving average of the empirical frequency distribution of the data. It smooths out the contribution of

TABLE 5.1: Some types of Kernel functions (Silverman, 1986)

Kernel	$K(x)$
Epanechnikov	$K(x) = 0.75(1 - x^2),\ \|x\| \leq 1$ $K(x) = 0,\ otherwise$
Triangular	$K(x) = 1 - \|x\|,\ \|x\| \leq 1$ $K(x) = 0,\ otherwise$
Gaussian	$K(x) = \frac{1}{(2\pi)^{1/2}} exp(-x^2/2)$
Rectangular	$0.5,\ \|x\| \leq 1$ $K(x) = 0,\ otherwise$

each observed data point over a local neighbourhood of that data point. The extent of this contribution is dependent upon the shape of the kernel function adopted and the width (bandwidth) accorded to it. Suppose the kernel function denoted as K and its bandwidth by h, the estimated univariate density at any point x is:

$$\hat{f}(x) = \sum_{i=1}^{n} \frac{1}{n \cdot h} K\left(\frac{x - x_i}{h}\right) \tag{5.1}$$

where there are n sample data x_i. K( ) is a kernel function that must integrate to 1, and h is the bandwidth that defines the locale over which the empirical frequency distribution is averaged. The kernel function K is usually chosen to be a smooth unimodal function with a peak at 0. There are various choices among kernels Table 5.1.

The Gaussian kernel function, a popular and practical choice, is used here:

$$K(x) = \frac{1}{(2\pi)^{1/2}} exp(-x^2/2) \tag{5.2}$$

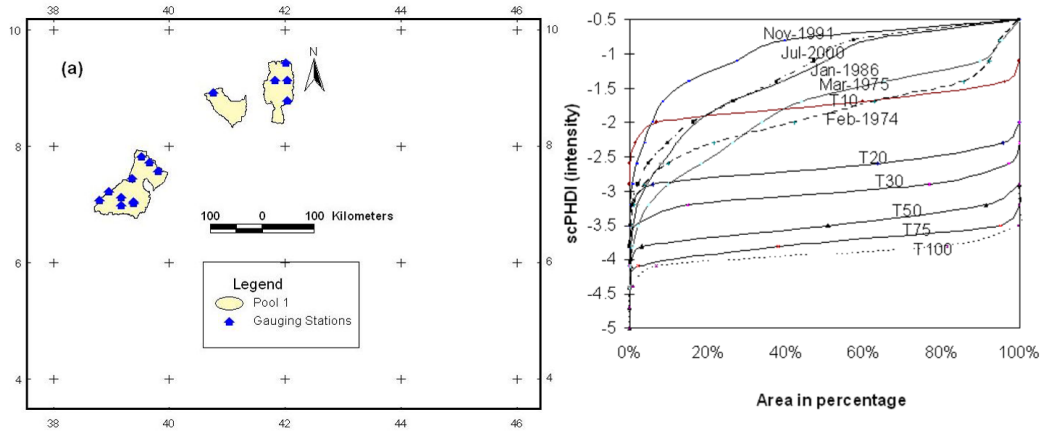
The choice of kernel bandwidth controls the smoothness of the probability density curve. Small values of h lead to very spiky estimates (not much smoothing) while larger h values lead to oversmoothing. A common method to choose the optimal bandwidth is to use the bandwidth that minimises the AMISE (Asymptotic Mean Integrated Squared Error). When a Gaussian kernel K is used, the evaluation of the optimal formula for h yields (Silverman, 1986)

$$h_{d,opt} = \left(\frac{4}{d+2}\right)^{1/(d+4)} \hat{\sigma} n^{-1/(d+4)} \tag{5.3}$$

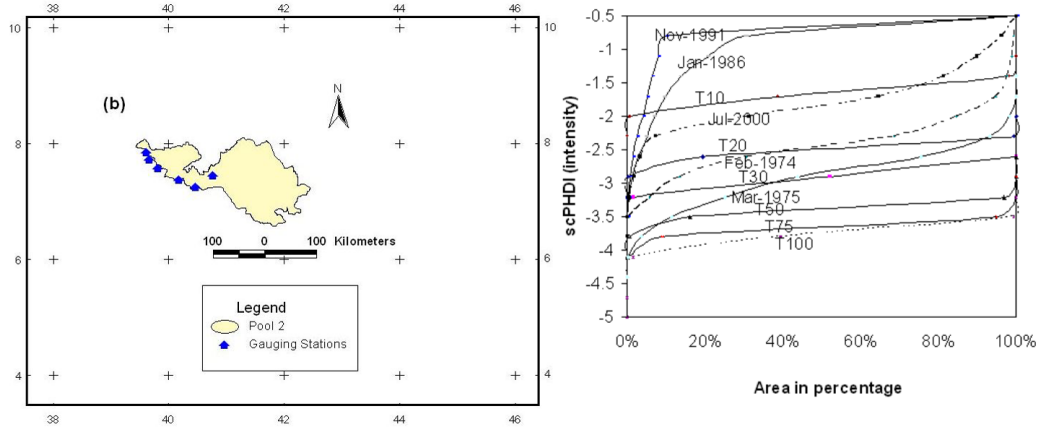
where $h_{d,opt}$=optimal bandwidth; $\hat{\sigma}$ denotes the standard deviation of the distribution in dimension d. In the univariate case this reduces to $h = 1.06\hat{\sigma}n^{-1/5}$.

5.4.2 Results and Discussion

Quantile estimates corresponding to various univariate return periods are computed at each of the stations considered. Then the values are spatially interpolated for each of the hydrologically homogeneous pools identified in the previous chapter. The results are shown in Figure 5.13 and 5.14. The integration of the area bounded by consecutive curves would yield the area weighted severity corresponding to each return period.



(a) Pool A

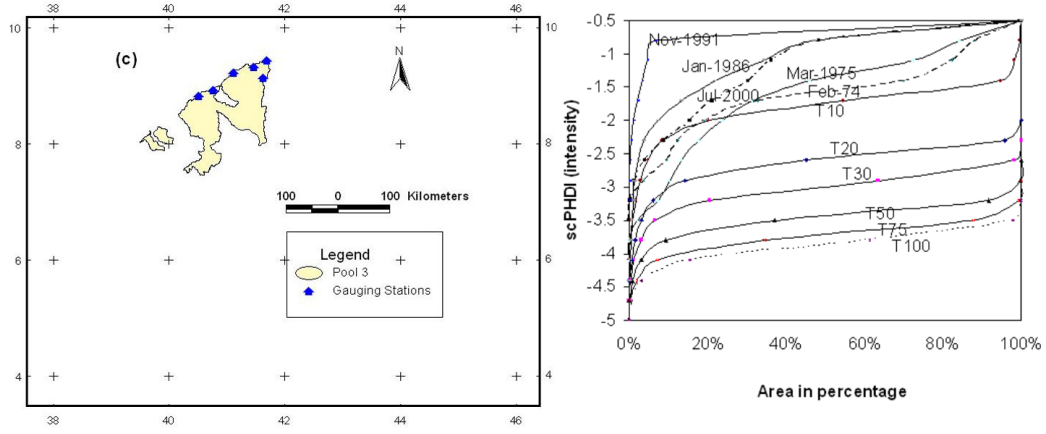


(b) Pool B

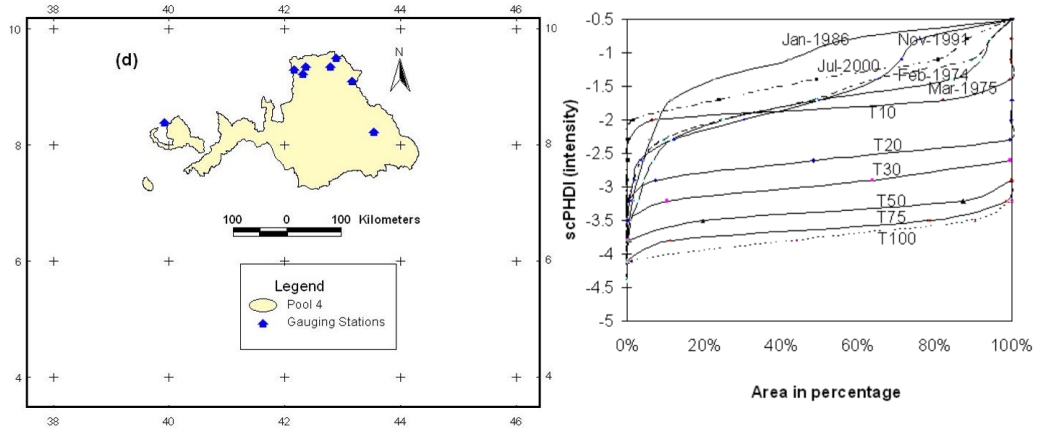
FIGURE 5.13: Severity-area-frequency curve of scPHDI at Pooling groups A and B

In scanning Figures 5.13 and 5.14 it can be observed that a drought of a comparable severity to that of 1991 drought have high frequency in most of the pooling groups. The 1991 drought has a similar coverage in pooling groups B and C. However, the notorious 1974/75 drought has higher return periods in Pooling group B than all the other regions.

It can be seen that the observed recent worst droughts show varied strengths within each pool. This entails to further differentiate each pooling group into smaller homogeneous unit. But further reducing the size of the pools has the disadvantage of reducing the available number of stations within a pooling group. Pooling group E which has the largest area coverage among the regions considered has only 4 operating stations with long records. Thus it is not considered in the Severity-area-frequency analysis.



(a) Pool C



(b) Pool D

FIGURE 5.14: Severity-area-frequency curve of scPHDI at Pooling groups C and D

Chapter 6

Paleohydrology

In this chapter main features of the proxy data used for streamflow data reconstruction and the methods are explained.

6.1 Background on Proxy Data

Proxy data gives a better opportunity to enhance data collected through instrumental records. Extending the record of climate variability into the distant past is necessary to better evaluate the modern instrumental record (Barnett et al., 1996). There are different sources of paleoclimatic/paleohydrologic data (Figure 6.1).

Among the various sources of paleoclimatic information available for the last thousand years, tree-ring chronologies have the advantage of being continuous, well replicated, exactly dated to the calendar year, and therefore easily comparable to instrumental records (Fritts, 1976).

6.2 Dendrochronologic Records

Dendrohydrology-tree ring reconstruction- utilizes tree ring data to study hydrological problems such as past variations in streamflows, lake levels and flood history. Tree growth can be used as proxy for streamflow because many of the same climatic factors that influence tree growth also influence annual stream flow, particularly precipitation and evapotranspiration. Hydroclimatic reconstructions are generated by calibrating tree

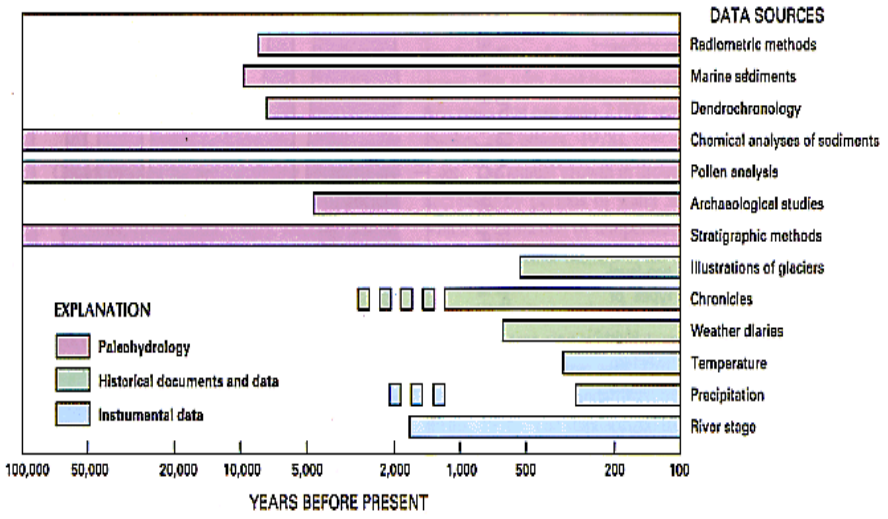


FIGURE 6.1: Sources of data used to reconstruct climatic or hydrologic conditions. (Modified from Liebscher 1987).

ring data with an instrumental time series, such as a gage record, typically using a linear regression model with tree ring chronologies as the predictor variables and the instrumental record to be estimated (Meko and Graybill, 1995; Meko et al., 2001). Reconstructions tend to express a narrower range of variability than do the instrumental data, a function of the statistical processing. Most often, but not always, dry extremes are more accurately expressed than wet extremes, so that the reconstruction is more reliable for dry extremes (Woodhouse and Meko, 2002). Core and disk samples of tree rings were collected using Incremental corer and disk sampler in Wabi Shebele basin. Streamflow reconstruction was carried out. Cross product, linear regression and reduction of error test were all statistically significant.

Sampling Sites

Sampling sites were selected in the upstream portion of the Wabi Shebele basin between 6° N to 9° 45'N latitude and 38° 45'E to 42° E longitude. Presence of old trees belonging to species which have datable rings is one of the criteria in selection of the trees. Wood cores 5.15 mm thick were cored using a Swedish increment borer 40 cm long and with 3 thread. CORIM Maxi with measuring range 0-450mm, accuracy 0.01mm and magnification up to 7X is used to measure to tree rings. Trees with a best combination of age and trunk health are cored mostly. Trees with a large stem diameter, large branches,

TABLE 6.1: The number and type of different tree species cored and disk sampled and experiences in the tropics on respective species.

Species	Core	Stem disk	Previous work (Tropics).
Podocarpus falcatus	5	6 (2)	(February and Stock, 1998) South Africa
Acacia tortilis	7	3(2)	(Maingi, 1998) Kenya
Eucalyptus globulus	9	2(2)	(Macfarlane et al., 2004) Australia
Juniperus procera/ Juniperus -	3/(-)	3(2)/3(2)	(Couralet et al., 2005) Ethiopia
Commiphora monoica	5	-	-
Erythrina brucei	2	-	-

dominant position among surrounding trees are some of the criteria to select. A number of core samples and disk samples were taken from Juniperus procera, Acacia tortilis, Eucalyptus globulus, Podocarpus falcatus, Erythrina brucei and Comifora monoica tree species Table 6.1. A total of 17 stem disks (with 2 replicates) and 31 cored tree ring samples from different tree species were collected.

6.3 Reconstructing Streamflows[1]

All increment cores were glued to wooden mounts, sanded and polished until the smallest rings were clearly visible. Ring counting gave a first estimate of tree age. Ring widths were measured to the nearest 0.01 mm. Cross-dating was ascertained by skeleton plots (Stokes and Smiley) and numerical techniques. A cubic smoothing spline with a 50% cutoff frequency of 20 yr was fit to each ring-width measurement series to remove non-climatic growth trends related to age, size, and the effects of stand dynamics (Cook et al., 1990). The application of the cubic smoothing spline in dendroclimatology has been described by Cook and Peters (1981). It ". . . can be thought of as a concatenation of cubic polynomial segments that are joined together at their ends. . . .". Splines are inherently superior to polynomials for approximating functions that are disjointed or episodic in nature because they are calculated piecewise. The index series for each sample were prewhitened with low-order autoregressive models to remove any residual

[1] A.A. Awass and G. Foerch (2007) Reconstructing Hydrological Drought from proxy data: the case of Wabi Shebele river basin. FWU Water Resources Publications, 6, 68-75

persistence not related to climatic variation that remained after standardization. The indices from individual cores were then averaged into a master chronology for each combination of site and species using a biweight robust estimate of the mean (Cook, 1985). Because visual cross-dating and computer-based quality control showed a strong similarity among the sites, the Juniperus procera ring-width series were combined to form a single chronology.

$$\bar{I}_t = \frac{\sum_{i=1}^{N} ln\,(w_t - y_t)_i}{N} \tag{6.1}$$

where $\bar{I}$ is average tree ring index (value of the master chronology), t is calendar year, w is ring width measurement in 0.01 mm precision, y is the growth trend, simulated by a cubic spline with 50% variance reduction at a frequency of one cycle per 20 years, i specimen, N is number of specimens available for year t, with $N > 5$.

A logarithmic transformation is made as shown in equation 6.1 to obtain a homoscedastic time series.

The rings in the species of Acacia lack clarity. This may be because of Acacia species possessing relatively high density wood (commonly more than 1.0 g/cm^3) and the presence of gum (Gourlay, 1995). In case of the Eucalyptus species it was very poorly identifiable. Growth of Eucalyptus globulus was largely independent of topsoil water content, giving it the potential to cause substantial dry-season groundwater depletion (Fritzsche et al., 2006). This may partly be explained due to deep root nature of eucalyptus species that enable them to reach a more permanent water table so making it more insensitive to current rainfall. The samples collected from Juniperus procera showed distinct rings. There are some missing rings in some of the samples and double ring pattern is also identifiable. This may be due to the bimodal nature of the rainfall. It is necessary to verify if the double rings occurred in those seasons where bimodal nature of the rainfall is acute. Most of the samples collected indicate young age trees less than 30 years. A relatively long data sample for the Adaba-Dodolla area was obtained from previous work at FEM, Wageningen (Couralet et al., 2005).

The lowest ring widths are observed at the beginning of the 1980's (Figure 6.2) which corresponds with the timing of the devastating drought in the country that claimed many lives and properties. Out of the set of tree ring widths measured a bootstrap

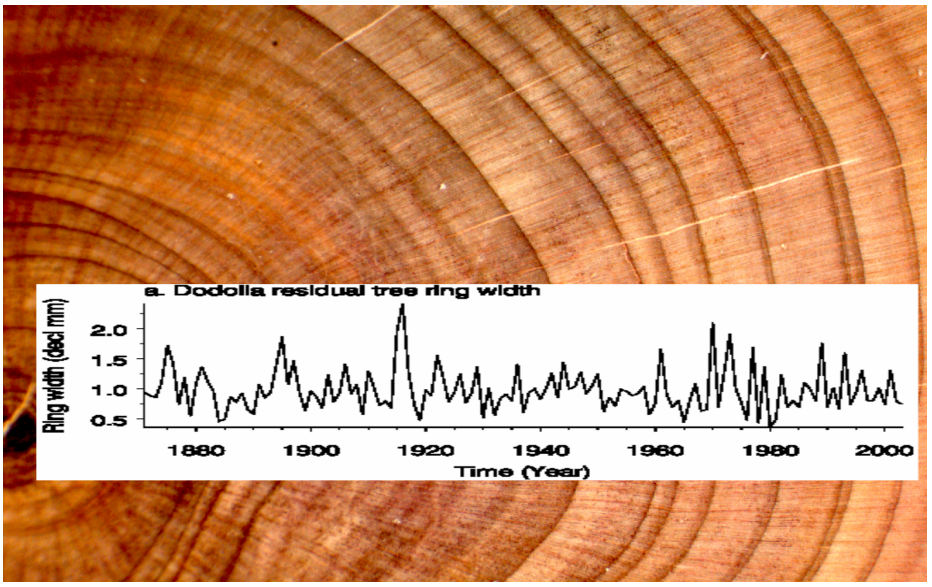


FIGURE 6.2: Residual tree ring series established from Juniperus procera at Adaba-Dodolla (data from Couralet et al. 2005)

sampling analysis was performed to generate more samples with replacement (Figure 6.3). Then the mean tree ring width was estimated to lie between 1.37 mm and 1.95 mm with significance of 0.05.

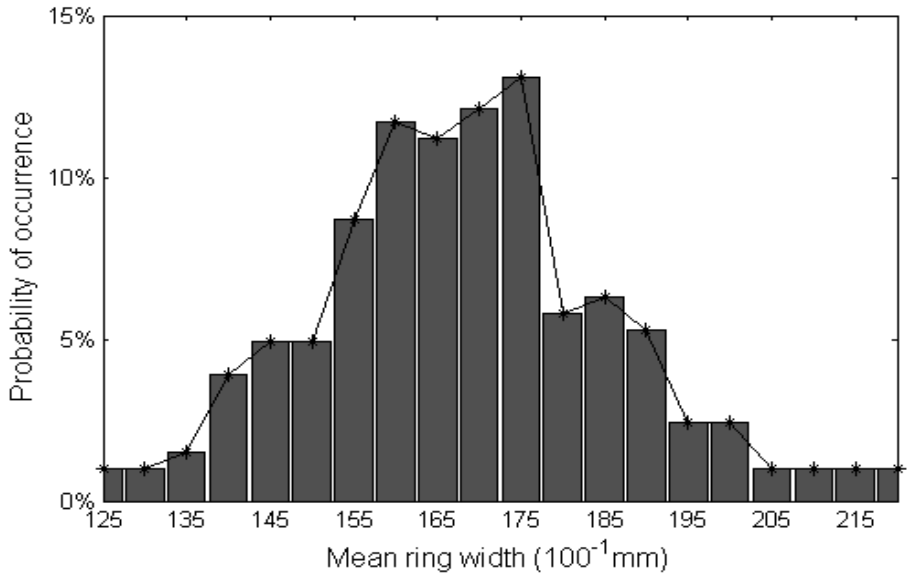


FIGURE 6.3: Probability of occurrence of the tree ring widths using Bootstrap sampling at Adaba- Dodolla (9 time series from Juniperus procera spp for common years 1945-2000)

Predominant cyclicity of the pattern in the tree ring series is noticed in a period of about 2, 3 and 7 years; the major peak being at around 3 years. There is a minor cyclicity

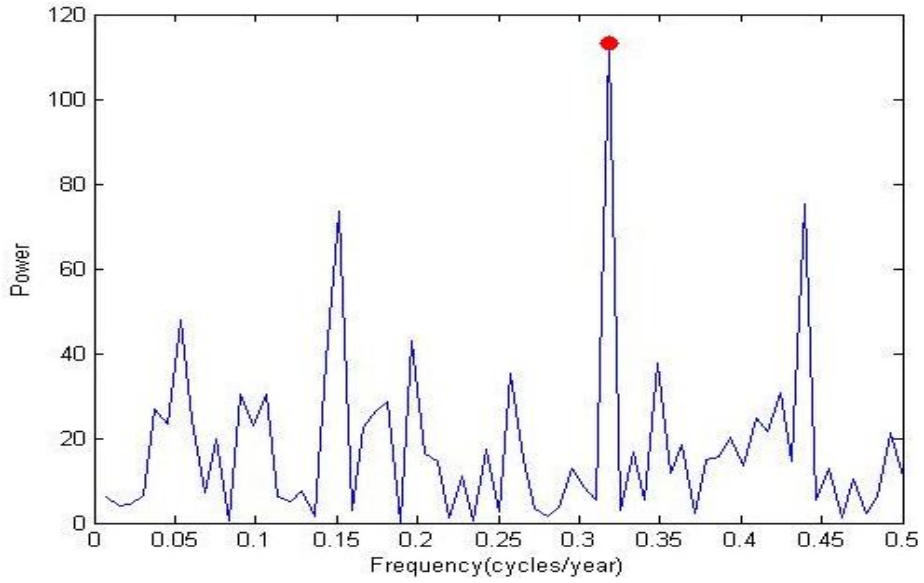


FIGURE 6.4: Periodogram of the residual tree ring widths at Adaba Dodolla in Wabi Shebele river basin.

at 20 years period (Figure 6.4). In order to identify the particular years in which this cyclicity are common, a wavelet analysis is performed using Morlet wavelet. It shows an increase in frequency in recent years. This is clearly discernible in the wavelet analysis (Figure 6.5).

A wavelet is a function with zero mean and that is localized in both frequency and time. One can characterize a wavelet by how localized it is in time ($\Delta\, t$) and frequency ($\Delta\, w$ or the bandwidth). The classical version of the Heisenberg uncertainty principle tells us that there is always a tradeoff between localization in time and frequency. Without properly defining $\Delta\, t$ and $\Delta\, w$, we will note that there is a limit to how small the uncertainty product $\Delta\, t\ \Delta\, w$ can be. A criticism of wavelet analysis is that the choice of wavelet is somewhat subjective. However as noted by Torrence and Compo (1998) the same is true for traditional transforms such as Fourier, Bessel, Legendre, etc.

Wavelet transform is a powerful way to characterize the frequency, the intensity, the time position, and the duration of variations in a climate data series (Jiang et al., 1997), which reveals the localized time and frequency information without requiring the time series to be stationary as required by the Fourier transform and other spectral methods. Jiang et al. (1997) use the wavelet transform with the 'Mexican hat' functions to analyze the runoff and the water level datasets. Usage of the wavelet transform in the study of climatic changes and hydrological changes and other fields is receiving an increasing

attention. Nakken (1999) applied the continuous wavelet transforms (CWTs) to detect the temporal changing characteristics of the precipitation and the runoff processes, and their correlations and separating roles of climatic changes caused by human activities on stream flow changes. In this research CWTs was used to detect and isolate patterns across temporal scales of the dendrochronologic records.

A physical signal should be independent of the choice of wavelet. Lau and Weng (1995) recommend that the analysing wavelet should bear some resemblance to the signal of interest. Since climate signals can be expected to be smooth and continuous, this leads naturally to the choice of non-orthogonal complex bases such as the Morlet and Paul functions (for a full discussion see (Lau and Weng, 1995)). The Morlet wavelet form is the most common and is selected for this analysis.

One particular wavelet, the Morlet, is defined as:

$$\psi_0(\eta) = \pi^{\frac{-1}{4}} e^{i\omega_0\eta} e^{\frac{-1}{2}\eta^2} \tag{6.2}$$

where ω_0 is dimensionless frequency and η is dimensionless time (it is is the product of the scale s and time t). When using wavelets for feature extraction purposes the Morlet wavelet (with $\omega_0 = 6$) is a good choice, since it provides a good balance between time and frequency localization.

The idea behind the continuous wavelet transform is to apply the wavelet as a bandpass filter to the time series. The wavelet is stretched in time by varying its scale (s), so that $\eta = st$, and normalizing it to have unit energy. For the Morlet wavelet (with $\omega_0 = 6$) the Fourier period (λ_{wt}) is almost equal to the scale ($\lambda_{wt} = 1.03s$).

The continuous wavelet transform of a discrete time series $x_n, \quad (n = 1, \ldots, N)$ with uniform time steps $\delta\, t$, is defined as the convolution of x_n with the scaled and normalized wavelet. So it is written as:

$$W_n^x(s) = \sqrt{\frac{\delta\, t}{s}} \sum_{n'=1}^{N} x_{n'} \psi_0 \left[(n' - n) \frac{\delta\, t}{s} \right] \tag{6.3}$$

where n=localized time index, s= wavelet scale, N=number of points in the time series. In practice it is faster to implement the convolution in Fourier space (see details in Torrence and Compo 1998). The wavelet power spectrum, defined as $|W_n(s)|^2$, is a convenient description of the fluctuation of the variance at different frequencies. By

TABLE 6.2: Frequency of the tree ring width and climatic signals every 20 years.

Period (years)	Climatic signals		Ring width,w	
	El Nino	La Nina	$w >$ mean of CI	$w <$ mean of CI
1900-1920	3	6	5	8
1920-1940	2	1	0	12
1940-1960	3	3	3	6
1960-1980	3	5	6	10
1980-2000	9	5	3	16
Total	20	20	17	52

N.B. The mean confidence interval (CI) of the ring width= $[1.37\,mm, 1.95\,mm]$.

varying the wavelet scale s and translating along the localized time index n, one can construct a picture showing both the amplitude of any features versus the scale and how this amplitude varies with time. At each scale s is normalized to have unit energy:

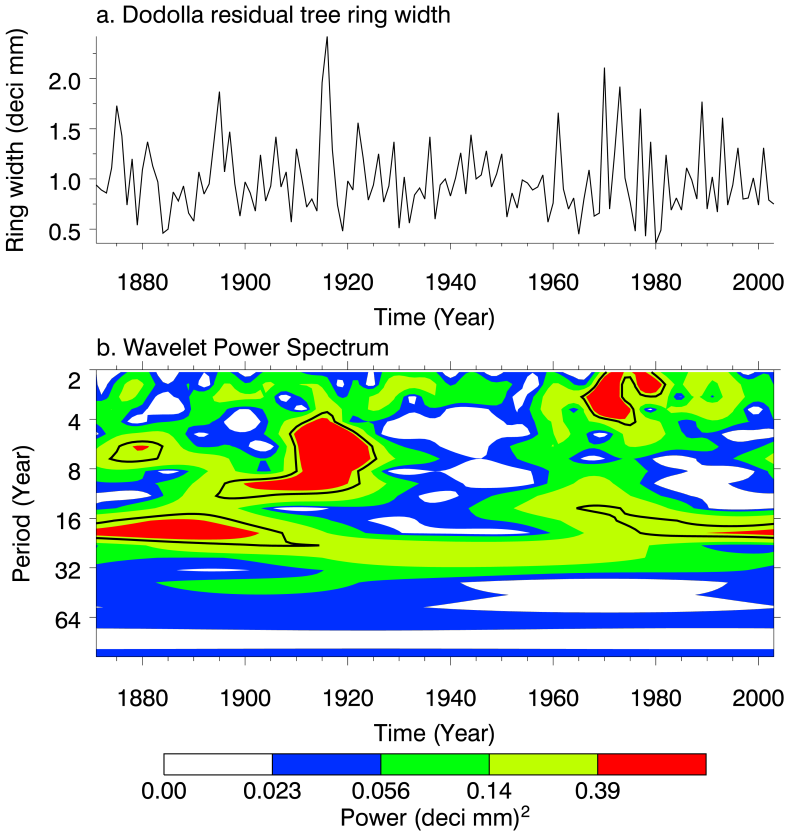


FIGURE 6.5: Wavelet power spectrum of the master tree ring width series at Dodolla area (Using Morlet wavelet)

Extreme events are occurring frequently in the last three decades. The frequency of extreme events and the smaller ring widths (i.e. less than the mean ring width) corresponds well (Table 6.2). The monthly SST anomaly (January-March) are significantly

correlated to annual streamflow of Imi in the order of r 0.78 for the last two decades. Similarly the monthly SOI has a negative correlation of 0.82 for same period. A composite index from the residual tree ring series, SST anomaly (January) and SOI (February) is developed using singular value decomposition. This index is used to simulate and extend the streamflow records at a downstream location, Imi gauging station. Extending the short streamflow data with least square regression may not be a good option for extreme events. An underestimation of the variance is seen in linear or log-linear regression techniques and may result in an underestimation of hydrologic extremes. It is also important to correct for transformational bias when regression analyses use the logarithms of data. Another approach to the least squares method of regression is to use method of maintenance of variance extension, Type II (MOVE 2) (Hirsch, 1982, Parrett and Johnson, 1994). The simulation shows a correlation of 0.84 for the period 2000 - 1974 significant at 0.05 (Figure 6.6). The correlation starts to reduce to 0.65 in the early 1970's.

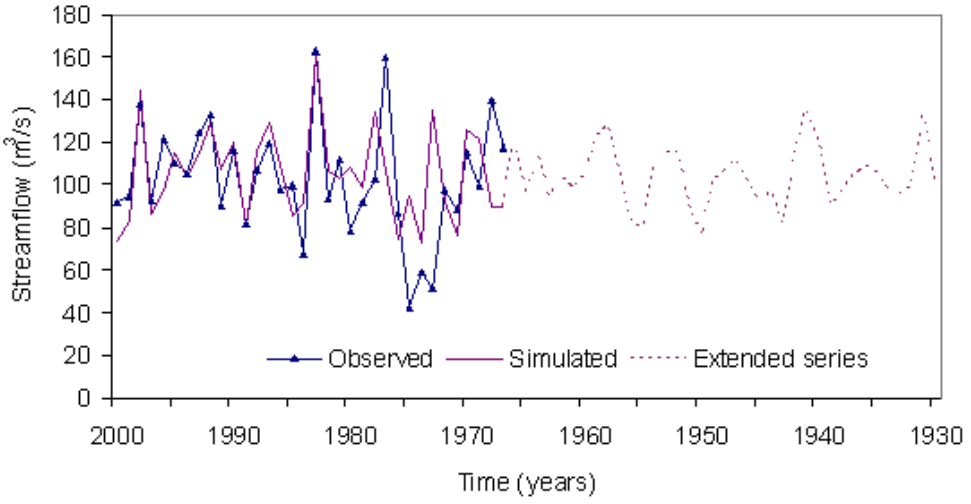


FIGURE 6.6: Observed and simulated streamflow series (2000 - 1974) and extended series (1974 - 1930) at Imi gauging station

6.3.1 Disaggregation of Annual Flows

The generated long series of annual streamflow is disaggregated to seasonal and then to monthly flows. Disaggregation is the simulation of the components of a vector of disaggregated variables given (i.e., conditional on) an aggregate variable. Statistically,

this implies that the joint probability distribution of the flow sequences at the different time periods needs to be preserved. There are several linear time series models (Valencia and Jr., 1973; Mejia and Rousselle, 1976; Stedinger et al., 1985; Santos and Salas, 1992; Koutsoyiannis, 2001) that have been developed over the years and are used widely for disaggregation. Summability of disaggregated flows and their mutual correlation structure (after some transformation) is preserved (Santos and Salas, 1992). Lane's condensed method is applied for the temporal disaggregation of the annual time series (Lane and Frevert, 1990). The model can be expressed by the following equation:

$$Y_{\upsilon,\tau} = A_\tau X_\upsilon + B_\tau \xi_{\upsilon,\tau} + C_\tau Y_{\upsilon,\tau-1} \tag{6.4}$$

where, $Y_{\upsilon,\tau}$ is an nm dimensional zero-mean vector of normally distributed seasonal flows and X_υ is an n-dimensional zero-mean vector of normally distributed annual flows. The noise $\xi_{\upsilon,\tau}$ is an nm dimensional vector of normally distributed noise terms with mean zero and the identity matrix as its variance - covariance matrices. These noise terms are independent in both time and space. A, B, and C are $(nm \rtimes n)$, $(nm \rtimes n)$ and $(nm \rtimes n)$ parameter matrices. υ, τ are indices corresponding to the year, and the season respectively.

The model parameter matrices A, B, and C can be estimated by using the method of moments (MOM) as (Lane and Frevert, 1990):

$$A_\tau = \{[M_{0,\tau}(YX) - M_{1,\tau}(Y)M^{-1}_{0,\tau-1}(Y)M^T_{1,\tau}(XY)] \tag{6.5}$$
$$[M_0(X) - M_{1,\tau}(XY)M^{-1}_{0,\tau-1}(Y)M^T_{1,\tau}(XY)]^{-1}\}$$
$$C_\tau = [M_{1,\tau}(Y) - A_\tau M_{1,\tau}(XY)]M^{-1}_{0,\tau}(Y) \tag{6.6}$$
$$B_\tau B^T_\tau = M_{0,\tau}(Y) - A_\tau M_{0,\tau}(XY) - C_\tau M^T_{1,\tau}(Y) \tag{6.7}$$

where

$$M_k(X) = E[X_\upsilon X^T_{\upsilon-k}]$$
$$M_{k,\upsilon}(Y) = E[Y_{\upsilon,\tau} Y^T_{\upsilon,\tau-k}]$$
$$M_{k,\upsilon}(YX) = E[Y_{\upsilon,\tau} X^T_{\upsilon-k}]$$
$$M_{k,\tau}(XY) = E[X_\upsilon X^T_{\upsilon,\tau-k}]$$

The temporal Lane model is a condensed type of the Mejia and Rouselles disaggregation model. The Valencia and Schaakes model does not preserve the covariance of the

first season of a year and any precedent season and was not considered. The original Lane model does not preserve the additivity of seasonal values but Lane proposed some adjustments. These adjustments are considered in disaggregating the annual values to monthly time scales. The following equation was used for the adjustment of temporal disaggregated data:

$$\hat{q}^*_{\upsilon,\tau} = \frac{\hat{q}_{\upsilon,\tau}\hat{Q}_\upsilon}{\sum_{t=1}^{\omega}\hat{q}_{\upsilon,\tau}} \tag{6.8}$$

where ω is the number of seasons, $\hat{Q}_\upsilon$ is the generated annual value, $\hat{q}_{\upsilon,\tau}$ is the generated seasonal value, $\hat{q}^*_{\upsilon,\tau}$ is the adjusted generated seasonal value.

6.3.2 Hydrological Drought Variables

Time series of the reconstructed monthly streamflow is transected by a constant threshold level to obtain the hydrologic drought variables of severity and duration.

6.3.2.1 Comparison with Instrumental Data

The severity and duration of drought identified from the generated series and historical instrumental records compare fairly well. Figure 6.7 provides the mean of the simulated parameters as a ratio of the historical records. The confidence intervals in the mean simulations show the uncertainty of the simulated data. The longest duration is better simulated than the severest drought in the historic instrumental data . The maximum drought is under-simulated. This can be attributed to the fact that reconstructions tend to underestimate the variability. Interestingly the Hurst coefficient is also well simulated (H=0.6385, recorded and H=0.6156, generated).

The magnitude of the severity of drought is dependent on the threshold level used. However, irrespective of the frequency of the drought, the severity increase as the threshold level is raised (Fig 6.8). This is in an agreement with the case mentioned for upstream station Dodolla and downstream station Gode (see Chapter 4). The strength among the various drought events is indistinguishable at a threshold level corresponding to more than 70% exceedance probabilities.

The results reported above for Imi is compared against a corresponding frequency for the area computed in severity-frequency-area curve for different pools in the basin (see Section 5.4). Here the scPHDI is assumed as a surrogate for hydrological drought from

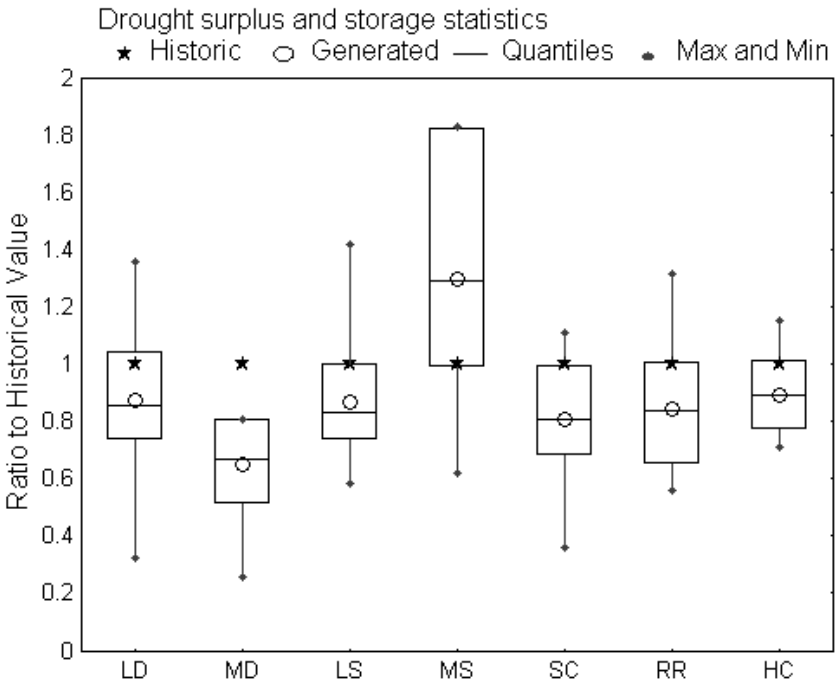


FIGURE 6.7: The generated and historical hydrological drought characteristics at Imi gauging station (LD- Longest drought, MD- Maximum deficit, LS- Longest surplus, MS- Maximum surplus, SC- Storage capacity, RR- Rescaled range, HC- Hurst coefficient).

instrumental records. The middle valley streches between Pool B and some areas in Pool D. From the climatic, morphologic and and physical charactristics of Imi it is more likely in Pool 2. Recent worst droughts in the severity-frequency-area curve of Pool B (Fig 5.13(b)) shows a similar pattern with the pattern captured in the severity of the generated hydrologic drought from proxy data. The 1974/75 drought is severest one in both cases with an order of magnitude for the return period between 30-50 years.

6.3.2.2 Univariate Drought Frequency

The severity and duration of drought have high correlation. Univariate modeling of these variables shows that the severity follows Weibull distribution where as the duration is better described by lognormal distribution (Table 6.3).

Two-parameter Weibull and two-parameter lognormal distributions are appropriate to model the severity and duration series respectively. The parameter estimates of these distributions with their standard of errors are provided in Table 6.4.

It is essential to model the dependence structure between the severity and duration of the hydrological drought variables in order to appropriately designate their joint frequency

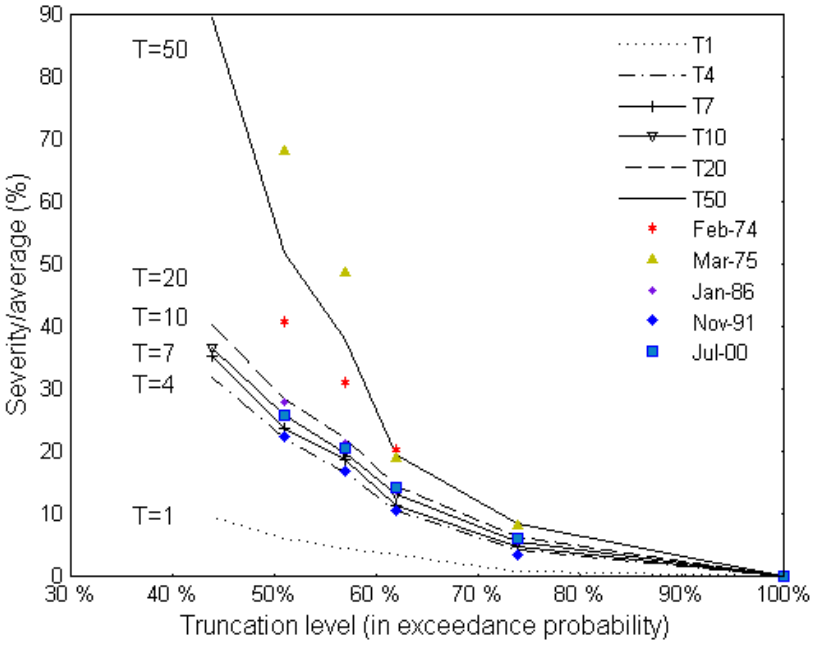


FIGURE 6.8: Frequency of severity of drought per average annual flow at different threshold levels for notable droughts in Imi gauging stations.

TABLE 6.3: Univariate maximum likelihood estimates of different distribution fitted to severity and duration series.

Distribution	Maximum Likelihood	
	Severity	Duration
Exponential	-3.9487	-288.757
Gamma	-3.7586	-272.183
Weibull	-3.5463	-274.465
Lognormal	-12.3910	-271.105
Log-logistic	-37.7698	-278.621

TABLE 6.4: Parameter estimates and standard errors of the fitted distributions to the severity and duration.

Parameters	Severity (Weibull)	Duration (Lognormal)
a (location parameter)	0.3882 (0.033)	0.9092 (0.065)
b (scale parameter)	1.0661 (0.075)	0.7513 (0.0463)
Mean	0.3786	3.2917
Variance	0.12626	8.2175

characteristics.

Chapter 7

Stochastic Simulation of Severity of Drought

In this chapter the temporal patterns of the hydrological drought is presented using stochastic analysis. The chapter is mainly based on Awass and Foerch (2008).

7.1 Trends of Drought

In this section whether drought in the Wabi Shebele is getting more frequent or more severe will be discussed. The trends in the inter-arrival time, frequency of drought event and severity of drought are tested using the reconstructed long time series in Chapter 6. The inter-arrival time represents the time lapse between the beginnings of one drought event to the next. Thus it includes the drought and the non-drought duration.The average inter-arrival time in the middle valley is about 6.14 months (Figure 7.1). These inter-arrival time in most cases consists of 3 months of drought and 3 months of recess (non-drought) that may be associated with the seasonality effect. There is no significant trend in the inter-arrival time. The drought that occurred at the beginning of the 1980s has a duration of about 20 months. This makes it the drought with the longest duration in the period 1930 to 2000. This has extended the inter arrival time appearing like an outlier. From Figure 7.1 it can also be discerned distinctly that there are about 10 major drought events which peak out in this 70 year duration. This matches with the common cyclicity of 7 years for this mega drought events.

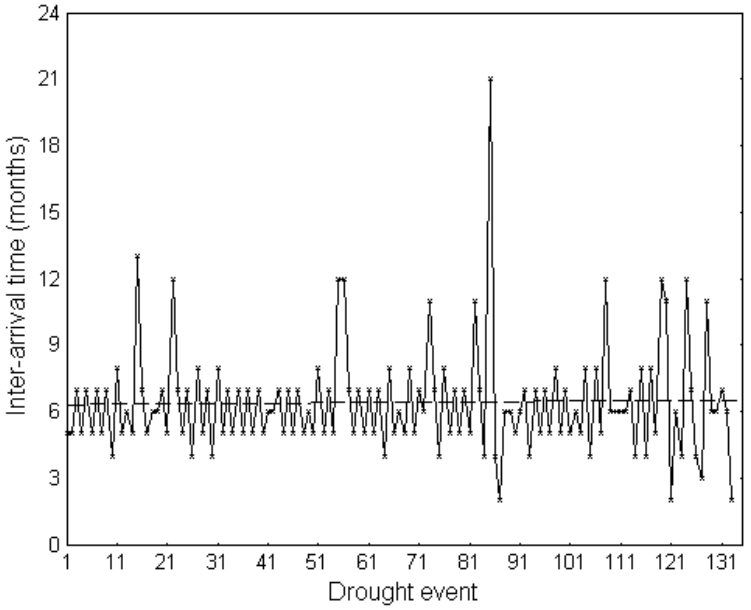


FIGURE 7.1: Inter-arrival time of drought in the middle valley of Wabi Shebele (months)

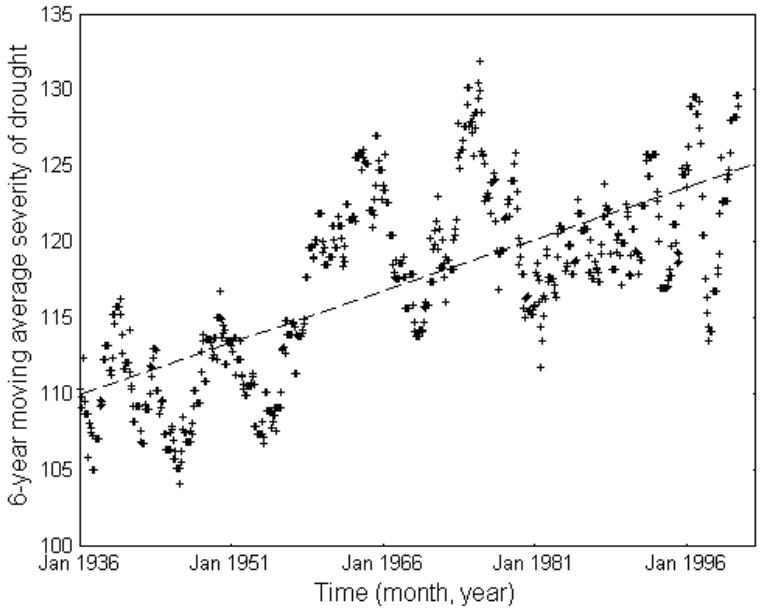


FIGURE 7.2: 6-year moving average of severity of drought in the middle valley of Wabi Shebele (in Mm^3)

Trend of severity of drought

A 6-year moving average of the severity of drought is computed. In addition to cyclicity in Figure 7.2 a significant trend is noticeable in the average of severity of drought. It was further worsened between the beginning of the 1960s to end of the 1970s. Non-linear

trends are also tested with statistical analysis that has supported the availability of a positive trend in the severity of drought in the period 1930-2000.

Trend of frequency of drought

A 6-year moving average analysis in the frequency of drought incidents in the middle valley shows there is no significant trend considering the whole duration of 70 years. However, it is observed that in the middle of the 1970s to the beginning of the 1980s there is an increase in frequency (Figure 7.3).

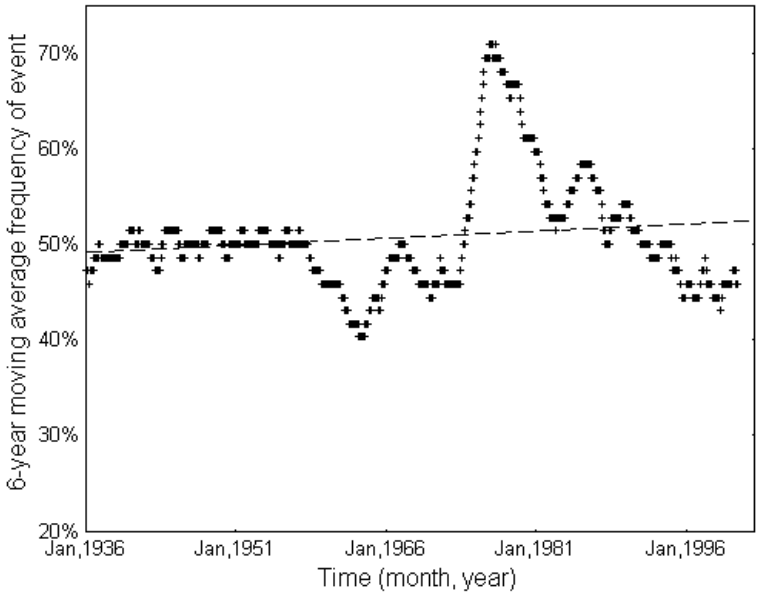


FIGURE 7.3: 6-year moving average of frequency of drought in the middle valley of Wabi Shebele

In the period considered there is an increase in severity of drought but the frequency is generally stationary(Figure 7.4). More severe and frequent events coupled have hit the area at the end of the 1970s.

7.2 Background

Drought occurs in virtually all climatic zones. However, its characteristics vary significantly from one region to another. Defining drought is therefore difficult; it depends on differences in regions, needs, and disciplinary perspectives. Hydrological drought is considered as the period during which streamflows are below the expected amount (threshold level) for a certain period of time in a given climatic region. The threshold

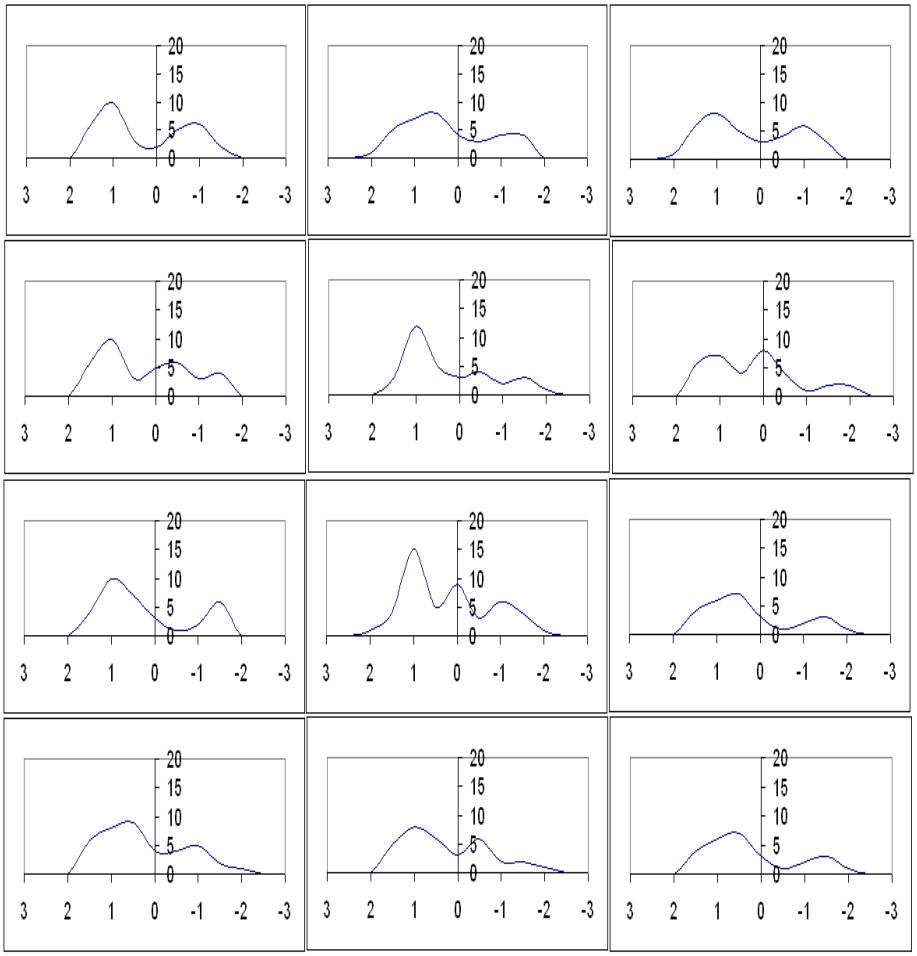


FIGURE 7.4: 6-year moving average of frequency of severity of drought plotted every 5 years in the middle valley of Wabi Shebele [abscissa axis=severity standardized within respective periods; ordinate axis=frequency]

level specifies some statistic of the drought variable and serves to divide the time series of the variable in question into deficit and surplus sections. The increasing demand on water resources requires better management of the water deficit situation, may it be unusual droughts, or yearly recurring low flows. It is not possible to avoid droughts. But drought preparedness can be developed and drought impacts can be managed.

Due to the randomness of the factors responsible for the occurrence and severity of drought, it can be considered as stochastic process. Statistical techniques dealing with the duration aspects of drought are reasonably well developed, whereas techniques for severity aspects are less satisfactory and require considerable improvement and refinement (Panu and Sharma, 2002). Drought severity is a crucial parameter, for instance, in sizing storage reservoirs towards combating droughts. In this study stochastic simulation of severity of hydrological drought in Wabi Shebele river basin using Box and

Jenkins methodology is carried out. This method provides systematic searching stages of identification, estimation, and diagnostic checking for an appropriate model (Box et al., 1994).

7.3 Data and Methodology

In order to simulate the severity of hydrological drought in Wabi Shebele river basin, the river basin was first delimited into hydrologically homogeneous pools. Different methods of delineating hydrologically homogeneous regions are available in literature like pattern recognition through neural networks, empirical orthogonal functions and L-moments (Hall and Minns, 1999; Kjeldsen et al., 1999; Tallaksen and Hisdal, 1999). Here the classification is done using clustering algorithm of hierarchical clustering using ward's method and verified using discriminant analysis (see Chapter 4). Hydrologically homogeneous pools in the basin are established using parameters mean rainfall, temperature, normalized digital vegetation index derived from satellite images and streamflow. Thus, five prominent homogeneous pools are developed (see Section 4.4.1). Time series analysis of severity of hydrological drought was carried out at two places one in the upstream semi-humid location at station 61004 near Dodolla (Pool A) and the other at the downstream arid location at station 63001 near Gode (Pool E) using Box and Jenkins methodology. Summary descriptive statistics of monthly flow data at the stream gauging stations of Dodolla (station 61004) and Gode (station 63001) is provided in Table 1. The coefficient of variation of the monthly streamflow is almost equal at the two stations.

Quite often, hydrologic phenomenon depicts cyclic and stochastic processes. Thus Fourier analysis is done to reveal the explained variance of the periodic component.

$$x_{\alpha,t} = M_t + e_i \tag{7.1}$$

Where $x_{\alpha,t}$ is the severity of hydrological drought of month t and year α. M_t is the periodic deterministic component reflected in the monthly mean values t = 1 for January and t = 12 for December and e_i is the stochastic component with mean zero and variance σ_e^2 where i = 1, 2, , N, and N is the number of observations. Harmonic analysis is used to identify the periodic (regular) variations in the time series. If we have N observations the

time series can be approximated by cosine and sine functions. In general, if the number of observations is N, the number of pairs of cosine and sine functions (j=number of harmonics) is equal to $\frac{N}{2}$. The periodic component may be expressed as:

$$M_t = A_0 + \sum_{j=1}^{6} \left(A_j cos\left(\frac{\pi\, jt}{6}\right) + B_j sin\left(\frac{\pi\, jt}{6}\right)\right) \tag{7.2}$$

Where the harmonic coefficients A_j and B_j are given by:

$$A_0 = \frac{1}{12}\sum_{t=1}^{12} M_t, \quad A_j = \frac{1}{6}\sum_{t=1}^{12} M_t cos\left(\frac{\pi\, jt}{6}\right), \quad B_j = \frac{1}{6}\sum_{t=1}^{12} M_t sin\left(\frac{\pi\, jt}{6}\right) \tag{7.3}$$

Since the harmonics are all uncorrelated, no two harmonics can explain the same part of the variance of the time series. We can add up the fractions to determine how many harmonics we need to explain most of the variations in the time series. The first two harmonics explain about 87% of the variance of the M_t series (Figure 7.5a) and 11-17% of the variance of the total series x ,t (Figure 7.5b). The other harmonics explain a negligible portion of both series (Figure 7.5). The total series may be written as:

$$X_{\alpha,t} = A_0 + A_1 cos\left(\frac{\pi\, t}{6}\right) + B_1 sin\left(\frac{\pi\, t}{6}\right) + A_2 cos\left(\frac{\pi\, t}{3}\right) + B_2 sin\left(\frac{\pi\, t}{3}\right) + e_i \tag{7.4}$$

As 82-89% of the variance of the total series is accounted by the stochastic component, it has a significant influence on the series.

7.3.1 Time Series Analysis

Identifying the nature of the phenomenon represented by the sequence of observations and forecasting (predicting) future values of the time series variable are among the prominent aims of time series analysis. Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once the pattern is established, we can interpret and integrate it with other data. In this study, time series of severity of hydrological drought is extracted after defining a threshold level for the streamflow series at both stations separately. The threshold level is defined in such a way as to consider both the supply and demand. The cumulative streamflow deficiency below the minimum average supply or demand of five consecutive years is regarded as

the severity of hydrological drought. Monthly streamflow data for the years 1982-2000 was obtained from hydrology section of Ministry of Water Resources in Ethiopia.

Autocorrelation plot of the original time series of severity of drought at both stages shows that they are non-stationary (Figure 7.6a). In particular, the need for regular differencing is indicated when the autocorrelations of the series tend to remain large for many successive lags, beginning at lag 1. Thus autoregressive integrated moving average (ARIMA) model, also known as Box Jenkins, is fitted to the time series of severity of hydrological drought. The ARIMA model is autoregressive to order p and moving average to order q and operates on the dth difference of x_t. Unlike exponential smoothing, ARIMA attempts to remove the trend before modelling the autocorrelation. One of the major advantages of the box and Jenkins methodology is its optimality and comprehensiveness of the family of models. A general weakness related to short time forecasting using ARIMA is the large number of data points needed for estimation. It is recommended that there are at least 50 observations in the input data (Wei, 1990).

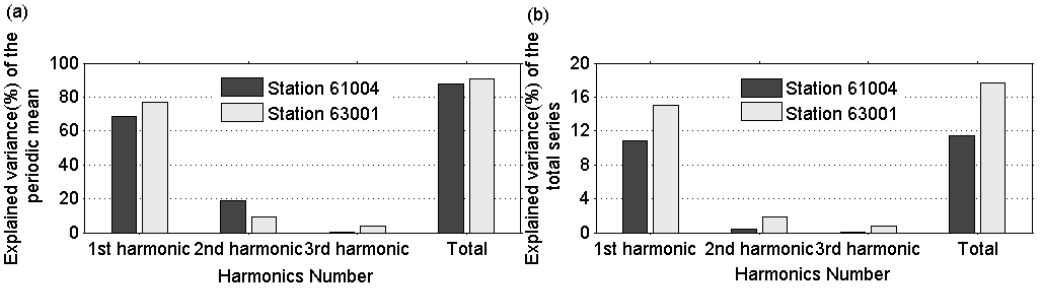


FIGURE 7.5: Explained variance in the monthly periodic mean and total series of severity of hydrological drought (228 months) in Wabi Shebele river basin.

The seasonal ARIMA model of order (p,d,q) (P,D,Q) can be expressed as:

$$(1-B)^d (1-B^s)^D \phi(B)\Phi(B)X_t = \theta(B)\Theta(B)a_t \tag{7.5}$$

Where B is the backward shift operator, s is the period of the season (in this case 12). $\phi(B)$, $\theta(B)$ are autoregressive and moving average operators of order p and q of regular series,$\Phi(B)$, $\Theta(B)$ are autoregressive and moving average operators of order P and Q of the seasonal series and a_t is the residual.

In equation 7.5 above, d is the order of the first differencing component and D is the order of the seasonal component. Differencing removes trend and seasonalities in a time

series and is used to obtain stationarity. The backward shift operator B can be defined as the lag operator:

$$BX_t = X_{t-1} \tag{7.6}$$

The autoregressive and moving average operators ($\phi(B)$, $\theta(B)$, $\Phi(B)\, and\, \Theta(B)$) of the regular and seasonal series can be represented as polynomials of order p, q, P, and Q respectively (see Equations 7.7 - 7.10).

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \tag{7.7}$$

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_p B^q \tag{7.8}$$

$$\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^P \tag{7.9}$$

$$\Theta(B) = 1 - \Theta_1 B - \cdots - \Theta_p B^Q \tag{7.10}$$

Detail description of ARIMA models can be referred in Box et al. (1994). For fitting seasonal autoregressive integrated moving average (SARIMA) model to the time series three stage procedures of model identification, estimation of model parameters and diagnostic checking of estimated parameters are carried out. Identification methods are rough procedures applied to a set of data to indicate the kind of representational model which is worthy of further investigation and the specific aim is to obtain some idea of the values of p, d, and q needed in the general linear ARIMA model and to obtain initial guesses of the parameters. After selecting an appropriate set of model parameters, the next step is to compute specific values for these parameters from the series data. These involve determining the optimum values for the selected parameters so that the sum of squares is minimized and the residuals are uncorrelated. Efforts have to be put to use as few parameters as necessary to obtain an adequate model. This step is referred as model diagnostic checking. The estimated model is used to generate forecasts and forecast confidence limits. In practice the model building process used by Box and Jenkins approach is iterative(Box et al., 1994). Seasonal patterns of time series can be examined via correlograms. Autocorrelation function (ACF) and partial autocorrelation function (PACF) should be used to gather information concerning the seasonal and non-seasonal AR and MA operators. ACF measures the amount of linear dependence between observations in a time series. Another useful method to examine serial dependencies is to examine the partial autocorrelation function (PACF) - an extension of autocorrelation, where the dependence on the intermediate elements (those within the lag) is removed.

The identification of the appropriate parametric time series model depends on the shape of the ACF and PACF. Failure of the ACF estimates other than integer multiple of s to damp out suggests that nonseasonal differencing is needed to produce stationarity. The sample PACF damps out at lags that are multiple of s. This suggests the incorporation of a seasonal MA component into the model. The failure of the sample PACF to truncate at other lags may imply a nonseasonal MA term is required (Figure 7.6b). The sample ACF reveals also the ACF truncates and is not significantly different from zero after lag 13 (Figure 7.6a). Thus a pure $MA(0,1,1)(0,1,1)^{12}$ is fitted to the severity of drought series at Gode. The model parameters are estimated using maximum likelihood estimation. A diagnostic check is carried out (section 5.1 and 5.2). The residual autocorrelation function (RACF) should be obtained to determine whether residuals are white noise.

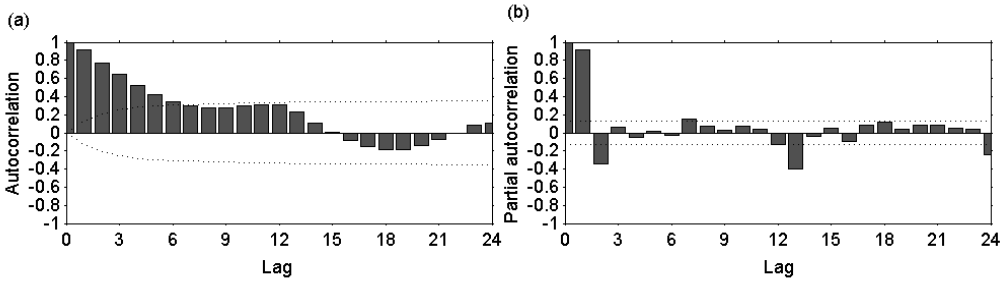


FIGURE 7.6: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the severity of hydrological drought time series at Gode station.

7.4 Results and Discussion

A basic tenet of model building is to keep the model as simple as possible but at the same time provide a good fit to the data being modelled (Hipel and McLeod, 1994). Different automatic selection criteria are available for balancing the apparently contradictory goals of good statistical fit and model simplicity. Among the automatic selection criterion used in model discrimination Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) are common(Hipel and McLeod, 1994). The diagnostic check stage determines whether residual are independent, homoscedastic and normally distributed.

7.4.1 Akaike Information Criterion (AIC)

In order to discriminate the most appropriate model Akaike Information Criterion is quite powerful. Mathematical definition of AIC is given as:

$$AIC = \frac{N}{n}(-2lnML) + 2k \tag{7.11}$$

Where In ML denotes maximized log likelihood function, N is number of observations of non-stationary series, n is number of differenced series (N-d) and k is the number of independently adjusted parameters. The first part of the right hand side of equation 7.11 explains how well the model fits and the second part accounts for simplicity of the model. Employment of the minimum AIC reinforces and complements the identification, estimation and diagnostic stages of model construction. AIC tends to overestimate the order of auto regression. Thus a Bayesian extension of the minimum Akaike criterion (BIC) is developed. One of the variants of BIC, Schwarz Bayesian Criterion (SBC) is employed here.

7.4.2 Schwarz Bayesian Criterion (SBC)

The Schwarz Bayesian Criterion (SBC) has a structure quite similar to that of AIC.

$$SBC = (-2lnML) + kln(n) \tag{7.12}$$

The model that gives the minimum AIC and SBC is selected as parsimonious. The performance of the candidate models is compared using sum of squares of errors (SSE), the mean absolute percentage error (MAPE), the white noise variance (WN variance), the natural logarithm of the likelihood function, Akaike's Final Prediction Error (FPE), the Akaike Information criterion (AIC), and Schwarz Bayesian Criterion(SBC) (Table 7.1). The white noise variance is the ratio of the sum of squares of errors divided by the number of observation.

In Table 7.1, it can be seen from the log likelihood function that as the number of parameters increases, the model fits better the observations. However, as it is observed from values of AIC and SBC, the model $(0,1,1)(0,1,1)^{12}$ is the best among the candidate models where as the ARMA model performs poor.

TABLE 7.1: Values of AIC and BIC for various model fits of hydrological drought severity at Gode station

Model	SSE	MAPE	WN variance	-2Log(Lik)	FPE	AIC	SBC
(1,0,1)(0,0,0)	1827664.96	112.52	8461.41	2568.44	8540.12	2574.44	2584.57
(0,1,1)(0,0,0)	1953498.55	416.39	9086.04	2569.90	9086.04	2573.90	2580.64
$(0,1,1)(0,1,1)^{12}$	1239742.65	528.24	6107.11	2363.07	6107.11	2369.07	2379.01
$(0,1,2)(0,1,1)^{12}$	1239473.81	529.24	6105.78	2363.07	6105.78	2371.07	2384.32
$(0,1,1)(1,1,1)^{12}$	1242140.56	536.62	6118.92	2362.76	6887.79	2370.76	2384.02

TABLE 7.2: Model parameters for the fitted severity of drought series at Gode station

Constant	-2.174			
Parameter	Value	Hessian standard error	Lower bound (95%)	Upper bound (95%)
MA(1)	0.144	0.070	0.007	0.281
SMA(1)	-0.876	0.065	-1.002	-0.749

Generally the value of SBC is more than AIC for all the fitted models. It appears that the SBC slightly better discriminates the models than AIC. Autocorrelation function of the residuals indicates the selected model fits adequately.

Table 7.2 gives the estimator for each coefficient of each polynomial, as well as the standard errors of the estimates of the parameters for the selected SARIMA model$(0,1,1)(0,1,1)^{12}$. Both the parameters corresponding to the coefficient of the moving average operator of the regular series, MA(1), and the seasonal series, SMA(1), are significantly different from 0 as the 95% confidence interval does not include 0. The range between the lower bound (95%) and upper bound (95%) is narrower in case of SMA (1) than MA (1) (Table 7.2). The confidence intervals are computed using the Hessian after optimization.

The model can be written in compact form:

$$(1-B)(1-B^{12})X_t = -2.174 + (1-0.144B)(1+0.876B^{12})a_t \tag{7.13}$$

The ARIMA $(0,1,1)(0,1,1)^{12}$ model fits well to the severity of drought at Gode station as it is witnessed by the plot of the simulated fit and the automatic selection criterion (Figure 7.7a and Table 7.1). Normality test and white noise test of the residuals is done through plot of ACF of the residuals and statistical analysis (Figure 7.7b). This confirms the adequacy of the model fitted to the available time series of severity of hydrological drought in the area.

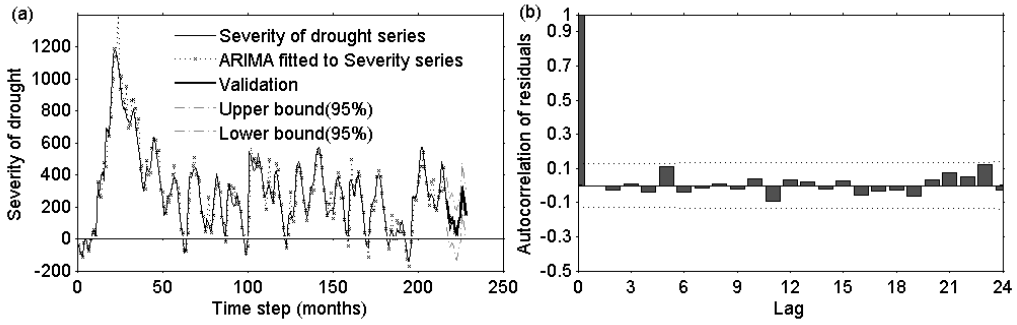


FIGURE 7.7: The fitted ARIMA $(0,1,1)(0,1,1)^{12}$ model to the severity of drought at Gode station and plot of autocorrelogram of the residuals.

7.5 Conclusion

Fourier analysis of the time series at both stations in Wabi Shebele river basin reveals that 82-89% of the variance of the total series is accounted by the stochastic component. The stochastic component of the total series of severity of drought at the upstream humid portion (station 61004) is higher than that of the downstream arid portion (station 63001). Stochastic simulation of hydrological drought is generally needed for reservoir sizing, for determining the risk of failure of water supply for irrigation systems, for determining the risk of failure of capacities of hydro-electric systems, for planning studies of future reservoir operation, for planning capacity expansion of water supply systems, and similar applications. Thus stochastic simulation of severity of hydrological drought using Box and Jenkins methodology is carried out. Different seasonal ARIMA candidate models are compared. It is found that the pure MA $(0,1,1)(0,1,1)^{12}$ outperforms all. The simulated results compare well with the historical records of severity of hydrological drought in the basin as confirmed by the residual white noise and normality test.

Chapter 8

Double Chain Markov Model

In this chapter a discrete-time dependency between successive observations of scPHDI is modeled using various forms of Double Chain Markov Models.

8.1 Background

The problem of forecasting future values of a random variable, from a stochastic point of view, is equivalent to the determination of the probability density function of future values conditioned by past observations. Once the conditional distribution is known, the forecast is usually defined as the expected value or a quantile of such distribution, and confidence intervals of the forecast values can be computed.

Analysis of short-term meteorological records shows evidence of a tendency for dry years to cluster together. This indicates that the sequence of dry years is not completely random. The Markov chain is a probabilistic model used to represent dependences between successive observations of a random variable. Drought characteristics of any phenomenon are dependent on the underlying generating mechanism, and can be modelled by a suitable stochastic process such as independent, first- and second-order Markov processes. This dependence can be measured by either the autocorrelation function or autorun function based on the conditional probabilities (Sen, 1998). The autocorrelation plot of scPHDI time series at Assassa meteorological station in Wabi Shebele is shown in Figure 8.1.

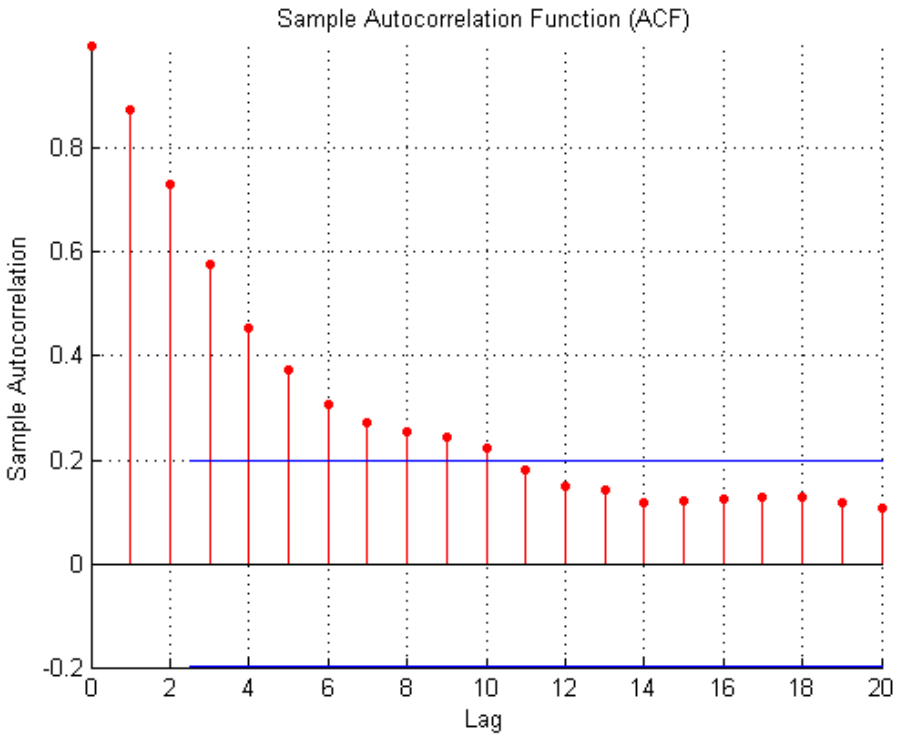


FIGURE 8.1: Plot of autocorrelation for scPHDI at Assassa station.

Basically, a Markov chain is a statistical model in which the current value (time t) of a variable X taking values in the finite set 1, . . .,M; is fully explained by the knowledge of the value taken by the same variable at time t-1. A Markov chain is characterized by a set of states, S, and by the transition probability, p_{ij}, between states. The transition probability p_{ij} is the probability that the Markov chain is at the next time point in state j, given that it is at the present time point in state i. The transition probability matrix C_1 summarizes the model giving the probability distribution of X_t for any possible value of X_{t-1}:

$$C_1 = [P(X_t = j|X_{t-1} = i)] = [p_{ij}] = \begin{array}{c} \\ t-1 \\ 1 \\ \vdots \\ M \end{array} \begin{array}{c} \begin{array}{ccc} & t & \\ 1 & \cdots & M \end{array} \\ \begin{vmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \cdots & p_{MM} \end{vmatrix} \end{array} \tag{8.1}$$

Each row of C_r is a probability distribution summing to one. Since the current value is fully determined by the knowledge of only one past period, this model is said to be of order 1. More generally, a Markov chain of order r , $r \geq 0$, is a model in which the

current value is explained by all lags up to t - r . The transition matrix is then of a larger size.

The special case $r = 0$ in which the current value is independent from the past is called the independence model. The transition matrix then becomes a single probability distribution:

$$C_0 = [P(X_t = j)] = (p_1 \cdots p_M) \tag{8.2}$$

8.2 Homogeneous Markov Model

The first-order Markov hypothesis says that the present observation at time t is conditionally independent of those up to and including time (t -2) given the immediate past time(t-1). Thus it can be written that:

$$\begin{aligned} P(X_t = i_0 | X_0 = i_t, \ldots, X_{t-1} = i_1) \\ = P(X_t = i_0 | X_{t-1} = i_1) \\ = qi_1 i_0(t) \end{aligned} \tag{8.3}$$

where $i_t, \ldots, i_0 \in 1, ..., M$. Suppose that the probability $qi_1 i_0(t)$ is time invariant then (t) it is replaced by $qi_1 i_0$ and it is known as a homogeneous Markov chain. Considering all combinations of i_1 and i_0, a transition matrix C can be constructed, each of whose rows sums to 1. The sample dimension and the number of states influence the accuracy of estimates; the number of parameters of the model depends on the number of states. In this study five drought categories, or states, were considered from a time series of scPHDI (i.e. non-drought, mild drought, moderate drought, severe drought and extreme drought). Table 8.1 provides the first order transition probability matrix of drought states using scPHDI at Assassa meteorological station at one month step.

Given that the model is in a known state S_i, the probability that it stays in that state for exactly d days can be answered using the model. This probability can be evaluated as the probability of the observation sequence O:

$$O = \underset{1}{S_i}, \underset{2}{S_i}, \underset{3}{S_i}, \ldots \underset{d}{S_i}, \underset{d+1}{S_j} \neq S_i,$$

TABLE 8.1: First order transition probability matrix of drought states using scPHDI at Assassa meteorological station at one month step

State	No drought	Mild	Moderate	Severe	Extreme
No drought	0.9454	0.0492	0.0055	0	0
Mild	0.3846	0.4615	0.1538	0	0
Moderate	0	0.1667	0.6667	0.1667	0
Severe	0	0	0.0962	0.8462	0.0577
Extreme	0	0	0	0.75	0.25

TABLE 8.2: First order transition probability matrix of drought states using scPHDI at Assassa meteorological station at four months step

State	No drought	Mild	Moderate	Severe	Extreme
No drought	0.8718	0.0851	0.0354	0.0075	0.0002
Mild	0.6445	0.1377	0.1432	0.0713	0.0033
Moderate	0.1914	0.1484	0.3078	0.3308	0.0216
Severe	0.018	0.0439	0.1909	0.6901	0.0571
Extreme	0.0046	0.0267	0.1623	0.7425	0.0639

TABLE 8.3: First order transition probability matrix of drought states using scPHDI at Assassa meteorological station at seven months step

State	No drought	Mild	Moderate	Severe	Extreme
No drought	0.8324	0.0909	0.0518	0.0237	0.0012
Mild	0.677	0.1024	0.1065	0.1069	0.0072
Moderate	0.3038	0.1046	0.2057	0.3588	0.0271
Severe	0.0724	0.0643	0.207	0.607	0.0493
Extreme	0.0466	0.057	0.203	0.6409	0.0525

given the model, which is

$$P(O|Model, x_1 = S_i) = (a_{ii})^{d-1}(1 - a_{ii}) = p_i(d)$$

where a_{ii} is the transition probability from state i to i, the quantity $p_i(d)$ is the discrete probability density function of duration d in state i. This exponential duration density is characteristic of the state duration in a markov chain. Based on $p_i(d)$, it is possible to calculate the expected number of duration in a state, $\bar{d}_i$, conditioned on starting in that state as:

$$\bar{d}_i = \sum_{d=1}^{\infty}(d \cdot p_i(d)) \tag{8.4}$$

$$= \sum_{d=1}^{\infty}(d \cdot (a_{ii})^{d-1}(1 - a_{ii})) = \frac{1}{1 - a_{ii}} \tag{8.5}$$

This is an entirely visible process since each observed output is exactly identified with one state of the process.

8.2.1 Nth order Markov Model

In some situations, the present depends not only on the first lag, but on the last r observations. Then there will be an rth-order Markov chain whose probabilities are:

$$\begin{aligned} & P(X_t = i_0 | X_0 = i_t, \ldots, X_{t-1} = i_1) \\ = & P(X_t = i_0 | X_{t-l} = i_r, \ldots, X_{t-1} = i_1) \\ = & q i_r i_0(t) \end{aligned} \tag{8.6}$$

The annual surface water or precipitation occurrences can be approximated by independent or simple first-order Markov chains, which usually fail to preserve the critical drought durations properly. Sen(1990) used second-order markov chain to model the critical drought duration.

Each possible combination of r successive observations of the random variable X is called a state of the model. The number of states is equal to m^r. In the case of a first-order Markov chain, each value taken by the random variable X is also a state of the model. In the first-order markov chain modeling of scPHDI above, there are 5 states (see Table 8.2 and 8.3). In higher order markov models the number of state also depends on the order r. Table 8.4 shows the second order transition probability matrix of the same data with 25 states (i.e. 5^2). In each row of the matrix R, there are (m - 1) independent probabilities, as the last transition probability equals one minus the other transition probabilities in that row. The total number of independent parameters to estimate is thus equal to $m^r(m-1)$. Unfortunately, as the order r of the chain and the number m of possible values increase, the number of independent parameters increases exponentially and rapidly becomes too large to be estimated efficiently, or even identifiably, with data sets of the sizes typically encountered in practice. In order to abate this drawback of Markov chains, the Mixture Transition Distribution Model (MTD) was introduced by Raftery (1985). Each element of a transition matrix is the probability of observing an event at time t given the events observed at times (t-r) to (t -1). In the MTD model, the effect of each lag upon the present is considered separately and the conditional probability is modeled by:

$$P(X_t = i_0 | X_{t-r} = i_r, \ldots, X_{t-1} = i_1) = \sum_{g=1}^{r} \lambda_g \cdot q i_g i_0. \tag{8.7}$$

TABLE 8.4: Second order transition probability matrix of drought states using scPHDI at Assassa meteorological station at one month step

t-2	t-1	No drought	Mild	Moderate	Severe	Extreme
No drought	No drought	0.9412	0.0535	0.0053	0.0000	0.0000
No drought	Mild	0.1818	0.7273	0.0909	0.0000	0.0000
No drought	Moderate	0.0000	0.0000	1.0000	0.0000	0.0000
No drought	Severe	0.0000	0.0000	0.0000	0.0000	0.0000
No drought	Extreme	0.0000	0.0000	0.0000	0.0000	0.0000
Mild	No drought	0.9091	0.0909	0.0000	0.0000	0.0000
Mild	Mild	0.3478	0.4348	0.2174	0.0000	0.0000
Mild	Moderate	0.0000	0.0000	0.8333	0.1667	0.0000
Mild	Severe	0.0000	0.0000	0.0000	0.0000	0.0000
Mild	Extreme	0.0000	0.0000	0.0000	0.0000	0.0000
Moderate	No drought	0.0000	0.0000	0.0000	0.0000	0.0000
Moderate	Mild	0.1667	0.8333	0.0000	0.0000	0.0000
Moderate	Moderate	0.0000	0.1515	0.6364	0.2121	0.0000
Moderate	Severe	0.0000	0.0000	0.2222	0.7778	0.0000
Moderate	Extreme	0.0000	0.0000	0.0000	0.0000	0.0000
Severe	No drought	0.0000	0.0000	0.0000	0.0000	0.0000
Severe	Mild	0.0000	0.0000	0.0000	0.0000	0.0000
Severe	Moderate	0.0000	0.1111	0.7778	0.1111	0.0000
Severe	Severe	0.0000	0.0000	0.1000	0.8714	0.0286
Severe	Extreme	0.0000	0.0000	0.0000	0.6667	0.3333
Extreme	No drought	0.0000	0.0000	0.0000	0.0000	0.0000
Extreme	Mild	0.0000	0.0000	0.0000	0.0000	0.0000
Extreme	Moderate	0.0000	0.0000	0.0000	0.0000	0.0000
Extreme	Severe	0.0000	0.0000	0.0000	0.6667	0.3333
Extreme	Extreme	0.0000	0.0000	0.0000	1.0000	0.0000

where the $qi_g i_0$ are the probabilities of an $m \times m$ transition matrix and λ_g is the weight parameter associated with lag g. This model has only $m(m-1)+(r-1)$ independent parameters. It is to be noted from this formula that each additional lag adds only one additional parameter. When the order is greater than 1, the MTD model is far more parsimonious than the corresponding fully parameterized Markov chain (See also Table 8.5). MTD model has also the advantage of consisting a small transition matrix and a vector of lag parameters which are easier to interpret. In the MTD model, the effect of each lag upon the present is considered separately. As such, it is a simplification of the Markov model, because interaction effects between lags are not considered.

8.3 Non-homogeneous Markov Model

The Markov chain is said to be non-stationary (or non-homogeneous) if the conditional probabilities depend on the time period (n) under consideration. The applicability of Markov chain hypothesis to model transitions of scPHDI values from one

TABLE 8.5: Maximum number of independent parameters for r-th order Markov and MTD models

Number of values m	Order l	Markov chain $M^r \times (m-1)$	MTD model $[m\times(m-1)]+(r-1)$
5	1	20	20
	2	100	21
	3	500	22
	4	2500	23

Remark: The actual number of independent parameters can be lower than the maximum number, because parameters estimated to be zero are not taken into account following the convention in counting degrees of freedom for models for categorical data.

drought class to another, it should be mentioned that such an assumption may not be valid in general. Indeed, under the non-homogeneous lag-1 Markov hypothesis, the lag-M transition probability matrix $\Pi(M)_\tau$, whose generic element (i, j) is given by $P[scPHDI_{v,\tau+M} \in C_j | scPHDI_{v,\tau} \in C_i]$, can be written as (Bremaud, 1999):

$$\Pi(M)_\tau = \Pi_\tau \Pi_{\tau+1} \ldots \Pi_{\tau+M-1} \tag{8.8}$$

where Π_τ is the lag 1 transition probability matrix whose generic element (i,j) is given by $P[scPHDI_{v,\tau+M} \in C_j | scPHDI_{v,\tau} \in C_i]$. The non-homogeneous formulation may be required for the scPHDI, since the homogeneous one would not be able to model the general seasonal pattern observed in transition probabilities.

8.4 Hidden Markov Model (HMM)

The above stochastic processes could be called an observable Markov model since the output of the process is the set of states at each instant of time, where each state corresponds to a physical (observable) event. Hidden Markov models include the case where the observation is a probabilistic function of the state i.e the resulting model is a doubly embedded stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observations.

In Markov chain, the output at time t depends directly on the output at time t-1. In Hidden Markov model, the outputs are conditionally independent. The conditional independence between outputs of a HMM is not always justified. It is entirely possible

to observe processes governed by a hidden markov chain in which successive observations are directly correlated (Berchtold, 1999).

8.5 Double Chain Markov Model (DCMM)

The Double Chain Markov Model (DCMM) is a generalization of both Markov chains and Hidden Markov Models. It is called double since it can be viewed as the superposition of two markov chains, a hidden chain governing the relation between outputs of an observed variable. Advantages of both markov models are conserved: the system is driven by an unobserved process, the successive outputs are directly correlated. The DCMM in discrete time combines two random variables: X_t whose state at time t is unknown for an observer external to the process, and Y_t which is observable . The model is fully described by the following elements (Berchtold, 1999):

- A set of hidden states, $S(X) = \{1, \ldots, M\}$.
- A set of possible outputs, $S(Y) = \{1, \ldots, K\}$.
- The probability distribution of the first hidden state, $\pi = \{\pi_1, \ldots, \pi_M\}$.
- A transition matrix between hidden states, $A = \{a_{ij}\}$, $i, j \in S(X)$.
- A set of transition matrices between successive outputs of Y_t given a particular state of X_t, $C = \{c_{ijs}\}$, $i, j \in S(Y)$, $s \in S(X)$. C can also be written in a more convenient way as $C = \{C^{(s)}\}$ with $C^{(s)} = [c_{ij}^{(s)}]$.

The complete identification of the DCMM requires the estimation of three sets of probabilities : $\{\pi, A, C\}$. The estimation of the model parameters is traditionally obtained by an Expectation-Maximization (EM) algorithm.

A homogeneous Markov chain can be viewed as a DCMM in which the external conditions represented by X never change. So, there is only one hidden state (M=1) and one transition matrix C. On the other hand, when there is $M > 1$ hidden states but each matrix $C^{(s)}$ has identical rows, the model reduces to a HMM.

The Markov chain approach applied to time series of the Palmer Index proved to be a useful tool for early warning aiming at drought management (Lohani and Loganathan,

1997). More recently, Ochola and Kerkides (2003) applied a Markov chain model to predict dry spells.

8.5.1 Maximum-likelihood Estimation

In the case of homogeneous Markov chains of order $f \geq 0$, the maximum likelihood estimator of the probability $p_{i_0|i_f,\ldots,i_1}$ of observing $Y_t = i_0$ given $Y_{t-f} = i_f, \ldots, Y_{t-1} = i_1$ is

$$p_{i_0|i_f,\ldots,i_1} = \frac{n_{i_f,\ldots,i_1,i_0}}{n_{i_f,\ldots,i_1,+}} \tag{8.9}$$

where $n_{i_f,\ldots,i_1,i_0}$ is the number of sequences of the form

$$Y_{t-f} = i_f, \ldots, Y_{t-1} = i_1, Y_t = i_0$$

in the data, and

$$n_{i_f,\ldots,i_1,+} = \sum_{i_0=1}^{K} n_{i_f,\ldots,i_1,i_0}$$

where K is the number of categories of the variable. The log-likelihood is then given by:

$$LL = \sum_{i_0=1}^{K} \cdots \sum_{i_f=1}^{K} n_{i_f,\ldots,i_1,i_0} \, log(p_{i_0|i_f,\ldots,i_1}) \tag{8.10}$$

Due to the structure of the DCMM, there is no direct formula to compute the log-likelihood. The problem is solved using an iterative procedure known as the forward-backward algorithm. The estimation of the model parameters is traditionally obtained by an Expectation-Maximization (EM) algorithm. The procedure used to compute DCMMs is to combine a genetic algorithm (to explore quickly the whole solution space) and an Expectation-Maximization (EM) algorithm (to find the exact parameter values of the optimum). Genetic algorithm is an iterative procedure computing simultaneously several possible solutions (the population). At each iteration, a new population is created by combining the members of the current population using the principles of genetics, that is selection, crossover and mutation. This new population has a better probability to come close to the global optimum than the previous population had. This method presents the advantage to allow the exploration of a large part of the solution space, with a very high probability to find the global optimum.

TABLE 8.6: Comparison of different Markovian models on scPHDI at Ogolcha station

Type and Order	No. of independent parameters	Log-likelihood	BIC
MC1	9	-128.197	307.57
MC2	16	-118.383	327.758
MC3	17	-97.887	**292.452**
MC4	22	-89.98	305.088
MTD2g	11	-123.018	308.592
MTD3g	14	-123.021	325.66
HMM1 *	11	-241.245	545.046
DCMM1-1*	21	-122.103	363.633
DCMM2-1*	25	-121.351	384.876
DCMM1-2*	35	-108.745	416.535
Independent (0)	4	-393.398	809.545

Different automatic selection criteria are available for balancing the apparently contradictory goals of good statistical fit and model simplicity. Among these methods generally based on the log-likelihood, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are now very popular.

BIC is defined as: BIC=-2 LL+ p log(n) where LL is the log-likelihood of the model, p is the number of independent parameters and n is the number of components in the log-likelihood. In both cases, the model with the smallest criterion value is chosen.

To have the same number of components in the log-likelihood of each model (295), it was conditioned on the first 5 data values and so did not include their contributions to the log-likelihood. Table 8.6 summarizes the results of the Markovian model (MC) of orders 1 to 4, the HMM, the DCMM and the independence model. The independence model is worse than any of the models, according to the BIC criterion, showing that there is dependence between successive observations. Among the fully parameterized Markov chains, the best result is achieved by the third-order Markov model (MC3) whose BIC is equal to 292.452. This stresses further that the scPHDI is not an AR1 process as its principal equation implies. Using the MTD reduced the number of independent parameters and slightly improved the log-likelihood of MTD2 against its corresponding Markov chain but it is not better in case of MTD3. On the other hand the higher order DCMM increased the number of independent parameters and subsequently the BIC performed less.

Chapter 9

Prediction of Risks of Hydrological Drought

Assessment of drought risk involves quantifying the frequency of drought and the likely damage. In this chapter assessment of relative risks are presented using univariate and multivariate frequency of drought and associated vulnerabilities in an area.

9.1 Support Vector Machines

Support vector machines seek to minimize an upper bound of the generalization error instead of the empirical error unlike in the other neural networks. The input vectors are mapped into high dimensional feature space through non-linear mapping function $\varphi(x)$. Monthly precipitation at time t and t-l, potential evapotranspiration at time t, and streamflow at time t for the years 1968-1994 are used in the analysis.

The linear function sets $f(x, \alpha) = (w \cdot \varphi(x)) + b$ in feature space are used for estimating the regression function. α is an element of a set of parameters, w determines the orientation of a discriminating hyperplane and b is a bias. w and b are parameters which are induced from the available training data corresponding to the decision function such that it generalizes well. Assuming that the training data set is given as:

$$\{(x_1, y_1), (x_2, y_2) \ldots (x_t, y_t)\} \qquad \text{where } x \in \Re^n \text{ is an N-dimensional input data vector and } y \in \Re \text{ is observed output value.} \tag{9.1}$$

The function is estimated by minimizing the regularized risk function (Vapnik, 1998):

$$\frac{1}{2}\|w\|^2 + C \cdot R_{emp}[f] \tag{9.2}$$

where $\|w\|$ is the norm of vector w. In equation 9.2, the first term $\frac{1}{2}\|w\|^2$ is called the regularized term. Minimizing this term will make a function as flat as possible. The second term $R_{emp}[f]$ is the empirical error measured by the loss function, and C is called the regularization constant which determines the deviation from the loss function. Here the ε-insensitive loss function (Figure 9.1) is used:

$$L_t(y_t, f(x_t)) = max\{0, |y - f(x)| - \varepsilon\}$$

The corresponding constraint optimization problem is:

$$min\frac{1}{2}\|w\|^2 + C\frac{1}{\imath}\sum_{i=1}^{t}\xi_i + \xi^* \cdot R_{emp}[f] \tag{9.3}$$

$$Subject\,to = \begin{cases} y_i - (w \cdot \varphi(x_i)) - b \leq \varepsilon + \xi_i^*, & i = 1, 2 \ldots t \\ (w \cdot \varphi(x_i)) + b - y_i \leq \varepsilon + \xi_i^*, & i = 1, 2 \ldots t \\ \varepsilon, \xi_i^* & i = 1, 2 \ldots t \end{cases}$$

where ε, ξ_i are slack variables.

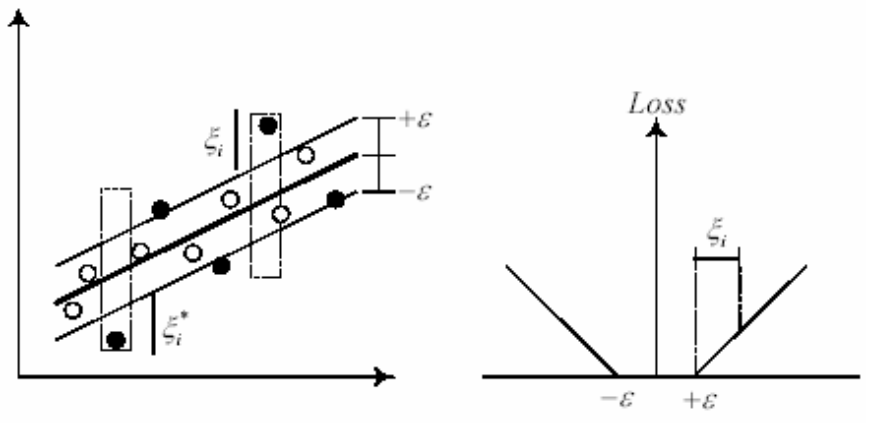


FIGURE 9.1: The soft margin loss setting for a linear SVM and ε-insensitive loss function.

9.1.1 Results and Discussion

Time series of streamflow records at station Wabi at the bridge from January 1968 to December 1981 is used for training. The data from January 1982 to December 1994 is used for validation. The results of the training and validation of the data set is provided in Figure 9.2. Flows that exceed the mean flow by more than twice the standard deviation are considered as high flows and are given less weight in the simulation. The variability in the monthly lowflow records is fairly well captured in the simulation. Radial basis function (RBF) is found to be a better kernel option in simulation with γ of 0.2 and σ square of 28. The streamflows are further truncated by a threshold level to obtain time series of severity and duration of drought. The univariate frequencies of the observed and simulated drought severity and duration events are compared. The frequency of the severity/average is much better simulated than the corresponding drought duration (Fig 9.3). The results signify that using support vector machines it is possible to sim-

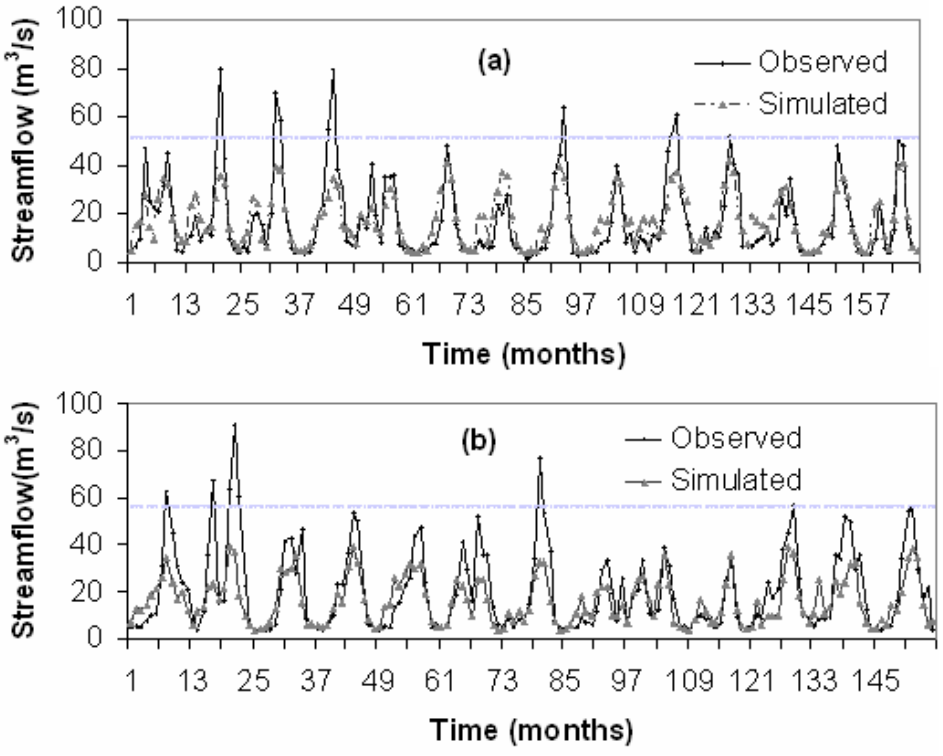


FIGURE 9.2: Time series of observed and simulated monthly streamflow records (a) training data set (Jan 1968 - Dec 1981) (b)validation data set (Jan 1982 - Dec 1994) at Wabi at the bridge

ulate the low flows using climatic variables and examine associated probable drought incidences. The length of severity and duration of drought time series used in the above

analysis is dependent on the length of the input climatic data which in most cases is of short span. Thus this frequency estimates may have the disadvantage that it may over/underestimate the frequencies.

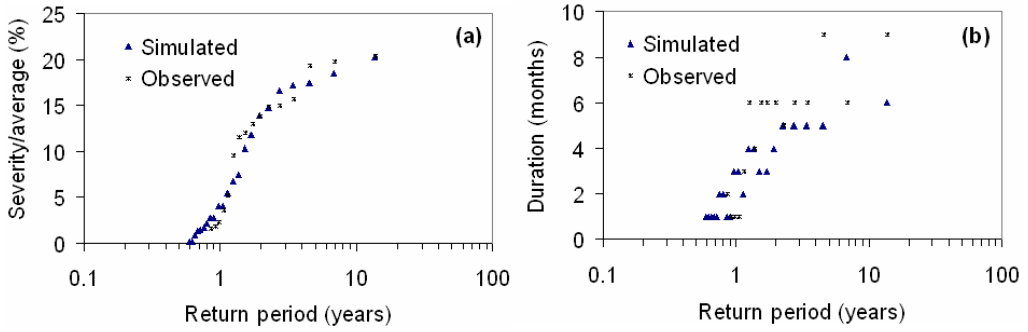


FIGURE 9.3: Observed and simulated return periods (log-scale) of (a) drought severity/average (b) drought duration at Wabi at the bridge

Separate analysis of drought characteristics cannot reveal the significant correlation relationships among drought characteristics. It is also essential to explore the bivariate characteristics of the drought to give a better picture of the events.

9.2 Copula Modelling

A copula is a multivariate distribution function that links marginal distributions together to form a joint distribution, and completely describes the dependence structure between the variables. Using a copula to build a multivariate distribution is flexible because no restrictions are placed on the marginal distributions. Copulas are new way of modeling the correlation structure between variables. Below it is focused on the bivariate case, though both the theory of copulas and the estimation methods presented here extend quite naturally to the general multivariate case.

Let two random variables x and y have a cumulative distribution function of $F_x(x)$ and $F_y(y)$ respectively and C denote the dependence function copula, then the bivariate distribution of x and y (i.e. $F_{xy}(x, y)$) is given as:

$$F_{xy}(x, y) = C(F_x(x), F_y(y)) \tag{9.4}$$

We will assume here that the distribution functions $F_x(x)$, $F_y(y)$ and $F_{xy}(x, y)$ are continuous, and sufficiently smooth for all required derivatives to exist. In the bivariate

case, differentiating equation 9.4 then the joint density is the product of the copula density, c and the marginal densities, $f_x(x)$ and $f_y(y)$.

$$f_{xy}(x,y) = f_x(x).f_y(y).c(F_x(x), F_y(y)) \tag{9.5}$$

As usual, we denote a cumulative distribution function (CDF) with an upper case letter and the probability density function (PDF) with a lower case letter. The different classes of copula include elliptical copula, extreme copulas, Archimedean copula and others. Elliptical copulas are copulas generated from elliptical multivariate distributions. Gaussian copula and t-copula belong to this class of copula. The other important class of copulas is the extreme value class (EV Copula). Let $u = F_x(x)$: marginal distribution of X and $v = F_y(y)$: marginal distribution of Y , a copula is said to be an EV copula if $\forall \quad t > 0$ and, the scaling property $C(u^t, v^t) = (C(u,v))^t$ holds $\forall (u,v) \in I^2$. The Gumbel-Hougard copula, Galambos copula, Husler and Reiss copula and BB5 copula are members of the EV class of copula.

Among the different classes of copulas, Archimedean copulas have wide range of applications. These copulas are very easy to construct and many parametric families belong to this class. These classes of copulas allow for a great variety of different dependence structures and all commonly encountered Archimedean copulas have a simple closed form expressions.

The one parameter Archimedean copula, denoted as C_ϕ , has the general form:

$$C_\phi(u,v) = \phi^{-1}(\phi(u) + \phi(v)) \qquad u, v \in (0,1] \tag{9.6}$$

where ϕ denotes a generator of the copula C_ϕ, which is in fact a distribution function. Let $\phi(\bullet)$ be the copula generator that is a convex decreasing function satisfying $\phi(1) = 0$; and $\phi^{-1}(\bullet)$ is equal to 0 when $v \geq \phi(0)$. The Archimedean copula representation permits reducing a multivariate formulation in terms of a single univariate function. The Ali-Mikhail-Haq copula, Clayton copula, Frank copula, Joe copula, BB1, BB2, BB6 copula are some of the members of the Archimedean class of copulas. The Gumbel-Hougard copula belongs to both extreme value copula and Archimedean copula. There are different choices for the generator so as to yield different families of copulas to model specific feature of dependence.

The dependence parameters are tied with the measures of association that capture more than a linear dependence unlike the well known correlation coefficient. Spearman's ρ and Kendal's τ are the two most commonly used measures of association. For the Archimedean copulas, Kendall's τ can be evaluated directly from the generator.

$$\tau_c = 1 + 4\int_0^1 \frac{\phi(t)}{\phi\prime(t)}\,dt \tag{9.7}$$

The simplest choice of generator $\phi = -ln(t)$ gives an independent copula family.

$$\begin{aligned} C_\phi(u,v) &= \phi^{-1}(\phi(u) + \phi(v)) \\ &= exp\{-(-(ln(u) - ln(v)))\} \\ &= exp(ln(uv)) = uv \end{aligned}$$

This is the case of the joint distribution of u and v when they are independent.

9.2.1 Families of Copulas

A collection of 22 one-parameter families of Archimedean copulas can be found in Nelsen (1999). In this study, the copulas investigated include Gumbel-Hougard family, Ali-Mikhail-haq family, Farlie-Gumbel-Morgenstern family, Frank Family, Clayton /Cook-Johnson Family and Galambos family.

Gumbel-Hougaard Family

This family is a member of both EV and Archimedean copula classes. Let u or v, be values of a uniformly distributed random variable varying from 0 to 1. Its generating function (with t=u or v) is expressed as $\phi(t) = (-ln(t))^\theta$. Substitution of $\phi(t)$ into Equation 9.6 yields:

$$\begin{aligned} C_\phi(u,v) = C_\phi[F_x(x), F_y(y)] &= F_{xy}(x,y) \\ = exp\{-[(-ln(u))^\theta + (-ln(v))^\theta]^{1/\theta}\} & \\ \tau = 1 - \theta^{-1} \quad \theta \in [1,\infty) & \end{aligned} \tag{9.8}$$

Where θ is a parameter of the generating function, τ is Kendall's coefficient of correlation. The relationship between Kendall's τ and the generating function shows that for the

Gumbel-Hougaard copulas, only the positive dependence structure of the bivariate data can be analyzed.

Ali-Mikhail-Haq Family

It is a family of Archimedean copula with the generating function expressed as (t=u or v):

$$\phi(t) = ln\frac{1-\theta(1-t)}{t} \tag{9.9}$$

$$C_\phi(u,v) = C_\phi[F_x(x), F_y(y)] = F_{xy}(x,y)$$

$$= \frac{uv}{1-\theta(1-u)(1-v)} \quad \theta \in [-1,1) \tag{9.10}$$

$$\tau = \left(\frac{3\theta-2}{\theta}\right) - \frac{2}{3}\left(1-\frac{1}{\theta}\right)^2 ln(1-\theta), \quad \tau \in [-0.182, 1/3) \tag{9.11}$$

The Ali-Mikhail-Haq copula can be applied only to a range of slightly negative to positive dependent random variables as by Kendall's τ.

Farlie-Gumbel-Morgenstern Family

This family of copula belongs to other class of copula than the common types of Elliptical, Archimedean, or EV class of copulas.

$$C_\phi(u,v) = uv(1+\theta(1-u)(1-v)) \quad \theta \in [-1,1] \tag{9.12}$$

$$\tau = \frac{2\theta}{9} \in [-2/9, 2/9] \tag{9.13}$$

Frank Family

It is a family of Archimedean copula with the generating function expressed as (with t=u or v):

$$\phi(t) = ln\left[\frac{exp(\theta\, t)-1}{exp(\theta)-1}\right] \tag{9.14}$$

$$\tau = 1 - \frac{4}{\theta}[D_1(-\theta)-1] \tag{9.15}$$

where D_1=first order Debye function D_k which for positive argument is defined as

$$D_k(\theta) = \frac{k}{x^k} \int_0^\theta \frac{t^k}{exp(t) - 1}\, dt \quad \theta > 0 \tag{9.16}$$

and the Debye function D_k with negative argument can be expressed as:

$$D_k(-\theta) = D_k(\theta) + \frac{k\theta}{k+1} \tag{9.17}$$

$$C_\phi(u, v) = C_\phi[F_x(x), F_y(y)] = F_{xy}(x, y)$$

$$= \frac{1}{\theta} ln \left[1 + \frac{[exp(\theta\, u) - 1][exp(\theta\, v) - 1]}{exp(\theta) - 1}\right] \quad \theta \neq 0 \tag{9.18}$$

Similar to the Ali-Mikhail-Haq copula, the Frank copula can be applied to deriving joint distributions of both negatively as well as positively dependent random variables. There is no restriction on the degree of correlation.

Clayton /Cook-Johnson Family

It is a family of Archimedean class of copula with a generating function of:

$$\phi(t) = t^{-\theta} - 1 \tag{9.19}$$

$$C_\phi(u, v) = C_\phi[F_x(x), F_y(y)] = F_{xy}(x, y)$$

$$= [u^{-\theta} + v^{-\theta} - 1]^{-1/\theta} \quad \theta \geq 0 \tag{9.20}$$

$$\tau = \frac{\theta}{\theta + 2} \tag{9.21}$$

Similar to the Gumbel-Hougaard copula, the Cook-Johnson copula is suitable only for positively correlated random variables.

Galambos Family

This is a family of EV class of copulas.

$$C_\phi(u, v) = uv[(-ln(u))^{-\theta} + (-ln(v))^{-\theta}]^{\frac{-1}{\theta}} \quad \theta \in [0, \infty) \tag{9.22}$$

$$\tau = \frac{\theta + 1}{\theta} \int_0^1 \left[\frac{1}{t^{1/\theta}} + \frac{1}{(1-t)^{1/\theta}} - 1\right]^{-1} dt \tag{9.23}$$

This family of copulas is suitable only for positively correlated random variables. Severity and duration of drought are positively correlated, thus the above families of copulas may be considered as candidates to model their joint distribution.

9.2.2 Identification of Copulas

9.2.3 Fitting Copulas to Hydrological Drought Variables

For fitting copulas to the bivariate hydrological drought variables of severity and duration, it is important to identify a copula and select best fitted parameters. Parametric, semi-parametric and non-parametric procedures are available for selecting appropriate values of dependence parameters in a copula based model.

9.2.3.1 Non-parametric Estimation

Given a random sample of bivariate observations (x_i, y_i) for $i = 1, \ldots, n$ and a joint distribution function F(x,y) with an associated Archimedean copula C ; one may want to identify the form of the generator . First it is important to define an intermediate (unobserved) random variable t_i = F(x_i, y_i) that has distribution function K(t) = Prob$(t_i > t)$. This distribution function is related to the generator of the Archimedean copula through the expression:

$$K(t) = K_\phi(t) = t - \frac{\phi(t)}{\phi\prime(t)} \tag{9.24}$$

The following procedure is adopted to identify the generating function (Genest and Rivest, 1993):

1. Find Kendall's tau using the usual (non-parametric or distribution-free) estimate

$$\tau_n = \binom{n}{2}^{-1} \sum_{j=1}^{i-1} Sign[(x_i - x_j)(y_i - y_j)] \tag{9.25}$$

Where n is the number of observations,

$$Sign = 1, \qquad if\,(x_i - x_j)(y_i - y_j) > 0$$
$$Sign = -1, \qquad if\,(x_i - x_j)(y_i - y_j) < 0$$

2. Construct a non-parametric estimate of K, as follows: a) first, define the pseudo-observations

$$t_i = (n-1)^{-1} \sum_{j=1}^{n} \text{if}[x_j < x_i \,\&\&\, y_j < y_i, 1, 0], \text{for i} = 1, \ldots, n \tag{9.26}$$

b) second, construct the estimate of K

$$K_n(t) = n^{-1} \sum_{i=1}^{n} \text{if}[t_i \leq t, 1\, 0] \tag{9.27}$$

where function If [condition, 1, 0] gives 1 if condition holds, and 0 otherwise.&& stands for logic operator "and".

3. Now construct a parametric estimate K_ϕ using equation 9.24. Illustratively,$\tau_n \rightarrow \theta n \rightarrow \phi n(t) \rightarrow K_{\phi\, n}(t)$, where subscript n denotes estimate.

The step 3 is to be repeated for every copula family that is needed to be compared. The best choice of generator then corresponds to the parametric estimate $K_{\phi\, n}(t)$ that most closely resembles the non-parametric estimate $K_n(t)$.

Comparing the appropriateness of the selected generating function can be done either by a distance measurement such as $\int_0^1 [K_{\phi\, n}(t) - K_n(t)]^2 dt$ or graphically by corresponding quantile-quantile (Q-Q) plots. Q-Q plots are used to determine whether two data sets come from populations with a common distribution. If the points of the plot, which are formed from the quantiles of the data, are roughly on a line with a slope of 1, then the generating function is satisfactory. Otherwise the copula function needs to be redefined.

9.2.3.2 Parametric Estimation

The parametric estimation procedure, often referred as Inference of margins (IFM), is a two step procedure as proposed by Joe and Xu (1996). Unlike the rank based methods, the marginal distributions are computed from the observed values. The estimate of θ is obtained through the maximization of the log-likelihood function of θ :

$$L(\theta; \alpha_1, \ldots, \alpha_d) = \sum_{i=1}^{n} log\, f(x_i, y_i; \alpha_1, \ldots, \alpha_d) \tag{9.28}$$

And the maximum likelihood estimator of θ is:

$$\hat{\theta} = arg\, max\, L(\theta)$$

The IFM method consists of doing d separate optimisations of the univariate likelihoods, followed by an optimisation of the multivariate likelihood as a function of the dependence parameter vector. It consists of the following two steps: 1. the log-likelihoods L_j of the d univariate marginals are separately maximised to get estimates of $\hat{\alpha_1}, \hat{\alpha_2}, \ldots, \hat{\alpha_d}$ respectively.

$$L_j(\alpha_j) = \sum_{i=1}^{n} log\, f_j(y_{ij}; \alpha_j), \quad j = 1, \ldots, d \tag{9.29}$$

2. the function $L(\theta, \hat{\alpha_1}, \hat{\alpha_2}, \ldots, \hat{\alpha_d})$ is maximised over θ to get $\hat{\theta}$. That is, under regularity conditions,$(\hat{\alpha_1}, \hat{\alpha_2}, \ldots, \hat{\alpha_d}, \theta)$ is the solution of

$$(\partial\, L_1/\partial\, \alpha_1, \ldots, \partial\, L_d/\partial\, \alpha_d, \partial\, L/\partial\, \theta) = 0 \tag{9.30}$$

9.2.3.3 Semi-parametric Estimation

In Semi-parametric procedure, unlike the IFM approach, which assumes that we have specified parametric univariate marginal distributions, we do not have to make any assumption on the distributions. When non-parametric estimates are contemplated for the marginals and the focus of the analysis is on the dependence structure, it requires the dependence parameter to be margin-free. We use the empirical data and transform the sample data into uniform variates.

This procedure consists of (a) transforming the marginal observations into uniformly distributed vectors using the empirical distribution function, and (b) estimating the copula parameters by maximizing a pseudo log-likelihood function.

So, given a random sample as before, we look for θ that maximizes the pseudo log-likelihood

$$L(\theta) = \sum_{i=1}^{n} log(c_\theta(F_n(x), G_n(y))), \tag{9.31}$$

in which F_n, G_n stands for re-scaled empirical marginal distributions functions, i.e.,

$$F_n(x) = \frac{1}{n+1}\sum_{i=1}^{n} \text{if}[x_i \leq x, 1, 0]$$
$$G_n(y) = \frac{1}{n+1}\sum_{i=1}^{n} \text{if}[y_i \leq y, 1, 0] \tag{9.32}$$

This re-scaling avoids difficulties from potential unboundedness of log(c (u, v)) as u or v tend to one. Genest et al. (1995) examined the statistical properties of the proposed estimator and proved it to be consistent, asymptotically normal and fully efficient at the independence case. The copula density c for each Archimedean copula can be acquired from the differentiation of equation 9.6.

It is essential to model the dependence structure between the severity and duration of the hydrological drought variables in order to appropriately designate their joint frequency characteristics. Copulas provide a method for describing the correlation structure between random variables that is independent of their individual marginals. The dependence between the random variables is accommodated by copula method through measure of dependence based on ranks, Kendall's τ. The Kendall's τ between the severity and duration of drought obtained from the elongated time series at Imi is found to be 0.7192. Among the copulas families considered, the obtained Kendall's value is outside the domain of Ali-Mikhail-Haq family and Farlie-Gumbel-Morgenstern copulas. Thus, these copulas are automatically excluded from further analysis.

The bivariate copula density plots illustrate that the choice of the functional form of the copula has a dramatic impact on the shape of the joint distribution, especially in the corners of the distributions (Figure 9.4). For example, notice the difference in the areas under the densities near the extreme corner points (0,0) and (1,1). The area under the densities at the low corner (0,0) represents the probability that both random variables take on an extreme low value. The area under the densities at the extreme high corner (1,1) represents the probability that both random variables take on an extreme high value. The Clayton copulas place a greater probability on extreme low/low events occurring compared to the other copulas considered in this analysis. The Gumbel-Hougaard and Galambos copulas place a greater probability on extreme high/high events occurring where as Frank copula places an approximately equal probability on extreme low/low events and extreme high/high events because the area underneath each of the respective

densities at the extreme low corner is approximately equal to the area underneath the density at the extreme high corner. It should be noted that while it is possible to compare the probabilities of extreme events occurring across the copulas, it is considerably more difficult to make such comparisons of the probabilities of other events occurring by simply looking at the density plots.

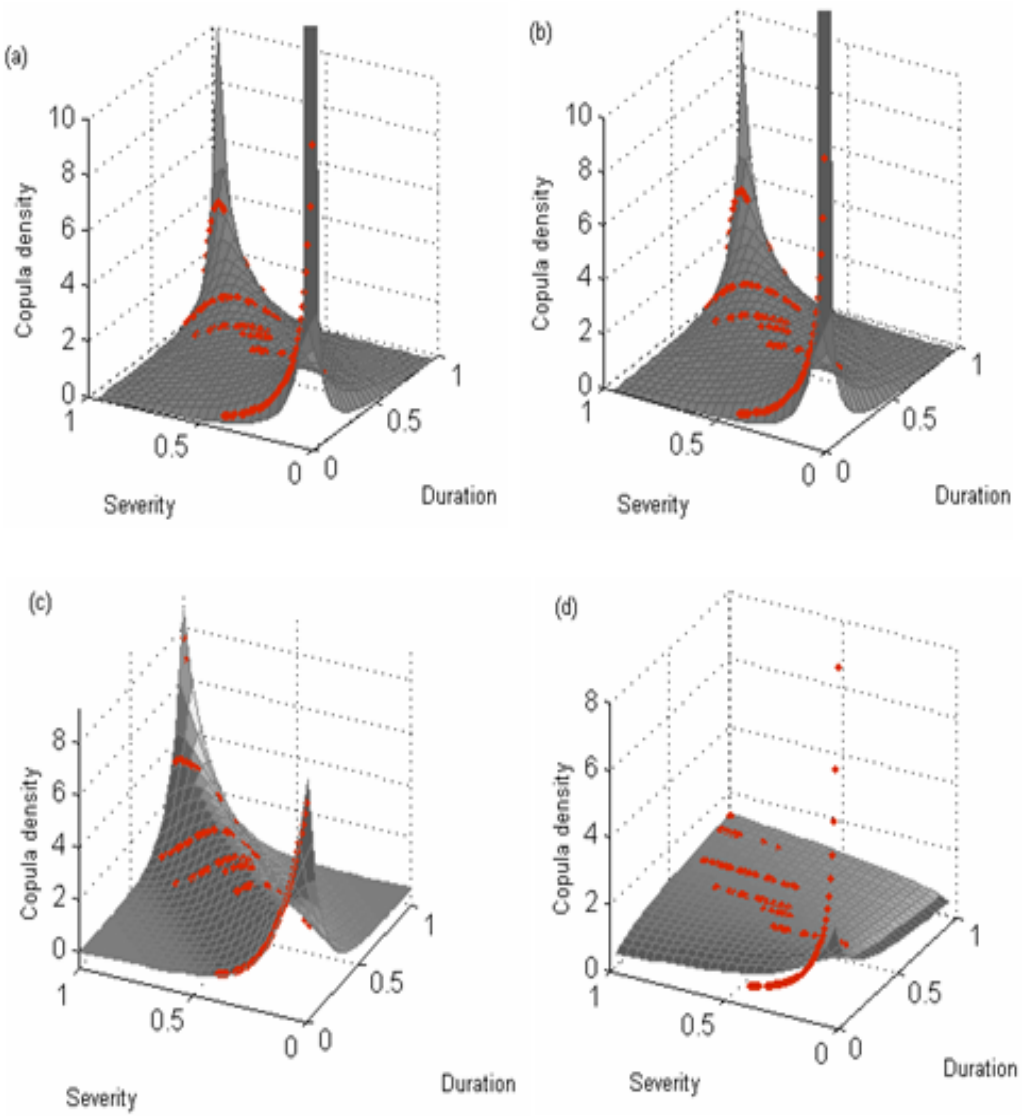


FIGURE 9.4: Copula density of theoretical and observed bivariate (severity and duration) drought series (a) Gumbel-Hougard copula (b) Galambos copula (c) Frank copula and (d) Clayton copula using semi-parametric procedure.

The copula parameters are estimated using three procedures: non-parametric, semi-parametric and parametric procedures. Graphical plots of the copula density estimates of the theoretical and observed bivariate drought series did not clearly show the appropriate copula to select. The plot in the case of the Gumbel-Hougard family and Galambos

TABLE 9.1: Non-parametric, semi-parametric and parametric estimates of copula dependence parameters.

Family	Gumbel-Hougard	Clayton	Frank	Galambos
Non-parametric procedure				
θ	3.561	5.122	12.346	
$d(K_\phi, K_n)$	0.0110	0.0132	0.0100	
Semi-parametric (Maximum pseudo-likelihood)				
θ	2.091	0.378	8.631	1.377
AIC	-134.5318	-50.7616	-160.3838	-135.0888
Parametric (Inference from margins)				
θ	1.810	1.615	5.410	1.111
AIC	-590.8850	-614.1150	-609.5292	-593.6926

family have similar characteristics where as the Clayton copula shows a distinct feature. The tail dependence is important in discriminating between different parametric copulas. It is necessary to numerically identify the suitability of the copulas to model the bivariate characteristics of the hydrological drought variables.

9.2.4 Results and Discussion

Goodness-of-Fit Statistics

The Akaike information criterion (AIC), developed by Akaike (1974), is used to discriminate among the models. AIC can be expressed as given in equation 7.11. AIC tries to tradeoff between reducing the number of parameters fitted and increasing the joint probability of occurrence. The adjustment to the maximized likelihood penalizes models that employ more parameters. The smaller the value of AIC, the better the model fits. In the family of copulas considered the number of parameters fitted is the same so maximum likelihood estimate alone can also be used instead of AIC. The copula parameter estimates obtained through the non-parametric, semi-parametric and parametric procedures and their corresponding goodness of fit for the different copulas is provided in Table 9.1.

Neither method is generally more convenient, but if there are outliers or if the marginal distributions are heavy tailed, it seems reasonable to choose the nonparametric approach. If we work with large data set, the likelihood estimator may be more precise. Although they usually perform well, the estimates of the association parameters derived by the IFM technique clearly depend on the choice of the marginal distributions, and thus always run the risk of being unduly affected if the models selected for the margins turn out to be inappropriate. The non-parametric procedure can be used to guide the

selection of a suitable parametric family of Archimedean copula. However, not every Archimedean copula allows the non-parametric or log-likelihood approach to estimating its parameter. Some of them do not provide closed form of equation 9.7 needed for the non-parametric method, or closed form of copula density function needed in log-likelihood function. Galambos family is not part of Archimedean class of copulas for which the closed form is defined. Thus the non-parametric procedure is not applied to this family of copula.

Another restriction comes from the inability of many copulas to capture the whole range of dependence. Only Frank copula can model the whole range of dependence. Clayton and Gumbel copulas are both extreme members of the Archimedean class of copulas. Interestingly, the asymmetry of the predicted values associated with Gumbels and Claytons model appear to be different. With Gumbels copula there is a clustering effect to the upper tail of the joint distribution, but this asymmetry is not as severe as in Claytons one-parameter system (Genest and Rivest, 1993).

In Table 9.1, it can be observed that in the non-parametric estimation procedure, the distance between the parametric and non-parametric estimates of K is smallest in case of Frank copula and largest in case of Clayton. However, there seems to be a disharmony in the goodness of fit statistic with AIC by parametric estimation procedure, which surprisingly promotes the Clayton followed by Frank copula. In the non-parametric and semi-parametric procedures, Frank copula performs better where as in the IFM procedure (parametric) Clayton copula appears more appropriate. The log-likelihood function of Clayton copula density is pretty sensitive to lower tail dependencies. The parameter estimates obtained using the different procedures show wide variability in case of Clayton. The performance of the EV copulas is not appreciably different. The Galambos copula marginally performs better than the Gumbel-Hougard copula. Since Frank copula is able to capture the whole range of dependence and its performance using the different estimation techniques, it is selected to model the severity and duration of the hydrological drought in the study area.

There is no significant outlier in the dataset. The prior specification of the marginal distributions leads to potential misspecification problems.The parameter estimate of Frank copula ($\theta = 8.63$) identified using semi-parametric procedure is used in the analysis. The joint probability distribution of severity and duration of drought using this parameter estimate of Frank copula is shown in Figure 9.5.

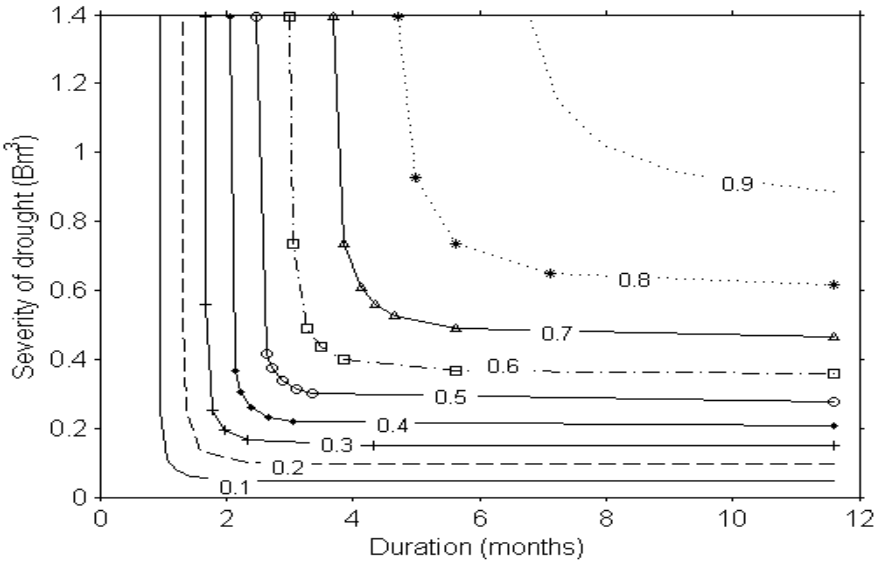


FIGURE 9.5: Joint probability distribution of severity and duration of drought using Frank copula.

Confidence Interval

Hydrologic uncertainty can be classified into three types: inherent, parameter, and model uncertainties. The occurrence of various hydrological events such as stream flow or rainfall events are considered as stochastic processes because of the observable natural, or inherent, randomness. The model uncertainty in many cases results from the lack of data and knowledge adequate to select the appropriate probability model or through the use of an over simplified model. Because of the lack of perfect hydrological information about these processes or events there exist informational uncertainties about the processes. These uncertainties are referred to as the parameter uncertainties and the model uncertainties. Parameter uncertainties reflect the variability in the determination of the parameters to be used in the model. Data uncertainties include a.)measurement inaccuracy and errors, b.) Inadequacy of the data gauging network, and c.)data handling a nd transcription errors. Here the confidence interval for the parameter estimated by semi-parametric procedure is presented.

The procedure for estimating a multivariate θ and computing associated approximate confidence region is described by Genest et al. (1995). Only the case for θ is real is presented below.

Letting $c_\theta(u,v) = \partial\, c_\theta(u,v)/\partial\theta$, Genest et al. (1995) show under mild regularity conditions that the root $\hat{\theta}_n$ of the equation:

$$\dot{\ell}(\theta) = \frac{\partial}{\partial\theta}\ell(\theta) = \sum_{i=1}^{n} \frac{\dot{c}_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)}{c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)} = 0 \tag{9.33}$$

is unique. Furthermore

$\hat{\theta}_n \approx N\left(\theta, \frac{v^2}{n}\right)$

where v^2 depends exclusively on the true underlying copula C_θ. A consistent estimate of v^2 is given by:

$$\hat{v}_n^2 = \hat{\sigma}_n^2/\hat{\beta}_n^2 \tag{9.34}$$

where

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(M_i - \bar{M})^2$$

and

$$\hat{\beta}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(N_i - \bar{N})^2$$

are sample variances computed from two sets of pseudo observations with means $\bar{M} = (M_1 + \cdots + M_n)/n$ and $\bar{N} = (N_1 + \cdots + N_n)/n$, respectively.

To compute the pseudo-observations M_i and N_i, one should proceed as follows:

- 1: Relabel the original data $(X_1, Y_1), \ldots, (X_n, Y_n)$ in such a way that $X_1 < \cdots < X_n$; as a consequence one then has $R_1 = 1, \ldots, R_n = n$.
- 2: Write $L(\theta, u, v) = logc_\theta(u, v)$ and compute L_θ, L_u, and L_v, which are the derivatives of L with respect to θ, u, and v, respectively.
- 3: For $i \in \{1, \ldots, n\}$, set

$$N_i = L_\theta\left(\hat{\theta}_n, \frac{i}{n+1}, \frac{S_i}{n+1}\right) \tag{9.35}$$

- 4: For $i \in \{1, \ldots, n\}$,let also

$$M_i = N_i - \frac{1}{n}\sum_{j=i}^{n} L_\theta\left(\hat{\theta}_n, \frac{j}{n+1}, \frac{S_j}{n+1}\right) L_u\left(\hat{\theta}_n, \frac{j}{n+1}, \frac{S_j}{n+1}\right) - \frac{1}{n}\sum_{S_j \geq S_i}^{n} L_\theta\left(\hat{\theta}_n, \frac{j}{n+1}, \frac{S_j}{n+1}\right) L_v\left(\hat{\theta}_n, \frac{j}{n+1}, \frac{S_j}{n+1}\right) \tag{9.36}$$

An approximate $100(1-\alpha)\%$ confidence interval for θ is then given by:

$$\hat{\theta}_n \pm z_{\alpha/2}\frac{v}{\sqrt{n}}$$

The 95% confidence interval for this estimate is found to be [8.4032, 8.8567].

Conditional probability of the duration of drought given that the severity is less than a certain amount (s) is given as:

$$p(D < d/S < s) = \frac{P(D < d, S < s)}{P(S < s)} = \frac{C(D < d, S < s)}{F(s)} \tag{9.37}$$

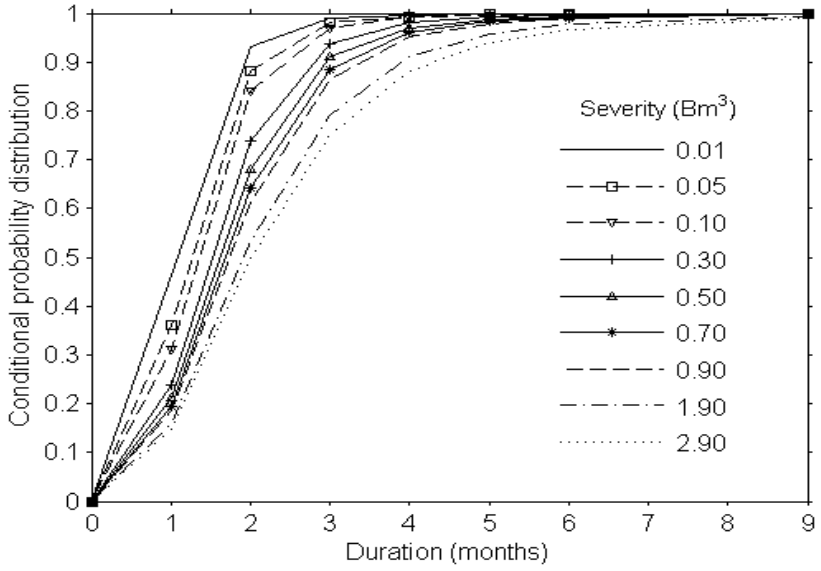


FIGURE 9.6: Conditional probability distribution of drought duration given that the severity is less than a certain amount, s.

At duration of 3 months the variation of the conditional probability distribution between the less severe and more severe conditions is the widest (Figure 9.6). Similarly for the conditional probability distribution of severity of drought given that the duration is less

than a certain amount (d) can be established as:

$$p(S < s/D < d) = \frac{P(D < d, S < s)}{P(D < d)} = \frac{C(D < d, S < s)}{F(d)} \tag{9.38}$$

The chance of occurrence of less severe drought events with duration 2 months and lower is quite high (Figure 9.7). Shiau (2003) proposed a methodology that categorizes the

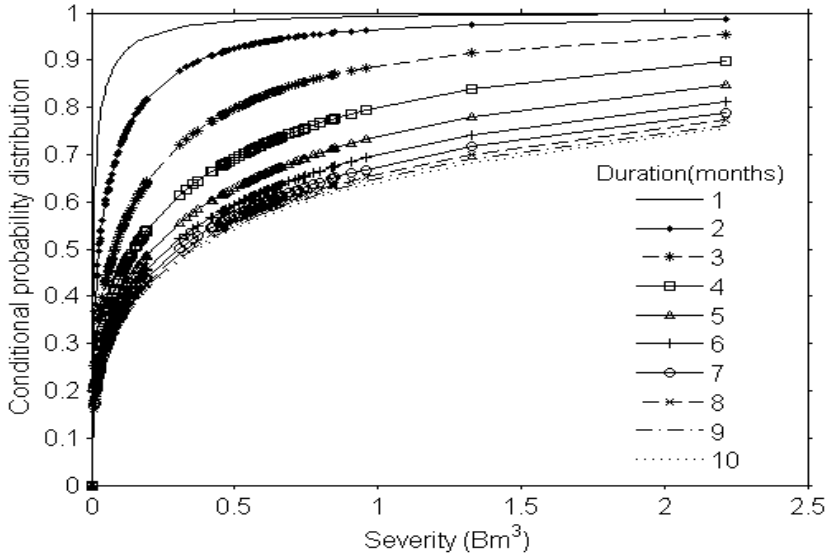


FIGURE 9.7: Conditional probability distribution of drought severity given that the duration is less than a certain amount, d.

return periods of bivariate distributed hydrologic events as joint and conditional return periods. The joint drought duration and severity return periods can be defined in two cases: return period for $D > d$ and $S > s$ and return period for $D > d$ or $S > s$. Both joint return period definitions for the copula-based drought events are described below.

$$\begin{aligned} T_{DS} = \frac{E(L)}{P(D \geq d, S \geq s)} &= \frac{E(L)}{1 - F_D(d) - F_S(s) + F_{DS}(d, s)} \\ &= \frac{E(L)}{1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))} \end{aligned} \tag{9.39}$$

$$T\prime_{DS} = \frac{E(L)}{P(D \geq d \, or \, S \geq s)} = \frac{E(L)}{1 - F_{DS}(d, s)} = \frac{E(L)}{1 - C(F_D(d), F_S(s))} \tag{9.40}$$

where T_{DS} denotes the joint return period for $D \geq d$ and $S \geq s$; T'_{DS} denotes the joint return period for $D \geq d$ or $S \geq s$. Figure 9.8 and Figure 9.9 show the joint drought duration and severity return period for ($D \geq d$ and $S \geq s$) and for ($D \geq d$ or $S \geq s$) respectively at Imi station.

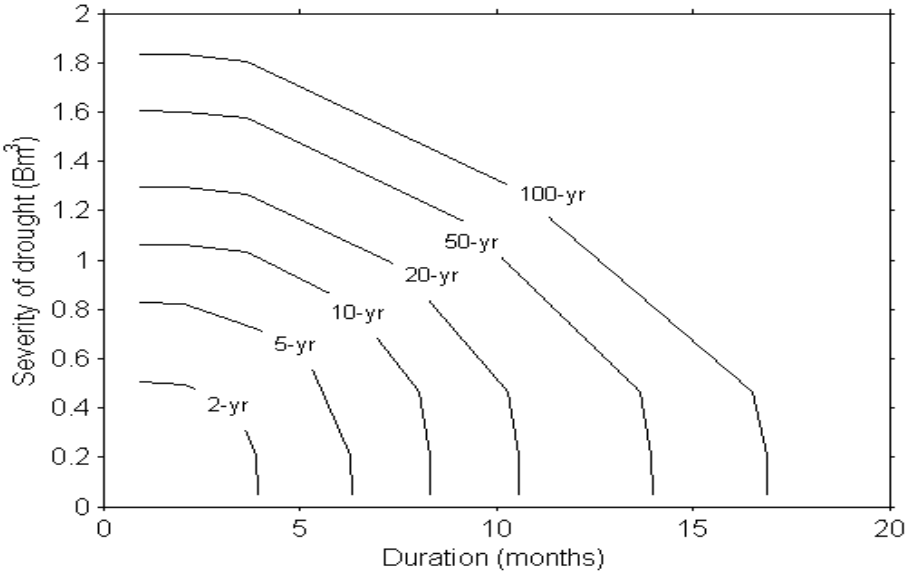


FIGURE 9.8: Joint drought duration and severity return period T_{DS} ($D \geq d$ and $S \geq s$)

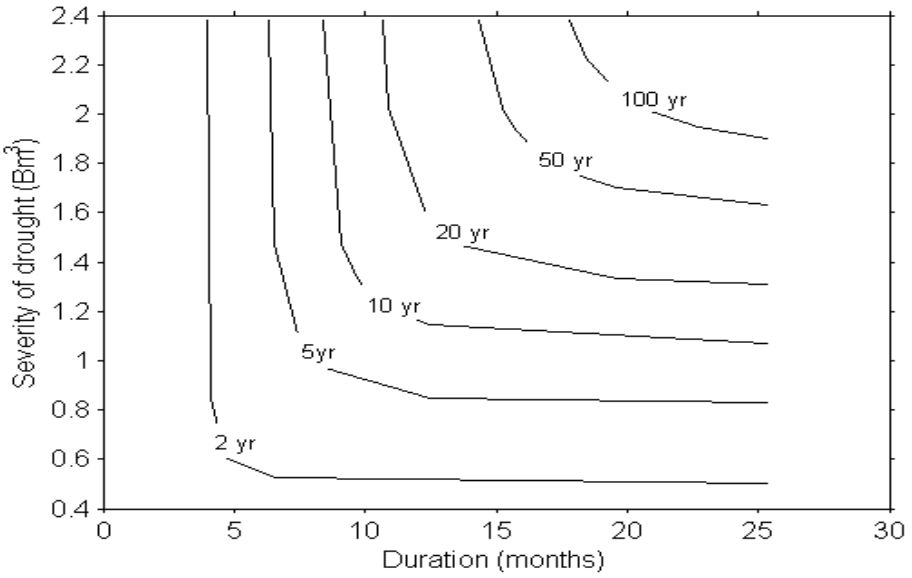


FIGURE 9.9: Joint drought duration and severity return period T'_{DS} ($D \geq d$ or $S \geq s$)

The conditional return period of drought duration given drought severity exceeding a certain threshold and the return period of drought severity given drought duration exceeding a certain threshold are both defined as follows:

$$T_{D|S\geq s} = \frac{T_S}{P(D \geq d,\, S \geq s)}$$
$$= \frac{E(L)}{[1 - F_S(s)][1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))]} \tag{9.41}$$

$$T_{S|D\geq d} = \frac{T_D}{P(D \geq d,\, S \geq s)}$$
$$= \frac{E(L)}{[1 - F_D(d)][1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))]} \tag{9.42}$$

where $T_{D|S\geq s}$ denotes the conditional return period for D given $S \geq s$; and $T_{S|D\geq d}$ denotes the conditional return period for S given $D \geq d$. Figures 9.10 demonstrates the conditional return periods of drought duration defined by Equations 9.41. The joint and conditional return periods of drought duration and severity can be used to evaluate the risk for malfunction of a specific water resources system.

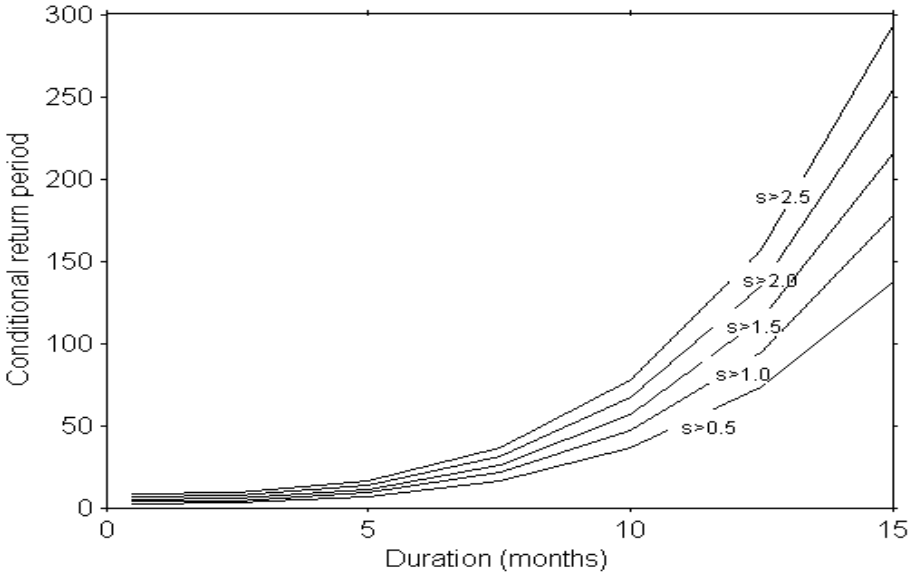


FIGURE 9.10: Conditional return period of drought duration given that the severity is less than a certain amount, s (Bm^3).

9.2.5 Conclusion

Hydrological drought events are extracted from the reconstructed time series of monthly streamflow records. Two-parameter Weibull and two-parameter lognormal distributions fit well the marginal distribution of severity and duration of the hydrological drought respectively. Since the marginal distributions are different, copulas are used to model the dependence structure between the variables. The nonlinear dependency between severity and duration of hydrological drought is studied using copula models. Different copula families and parameter estimation techniques are evaluated. The goodness of fit is tested graphically and using Akaike Information criterion (AIC). Among the different families of copulas, it is hard to detect the most appropriate copula graphically. According

to goodness of fit statistic it is found numerically that Frank copula performs better. The kendall-tau parameter of 0.7192 is outside the domains of the Ali-Mikhail-Haq and Farlie-Gumbel-Morgenstern family so these families are not considered in the evaluation of the copulas. Joint and conditional probabilities of severity and duration of drought in the area is specified from the derived relationship to provide empirical insight on the nature of the extreme events.

9.3 Drought Risks

Risk is generally defined as the probability of a specific undesired event (Andrews and Moss, 2002). The risk that a particular drought event may occur in a given time period n may be of interest in planning water resources systems. For example, if a water supply system has been designed to cope with a certain level of droughts, one may want to determine the risk that drought events exceeding such an expectation may occur during the time period n. Such a risk can be determined from the historical or from the generated samples (Salas et al., 2005).

$$R_{T,n} = 1 - \left(1 - \frac{1}{T}\right)^n$$

This method considers the random nature of only the natural events, such as flood or droughts, and accounts for their statistical characteristics through frequency analysis. Alternatively, risk or expected loss can be defined quantitatively as the product of the consequences of a specific incident and the probability over a time period or frequency of its occurrence. In flood studies the risk is often understood as a product of the probability of the flooding and a measure of its adverse consequences. Let $f_{T,n}(x)$ be the density function of the design exceedance, derived from n years of observations, $X_{T,n}$ and $x_T^\star$ the value which corresponds to the true parameter values. Given the deviation $(X_{T,n} - x_T^\star)$, the loss due to over or under-design may be expressed by a loss function $W(X_{T,n} - x_T^\star)$. The hydraulic risk is defined as the expected loss(Nachtnebel and Konecny,1987):

$$R_{T,n} = \int_0^\infty W(x - x_T^\star) f_{T,n}(x) dx$$

Generally the damages can be direct losses, indirect tangible losses and intangible losses. Not all damages resulting from a natural hazard can be quantified e.g. indirect effects

due to flooding. The impacts due to drought are mainly non-structural. In case of risk due to failure of structures a number of methods have been proposed by different investigators, including (but not limited to) the methods of: return period, safety factor, direct integration, Monte Carlo simulation, reliability index, mean-value first-order second moment (MFOSFM) analysis, and advanced first-order second moment(AFOSM) analysis. A general comparison of these methods is given in Duckstein et al. (1987).

Societal risk from a natural hazard is determined not only by the degree of exposure or frequency of the natural hazard but also by the vulnerability of society at that moment in time. Vulnerability is dynamic in response to changes in the economic, social, and environmental characteristics of the locale or region. Thus risk of drought indicates the potential adverse effects of drought as a product of both the frequency and severity of the hazard and corresponding vulnerability (Wilhite and Buchanan-Smith., 2005). In this study the latter definition of risk is adopted.

Vulnerability implies characteristics of populations, activities, or the environment that make them susceptible to the effects of drought. The degree of vulnerability depends on the environmental and social characteristics of the region and is measured by the ability to anticipate, cope with, resist, and recover from drought.

Because of the number of affected groups and sectors associated with drought, the geographic size of the area affected, and the difficulties in quantifying environmental damages and personal hardships, the precise determination of the financial costs of drought is a formidable challenge. These costs and losses are also quite variable from one drought year to another in the same place, depending on timing, intensity, and spatial extent of the droughts. Besides, it is difficult to disentangle the risk due to hydrological drought and other categories of drought. A number of interrelated impacts from human life, environmental disaster, others can be enumerated.

The impacts of drought are commonly classified as economic, environmental, and social. A comprehensive list of the impacts associated with drought has been described by Wilhite and Buchanan-Smith. (2005).

Drought produces a complex web of impacts that spans many sectors of the economy and reaches well beyond the area experiencing physical drought. This complexity exists because water is integral to societys ability to produce goods and provide services.

- Economic Impacts: Droughts cause hardship on many different sectors of an areas economy. Many economic impacts occur in agriculture and related sectors, including forestry and fisheries, because of the reliance of these sectors on surface and subsurface water supplies. In addition to obvious losses in yields in crop and livestock production, drought is associated with increase in insect infestations, plant disease and wind erosion. Rivers and lakes drop to low levels during drought, while turbidity and salinity increase, affecting fish habitat. Water based recreation may also decrease. The incidence of forest and range fires increases substantially during extended droughts, which in turn places both human and wildlife populations at higher levels of risk.

- Social Impacts: Drought affects human health, both physically and emotionally, in both rural and urban areas. The impacts of drought includes public safety from forest and range fires, health related low flow problems, reduced quality of life , and inequalities in the distribution of impacts and disaster relief. Many of the impacts identified as economic and environmental have societal impacts as well. Migration is one major problem usually to urban areas within the stressed area, or to regions outside the drought area. When the drought has abated, the migrants seldom return home, depriving rural areas of valuable human resources. The drought migrants place an increasing pressure on the social infrastructure of the urban areas, leading to increased poverty and social unrest.

- Environmental Impacts: Environmental losses are the result of damages to plant and animal species, reduction and degradation of wildlife habitat, air and water quality;degradation of visual and landscape quality; loss of biodiversity; and soil erosion. Some of the effects are short term and conditions quickly return to normal following the end of the drought. Other environmental effects linger for some time or may even become permanent. In most cases environmental losses are difficult to quantify.

Risk of drought is a combined effect of the frequency of drought and the vulnerability of the area to it. As the first variable is quite naturally induced, it is important to focus on the vulnerability to mitigate the impacts of these hazard. Vulnerability by itself is a function of the exposure, sensitiveness and cope up mechanisms in place(Wilhite and Buchanan-Smith., 2005). Often drought results from a number of composite factors

so the spatial nature of the risk can be modelled using overlay of variable layers on meteorological, hydrological, and physical data of the area. The ability of the people and the land to withstand rainfall, hydrological and even agricultural droughts depends to a large extent on their status preceding the event or series of events of rain shortfalls. This status is, in turn, determined by a large number of factors. Some of the data employed involves remote sensed data. In this work the hydrological drought risk is identified as:

$$R_{i,t} = P_{i,t} \cdot \sum_{j=1}^{n} w_j f_{i,j,t} \tag{9.43}$$

where

$R_{i,t}$=Risk in subwatershed i at time t,

$P_{i,t}$=probability of drought incident in subwatershed i at time t,

w_j=weight of factor j,

$f_{i,j,t}$=factor j in subwatershed i at time t(e.g. stream density)

In the absence of proper coping mechanisms in place, as is common in many developing countries, the risks of drought may go as far as claiming the lives of humans and animals. Recently in Ethiopia drought insurance pilot project is introduced by WFP. In 2006, an index was designed for an insurance contract transacted by an international reinsurer, AXA Re Paris. During 2007, the index was refined and a livelihood risk management framework was developed for the country with the active support of World Bank, USAID and DFID. The regional drought index, called the Livelihood Early Assessment and Protection (LEAP) Index, is currently under testing stage(Hess, 2008). The index is tied only to agricultural drought. It is calculated on the basis of the yield reduction of the dominant crops of an area and the weighted historical maximum population in need. It should be noted that historical population in need of assistance in a certain year may not disclose the actual proportion affected by a drought incident alone. A number of factors may predominantly affect the vulnerability of an area to drought including the physical characteristics of the area (like drainage density, slope,ground water availability, land use/cover), the social characteristics (like population pressure) and infrastructure (like road network). Data on some of the factors are not readily available. The selection of indicators is based on their applicability and sensitivity to vulnerability of drought. While measures of hydroclimatic variables are likely to be

universally appropriate, identical suite of indicators would not necessarily be applicable for other geographic locations. The following factors are considered in this study:

- stream density,
- population density,
- road density,
- ground water availability (interms of static water level and safe yield)
- slope of the catchment and
- the severity of drought captured in NDVI.

No single variable is sufficient to portray the complexity of the vulnerability of an area to drought.The relation of each of these factors to drought is discussed in section 4.2. The whole of the population in an area would not be equally vulnerable to drought. The poor are more vulnerable in this regard. In this study the relative risks of drought in each of the subwatersheds in Wabi Shebele river basin is investigated using multicriteria analysis.

The procedure involves:

1. Scale all the factors between 0 and 1 to avoid variations in units and magnitude. e.g. Areas with high stream density are expected to have relatively low vulnerability. In regional analysis usage of different variables with varying units/scales of measure is a common problem. Thus the process of making the data scale-free is essential to make the indicators comparable.
2. Put weights to each of the factors relative to the other in its strength to affect vulnerability. In the absence of any evidence to support prioritization,it is difficult to give the weights to the factors. Thus sensitivity analysis of these weights are included in the analysis.
3. Sum the weighted factors at each of the subwatersheds.
4. Identify the expected probability of hydrological drought in the subwatersheds. Here the time series of scPHDI is used. Spatial distribution of average frequency of hydrological drought is provided in Figure 9.11

5. Compute the risk as per equation 9.43.

6. Standardize the risks obtained above at each of the subwatersheds and classify the standardized risks,$R_{std,i}$ as follows:

$$R_{std,i} = \begin{cases} \geq 3\sigma_{R_{i,t}}, & \text{Extreme relative risk;} \\ 2\sigma_{R_{i,t}} \leq x \leq 3\sigma_{R_{i,t}}, & \text{Severe relative risk;} \\ \sigma_{R_{i,t}} \leq x \leq 2\sigma_{R_{i,t}}, & \text{Moderate relative risk;} \\ -\sigma_{R_{i,t}} \leq x \leq \sigma_{R_{i,t}}, & \text{Mild relative risk;} \\ \text{otherwise,} & \text{Low relative risk.} \end{cases}$$

where $\sigma_{R_{i,t}}$ is the standard deviation of the risks in the river basin.

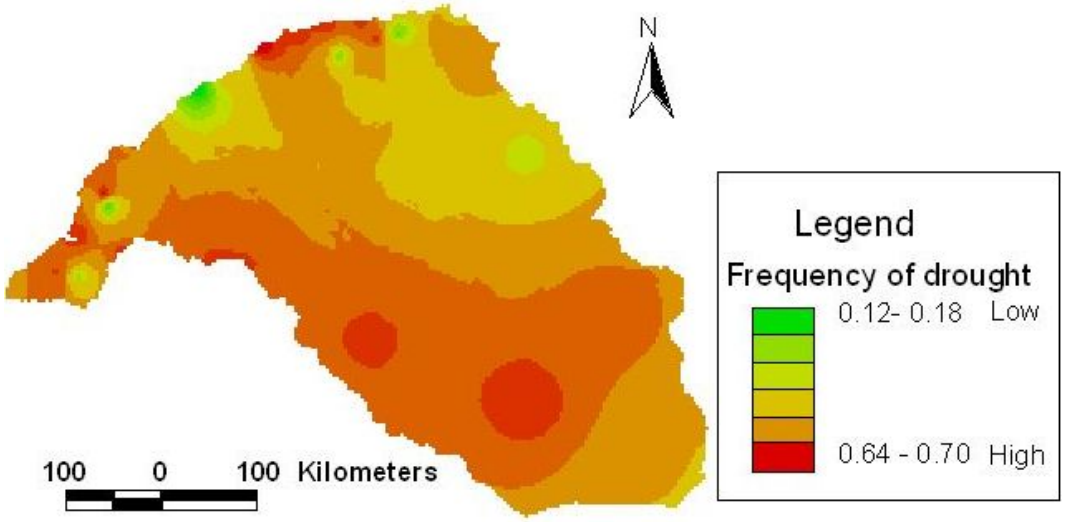


FIGURE 9.11: Area distribution of average frequency of hydrological drought.

It is possible to identify time series of land use and land cover information from Normalized digital vegetation index (NVDI) maps derived from remotely sensed data. This NDVI will also benefit in identifying homogeneous regions. Severest deviation of monthly NDVI from their respective average monthly NDVI values for the years 1982-2000 in Wabi shebele river basin is shown in Figure 9.12.

The impact of human societies on natural resources does not depend solely on the demographic density, and the notions of "load capacity" and "critical threshold" must be handled with great care. Many examples demonstrate that these criteria can vary enormously, depending upon the strategies and the technologies used by the people. A large

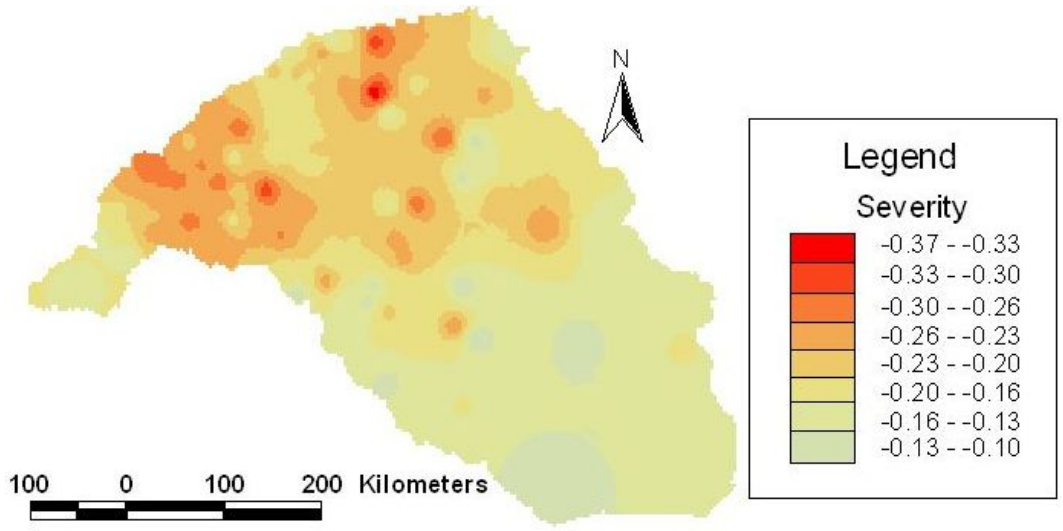


FIGURE 9.12: Severest deviation of monthly NDVI from their respective average monthly NDVI values for the years 1982-2000.

number of variables together form the characteristic/defining features for the classification of blocks that are amenable for similar drought-proofing interventions. The spatial distribution of vulnerability to drought in Wabi Shebele using the factors enumerated above is provided below (Figure 9.13). Applying a sensitivity analysis helps to identify

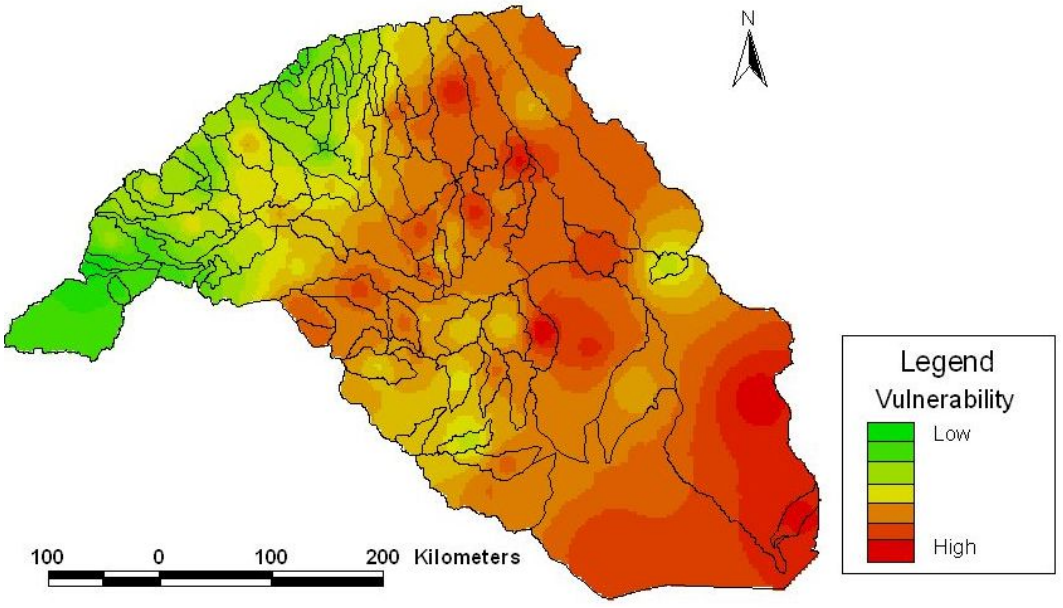


FIGURE 9.13: Areal distribution of relative average vulnerability (using equal weighted factors) in Wabi Shebele river basin.

the variables contributing most to uncertainties. Principal component analysis is done

TABLE 9.2: Weight allocation in the different trials to the various variables of vulnerability in the Wabi Shebele basin

Factor	Weight					
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Stream density	0.1905	0.2381	0.0476	0.2381	0.1429	0.2857
Slope	0.2857	0.2857	0.2381	0.0952	0.0952	0.0952
Road density	0.1429	0.1429	0.2857	0.2857	0.1905	0.2381
Population density	0.0952	0.1905	0.1905	0.1905	0.2381	0.0476
$NDVI_{severity}$	0.0476	0.0476	0.1429	0.1429	0.0476	0.1905
SWL	0.2381	0.0952	0.0952	0.0476	0.2857	0.1429

on the variables and the first 6 principal components explain more than 87% of variability among the factors. The different variables considered to assess the vulnerability are given weights on the relative magnitude of the elements of the eigen vectors (Table 9.2). In Trial 1 higher weightage is provided to the slope, followed by the static water level, stream density, road network, population density and severity of NDVI. The composite index of vulnerability is computed for each case as the weighted sum of the factors. The first trial is essentially the same in affecting the vulnerability as is the case with equal weights condition (see Figure 9.13 and 9.14a). Large areas with high vulnerability are obtained using the weight assigned in the first trial. Very high vulnerability in localised areas is noticeable in the sixth trial Figure 9.14. Areal distribution of relative average risks using weights assigned from corresponding trials is provided in Figure 9.15. It should be noted that the risk in this section gives a more generalized view using multi-criteria analysis. The spatial extent and variability of the phenomenon are so great as to defy exact prediction and measurement at every point.

Risk Management: The opposite of crisis management, where a proactive approach is taken well in advance of drought so that mitigation can reduce drought impacts, and so relief and recovery decisions are made in a timely, coordinated, and effective manner during a drought. This process has the potential to lead to the identification of effective and appropriate drought risk reduction activities rather than ad-hoc responses or unresearched mitigation plans that may have little effect on reducing drought impact in the future.

It is important to identify and rank the priority of relevant drought impacts; examine the underlying environmental, economic, and social causes of these impacts; and then choose actions that will address these underlying causes. In order to cushion the devastating effects of drought, a concerted effort needs to be done to reduce vulnerability to it. As

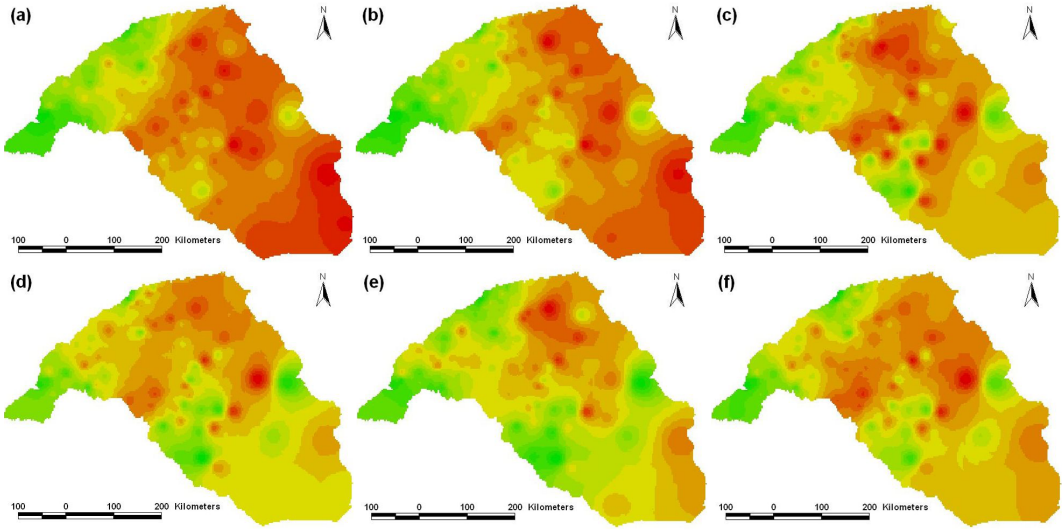


FIGURE 9.14: Sensitivity of the vulnerability to the different contributing factors in (a) Trial 1 (b) Trial 2 (c) Trial 3 (d) Trial 4 (e) Trial 5 (f) Trial 6

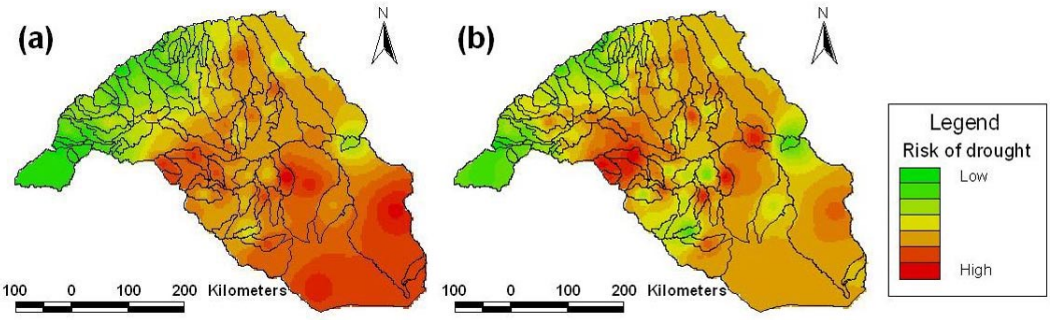


FIGURE 9.15: Areal distribution of relative average risks using weights assigned in (a) Trial 1 (b) Trial 6.

the saying goes "A stitch in time saves nine".

9.3.1 Implications for Water Management

Investigating a history of drought frequency, severity and duration for a region provides a greater understanding of the region's drought characteristics and the probability of drought recurrence at various levels of severity. This type of information is beneficial in the development of drought response and mitigation strategies and preparedness plans. These plans need to be kept updated as conditions change. Without knowing just how severe a drought may become, it is difficult to know what actions are best to take, if any, when a drought is likely to have begun. Water managers must take appropriate action to reduce future losses and not increase them by imposing unnecessary restrictions

when they are not needed(Wright et al., 1986). Water resources management measures involve a certain degree of uncertainty and risk because both water supplies and demands have inherent components of randomness. Some of the factors for this randomness can be attributed to the spatial and temporal hydro-climatic variability. Hence, drought planning and management activities must involve predicting potential droughts and forecasting the evolution of ongoing droughts. The basis of any drought management plan is a robust system of indicators that can identify and diagnose anomalies in water availability and can provide the basis for early detection of drought episodes(Tallaksen and van Lanen, 2004).

The nomadic and pastoral settlement program in Ethiopia and Somalia requires appropriate drought early warning and prevention mechanisms(Lautze et al., 2003). Livestock is an important livelihood source, especially in drought afflicted agro-ecologies like lower part of Wabi shebele river basin. A gradual shift from rainfed agriculture to irrigated schemes in the upper portion of the basin entails all riparian countries to give due focus to analysis of drought in its whole domain. The stochastic component of the total series of severity of drought at the upstream humid portion tends to be higher than that of the downstream arid portion. The flow at upstream location may be much influenced with the stochastic nature of rainfall in the highlands. At the downstream portion of Wabi shebele river no appreciable flow joins the main water course. The settlement in the basin shows that denser population is found in the upstream highlands of the basin. A majority of the population in the highlands depend on agriculture while the lowlanders are pastoralists. Conservation reservoirs can retain excess water from periods of high flows for use during periods of drought. Thus, natural streamflow during the drought period is sometimes significantly increased by artificial ways and the recorded or observed streamflow drought gets a different shape. The only existing dam in Wabi shebele river basin is the Melka Wakena hydropower dam located at upstream portion of the basin. The need for irrigation development in the basin has been recognised in the last two decades. Thus, about 72 small scale Irrigation schemes have been constructed. Currently there is a medium scale irrigation diversion scheme at Gode(MoWR, 2003). Performing simulations of alternative actions given supply and demand situations experienced in the past can help identify the possible range of likely impacts associated with any particular set of trigger values and associated decisions. Stochastic simulation of hydrological drought is also essential in the design and operation of reservoirs.

Crisis-oriented drought response efforts have been largely ineffective, poorly coordinated, untimely, and inefficient in terms of the resources allocated. One way to make drought management easier is to establish a sequence of increasingly strict conservation measures based on a sequence of drought triggers, and seek the public's approval. The current water management situation in the basin is hardly coordinated with a multiplicity of projects and initiatives under the auspices of individual regions, and a variety of NGOs, private sector actors, and international funding agencies. Increased cooperation and share of information on evolving drought events at basin level can reduce possible risks.

Chapter 10

Summaries and Conclusions

In this chapter the main summaries of the research and important conclusions drawn are stated. It also provides some recommendations, poses open questions and suggests areas of future research.

10.1 Conclusions

In this research efforts have been made to analyse the hydrological drought with due emphasis to ungauged catchments. The analysis of the severity of drought embodies a multitude of methodologies. Identification of severity of drought and unveiling its intrinsic characteristics is illustrated. Stochastic simulation of hydrological drought, streamflow reconstruction and extracting the multivariate features of this extreme hydrological event from proxy data are among the major items achieved. Associated relative risks of drought are also scrutinized. The case study is demonstrated in Wabi Shebele river basin in Ethiopia. The approaches and conclusions drawn here have a more wider application in other similar river basins.

It is essential to appropriately understand the spatial and temporal characteristics of drought in this water deficit transboundary river basin to deny famine a future in the region. Investigation of the hydrological drought is made from perspectives of various data sources like streamflow records, meteorological records, and proxy data that comprises tree rings and climatic indices like ENSO. Quite often, absence of recorded long

time streamflow data hinders a reliable drought analysis and understanding of the phenomenon in the past. Signatures of water stress are imprinted on tree rings. In this study proxy data from riparian tree rings and climatic indices are used for simulation and extension of instrumental records. Frequency of smaller rings in the residual tree ring series for the Adaba-Dodlla area appears to be increasing in recent years. Extremes of climatic signals may have contributed to smaller ring widths in the region. NDVI can be used as an indicator of the biosphere response to climate variability at a range of time scales. A composite index from proxy data of tree rings and climatic indices (ENSO) is developed. This index is used to simulate and extend the streamflow records at a downstream location, Imi gauging station. Long time series of annual streamflow are reconstructed from proxy data. This composite index from the proxy data explains about 70% of the variation in the streamflow at Imi gauging station in the last two and half decades.

Temporal disaggregation of reconstructed annual series to monthly series is carried out using LANEs condensed disaggregation model. Disaggregating the annual data to seasonal time scales enabled to examine long time series of the hydrological drought patterns. Extracting hydrological drought entailed defining a possible threshold levels. Threshold levels of low exceedance probability are found to be appropriate for the study area unlike the commonly used higher exceedance probabilities in temperate climate. It is observed that higher truncation levels produce higher severity irrespective of the return period. It was shown that the frequency of some notable recent droughts as revealed with data generated from proxy records well matches that of instrumental data. This underlines the importance of these proxy records for data augmentation.

Due to the randomness of the factors responsible for the occurrence and severity of drought, it can be considered as stochastic process. Stochastic simulation of hydrological drought is performed using SARIMA models from time series of instrumental monthly streamflow records in the study area. Fourier analysis of the time series at two stations reveals that 82-89% of the variance of the total series is accounted by the stochastic component. Thus stochastic simulation of severity of hydrological drought using Box and Jenkins methodology is carried out. Different seasonal ARIMA candidate models are compared. It is found that the pure MA $(0,1,1)(0,1,1)^{12}$ outperforms all. The simulated results compare well with the historical records of severity of hydrological drought in the basin as confirmed by the residual white noise and normality test. There

should be greater emphasis on monitoring and drought management during normal years than during the actual drought periods. The stochastic component of the total series of severity of drought at the upstream humid portion is higher than that of the downstream arid. The flow at upstream location may be much influenced with the stochastic nature of rainfall in the highlands.

Observed hydrological processes result from the joint action of different hydrometeorological processes, the dominance of which varies through time. This interplay can be supposed to be imprinted in the temporal evolution of the frequency content of observed processes such as discharge. Multivariate characteristics of drought is important to show the multi faced patterns displayed by this devastating phenomenon. Copula modeling has been able to capture the joint probability structure of severity and duration of drought. Two-parameter Weibull and two-parameter log-normal distributions fit well the marginal distribution of severity and duration of the hydrological drought computed from the reconstructed time series of monthly streamflow records. Since the marginal distributions of the hydrological drought variables are different, copulas are used to model the dependence structure between them. The nonlinear dependency between severity and duration of hydrological drought is studied using copula models. Different copula families and parameter estimation techniques are evaluated. The goodness of fit is tested graphically and using Akaike Information criterion (AIC). Among the different families of copulas, it is hard to detect the most appropriate copula graphically. According to goodness of fit statistic it is found numerically that Frank copula performs better. The Kendall-tau parameter of 0.7192 is outside the domains of the Ali-Mikhail-Haq and Farlie-Gumbel-Morgenstern family so these families are not considered in the evaluation of the copulas. Joint and conditional probabilities of severity and duration of drought in the area is specified from the derived relationship to provide empirical insight on the nature of the extreme events.

Estimation of continuous time series of base flow index which specifically characterize the high frequency, low amplitude base flow regime are done at gauged catchments in Wabi Shebele basin. The relative strengths in association between the climatic, morphometric and geologic features of the catchment to the base flow estimates are weighted and a plausible relationship is produced. There is a strong negative correlation between BFI and NDVI which is even more pronounced during the dry seasons in the basin. Values of base flow index determined for a network of stream flow gauges are matched to the

composite morphometric and climatic data using spatial and regression analyses. An exponential variant model fitted to the flow duration curves is used as an aid in deriving a relationship between BFI and Q_{70}. The developed relationship can be used for fairly estimating the base flows in the ungauged portion of the river basin considered. However, in view of the tremendous spatio-temporal heterogeneity of climatic and landscape properties extrapolation of response information or knowledge from gauged to ungauged basins remains fraught with considerable difficulties and uncertainties.

It is aimed to determine whether droughts have distinct spatial patterns regarding temporal evolution and variability in the study area in which the geographic and climatic characteristics have a high degree of diversity. These issues have a high theoretical and applied interest. Confirmation of large spatial variability of this meteorological/hydrological phenomenon would imply the need to develop early drought warning plans at a local level. Catchment characteristics can be related to low flow characteristics thus are used to delineate the hydrologically homogeneous pools. Delineating the basin on its morphometric, climatic, geologic features and NDVI characteristics is advantageous that it enables to allocate ungauged areas to the pools. Various techniques are used to validate the homogeneity of the identified pools. In order to analyse the spatial patterns of drought it was imperative to rely on data predominantly available throughout the basin. The Palmer hydrological severity index is used. It is shown that the parameters of PDSI are found to be different than that identified by Palmer. Thus scPHDI is used as a compromise. It is also shown that the scPHDI is not a first order autoregressive process as its principal equation indicates. The design of early drought warning plans would be a priority management task in areas with more frequent and intense droughts. Severity-area-frequency analysis of drought in the area using nonparametric kernels shows high variability of drought events within the pools formed. Wavelet analysis of the spatial scPHDI further strengthened that there is more variation along the longitude.

A more generalized view of relative drought risk in the study area is provided using multicriteria analysis. Risk of drought is a combined effect of the frequency of drought and the vulnerability of the area to it. As the frequency of drought is quite naturally induced, it is important to focus on the vulnerability to mitigate the impacts of this hazard. The spatial extent and variability of the phenomenon are so great as to defy exact prediction and measurement at every point. No single variable is sufficient to portray the complexity of the vulnerability of an area to drought. Although droughts

are natural hazards, society can reduce its vulnerability and therefore lessen the risks associated with drought events. The impacts of droughts, like those of other natural hazards, can be reduced through planning and preparedness. Developing a history of drought frequency, severity, duration and impacts for a region provides a greater understanding of the region's drought characteristics and the probability of drought recurrence at various levels of severity. Employing these information to develop timely updated drought response and mitigation strategies and preparedness plans may reduce the unfortunate sequel of drought.

10.2 Recommendations

In the reconstruction of hydrological drought from proxy data, further research from stem disks of old trees from different tree species may offer better evidence. It is also important to establish appropriate relationship between early wood and late wood ring widths vis-a-vis the seasonality of precipitation signals. Temporal satellite images of high resolution may help in this respect.

The truncation level which is used to objectively demarcate the onset and end of drought is subjective. There is a need to substantiate the truncation level of drought explicitly taking into consideration both supply and demand of water.

Quantifying the monetary value of the damage attributable to hydrological drought risk is important to give due mitigation measures and implement efficient mitigation measures. Because of the number of affected groups and sectors associated with drought, the geographic size of the area affected, and the difficulties in quantifying environmental damages and personal hardships, the precise determination of the financial costs of drought is a formidable challenge.

References

Acreman, M. and Wiltshire, S. (1989). The regions are dead; long live the regions. method of identifying and dispensing with regions for flood frequency analysis. In Roald, T., Nordseth, K., and Hassel, K. A., editors, *FRIENDS in Hydrology-Proc. Bolkesj Symp., April 1989*, pages 175–188, Wallingford, UK. IAHS Publ. 187.

Allen, R., L.S.Pereira, D.Raes, and Smith, M. (1998). Statistical analysis of weather data sets. In FAO, editor, *Crop evapotranspiration - Guidelines for computing crop water requirements*. FAO, Rome.

Alley, W. (1984). The palmer drought severity index. limitations and assumptions. *Journal of Climate*, 23:1100–1109.

Andrews, J. and Moss, T. (2002). *Reliability and Risk Assessment*. Professioanl Engineering publishing Ltd, London.

Attia, B. and Abulhoda, A. (1992). The ENSO phenomenon and its impact on the Nile's hydrology. In Abu-Zeid, M. and Biswas, A., editors, *Climate Fluctuations and Water Management.*, pages 71–79. Butterworth Heinemann.

Awass, A. A. and Foerch, G. (2006a). Assessing the threshold levels used to detect hydrological and environmental drought in semi- arid areas. In *Proceedings of Man and River systems II December 4-6*, Paris, France.

Awass, A. A. and Foerch, G. (2006b). Catchment characteristics as predictors of base flow index. In *Proceedings of TROPENTAG 2006*, Bonn, Germany.

Awass, A. A. and Foerch, G. (2008). Stochastic simulation of the severity of hydrological drought. *Water and Environment Journal*, 23:2–10.

Barnett, T., Santer, B., Jones, P., Bradley, R., and Briffa, K. (1996). Estimates of low frequency natural variability in near-surface air temperature. *The Holocene*, 6:255–263.

Bayazit, M. and Onoz, B. (2002). LL-moments for estimating low flow quantiles. *Hydrological sciences J.*, 47(5):707–720.

Beran, M. and Rodier, J. (1985). *Hydrological aspects of drought-A contribution to the International Hydrological Program.* UNESCO/WMO, Geneva.

Berchtold, A. (1999). The double chain markov model. *Commun. Statist. Theory Methods*, 28:2569–2589.

Bonaccorso, B., Bordi, I., Cancielliere, A., Rossi, G., and Sutera, A. (2003). Spatial variability of drought: an analysis of the SPI in Sicily. *Water Resour Manag*, 17:273296.

Box, G., Jenkins, G., and Reinsel, G. (1994). *Time Series Analysis, Forecasting and Control, 3rd ed.* Prentice Hall, Englewood Clifs, NJ.

Bremaud, P. (1999). *Markov chains, Gibbs fields, Montecarlo simulation and queues.* Springer, The Netherlands.

Calow, R., MacDonald, A., and Nicol, A. (2000). Planning for groundwater drought in Africa: Towards a systematic framework for assessing water security in Ethiopia. BGS Technical Report WC/00/13, British Geological Survey, UK.

Cao, L. (2003). Support vector machines experts for time series forecasting. *Neurocomputing*, 51:321–339(19).

Cheryachukin, A. and Sitnin, O. . V. (2000). Melka wakana hydroproject on the Wabi shebele river in Ethiopia. *Hydrotechnical Construction*, 34(8-9):111–115.

Cook, E. (1985). *A time series analysis approach to tree ring standardization.* PhD thesis, Laboratory of tree ring research, Tucson.

Cook, E., Briffa, K., Shiyatov, S., , and Mazepa, V. (1990). Tree-ring standardization and growth-trend estimation. In Cook, E. and Kairiukstis, L., editors, *Methods of Dendrochronology*, pages 104–123. Kluwer Publ., Dordrecht.

Cook, E. and Peters, K. (1981). The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. *Tree-Ring Bulletin*, 41:45–53.

Couralet, C., Sass-Klaassen, U., Sterck, F., Bekele, T., and Zuidema, P. (2005). Combining dendrochronology and matrix modelling in demographic studies: an evaluation for Juniperus procera in Ethiopia. *Forest Ecology and Management*, 216(1/3):317–330.

Cunnane, C. (1989). *Statistical distributions for flood frequency analysis-Operational hydrology report no. 33.* WMO, Geneva.

Demuth, S. and Stahl, K. (2001). ARIDE assessment of the regional impacts of drought in europe. File report, EU contract, ENV4-CT-97-0553, Institute of Hydrology, University of Freiburg, Freiburg, Germany.

Doesken, N., McKee, T. B., and Kleist, J. (1991). Development of a surface water supply index (SWSI) for the Western United States. climatology report. Climatology Report no. 91-3, Colorado State University, Ft. Collins, CO.

Dracup, J. A., Lee, K. S., and Paulson, E. D. (1980). On the definition of droughts. *Water Resour. Res.*, 16(2):297–302.

Duckstein, L., Ribeiro, F. R., Wasti, S. T., Çetin Yilmaz, and Plate, E. J. (1987). *Engineering Reliability and Risk in water resources.* Kluwer Academic publishers, Dordrecht, Martinus Nijhoff.

Eder, B., Davis, J., and Monahan, J. (1987). Spatial and temporal analysis of the palmer drought severity index over the South-Eastern United States. *J. Climatology*, 7:31–51.

Edwards, D. and McKee, T. (1997). Characteristics of 20th century drought in the United States at multiple time scales.climatology report number 972. Technical report, Colorado State University, Fort Collins, CO.

Eltahir, E. (1996). El Nino and the natural variability in the flow of the Nile river. *Water Resources Research*, 32(1):131–137.

February, E. C. and Stock, W. D. (1998). An assessment of the dendrochronological potential of two podocarpus species. *The Holocene*, 8(6):747–750.

Fleig, A., Tallaksen, L. M., Hisdal, H., and Demuth, S. (2005). A global evaluation of streamflow drought characteristics. *Hydrology and Earth System, Sciences Discussions*, 10:535–552.

Frick, D., Bode, D., and Salas, J. (1990). Effect of drought on urban water supplies. I: Drought analysis. *J. Hydraul. Eng.*, 116(6):733–753.

Fritts, H. (1976). *Tree Rings and climate*. Academic Press, New York.

Fritzsche, F., Abate, A., Fetene, M., Beck, E., Weise, S., and Guggenberger, G. (2006). Soilplant hydrology of indigenous and exotic trees in an Ethiopian montane forest. *Tree Physiology*, 26:1043–1054.

Gelder, P. V. (2000). *Statistical Methods for the Risk-Based Design of Civil Structures*. PhD thesis, Civil Engineering department, Delft, The Netherlands.

Genest, C., Ghoudi, K., and Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distribution. *Biometrika*, 82(3):543–552.

Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate archimedean copulas. *J. Am. Stat. Assoc.*, 88:1034–1043.

Gonzalez, J. and Valdes, J. B. (2003). Bivariate drought recurrence analysis using tree ring reconstructions. *J. of Hydrologic Engineering*, 8(5).

Gourlay, I. (1995). The definition of seasonal growth zones in some African Acacia species - a review. *IAWA Journal*, 16(4):353–359.

Greenwood, J., Landwehr, J., Matalas, N., and Wallis, J. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.*, 15:1049–1054.

Guo, S., Kachroo, R., and Mngodo, R. (1996). Nonparametric kernel estimation of low flow quantiles. *Journal of Hydrology*, 185:335–348.

Gustard, A. and Gross, R. (1989). Low flow regimes of Northern and Western Europe. In Roald, L., Nordseth, K., and Hassel, K., editors, *FRIENDS in Hydrology*, pages 205–213. IAHS pub. No. 187, Bolkesjo, Norway.

Guttman, N. (1999). Accepting the Standardized Precipitation Index: a calculation algorithm. *Journal of the American Water Resources Association*, 35:311–322.

Haile, T. (1988). Causes and characteristics of drought in Ethiopia. *Ethiopian Journal of Agricultural Science*, 10:85–97.

Hall, M. J. and Minns, A. (1999). The classification of hydrologically homogeneous regions. *Hydrological sciences J.*, 44(5):693–704.

Hayes, M., Svoboda, M., Wilhite, D., and Vanyarkho, O. (1999). Monitoring the 1996 drought using the standardized precipitation index. *Bulletin of the American Meteorological Society*, 80(3):429–438.

Heddinghaus, T. and Sabol, P. (1991). A review of the palmer drought severity index and where do we go from here. In *Proceedings of the 7th Conference on Applied Climatology*, pages 242–246, Salt Lake City. American Meteor. Soc.

Helsel, D. and Hirsch, R. (1992). *Statistical methods in water resources. Studies in Environmental Science, no. 49.* Elsevier Publishers, New York.

Hess, U. (2008). Early warning component - LEAP principles. In *Proceedings of the joint Food Security Coordination Bureau (FSCB) / World food Programme (WFP) workshop on Livelihood Early Assessment and Protection (LEAP): its potential application, benefits and limits*, Addis Ababa, Ethiopia. World food Programme.

Hipel, K. and McLeod, A. (1994). *Time series modelling of water resources and environmental systems.* Elsevier Science B.V., Amsterdam.

Holben, B. and Justice, C. (1981). An examination of spectral band ratioing to reduce the topographic effect on remotely sensed data. *International Journal of Remote Sensing*, 2:115–133.

Holmes, M., Young, A., Gustard, A., and Grew, R. (2002). A region of influence approach to predicting flow duration curves within ungauged catchments. *Hydrology and Earth System Sciences*, 6(4):721–731.

Hosking, J. (1986). The theory of probability weighted moments. Research report RC12210, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. statist. Soc.*, 52:105–124.

Hosking, J. (1994). The four-parameter kappa distribution. *IBM J. Res. Develop.*, 38(3):251–258.

Hosking, J. and Wallis, J. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29:271–281.

Hosking, J. and Wallis, J. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge, U.K.

Hosking, J., Wallis, J., and Wood, E. (1985). Estimation of the Generalised Extreme Value distribution by the method of probability weighted moments. *Technometrics*, 27(3):251–261.

Hudson, H. and Hazen, R. (1964). Droughts and low streamflow. In Chow, V., editor, *Handbook of applied hydrology,Section 18*. McGraw-Hill, New York.

Hughes, D., Hannart, P., and Watkins, D. (2003). Continuous base flow separation from time series of daily and monthly stream flow data. *Water SA*, 29(1).

IH (1999). *Flood Estimation Handbook*. Institute of Hydrology, Wallingford, UK.

Jiang, J., Zhang, D., and Fraedrich, K. (1997). Historic climate variability of wetness in East China (960-1992): a wavelet analysis. *International Journal of Climatology*, 17(9):969–981.

Joe, J. and Xu (1996). The estimation method of inference functions for margins for multivariate models. Technical report no. 166, Department of Statistics, University of British Columbia.

Karl, T. (1986). The sensitivity of the palmer drought severity index and palmers z-index to their calibration coefficients including potential evapotranspiration. *J Clim Appl Meteorol*, 25:77–86.

Keyantash, J. and Dracup, J. (2002). The quantification of drought: an evaluation of drought indices. *Bulletin of American Meteor Soc*, pages 1167–1180.

Keyantash, J. and Dracup, J. (2004). An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage. *Water Resources Res.*, 40(9).

Khaliq, M., Ouarda, T., Ondo, J.-C., Gachon, P., and Bobee, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydrometeorological observations: A review. *Journal of Hydrology*, 329:534– 552.

Kim, T., Valdes, J., and C.Yoo (2003). Nonparametric approach for. estimating return periods of droughts in arid regions. *J Hydrologic Eng ASCE*, 8(5):237–246.

Kim, T.-W. and Valdes, J. B. (2002). Frequency and spatial characteristics of droughts in the Conchos river basin, Mexico. *Water International*, 27(3):420–430.

Kjeldsen, T., Lundorf, A., and Rosbjerg, D. (1999). Regional partial duration series modeling of hydrological droughts in Zimbabwean rivers using a two component exponential distribution. In *Hydrological Extremes: Understanding, predicting, mitigating. Proceedings of IUGG 99 symposium July 1999*, HSI Birmingham. IAHS publ. 255.

Koutsoyiannis, D. (2001). Coupling stochastic models of different time scales. *J. Water Resour. Res.*, 37(2):379–391.

Krokli, B. (1989). Low flow analysis. In *FRIENDS in Hydrology*, pages 443 – 451. IAHS Publication No. 187.

Lane, W. and Frevert, D. (1990). *Applied stochastic techniques. personal computer version 5.2, Users manual.* Earth sciences division, USBR, Denver, Colorado.

Lau, K. and Weng, H. (1995). Climate signal detection using wavelet transformation: How to make a time series sing. *Bulletin of the American Meteorological Society*, 76:2391–2402.

Lautze, S., Aklilu, Y., Raven-Roberts, A., Young, H., Kebede, G., and Leaning, J. (2003). Risk and vulnerability in ethiopia: learning from the past, responding to the present, preparing for the future. Technical report, USAID, Addis Ababa.

Lawrimore, J., Jr, R. H., Svoboda, M., Swail, V., and Englehart, P. (2002). Beginning a new era of drought monitoring across North America. *Bulletin of American Meteor Soc*, pages 1191–1192.

Lettenmaier, D., Wallis, J., and Wood, E. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resour. Res.*, 23(2):313–323.

Liebscher, H. (1987). Paleohydrologic studies using proxy data and observations. In Solomon, S., Beran, M., and Hogg, W., editors, *Proceedings of the symposium on the influence of climate change and climatic variability on the hydrologic regime and water resources*, pages 111–121, Vancouver, Canada. IAHS.

Lohani, V. and Loganathan, G. V. (1997). An early warning system for drought management using the Palmer Drought Index. *Journal of the American Water Resources Association*, 33(6):1375–1386.

Macfarlane, C., Adams, M., and White, D. (2004). Productivity, carbon isotope discrimination and leaf traits of trees of Eucalyptus globulus labill. in relation to water availability. *Plant Cell and Environment*, 27(12):1515–1524.

Maingi, J. (1998). *Land Use and Vegetation Change in Response to River Basin Development in the Lower Tana River Basin of Eastern Kenya.* PhD thesis, Laboratory of tree ring research, Tucson.

Malo, A. and Nicholson, S. (1990). A study of rainfall and vegetation dynamics in the African Sahel using normalized difference vegetation index. *Journal of Arid Environments*, 19:1–24.

Mazvimavi, D., Meijerink, A., Savenije, H., and Stein, A. (2005). Prediction of flow characteristics using multiple regression and neural networks: a case study in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(11-16):639–647.

McKee, T., Doesken, N., and Kleist, J. (1993). The relation of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, pages 179–184, Boston. American Meteor. Soc.

McKee, T., Doesken, N., and Kleist, J. (1995). Drought monitoring with multiple time scales. In *Proceedings of the 9th Conference on Applied Climatology*, pages 233–236, Dallas. American Meteor. Soc.

Mejia, J. and Rousselle, J. (1976). Disaggregation models in hydrology revisited. *Water resources research*, 12:185–186.

Meko, D. and Graybill, D. (1995). Tree-ring reconstruction of upper Gila river discharge. *Water Resour Bull*, 31:605–616.

Meko, D., Therrell, M., Baisan, C., Hughes, M., and Graybill, D. (2001). Sacramento river fow reconstructed to A.D. 869 from tree rings. *JAWRA*, 37(4):1020–1037.

MoWR (2003). Wabi shebele river basin integrated master plan study project vol I-XII. Technical report, Ministry of Water Resources, Addis Ababa.

Moy, L. and Kapadia, A. (1995). Predictions of drought length extreme order statistics using run theory. *Journal of Hydrology*, 169:95–110.

Nakken, M. (1999). Wavelet analysis of rainfall-runoff variability isolating climatic from anthropogenic patterns. *Environ. Modelling Software*, 14:283–295.

NDMC (2005). Understanding and defining drought, National Drought Mitigation Center. URL: http://www.drought.unl.edu/whatis/concept.htm: accessed June 2005.

Nelsen, R. (1999). *An introduction to copulas*. Springer, New York.

NMSA (1996). Assessment of drought in Ethiopia. Meteorological research report series, National Meteorological Services Agency of Ethiopia, Addis Ababa , Ethiopia.

Ntale, H. K. and Gan, T. Y. (2003). Drought indices and their application to East Africa. *Int. J. Climatol.*, 23:1335–1357.

Ochola, W. and Kerkides, P. (2003). A markov chain simulation model for predicting critical wet and dry spells in Kenia: Analysing rainfall events in the Kano plains. *Irrigation and Drainage*, 52(4):327–342.

Ogallo, L. (1994). Interannual variability of the east african monsoon wind systems and their impact on East African climate. TD 629:99-104, WMO.

Palmer, W. (1965). Meteorological drought. Technical report, US Weather Bureau Research Paper 45, USA.

Panu, U. and Sharma, T. (2002). Challenges in drought research: some perspectives and future directions. *Hydrological Sciences J.*, 47:19–30.

Peters, E., Torfs, P., van Lanen, H., and Bier, G. (2003). Propagation of drought through groundwater - a new approach using linear reservoir theory. *Hydrological Processes*, 17(15).

Pike, R. and Wilson, S. (1971). Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Bull. Geologic Soc. Am.*, 82:1079–1084.

Post, D. (2005). A new method for estimating flow duration curves: an application to the Burdekin river catchment, North Queensland, Australia. In *Proceedings of the International Environmental Modelling and Software Society*, Osnabruck, Germany. International Environmental Modelling and Software Society.

Raftery, A. (1985). A model for high-order markov chains. *Journal of the Royal Statistical Society, B-series*, 47(3):528–539.

Reynolds, C., Jackson, T. J., and Rawls, W. J. (1999). Estimating available water content by linking the FAO soil map of the world with global soil profile databases and pedo-transfer functions. In *Proceedings of the AGU 1999 Spring Conference May 31-June 4*, Boston, MA.

Rogers, P., Bhatia, R., and Huber, A. (1998). Water as a social and economic good: How to put the principle into practice. TAC background papers 2, Global Water Partnership, Stockholm, Sweden.

Rouhani, S. and Cargile, K. (1989). A geostatistical tool for drought management. *J. Hydrology*, 108:257–266.

Rouse, J., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C. (1974). Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. Type iii final report, Greenbelt, MD: NASA/GSFC.

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. *Journal of Water Resources Planning and Management(ASCE)*, 131(5):383–393.

Santos, E. and Salas, J. (1992). Stepwise disaggregation scheme for synthetic hydrology. *J. Hydraul. Eng.*, 118:765–1784.

SC-UK, DPPB, and Partners (2001). Degahbur agropastoral livelihood zone, a household economy approach. Base line study, Save the Children-UK and Disaster prevention and preparedness Burea, Ethiopia.

Seleshi, Y. (1998). *Stochastic predictions of summer rainfall amounts over the North East African highlands and over India.* PhD thesis, V.U.B.-Hydrology, Belgium.

Sen, Z. (1980). Statistical analysis of hydrological critical droughts. *J. Hydraul. Div., ASCE*, 106:99–115.

Sen, Z. (1998). Probabilistic formulation of spatio-temporal drought pattern. *Theor. Applied Climatology*, 61:197–206.

Shafer, B. and Dezman, L. (1982). Development of a surface water supply index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. In *Proceedings of the Western Snow Conference*, pages 164–175, Fort Collins, Colorado. Colorado State University.

Sharma, T. (2000). Drought parameters in relation to truncation levels. *Hydrological Processes*, 14(7):1279–1288.

Shiau, J. (2003). Return period of bivariate distributed hydrological events. *Stochastic Environmental Research and Risk Assessment*, 17(1-2):42–57.

Shiau, J. and Shen, H. (2001). Recurrence analysis of hydrologic droughts of differing severity. *J. Water Resource Planning and Management*, 127(1):30–40.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

Sloto, R. and Crouse, M. (1996). HYSEP: A computer program for streamflow hydrograph separation and analysis. Water resources investigation report 96-4040, U.S. Geological Survey.

Smakthin, V. (2001). Low flow hydrology: a review. *Journal of Hydrology*, 240(3-4):147–186.

Smakthin, V. and Hughes, D. (2004). *Review, automated estimation and analysis of drought indices in South Asia-Working paper (83)*. IWMI, Colombo, Sri Lanka.

Smith, L., Turcotte, D., and Isacks, B. L. (1998). Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrolological Processes*, 12:233–249.

Stedinger, J. (1980). Fitting Log normal distributions to hydrologic data. *Water Resources Research*, 16(3):481–490.

Stedinger, J., Pei, D., and Cohn, T. A. (1985). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water resources research*, 21(5):665–675.

Szalai, S., Szinell, C., and Zoboki, J. (2000). Drought monitoring in Hungary. In *Early warning systems for drought preparedness and drought management*. World Meteorological Organization, Lisboa.

Tallaksen, L. and Hisdal, H. (1997). Regional analysis of extreme streamflow drought duration and deficit volume. In *Proceedings of the Symp. Regional Hydrology: Concepts and Models for Sustainable Water Resource Management (FRIEND '97)*, pages 141–150, Postojna, Slowenia. International Association of Hydrological Sciences.

Tallaksen, L. and Hisdal, H. (1999). Methods for regional classification of streamflow drought series: The EOF method and L-moments. Technical Report 2, ARIDE project.

Tallaksen, L. and van Lanen, H. (2004). *Hydrological Drought - Processes and Estimation Methods for Stream flow and Groundwater. Developments in Water Science, 48.* Elsevier Science B.V., Amsterdam.

Tate, E. and Gustard, A. (2000). Drought definition: a hydrological perspective. In J.V.Vogt and F.Somma, editors, *Drought and Drought Mitigation in Europe*, pages 23–48. Kluwer Academic Publishers, the Netherlands.

Torrence, C. and Compo, G. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78.

UNEP (1992). *World Atlas of Desertification*. United Nations Environment Programme, London.

Valencia, D. and Jr., J. S. (1973). Disaggregation processes in stochastic hydrology. *Water resources research*, 9(3):580–585.

van der Schrier, G., Briffa, K. R., Jones, P. D., and Osborn, T. J. (2006). Summer moisture variability across Europe. *J. Climate*, 19:2818–2834.

van Lanen, H. A. J. and Peters, E. (2000). Definition, effects, and assessment of groundwater drought. In Vogt, J. V. and Somma, F., editors, *Drought and Drought Mitigation in Europe*, pages 49–61. Kluwer Publ., Dordrecht.

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley, New York.

Wang, Q. (1996). Direct sample estimators of L-moments. *Water resources research*, 32:3617–3619.

WAPCOS (1995). Water resources development master plan for Ethiopia. Final report, volume I to X. prepared for EVDSA, Water and Power Consulting Services I Ltd, Addis Ababa, Ethiopia.

Watanachaturaporn, P. and Arora, M. K. (2004). SVM for classification of multi- and hyperspectral data. In Varshney, P. K. and Arora, M. K., editors, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. Springer-Verlag.

Wei, W. (1990). *Time Series Analysis - Univariate and Multivariate Methods*. Addison & Wesley, UK.

Wells, N., Goddard, S., and Hayes, M. (2004). A self-calibrating Palmer Drought Severity Index. *Journal of Climatology*, 17:2335–2351.

Wilhite, D. and Buchanan-Smith., M. (2005). Drought as hazard: Understanding the natural and social context. In Wilhite, D., editor, *Drought and Water Crises: Science, Technology, and Management Issues.*, pages 3–29. CRC Press, Boca Raton, FL.

Wilhite, D. and M.H.Glantz (1987). Understanding the drought phenomenon: the role of definitions. In Wilhite, D., Easterling, W., and D.A.Wood, editors, *Planning for Drought.*, pages 11–27. Vestview Press, Boulder, CO.

Wiltshire, S. (1985). Grouping basins for regional flood frequency analysis. *Hydrological Sciences Journal*, 30(1):151–159.

Wiltshire, S. (1986). Identification of homogeneous regions for flood frequency analysis. *J.Hydrology*, 84:287–302.

Woodhouse, C. and Meko, D. (2002). Introduction to tree-ring based streamflow reconstructions. *Southwest Hydrology*, 1:14–15.

Wright, J., Houck, M., Diamond, J., and Randall, D. (1986). Drought contingency planning. *Civil Engineering Systems*, 3(4):210–221.

Yevjevich, V. (1967). *An Objective Approach to Definitions and Investigations of Continental Hydrologic Drought, Hydrology Paper, N. 23.* Colorado State University, Fort Collins, CO.

Zecharias, Y. and Brutsaert, W. (1988). Recession characteristics of groundwater outflow and base flow from mountainous watersheds. *Water Resources Research*, 24(10):1651–1658.

Zrinji, Z. and Burn, D. (1996). Regional flood frequency with hierarchical region of influence. *J.Water Res. Planning-ASCE*, 122(4):245–252.

Appendix A

Characteristics of the river basin

Location

The Wabi Shebele river basin is a transboundary river basin that starts in Ethiopia and flows through Somalia to the Indian Ocean (Figure A.1).

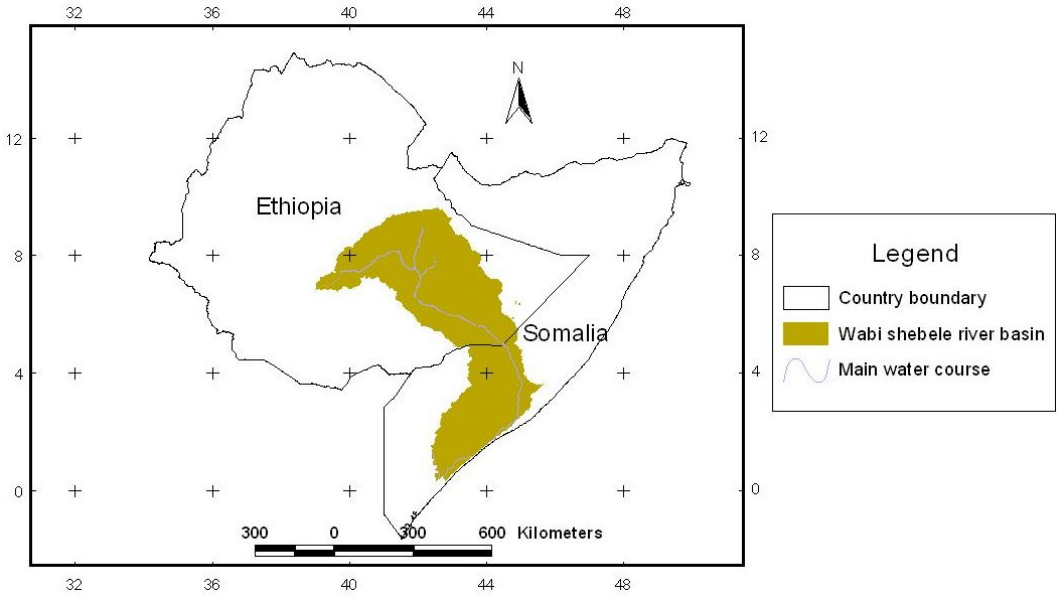


FIGURE A.1: Location of the whole Wabi Shebele river basin in East Africa.

Climate

The mean annual precipitation shows wide variation. The North western portion receives high precipitation compared with very small precipitation received at the far downstream areas. The coefficient of variability is very high in the arid downstream portion (Figure A.2). Monthly average precipitation at the North Western portion of the basin shows that there is a decrease in the amount in the later half of the 90's (Figure A.3).

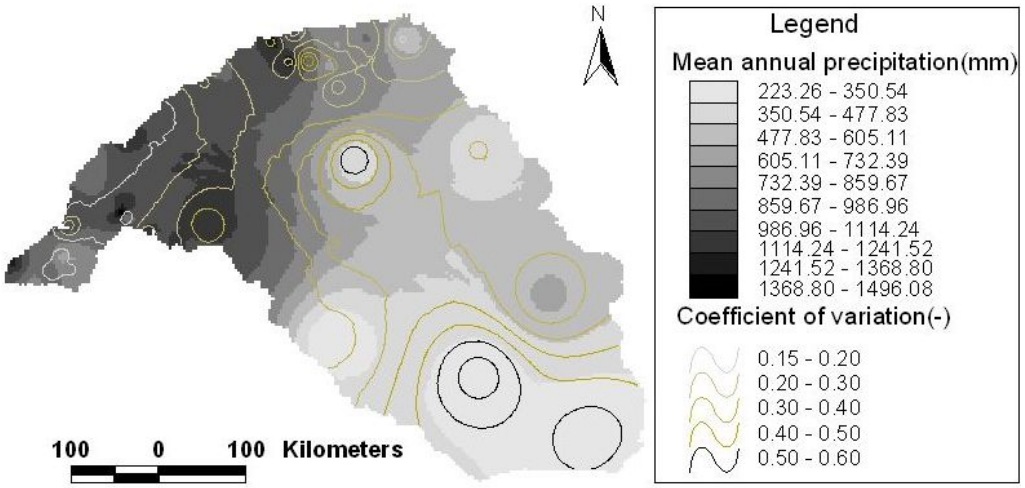


FIGURE A.2: Mean annual precipitation(mm) and coefficient of variation.

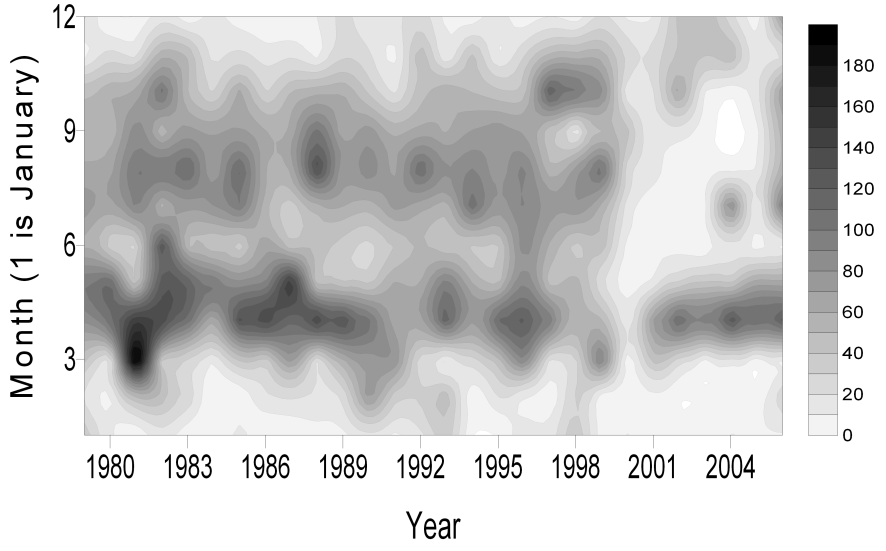


FIGURE A.3: Monthly average precipitation at the North Western portion of Wabi Shebele

This seasonal reduction in precipitation is markedly high in the main rainy season.

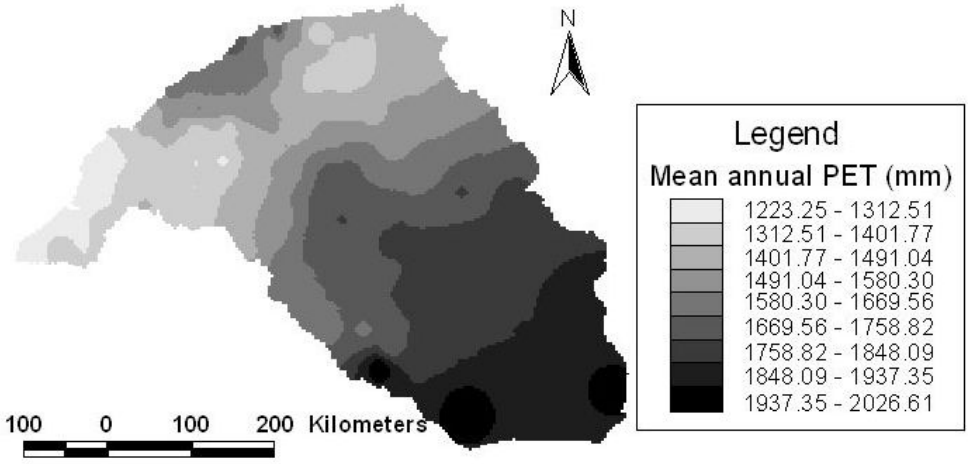


FIGURE A.4: Mean annual potential evaporation(mm).

Settlement

The population density averaged at sub-basin level in Wabi Shebele river basin (Figure A.5).

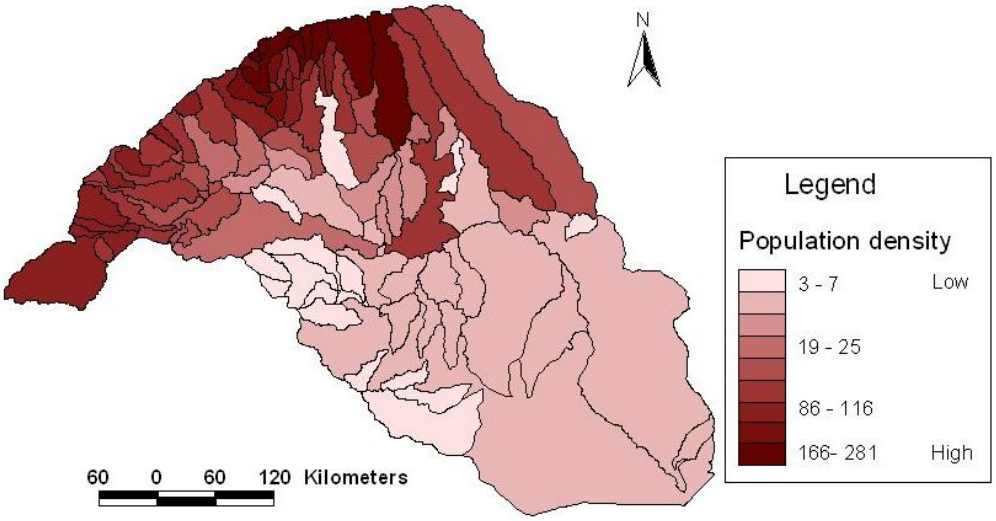


FIGURE A.5: Average population density at each of the subbasins in Wabi Shebele river basin.

The population settlement is higher in the humid upstream portions of the river basin.

Relief

The lower arid portion of the basin has almost a flat slope. The road network density is low.

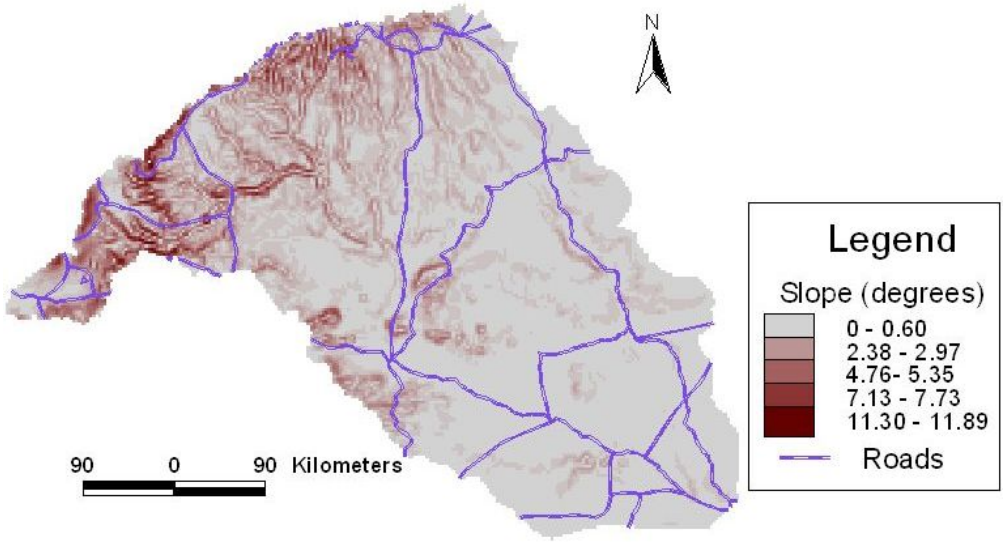


FIGURE A.6: Average slope and major road network in Wabi Shebele river basin.

Hydrogeology

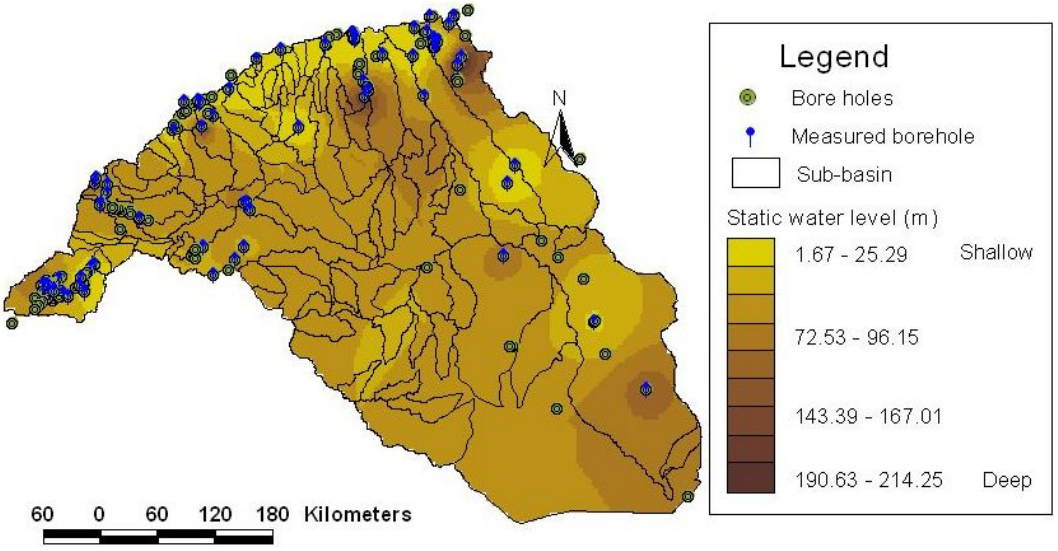


FIGURE A.7: Distribution of boreholes and those with data on safe yield and static water level (in metres) from surface

TABLE A.1: Some attributes of the Meteorological stations in and around Wabi shebele river basin

Station code	Nearby town	Installed year	Latitude (degrees)	Longitude (degrees)	Altitude (metres above msl)	Class
39070101	Adaba	1955	7.01	39.24	2420	4
40090014	Afdem	1962	9.28	40.59	1030	4
39070064	Agarfa	1979	7.16	39.49	2550	1
42090214	Alemaya	1954	9.26	42.01	2125	4
39070173	Arata	1974	7.57	39.04	1820	3
39070011	Assassa	1960	7.07	39.11	2350	1
40090243	Assebe Tefferi	1980	9.04	40.52	1900	3
42090043	Babile	1968	9.13	42.2	1600	3
41090194	Bedeno	1968	9.08	41.38	2050	4
40080244	Bedessa	1962	8.55	40.46	1820	3
39070744	Bekoji farm	1978	7.32	39.15	2810	1
41090021	Deder	1954	9.19	41.28	2350	1
43080053	Degehabur	1968	8.13	43.33	1070	1
42040011	Delo Odo	1972	4.12	42.02	605	1
40070034	Delo Sebro	1968	7.15	40.28	2200	4
41090222	Dire Dawa	1952	9.36	41.51	1260	1
39060131	Dodolla Edo	1954	6.59	39.11	3000	3
42080054	Fedis	1968	9.08	42.03	1700	4
40070044	Gassera	1968	7.22	40.11	2320	4
40080021	Gelemso	1969	8.49	40.31	1940	4
40070011	Ginnir	1955	7.08	40.42	1750	4
43050021	Gode	1967	5.54	43.35	295	1
41090061	Grawa	1988	9.08	41.5	2100	4
39080383	Gunna	1850	8.23	39.56	1800	4
42090164	Gursum	1969	9.21	42.23	1900	3
42090031	Harar	1980	9.18	42.1	2100	3
41090284	Hirna	1968	9.13	41.07	2050	3
39070253	Hunte	1981	7.03	39.24	2380	1
42060034	Hurso	1981	9.37	41.39	1200	1
39080364	Huruta	1974	8.09	39.22	2000	1
42060011	Imi	1984	6.28	42.09	1100	4
39070144	Indito	1968	7.34	39.49	2480	4
40070224	Jarra	1850	7.27	40.46	1960	4
42090011	Jijiga	1952	9.2	42.47	1775	1
43090013	Kebri Beyah	1968	9.06	43.1	1600	3
44060011	Kebri Dehar	1968	9.06	43.1	1600	4

Continued Table A.1

Station code	Nearby town	Installed year	Latitude (degrees)	Longitude (degrees)	Altitude (metres above msl)	Class
44050023	Kelafo	1954	5.36	44.07	250	3
39080074	Keleta	1962	8.19	39.25	1500	4
38070013	Koffele	1952	7.04	38.48	2620	3
38070384	Kora	1962	9.07	40.32	1530	4
41090314	Kulubi	1960	9.26	41.42	2000	4
39080121	Kulumsa	1963	8.08	39.08	2200	1
39070073	Mararo	1968	7.27	39.22	2940	4
40080304	Matchara	1966	9.3	42.54	1700	4
42080044	Midagalola	1965	8.47	42.03	1300	4
38070373	Munesa Mission	1958	7.35	38.54	2510	3
44050054	Mustahil	1968	5.15	44.44	250	4
39080153	Ogolcha	1967	8.04	39.02	1760	3
39070111	Robe Arsi	1968	7.51	39.37	2400	1
40070392	Robe Bale	1973	7.08	40	2480	1
39070021	Sagure	1958	7.46	39.09	2480	4
39070084	Sedika	1968	7.43	39.4	2545	4
38070044	Shashemene	1954	7.12	38.36	2080	3
40070061	Sheneka	1979	7.15	40	2400	1
38070274	Sire	1953	7.13	38.58	2390	3
40060014	Sofomor	1987	6.54	40.5	1680	4
39070031	Ticho	1959	7.49	39.32	2550	3
42040024	Weldia	1987	4.2	42.04	1001	4

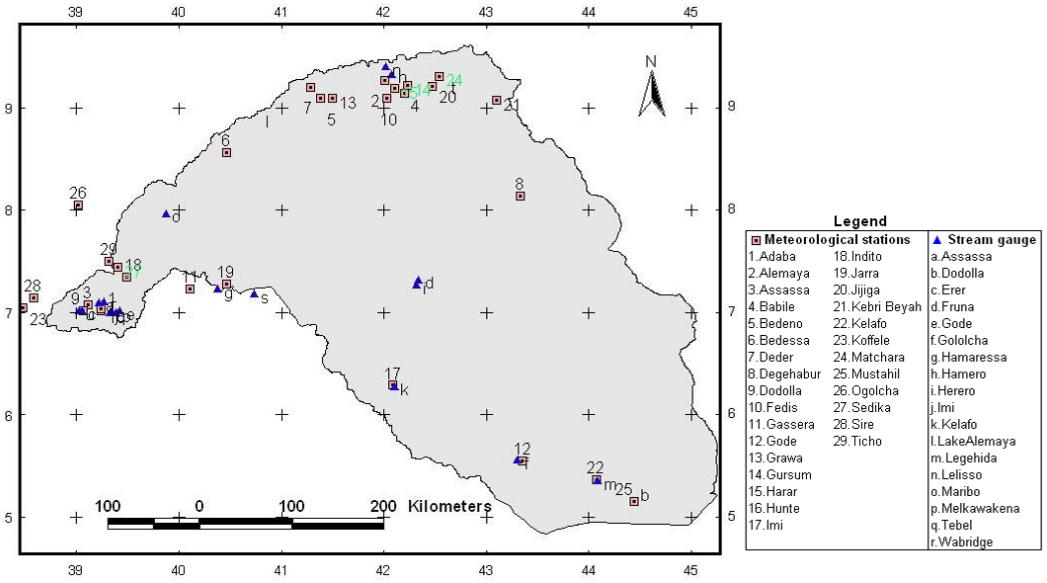


FIGURE A.8: Location of some of the hydrometeorological gauging stations

TABLE A.2: Some attributes of the Hydrological stations in Wabi shebele river basin

River/lake	Nearby town	Latitude (degrees)	Longitude (degrees)	Installed year	SG[a]	WLR[b]	Area Km^2
Lelisso	Adaba	7.000	39.383	10-1-67	x	x	135
Maribo	Adaba	7.000	39.333	24-1-67	x	x	185
Ukuma	Dodola	7.017	39.050	25-1-67	x		137
Wabi	@ Bridge	7.017	39.033	30-1-67	x	x	1035
Assassa	Assassa	7.100	39.217	25-2-67	x		68.1
Wolkessa	Azazera Kerey	7.833	39.550	8-3-67	x		27
Ulul	Azazera Kerey	7.833	39.550	9-3-67	x	x	21
Harerghe	Assa Osman	7.767	39.550	25-6-67	x	x	41
Wabi	Melka wakena	7.11	39.27	30-1-67			4452
Robe	Robe	7.850	39.633	8-3-69			175
Herero	@ Herero	7.000	39.317	24-1-67	x		133
Maribo	Kara Birole	6.867	39.367	5-7-81	x	x	200
Alkeso	Alkeso	7.867	39.550	27-05-98	x		49.6
Einemor	Ticho	7.817	39.517	26-05-98	x		49.4
Medhaidu	Bedessa	8.900	40.767	04-07-98	x	x	
L.Alemaya	@ Alemaya	9.400	42.017	27-3-75	x		50
L.Adele	@ Adele	9.433	41.950	25-5-80	x		48
Hamaressa	Harer	9.333	42.083	26-11-80	x		56
Bissidimo	@ Bissidimo	9.333	42.133	25-11-80	x		147
Hirna	Hirna	9.217	41.100	2-12-80	x		25
Dawe	Gara Muleta	9.333	41.800	12-2-81	x		344
Tebel	Ginir	7.183	40.733	27-1-83	x		79
Upper Erer	@ Erer	9.233	42.250	4-2-83	x		469
Upper Fafen	@ Bridge	9.233	42.600	19-6-83	x		900
Jawes	Bedessa	8.900	40.783	18-1-84	x		21.5
Jijiga	@ Jijiga	9.350	42.800	11-3-85	x		947.5
Ramis	Ramis	7.85	41.34		x		10384
Dungata	Dungata	8.42	40.42		x		5534
Kollu	Bedessa	8.900	40.783	27-5-87	x		21
Erer	Hamero	7.32	42.34		x		15132
Gololcha	Gololcha	7.23	40.38		x		7163
Daketa	Hamero	7.29	42.28		x		15188
Wabi	Imi	6.28	42.1	1-09-67	x		91600
Wabi	Legehida	7.97	39.87		x		20473
Wabi	Hamero	7.27	42.32	1-09-67	x		63644
Wabi	Gode	5.56	43.3	1-09-67	x	x	124108
Wabi	Kelafo	5.36	44.08	1-09-67	x		139100
Wabi	Burkur	5.25	44.44		x		144000

[a]SG: Staff Gauge
[b]WLR: Water Level Recorder

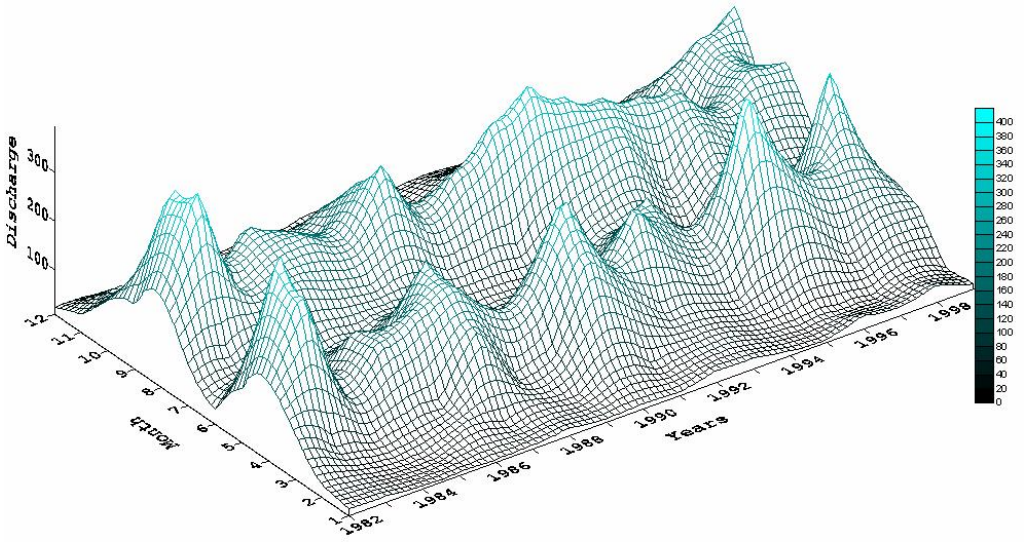


FIGURE A.9: Mean monthly stream flow at Imi gauging station (m^3/s) for the period 1982-2000

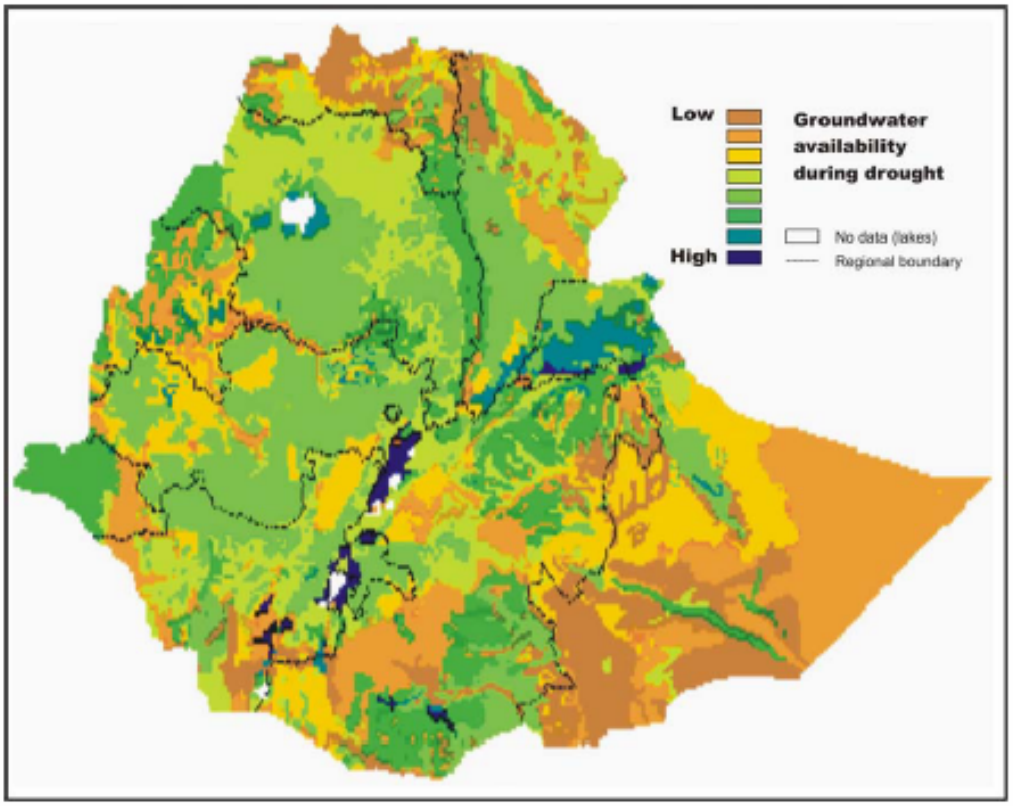


FIGURE A.10: Summary map - groundwater availability during drought(Calow et al., 2000)

Appendix B

Method of L-Moments

The L-moments are derived from a linear combination of the probability weighted moments (PWMs) and represent an alternative set of scale and shape statistics of a data sample or a probability distribution (Hosking, 1990). Their main advantages over conventional (product) moments are that they are able to characterise a wider range of distributions and (when estimated from a sample) are less subject to bias in estimation and more robust to the presence of outliers in the data

Probability weighted moments are defined as (Greenwood et al., 1979):

$$\beta_r = E[xF(x)^r],$$

which can be written as

$$\beta_r = \int_0^1 x(F)F^r dF,$$

Where β_r is the rth order PWM and $F(x)$ is the cumulative distribution function (CDF) of x, $x(F)$ is the inverse CDF of x evaluated at the probability F, and r = 0, 1, 2, . . ., is a nonnegative integer. When r = 0, β_0 is equal to the mean of the distribution $\mu = E[x]$.

For any distribution the rth L-moment λ_r is related to the rth PWM (Hosking, 1990) through:

$$\lambda_{r+1} = \sum_{k=0}^{r} \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \tag{B.1}$$

Let x be a real valued random variable with cumulative distribution function F(x) and

let $x_{1:n} \leq x_{1:n} \leq \ldots \leq x_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of x. The first four L-moments are related to the PWMs using

$$\lambda_1 = E\{x_{1:1}\} = \beta_0 \tag{B.2}$$

$$\lambda_2 = \frac{1}{2}E\{x_{2:2} - x_{1:2}\} = 2\beta_1 - \beta_0 \tag{B.3}$$

$$\lambda_3 = \frac{1}{3}E\{x_{3:3} - 2x_{2:3} + x_{1:3}\} = 6\beta_2 - 6\beta_1 + \beta_0 \tag{B.4}$$

$$\lambda_4 = \frac{1}{4}E\{x_{4:4} - 3x_{3:4} + 3x_{2:4} - x_{1:4}\} = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{B.5}$$

L-moment ratios are the quantities:

$$\tau = \lambda_2/\lambda_1 \; and \; \tau_r = \lambda_r/\lambda_2$$

which are analogues to the traditional ratios, i.e. τ is the L-coefficient of variation (L-CV), τ_3 is the L-skewness and τ_4 is the L-kurtosis.

Parameters are estimated by equating the sample L-moments with the distribution L-moments. In practise, the L-moments must be estimated from a finite sample. The unbiased sample estimates of the PWM for any distribution can be computed from;

$$b_r = n^{-1}\sum_{i=0}^{n}\frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)}x_i, \qquad r = 0, 1, \cdots, n-1. \tag{B.6}$$

where x_i is an ordered set of observations $x_1 \leq x_2 \leq x_3 \leq \ldots x_n$.

L-moments and L-moment ratios are more convenient than probability-weighted moments, because they are more easily interpretable as measures of distributional shape. In particular, λ_1, is the mean of the distribution, a measure of location; λ_2, is a measure of scale; and τ_3 and τ_4, are measures of skewness and kurtosis, respectively.

Probability distributions

Modified from Tallaksen and van Lanen (2004).

Weibull distribution

Definition

Parameters: $\xi(location), \alpha(scale), \kappa(shape)$

Range: $\alpha > 0, \kappa > 0, \xi < x < \infty$

Special case: Exponential distribution for $\kappa = 1$

$$f(x) = \frac{\kappa}{\alpha}\left(\frac{x-\xi}{\alpha}\right)^{\kappa-1} exp\left[-\left(\frac{x-\xi}{\alpha}\right)^{\kappa}\right] \tag{B.7}$$

$$F(x) = 1 - exp\left[-\left(\frac{x-\xi}{\alpha}\right)^{\kappa}\right] \tag{B.8}$$

$$x_p = \xi + \alpha[-ln(1-p)]^{1/\kappa} \tag{B.9}$$

The Weibull distribution is a reverse Generalized Extreme Value distribution with parameters:

$\xi_{GEV} = \xi_{WEI} - \alpha_{WEI}, \quad \alpha_{GEV} = \frac{\alpha_{WEI}}{\kappa_{WEI}}, \quad \kappa_{GEV} = \frac{1}{\kappa_{WEI}}$

where subscripts GEV and WEI refer to Generalized Extreme Value and Weibull distributions respectively.

Moments

$$\mu = \xi + \alpha\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{B.10}$$

$$\sigma^2 = \alpha^2\left[\Gamma\left(1+\frac{2}{\kappa}\right) - \left[\Gamma\left(1+\frac{1}{\kappa}\right)\right]^2\right] \tag{B.11}$$

$$\gamma_3 = \frac{\Gamma\left(1+\frac{3}{\kappa}\right) - 3\Gamma\left(1+\frac{1}{\kappa}\right)\Gamma\left(1+\frac{2}{\kappa}\right) + 2\left[\Gamma\left(1+\frac{1}{\kappa}\right)\right]^3}{\left[\Gamma\left(1+\frac{2}{\kappa}\right) - \left[\Gamma\left(1+\frac{1}{\kappa}\right)\right]^2\right]^{3/2}} \tag{B.12}$$

where $\Gamma(\cdot)$ is the Gamma function.

$$\Gamma(n) = \int_0^\infty e^{-t}t^{n-1}dt \qquad n > 0 \tag{B.13}$$

L-moments

$$\lambda_1 = \xi + \alpha\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{B.14}$$

$$\lambda_2 = \alpha(1 - 2^{-1/\kappa})\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{B.15}$$

$$\tau_3 = 3 - \frac{2(1 - 3^{-1/\kappa})}{1 - 2^{-1/\kappa}} \tag{B.16}$$

Moment estimates

If ξ is known, the moment estimate of κ can be obtained by combining equations B.10 and B.11, which is solved using Newton-Raphson iteration. The moment estimate of α is then given by:

$$\hat{\alpha} = \frac{\hat{\mu} - \xi}{\Gamma\left(1 + \frac{1}{\kappa}\right)} \tag{B.17}$$

If ξ is unknown, the moment estimate of κ can be obtained from the skewness estimator (equation B.12) using Newton-Raphson iteration.

Moment estimates of ξ and α are given by:

$$\hat{\alpha} = \frac{\hat{\sigma}}{\sqrt{\Gamma\left(1 + \frac{2}{\hat{\kappa}}\right) - \left[\Gamma\left(1 + \frac{1}{\kappa}\right)\right]^2}}, \quad \hat{\xi} = \hat{\mu} - \hat{\alpha}\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{B.18}$$

L-moment estimates

If $\hat{\xi}$ is known, L-moment estimates of α and κ are given by:

$$\hat{\kappa} = -\frac{ln\,2}{ln\left(1 - \frac{\hat{\lambda}_2}{\hat{\lambda}_1}\right)}, \quad \hat{\alpha} = \frac{\hat{\lambda}_1 - \xi}{\Gamma\left(1 + \frac{1}{\kappa}\right)} \tag{B.19}$$

If ξ is unknown, the shape parameter can be estimated from the approximate formula for estimation of the shape parameter of the GEV distribution using $-\tau_3$ and $\kappa_{GEV} = 1/\kappa$. L-moment estimates of ξ and α are then given by:

$$\hat{\alpha} = \frac{\hat{\lambda}_2}{\left(1 - 2^{-1/\kappa}\right)\Gamma\left(1 + \frac{1}{\hat{\kappa}}\right)}, \quad \hat{\xi} = \hat{\lambda}_1 - \hat{\alpha}\Gamma\left(1 + \frac{1}{\hat{\kappa}}\right) \tag{B.20}$$

Maximum likelihood estimates

If ξ is known, the maximum likelihood estimate of κ can be obtained by solving:

$$\frac{1}{\kappa} = \frac{\sum_{i=1}^{n}(x_i - \xi)^{\kappa} ln(x_i - \xi)}{\sum_{i=1}^{n}(x_i - \xi)^{\kappa}} - \frac{1}{n}\sum_{i=1}^{n} ln(x_i - \xi) \tag{B.21}$$

using Newton-Raphson iteration. The maximum likelihood estimate of α is subsequently obtained from:

$$\hat{\alpha} = \left[\frac{1}{n}\sum_{i=1}^{n}(x_1 - \xi)^{\hat{\kappa}}\right]^{1/\hat{\kappa}} \tag{B.22}$$

Generalized pareto (GP) distribution

Definition

Parameters: $\xi(location), \alpha(scale), \kappa(shape)$

Range: $\alpha > 0$,$\xi \leq x < \infty$ for $\kappa < 0$, $\xi \leq x \leq \xi + \alpha/\kappa$ for $\kappa > 0$

Special case: Exponential distribution for $\kappa = 0$

$$f(x) = \frac{1}{\alpha}\left[1 - \kappa\frac{x-\xi}{\alpha}\right]^{1/\kappa - 1} \tag{B.23}$$

$$F(x) = 1 - \left[1 - \kappa\frac{x-\xi}{\alpha}\right]^{1/\kappa} \tag{B.24}$$

$$x_p = \xi + \frac{\alpha}{\kappa}\left[1 - (1-p)^{\kappa}\right] \tag{B.25}$$

Moments

$$\mu = \xi + \frac{\alpha}{1+\kappa} \tag{B.26}$$

$$\sigma^2 = \frac{\alpha^2}{(1+\kappa)^2(1+2\kappa)} \tag{B.27}$$

$$\gamma_3 = \frac{2(1-\kappa)\sqrt{1+2\kappa}}{1+3\kappa} \tag{B.28}$$

L-moments

$$\lambda_1 = \xi + \frac{\alpha}{1+\kappa} \tag{B.29}$$

$$\lambda_2 = \frac{\alpha}{(1+2\kappa)(2+\kappa)} \tag{B.30}$$

$$\tau_3 = \frac{(1-\kappa)}{(3+\kappa)} \tag{B.31}$$

Moment estimates

If ξ is known, the moment estimate of α and κ are given by:

$$\hat{\kappa} = \frac{1}{2}\left[\left(\frac{\hat{\mu}-\xi}{\sigma}\right)^2 - 1\right], \quad \alpha = (\hat{\mu}-\xi)(1+\hat{\kappa}) \tag{B.32}$$

If ξ is unknown, κ can be obtained from the skewness estimator (equation C.6) using Newton-Raphson iteration scheme.

Moment estimates of ξ and α are subsequently obtained from:

$$\hat{\alpha} = \hat{\sigma}(1+\hat{\kappa})\sqrt{1+2\kappa}, \quad \hat{\xi} = \hat{\mu} - \frac{\hat{\alpha}}{1+\hat{\kappa}} \tag{B.33}$$

L-moment estimates

If $\hat{\xi}$ is known, L-moment estimates of α and κ are given by:

$$\hat{\kappa} = \frac{\hat{\lambda}_1 - \xi}{\hat{\lambda}_2} - 2, \quad \hat{\alpha} = (\hat{\lambda}_1 - \xi)(1+\hat{\kappa}) \tag{B.34}$$

If ξ is unknown, L-moment estimates are given by:

$$\hat{\kappa} = \frac{1-3\hat{\tau}_3}{1+\hat{\tau}_3}, \quad \hat{\alpha} = \hat{\lambda}_2(1+\hat{\kappa})(2+\hat{\kappa}), \quad \hat{\xi} = \hat{\lambda}_1 - \frac{\hat{\alpha}}{1+\hat{\kappa}} \tag{B.35}$$

Maximum likelihood estimates

The log likelihood function reads:

$$L = -nln\alpha + \frac{1-\kappa}{\kappa}\sum_{i=1}^{n} ln\left[1 - \frac{\kappa}{\alpha}(x_i - \xi)\right] \tag{B.36}$$

If ξ is known, the maximum likelihood estimates can be obtained by solving:

$$\frac{\partial L}{\partial \alpha} = 0, \quad \frac{\partial L}{\partial \kappa} = 0, \tag{B.37}$$

using a modified Newton-Raphson iteration scheme (Hosking and Wallis, 1997).

Log-Normal (LN) distribution

Definition

Parameters: $\xi(location), \mu_y\,(mean), \sigma_y\,(standard\,deviation)$

Range: $\sigma_y > 0$,$x > \xi$

If x is distributed according to a Log-Normal distribution, then $y = ln(x-\xi)$ is normally distributed. The parameters μ_y and σ_y are the population mean and variance of y.

$$f(x) = \frac{1}{(x-\xi)\sigma_y\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{ln(x-\xi)-\mu_y}{\sigma_y}\right)^2\right] \tag{B.38}$$

$$F(x) = \Phi\left(\frac{ln(x-\xi)-\mu_y}{\sigma_y}\right) \tag{B.39}$$

$$x_p = \xi + exp(\mu_y + \sigma_y\Phi^{-1}(p)) \tag{B.40}$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the cumulative distribution function and the quantile function of the standard Normal distribution.

Moments

$$\mu_x = \xi + exp\left[\mu_y + \frac{1}{2}\sigma^2\right] \tag{B.41}$$

$$\sigma_x^2 = (exp[2\mu_y + \sigma_y^2])(exp(\sigma_y^2) - 1) \tag{B.42}$$

$$\gamma_3 = 3\phi + \phi^3, \quad \phi = \sqrt{exp(\sigma_x^2) - 1} \tag{B.43}$$

L-moments

$$\lambda_{1,y} = \mu_y \tag{B.44}$$

$$\lambda_{2,y} = \frac{\sigma_y}{\sqrt{\pi}} \tag{B.45}$$

Moment estimates

If ξ is known, moment estimates of μ_y and σ_y are given by the sample mean and standard deviation of the logarithmic transformed data $\{y_i = ln(x_i - \xi), \quad i = 1, 2, \ldots, n\}$.

If ξ is unknown it can be estimated based on a lower bound quantile estimator of ξ (Stedinger, 1980) or based on the sample moments in real space $\{x_i, i = 1, 2, \ldots, n\}$ where a bias correction of the sample skewness is adopted.

L-moment estimates

If ξ is known, moment estimates of μ_y and σ_y can be estimated from the sample L-moments of the logarithmic transformed data $\{y_i = ln(x_i - \xi),\ i = 1, 2, \ldots, n\}$:

$$\hat{\mu}_y = \hat{\lambda}_{1,y}, \quad \hat{\sigma}_y = \sqrt{\pi}\hat{\lambda}_{2,y} \tag{B.46}$$

Maximum likelihood estimates

If ξ is known, maximum likelihood estimates of μ_y and σ_y are given by:

$$\hat{\mu}_y = \frac{1}{n}\sum_{i=1}^{n} ln(x_i - \xi), \quad \hat{\sigma}_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[ln(x_i - \xi) - \hat{\mu}_y]^2} \tag{B.47}$$

If ξ is unknown, the maximum likelihood estimate of ξ can be obtained by solving:

$$\frac{\partial L}{\partial \xi} = \frac{\partial}{\partial \xi}\left(\sum_{i=0}^{n} ln(\sqrt{2\pi}\sigma(x_i - \xi)) + \frac{1}{2}\sum_{i=0}^{n}\left(\frac{ln(x_i - \xi) - \mu}{\sigma}\right)^2\right) = 0 \tag{B.48}$$

using a bisection iteration scheme. The parameter estimates of μ_y and σ_y are subsequently obtained from equation B.47.

Four-parameter Kappa distribution

Definition

Parameters: $\xi(location)$, $\alpha(scale)$, $\kappa(shape)$, $h(shape)$

Range: $\alpha > 0$

Special case: Several established distributions are special cases of the four-parameter kappa distribution. When the parameter h = 1, the distribution yields a generalized Pareto distribution; h = 0, a GEV distribution; and h = - 1 a generalized logistic distribution. An exponential distribution arises when h = 1 and k = 0, a Gumbel distribution when h = 0 and k = 0, a logistic distribution when h = -1 and k = 0, and a uniform distribution when h = 1 and k = 1. When h = 0 and k = 1, the four-parameter kappa distribution is a reverse exponential distribution; i.e., 1 - F(-x) is the cumulative distribution function of an exponential distribution.

$$F(x) = \begin{cases} \{1 - h[1 - \kappa(x - \xi)/\alpha]^{1/\kappa}\}^{1/h}, & \text{if } \kappa \neq 0, h \neq 0, \\ exp\{-[1 - \kappa(x - \xi)/\alpha]^{1/\kappa}\}, & \text{if } \kappa \neq 0, h = 0, \\ \{1 - h\, exp[-(x - \xi)/\alpha]\}^{1/h}, & \text{if } \kappa = 0, h \neq 0, \\ exp\{-exp[-(x - \xi)/\alpha]\}, & \text{if } \kappa = 0, h = 0, \end{cases} \tag{B.49}$$

Since the three other forms of the CDF of the four-parameter kappa distribution are the limiting forms of the first case in equation B.49 as k or h approaches 0, it is convenient to write the CDF simply as expressed in equation B.49.

Thus the probability density function, f(x) is:

$$f(x) = \alpha^{-1}[1 - \kappa(x - \xi)/\alpha]^{(1/k)-1}[F(x)]^{1-h} \tag{B.50}$$

and the quantile function (inverse cumulative distribution function), x(F):

$$x(F) = \xi + \alpha/\kappa \left[1 - \left(\frac{1 - F^h}{h}\right)^{\kappa}\right] \tag{B.51}$$

To enable the four parameters of the distribution to be estimated from the first four L-moments, the parameter space must be restricted, so that only one set of parameters corresponds to a given set of the first four L-moments. This corresponds to the following conditions on the parameters (Hosking, 1994):

(a) $K > -1$;

(b) if $h < 0$, then $hk > -1$;

(c) $h > -1$;

(d) $k + 0.725h > -1$.

Conditions a and b ensure the existence of the L-moments; conditions c and d ensure the uniqueness of the parameters, given the four L-moments, β_r.

L-moments

If $\kappa \neq 0$,the β_r are given as:

$$r\beta_{r-1} = \begin{cases} \xi + \alpha/\kappa \left[1 - \frac{r\Gamma(1+\kappa)\Gamma(r/h)}{h^{1+\kappa}\Gamma(1+\kappa+r/h)}\right], & \text{if } h > 0, \kappa > -1, \\ \xi + \alpha/\kappa[1 - r^{-\kappa}\Gamma(1+\kappa)], & \text{if } h = 0, \kappa > -1, \\ \xi + \alpha/\kappa \left[1 - \frac{r\Gamma(1+\kappa)\Gamma(-\kappa-r/h)}{h^{1+\kappa}\Gamma(1-r/h)}\right], & \text{if } h < 0, -1 < \kappa < -1/h, \end{cases} \tag{B.52}$$

Similarly, if $\kappa = 0$ then β_r can be obtained from:

$$r\beta_{r-1} = \begin{cases} \xi + \alpha[\gamma + log\, h + \Psi(1 + r/h)], & \text{if } h > 0, \\ \xi + \alpha(\gamma + log\, r), & \text{if } h = 0, \\ \xi + \alpha[\gamma + log(-h) + \Psi(-r/h)], & \text{if } h < 0, \end{cases} \tag{B.53}$$

where $\gamma = 0.5772$...is Euler's constant, and Ψ is the digamma function, $\Psi(x) = \frac{d[log\Gamma(x)]}{dx}$.

L-moment estimates

L-moments of the four-parameter kappa distribution can be provided as in equation B.52 or equation B.53. Thus estimation of the parameters of the four-parameter kappa distribution requires the solution of these equations given the first four sample L-moments (β_0, β_1, β_2 *and* β_3). The L-moment ratios τ_3 and τ_4 are functions only of the shape parameters k and h. No explicit solution is possible but the equations can be solved by Newton-Raphson iteration.

Appendix C

Maintenance of variance extension type 2

The Maintenance of Variance Extension, Type 2 (MOVE.2) method for infilling is described by Hirsch (1982) and Alley and Burns (1983) for infilling missing flow data. This technique has the ability to maintain the variance of infilled data sets. The flow values at the reference station are denoted as "x(i)" where "i" is an index of time. The short observed data at the station to be filled are denoted as "y(i)". The observed events for the two sequences are represented as:

$x(1), \cdots, x(N_1)$, $x(N_1+1), \cdots, x(N_1+N_2)$

$y(1), \cdots, y(N_1)$

where N_1 represents the number where the two sequences have data and N_1+N_2 represents the total number of values. The following notations identify the different parts of the two series.

$x = x(1), \cdots, x(N_1)$, $x(N_1+1), \cdots, x(N_1+N_2)$

$x_1 = x(1), \cdots, x(N_1)$

$x_2 = x(N_1+1), \cdots, x(N_1+N_2)$

$y_1 = y(1), \cdots, y(N_1)$

It is not necessary for the two sequences to begin or end simultaneously, nor for the

observations be consecutive (Hirsch, 1982). The missing values are estimated by the relationship described below. Note that "m()" and "$S^2()$" represent the mean and variance of the series in the parentheses respectively. The variable denoted as "r" represents the product moment correlation coefficient of "x_1" and "y_1". The strength of the MOVE.2 method is that the mean and variance estimates for "x" are based on all "$N_1 + N_2$" observations, and the mean and variance estimates for "y" (i.e., "$m\prime(y)$" and "$S\prime^2(y)$" respectively) are based on the historical values of "y" and on information transferred from the "x" sequence of data.

$$y\prime(i) = m\prime(y) + \frac{S\prime(y)}{S(x)}[x(i) - m(x)] \tag{C.1}$$

$$m\prime(y) = m(y_1) + \frac{N_2}{N_1 + N_2} r \frac{S(y_1)}{S(x_2)}(m(x_2) - m(x_1)) \tag{C.2}$$

$$S\prime^2(y) = \frac{1}{N_1 + N_2 - 1}[(N_1 - 1)S^2(y_1) + (N_2 - 1)r^2 \frac{S^2(y_1)}{S^2(x_1)} S^2(x_2) +$$

$$(N_2 - 1)a^2(1 - r^2)S^2(y_1) + \frac{N_1 N_2}{N_1 + N_2} r^2 \frac{S^2(y_1)}{S^2(x_1)}(m(x_2) - m(x_1))^2] \tag{C.3}$$

$$a^2 = \frac{N_2(N_1 - 4)(N_1 - 1)}{(N_2 - 1)(N_1 - 3)(N_1 - 2)} \tag{C.4}$$

Glossary

Autocorrelation A measure of the linear relationship between two separate instances of the same random variable as distinct from correlation which refers to the linear relationship between two distinct random variables., 29

Conditional probability Conditional probability describes the probability of an event occurring given that another event has already occurred., 142

Copula A copula is a multivariate joint distribution defined on the n-dimensional unit cube $[0, 1]^n$ such that every marginal distribution is uniform on the interval [0,1]., 128

Crisis-oriented Crisis-oriented drought management focuses on the impacts as a consequence of drought. It lacks proper preplanning to mitigate the impacts., 155

Cross-dating Cross-dating is a basic technique in dendrochronology that ensures each individual tree ring is assigned its exact year of formation. This is accomplished by matching patterns of wide and narrow rings between cores from the same tree and between trees from different locations., 91

Dendrochronology Dendrochronology is the study and method of scientific dating based on the analysis of tree ring growth patterns., 89

Drought Drought is a sustained and regionally extensive occurrence of natural water availability(e.g. precipitation, soil moisture, groundwater and streamflow) below a certain threshold level., 1

Generating function A generating function of a random variable is an expected value of a certain transformation of the variable., 130

Homogeneous pool Hydrologically homogeneous pool is considered to be delineated by those watersheds that produce the same high or low flow characteristics and pass the statistically homogeneity test., 48

Hovmoller plot Hovmoller plot is a diagram commonly used for plotting meteorological variables. The axes of a Hovmller diagram are latitude/longitude (x-axis) and time (y-axis) with the value of some field represented by color or shade., 77

Joint probability A statistical measure where the likelihood of two events occurring together and at the same point in time are calculated., 128

Kernel density Kernel density is a nonparametric way of estimating the probability density function of a random variable., 84

Partial autocorrelation The Partial autocorrelation measures the linear correlation of a series, Yt with itself, lagged by k observations, Yt+k under elimination of the influence of the variables lying between them., 110

Proxy data Proxy data is data that paleoclimatologists gather from natural recorders of climate variability, e.g., tree rings, ice cores, fossil pollen, ocean sediments, coral and historical data., 89

Risk-oriented Risk-oriented drought management unlike crisis-oriented has a long term goal of reducing the impacts of drought through the adoption of drought management plans., 153

Stream order Stream order is a measure of the position of a stream in the hierarchy of tributaries. Orders range from small streams with no branches (1st Order) up to large streams with lots of tributaries., 10

Support vector machine A type of statistical learning method applied to both classification and regression problems. The formulation embodies the structural risk minimisation principle that minimises an upper bound on the generalisation error, as opposed to empirical risk minimisation which minimises the error on the training data., 63

Transition probability The transition probability is the probability of transitioning from one state to another in a single or multiple steps., 115

Uncertainty The lack of certainty, A state of having limited knowledge where it is impossible to exactly describe existing state or future outcome, more than one possible outcome., 92

Ungauged catchment Ungauged catchment is the one that has records of hydrological observations of too short, of too poor a quality, or even nonexistent., 1

Wavelet analysis Wavelet analysis is method for analyzing nonstationarity of a time series by decomposing it into time/frequency space simultaneously. It provides information on both the amplitude of any periodic signals within the series, and how this amplitude varies with time., 93

Index