RAINFALL STUDIES FOR THE HIGHLANDS OF ERITREA

By

Mehari Tesfazghi Mebrhatu

Submitted in partial fulfilment of the requirements for the degree of
Master of Science in Agrometeorology

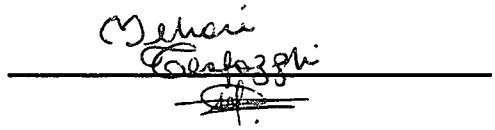
In the Faculty of Natural and Agricultural Sciences
University of the Free State

Promotor: Dr. J.C. Venter
Co-Promotor: Prof. Dr. Sue Walker

Bloemfontein
May 2003

DECLARATION

I declare that this thesis is my own work and to the best of my knowledge contains no work submitted previously as a dissertation or thesis for any degree or diploma at any other University



University of the Free State, 2003

ABSTRACT

RAINFALL STUDIES FOR THE HIGHLANDS OF ERITREA

by

Mehari Tesfazghi Mebrhatu

(MSc. in Agrometeorology, University of the Free State)

Long-term rainfall records of the highlands of Eritrea were examined to reveal the short and long-term patterns. Monthly and annual spectra have been analysed to provide a preliminary assessment in terms of homogeneity, spell length, water balance, drought and spatial variability.

Three methods of spatial interpolation were evaluated for mapping continuous spatial rainfall values for mean annual rainfall totals. Interpolation methods examined in this study include inverse distance weighted (IDW), Spline and Kriging. A statistical assessment of the continuous surface created indicates that there is little difference between the estimating ability of the three spatial interpolation methods, but with Kriging performing better overall than Spline or IDW.

Monthly means provide little information on many properties of the rainfall that are relevant to the wide variety of rainfall-related activities and decision making. A stochastic rainfall generator model based on a first-order Markov chain was developed for generation of artificial daily rainfall data from monthly totals. Intensive statistical testing of the performance of the model reveals that the model can be used with confidence to generate a pseudo daily rainfall dataset for the peak rainy months (July-August), for the highlands of Eritrea.

A deterministic model was developed to investigate how the global rainfall predictors correlate to the two peak rainy months (July-August), which contribute 65% of the total rainfall in the highlands of Eritrea. The main aim of looking at these relationships is to develop a simple statistical model for forecasting rainfall amount. In a preliminary step, in order to identify the most influential rainfall predictor a correlation matrix and stepwise regression of 11 predictors with different lags were analysed. The influence of the southern Indian Ocean Sea Surface Temperatures was identified as the most influential predictor for the highland of Eritrea. The model was validated using the jack-knife skill test, Chi-squared and deviation based test giving a promising result.

UITTREKSEL

Reënval Studies vir die Hooglande van Eritrea

deur

Mehari Tesfazghi Mebrhatu

(MSc. in Landbouweerkunde, Universiteit van die Vrystaat)

Die langtermyn reënvalverslag van Eritrea se Hooglande is ondersoek om beide die kort- en langtermyn patrone aan te dui. Maandelikse en jaarlikse spektra is geanaliseer om 'n voorlopige beraming in terme van homogeniteit, durasie lengte, waterbalans, droogte en ruimtelike veranderlikheid te voorsien.

Drie metodes van ruimtelike interpolasie is geëvalueer om kontinue ruimtelike reënvalwaardes vir gemiddelde jaarlikse reënval totale te karteer. Interpolasie metodes ondersoek tydens hierdie studie sluit in "Inverse distance weighted" (IDW), Spline en Kriging. 'n Statistiese analise van die kontinue oppervlak wat geskep is, dui min verskille tussen die ramingsvermoë van genoemde drie ruimtelike interpolasie metodes, maar Kriging het oor die algemeen beter gevaar as Spline of IDW.

Maandelikse gemiddeldes voorsien min inlgiting oor talle kenmerke van reënval relevant tot die wye verskeidenheid reënval-verwante aktiwiteite en besluitneming. 'n Stogastiese weergenererende (gebaseer op 'n eerste-orde Markov ketting) model is ontwikkel vir generasie van kunsmatige daaglikse reënvaldata vanaf maandelikse totale. Dit blyk uit intensiewe statistiese toetsing van die werkverrigting van die model dat genoemde model met vertroue gebruik kan word om 'n pseudo daaglikse reënval datastel vir die Hooglande van Eritrea te genereer.

'n Deterministiese model is ontwikkel om ondersoek in te stel oor hoe die globale reënval voorspellers verwant is aan die twee spits reënmaande (Julie-Augustus). Laasgenoemde twee maande dra tot 65 % van die totale reënval in die Hooglande van Eritrea by. Die hoofdoel met die ondersoek van hierdie verhoudings is om 'n eenvoudige statistiese model vir voorspelling van reënval hoeveelheid te ontwikkel. In 'n poging om die sterkste reënvalvoorspeller te identifiseer, is as voorlopige stap, 'n korrelasie matriks en stapsgewyse regressie van 11 voorspellers met verskillende sloertye geanaliseer. Die invloed van die suidelike Indiese Oseaan See Oppervlak Temperature is geidentifiseer as die sterkste voorspeller vir die Hooglande van Eritrea. 'n Model is gevalideer deur gebruik te maak van die "Jack-knive" vaardigheidstoets, Chi-kwadraat en afwykingsgebaseerde toetse. Die resultate was belowend.

ACKNOWLEDGEMENT

I would like to thank God Almighty for providing whenever there was a need.

I wish to express my sincere gratitude and deepest appreciation to:

- Dr J. C. Venter, my advisor, who introduced me to the fascinating world of climatic analysis and modelling. His stimulating lectures provided a foundation for much of this thesis.
- Prof. S. Walker, my co-supervisor, for her patient guidance, vision, critical comments, unfailing encouragement and for editing the manuscript.
- Dr. M. Tsubo, Department of Agrometeorology at the University of the Free State, for his incredible assistance and editing of the manuscript.
- Prof. A. De Waal and Dr. J. Van Zyl, Department of Mathematical Statistics at the University of the Free State, for their valuable comments.
- Prof. R. D. Van Buskirk (Lawrence Berkeley National Laboratory), Dr. J. Ian Stewart (President, WHARF), Dr. W. Landman (South African Weather Services) and Prof. M. Jury (Zululand University) for their provision of the rainfall and the Sea Surface Temperatures data.
- Dr. C. H. Barker, Department of Geography at the University of the Free State, for helping with the different techniques used in GIS.
- Mr. H. Ogindo (Ph.D. student in Agrometeorology at the University of the Free State), Mr. Weldemichael Abrha (Ministry of Agriculture Eritrea) and Mr. Johan Van Den Berg (Envirovision-South Africa) for their technical assistance.
- Government of Eritrea for providing the funds and University of Asmara for allowing me to come and study.
- My parents and friends who have given emotional support and assistance.

"Thus saith the LORD, let not the wise man glory in his wisdom, neither let the mighty man glory in his might, let not the rich man glory in his riches: but let him that glorieth glory in this, that he understandeth and knoweth me, that I am the LORD which exercise loving-kindness, judgement, and righteousness, in the earth: for in these things I delight, saith the LORD." - Jeremiah 9: 23-24 (KJV)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

μ	Mean of the precipitation (subscripts 1 and 2 representing mean of the smaller and larger exponential distribution) (chapter 2)
	Attenuation coefficient (chapter 4)
β	Scale parameters of the gamma distribution (chapter 5)
	Coefficient of multiple regression model (chapter 6)
α	Shape parameters of the gamma distribution
π	Unconditional probability of a wet day
$\Gamma(\alpha)$	The gamma function
σ^2	Year-to-year variance of monthly precipitation
ε	Weighting parameter for mixed exponential distribution
a	y-intercept (chapter 3)
b	Slope (chapter 3)
	Exponent divergence into k-dimensional space for inverse distance weighting (chapter 4)
c	Smoothing parameter for inverse distance weighting
CDF	Cumulative distribution function
D	The index of agreement (chapter 4 and 6)
	Vertical deviation between two curves (measured and generated) (chapter 5)
d	The persistence parameter for a first order Markov chain precipitation occurrence
ENSO	El Niño Southern Oscillation
gr	Generated daily rainfall (mm) (final)
I	Input variable (rainfall predictors)
i_2	The days of year on which the rainy season starts
i_3	The days of year on which the rainy season ends
IDW	Inverse distance weighting
m	Site rainfall dependent parameters
MAE	Mean absolute error
N	Number of days in the time series
n	Number of the paired set data (errors) (chapter 4)
	Julian day numbers (chapter 5)

	The total number of generated data (chapter 5)
ng	Number of years to be generated
o$	Name of rainfall station (e.g. Asmara1)
PDF	Probability density function
PDO	Pacific Decadal Oscillation
P_{dw}	Probability of wet day following a dry day
P_{ww}	Probability of wet day following wet day
r	Distance between the data point and interpolated point (chapter 4)
	Generated daily rainfall (mm) (provisional) (chapter 5)
R_i	Rainfall amount monthly total (subscripts j and j-4 representing the Current rainfall amount for July-August and total rainfall amount of the previous year for November-December).
RMSE	Root mean square error
RMSEs	Systematic root mean square error
RMSEu	Unsystematic root mean square error
RND	Random number generator function
S	Site dependent "scale" parameters
SIn	South Indian Ocean sea surface temperature
S_n	The nth partial sum
SOI	Southern Oscillation Index
SPI	the standardised precipitation index
SST	Sea surface temperature
u	Uniformly distributed random number (chapter 2)
	Ungrouped data-set (chapter 3)
w	Weight of the data point and interpolated point
x	Daily precipitation amount (chapter 2)
	Coded time for least square (chapter 3)
	Measured monthly or annual value (chapter 4)
xr	The first seed (random number)
y	Rainfall amount (chapter 3)
	Estimated monthly or annually value (chapter 4)
zi	Data point of mean annual rainfall amount
zj	Interpolated point of the weighted average

CHAPTER 1

INTRODUCTION

1.1 Rationale

One of the most severe and important constraints on Eritrean development is the availability of water, either in the form of direct rainfall, surface water flow or ground water resources. Water is a scarce and precious resource in Eritrea. If planning of water resources development is to be successful, improved estimates of probable rainfall and its characteristics and timing are crucial. The analysis of rainfall records can contribute much towards the understanding of various rainfall properties such as quantity, distribution and intensity, which together make up the overall rainfall pattern. Further, by having sufficient data on the correlation of rainfall with different rainfall predictors in different regions of the ocean or the atmosphere, it is possible to understand more about the rainfall pattern across the country. Acquiring sufficient accurate rainfall data is often the most critical stumbling block to studying the rainfall pattern. The historical records for the location of interest (the central and southern highland zones of Eritrea) are of insufficient duration. The lack of sufficient data strongly hampered investigations on the spatial evolution of climatic anomalies. To address the need for readily available rainfall data for these locations, a weather generator will be used. This weather generator is used to generate rainfall time series of synthetic daily rainfall data of any length for the site of interest.

Stochastic weather models ("weather generators") are statistical models that can be used as random generators of which the output resembles the weather data to which they have been trained. These models are useful in a number of applications where the observed climate record is inadequate with respect to length, completeness or spatial coverage.

1.2 Geography of Eritrea

Eritrea lies between latitudes 12°40' and 18°02' N and longitudes 36°30' and 43°20' E. The land surface area is 121,837 km^2 (FAO, 1994). The total annual rainfall tends to increase from north to south. It varies from less than 200 mm in the north-western lowlands to more than 700 mm in the south-western lowlands. A small area on the eastern escarpment known as the "green belt" receives on average over 900 mm per year. Most parts of Eritrea receive rainfall from the south-western Monsoon winds during the spring and summer months (April to October) (FAO, 1994).

Eritrea is one of those countries in Sub-Saharan Africa which has for years suffered natural hazards such as adverse weather conditions, serious soil depletion and drought which commonly constrain the productivity of African agriculture. It is an arid to semi-arid agricultural and agro-pastoralist country. It is strategically located along the Red Sea on the north-eastern part of the African continent, usually referred to as the Horn of Africa. Eritrea's neighbouring countries are Sudan on the West and North, Ethiopia on the South and Djibouti on the south-east (Figure 1.1). It has a coastal line extending 1000 km and some 112 islands in the Red Sea (SG-2000, 1997). Currently Eritrea has a population of about 3.5 million. Agriculture is the dominant sector of the economy and contributes around 50 percent of the gross domestic product (GDP).

Eritrea has three distinct physiographic zones: the Red Sea coastal plain, the central highland and the western plains. The Red Sea Coastal Plain comprises a narrow strip of land lying between the sea and the eastern escarpment of the central highlands in the north and the wider Denkel plain that lies between the sea and the Ethiopian border in the south-east. Both these areas are sandy and desert-like and are characterised by low hills interspersed with gently sloping land. The central highlands comprise the heavily eroded plateau that runs about 350 km south to north. Thus land higher than 1500 m constitutes a band 160 km wide in the south and only about 30 km wide in the north. The plateau slopes to the west, falling away in a series of broad steps characterised

by extensive elevated plains, which ultimately merge with the western plains. The western plains lying at an altitude of up to 600 m, represent the eastern extremity of the Sudan plains and their surfaces consist of alluvial sediments and aeolian sands (FAO, 1994).

Altitude is the major factor determining temperature. The climate ranges from hot arid adjacent to the Red Sea, to temperate sub-humid in isolated micro-catchments within the eastern escarpment of the highlands. It is mainly influenced by the arid and semi-arid climates of the Sahara desert in the north-west and the Arabian Desert across the Red Sea. About 72% of the territory is classified as very hot or hot (with mean annual temperature greater than or equal to 24°C) while not more than 14% is classified as mild or cool (with mean annual temperature below 21.5°C) (FAO, 1994).

The rainfall in Eritrea is characterised by two different mechanisms: tropical summer rains on the high plateau's and local cool-season rains on the hot coast of the Red Sea, overlapping in an isolated area of all year rains on the steep slopes of the plateau - mirrored on the eastern (Arabian) side of the rift (Flohn, 1987). The rainfall is mainly convective and thus it is localised and its spatial and temporal distribution as well as the intensity of rainfall is highly variable. The spatial variation of the rainfall is extreme, with regions receiving 1000 mm of annual rainfall as opposed to about 200 mm of annual rainfall separated by as little as 15 km (FAO, 1994). Like the rest of Sahelian Africa, Eritrea receives its rainfall from the south-west monsoon from April/May to September/October (FAO, 1994). "Short rains" fall during April/May and the "main rains" start in July, with the heaviest precipitation being during July and August. From November through March, the coastal plains and the southern part of the eastern escarpment of the central highlands receive winter rainfall, that is borne by northern and north-eastern continental air-streams that carry little moisture until crossing the Red Sea (FAO, 1994). The coastal areas receive occasional rainfall during summer due to migration of strong storms from the Yemen highgrounds (van Buskirk & Araia, 1994).

The prevailing wind over Eritrea is north-easterly in winter and south-westerly during summer. This reversal of wind is caused by the monsoon system (Mohammedberhan, 1998). In general, in the highlands low level easterlies change into westerlies with height. The zonal and meridional winds during a good rainy season are of the same order of magnitude as in the upper troposphere. During a bad rainy season zonal winds are stronger than meridional winds in the upper troposphere, particularly at 200 hPa level (NMSA, 1996).

1.3 Objectives

The specific objectives of this study are:

i) To perform both spatial and temporal analyses for long and short term rainfall data. (Chapters 3 and 4).

ii) To develop a stochastic model for generating pseudo daily rainfall data from measured monthly rainfall data (Chapter 5).

iii) To develop a statistical model for seasonal rainfall forecasts (Chapter 6).

1.4 Data

Raw rainfall data was gathered from various sources: Asmara Ministry of Agriculture, National Meteorological Service Asmara, Ethiopian National Meteorological Services Agency and personal contacts. There is only one station out of 22 (Figure 1.1) with a daily long-term data set. All stations lie between approximately 14°N-16°N and 38°E-40°E (Table 1.1). Sea Surface Temperatures data, Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO) and sunspot data were collected from different sources of information: Cpc (2002), South African Weather Service and personal contacts.

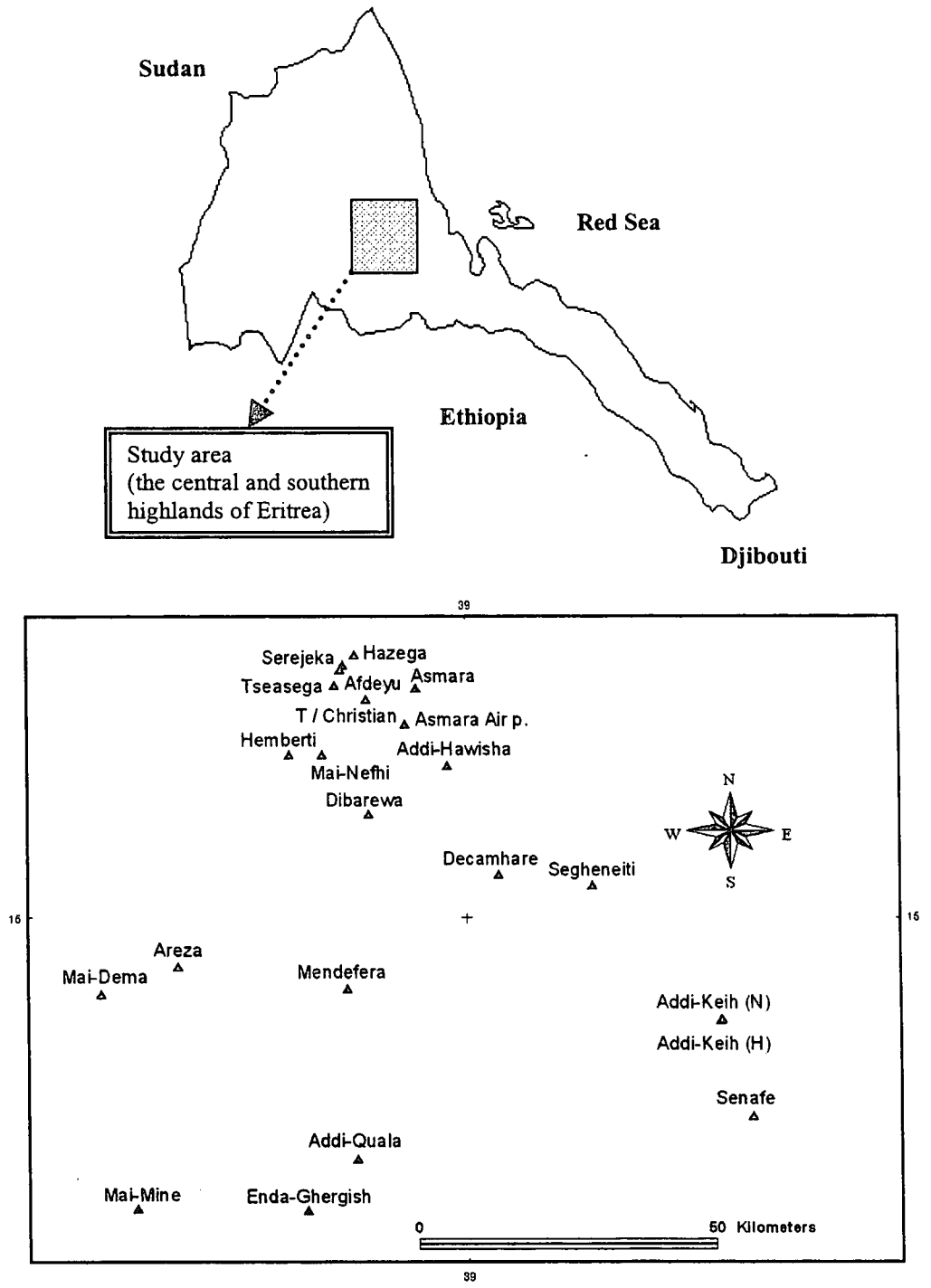


Figure 1.1 **Map of Eritrea. The box shows the site of interest that is the central and southern highlands of Eritrea. *Inset* shows the distribution of the stations throughout the central and southern highlands of Eritrea.**

Table 1.1 Geographical location of 22 Stations in the Central and Southern highland Zones of Eritrea.

Site	Altitude (m)	Latitude (N)	Longitude (E)	Zone	Data Type	Data Available	Years of Data
Addi Keih (H)	~2269	14°50'45"	39°23'00"	Southern	Decadal	1994-2001	8
Addi Keih (N)	~2280	14°50'45"	39°23'00"	Southern	Decadal	1992-2001	10
Addi Quala	~2010	14°38'00"	38°50'00"	Southern	Decadal	1992-2001	10
Addi-Hawisha	~2320	15°14'00"	38°58'20"	Central	Decadal	1992-2001	10
Afdeyu	~2440	15°10'00"	38°52'10"	Central	Decadal	1993-2001	9
Areza	~2000	14°55'30"	38°34'00"	Southern	Decadal	1992-2001	10
Asmara	~2150	15°21'00"	38°55'30"	Central	Monthly	1913-2000	88
Asmara					Daily	1943-1988	45
Asmara Air P.	~2160	15°17'45"	38°54'30"	Central	Decadal	1993-2001	9
Decamhare	~2060	15°04'00"	39°03'00"	Southern	Decadal	1992-2001	10
Dibarewa	~1880	15°15'15"	38°45'00"	Southern	Decadal	1992-2001	10
Enda-Ghergish	~1440	14°33'20"	38°45'30"	Southern	Decadal	1992-2001	10
Hazega	~2100	15°24'00"	38°50'00"	Central	Decadal	1994-2001	8
Hemberti	~2080	15°15'00"	38°44'00"	Central	Decadal	1992-2001	10
Mai-Dema	~1500	14°53'00"	38°27'10"	Southern	Decadal	1993-2001	9
Mai-Mine	~1800	14°33'30"	38°30'20"	Southern	Decadal	1994-2001	8
Mai-Nefhi	~2160	15°15'00"	38°47'00"	Central	Decadal	1992-2001	10
Mendefera	~2060	14°53'30"	38°49'10"	Southern	Monthly	1941-2001	60
Segheneiti	~2109	15°03'00"	39°11'30"	Southern	Decadal	1993-2001	9
Senafe	~2320	14°42'00"	39°25'45"	Southern	Decadal	1992-2001	10
Serejeka	~2300	15°09'30"	38°51'10"	Central	Decadal	1994-2001	8
T/Christian	~2100	15°20'00"	38°51'00"	Central	Decadal	1994-2001	8
Tseazega	~2160	15°21'15"	38°48'10"	Central	Decadal	1993-2001	10

CHAPTER 2

LITERATURE REVIEW

2.1 Stochastic Weather data Generator Models

One of the cornerstones of many types of modelling activity is weather data generation, where a statistical model is used to generate a time series of the climatic variables of interest (Jones & Thornton, 1993). Weather data generators are time-series models, usually for daily weather data. They can be viewed as a means to summarise climatic conditions in terms of a few parameters or, alternatively, as elaborate random number generators of which the output resembles real data (Hantel & Acs, 1998; Wilks, 2002). As Wilks and Wilby (1999) state, weather data generators are not weather forecasting algorithms, and thus are quite different from deterministic weather models. There are many reasons for using generated or simulated rainfall data. The historical record may be inadequate and need to be supplemented; or gaps in the historical records may need to be filled.

Weather data generators can produce time series of synthetic daily weather data of any length (Jones & Thornton, 1993; Wilks, 1998). They are also used to interpolate observed data to produce synthetic weather data at new sites. They have recently been employed in the construction of climate change scenarios (Wilks, 1999a). Any generator should be tested to ensure that the data that it produces is satisfactory for the purposes for which was designed (Semenov & Barrow, 1997). The accuracy required would depend on the desired application of the data and may vary considerably for different climates. Stern and Coe (1982) have illustrated the utility and importance of modelling the daily rainfall process for agricultural planning in east and west Africa. Stochastic weather data generators are used in a wide range of studies, such as hydrological applications (Heneker *et al.*, 2001), environmental management, analysis of water resources, water engineering design, design and operation of water

resources systems (Srikanthan & McMahon, 2001) and conservation planning (Hayhoe, 1998). They can also be used to quantify the daily, monthly and annual statistics of rainfall, the risk of storms and the probabilities of droughts of various duration and intensity (McNeill *et al.*, 1993). The synthetic daily weather series can serve as an input to crop simulation models (Riha *et al.*, 1996) and ecological models (Kittle *et al.*, 1995). Another common use of stochastic weather models is the estimation of missing meteorological data. A weather data generator can serve as a computationally inexpensive tool to produce multiple-year climate change scenarios at the daily time-scale, including changes in the mean and climate variability (Katz, 1996a; Semenov & Barrow, 1997; Wilks, 1999b). Hartkamp *et al.*, (2003) compared the performance of three weather data generators, MARKSIM, SIMMETEO and WGEN, with observed daily weather data for one of the major maize growing regions in Northwest Mexico. He recommended the weather generator SIMMETEO to be used in modelling applications at single point locations.

Generating daily rainfall data sequences uses a separate process for the rainfall occurrence and amount (Buishand, 1977, 1978). To use this technique a wet day should be suitably defined. When all days with positive rainfall amounts are called wet there may be some problems since fog or dew or light drizzle can also lead to small values being identified enormously. On the other hand, small rainfall amounts are also often registered as zero. To solve these problems wet and dry days should be well defined. A wet sequence is defined as a sequence of one or more days where each day has an amount equal to or above a certain threshold value. Similarly, a dry day has a value less than the threshold amount. The amount used can vary according to specific project requirements – values that have been used include for example 0.1mm (McCaskill, 1990a, 1990b), 0.25 mm (Katz, 1977; Zucchini *et al.*, 1992), 0.8 mm (Kottagoda & Horder, 1980). Thus, a daily rainfall model can be broadly described in two stages, the first reproducing the occurrence of wet days and the second reproducing the rainfall amounts (Cole & Sherrif, 1972). Based on the stochastic model for precipitation occurrence and intensity, the variance of the

monthly total precipitation is related to the characteristics of daily precipitation according to the following formula (Mearns *et al.*, 1997):

$$\sigma_I^2 \cong N\pi\alpha\beta^2[1+\alpha(1-\pi)\frac{(1+d)}{(1-d)}] \qquad (2.1)$$

where: σ^2_I = year-to-year variance of monthly precipitation;

N = number of days in the time series;

π = unconditional probability of a wet day $= \frac{P_{dw}}{(P_{wd}+P_{dw})}$;

α, β = shape, scale parameters of the gamma distribution;

d = the persistence parameter for a first order Markov chain precipitation occurrence , ($d=P_{ww}-P_{dw}$).

P_{ww} Probability of wet day following wet day

P_{dw} Probability of wet day following a dry day

The expression Nπαβ is the mean monthly total precipitation. Any model that can describe the daily precipitation occurrence process well should also be able to represent the distribution of dry and rainy sequences well (Chin, 1977).

2.1.1 Precipitation occurrence models

Analysis of daily precipitation time series always tends to reveal the existence of stochastic dependence. There is a tendency for rainy days and dry days to cluster and to form sequences respectively (Chin, 1977). This reality of meteorological persistence can best be described by a Markov chain model of proper order corresponding to the order of conditional dependence of the physical phenomena. Even though it is beyond the scope of this study, the alternative renewal process (spell-length model) proposed by Geen (1964) is an alternative to Markov chain models for simulating precipitation occurrence. Markov chain models with different levels of order are discussed below.

A Markov model provides a very convenient and efficient means of generating sequences of random numbers that resemble the corresponding real weather data. For each simulated day, a uniformly distributed random number **u** is drawn from the interval [0,1] (e.g., Brately *et al.*, 1987). Once the random number **u** has been generated, the sequence of wet or dry days is determined as

follows (Wilks & Wilby, 1999): If the previous day (t-1) was dry, then day t is simulated to be wet if $\mathbf{u} \leq P_{dw}$, and otherwise it is also dry. If the previous day was wet, then the current day is simulated to be wet if $\mathbf{u} \leq P_{ww}$, and is dry otherwise. Thus the two conditional probabilities are defined as:

P_{dw} = Probability {today is wet | yesterday was dry} (2.2a)

P_{ww} = Probability {today is wet | yesterday was wet} (2.2b)

Two of the most attractive features of Markov chain models are the ease with which seasonality is accommodated and availability of effective statistical inference procedures for parameter estimation and model selection (Stern & Coe, 1984). Many authors have used different orders of Markov chains to model the daily occurrence of precipitation.

2.1.1.1 The First-order Markov chain

Gabriel and Neumann (1962) proposed the use of a first-order Markov chain for precipitation occurrence at Tel Aviv, Israel. In using a first-order Markov chain one uses the approximation that the 'memory' of the process has a duration of 1 day. In other words one assumes that, for the purposes of predicting whether day t will be wet or dry, knowing the state on day t-1 is equivalent to knowing the state on all days preceding t (Zucchini & Adamson, 1984). This model can be fully defined by the two conditional probabilities (Katz, 1977; Richardson, 1981; Wilks, 1989, 1992, 1999a) which are called transition probabilities (Wilks, 2002). Since there are only two possible states on a given day, the two complementary transition probabilities are $P_{dd}=1-P_{dw}$ (dry day following a dry day) and $P_{wd}=1-P_{ww}$ (dry day following a wet day). Zucchini and Adamson (1984) demonstrated that a first order Markov chain provides an adequate description of the occurrence of wet and dry sequences of days in the complete range of Southern Africa conditions. A first-order Markov chain has been found to be an adequate model for precipitation occurrence in many different regions (Lowry & Guthrie, 1968; Cole & Sherrif, 1972; Smith & Schreiber, 1973; Chin, 1977; Woolhiser & Pegram, 1979; Phien & Warakittimalee, 1981; Richardson, 1981; Roldan & Woolhiser, 1982; Wilks, 1989, 1992, 1999a).

In spite of its reported applicability, there are also numerous instances in which the simple Markov chain is not able to depict the daily precipitation occurrences properly (Buishand, 1978; Racsko *et al.*, 1991). Different researchers address this deficiency by increasing the length of the simple Markov model's 'memory' (e.g., Dennett *et al.*, 1983; Singh & Kripalani, 1986; Jones & Thornton, 1997; Wilks, 1999a). But one has to be aware that increasing the order has to be done at the cost of increasing complexity and the number of parameters in the model. Coe and Stern (1982) stated that higher orders are not often necessary. If they fit reasonably well, low-order chains are preferable for two reasons:-

i) The number of parameters to be estimated is kept to a minimum, so that better estimates are obtained

ii) The subsequent use of the fitted model to calculate other quantities, such as the probabilities of long dry spells, is simpler.

Only when the dry spells are not adequately modelled by the first-order Markov model it is possible to improve the statistics of the simulated dry spell using 'hybrid-order' Markov models, in which the Markov 'memory' extends further back in time for the dry spell only (Stern & Coe, 1984; Wilks, 1999a). If a first-order, two-state Markov chain is to be fitted then the data matrix may be condensed into a **2 * 2 * t** matrix (Katz, 1974; Coe & Stern, 1982). The four entries for each day are the number of rain days following rain days, the number of rain days following dry days, the number of dry days following rain days and the number of dry days following dry days. If the full year is analysed then day 366 (31 December) for the previous year is used to give day 1. In non-leap years, when day 60 (29 February) does not exist, the state 1 of day 59 is used to calculate day 61 (1 March).

2.1.1.2 The Second-order Markov chain

A second-order Markov chain uses the wet/dry state on both the preceding day and the day before that, such that eight transition probabilities P_{ijk} must be defined. Here each of the indices **i**, **j**, **k** may be either wet or dry. Hence, P_{wdw} would be the probability of a wet day given that the previous day was dry and the day before that was wet. When a second-order Markov chain is fitted, four

separate functions have to be estimated, **2 * 2 * 2 * t** (Chin, 1977; Wilks & Wilby, 1999). These functions contain the number of days, at time **t**, in state **j** with the previous day in state **i** and the day before that in state **k**.

2.1.1.3 The Third-order Markov chain

Third-order Markov chains can be similarly defined, i.e. taking into account events occurring over the previous 3 days. However, the number of parameters required increases exponentially as the order increases. On the basis of a hypothesis testing procedure, Chin (1977) made a comparison among 1st, 2nd and 3rd order. The first- and second-order Markov chains failed to follow the fluctuations in the distribution of dry spells at Macon, Georgia, while the third-order model not only gave a realistic representation of these fluctuations, but also the best description of the precipitation occurrence process in the winter months of January and February. Jones and Thornton (2000) have generated daily weather data for Latin America and Africa, based on a stochastic weather generator that uses a third-order Markov process. A third-order Markov chain was tested for three sites in Central and South America and in the African Sahel, with positive significant results (Jones & Thornton, 1993).

2.1.1.4 Hidden State Markov (HSM)

The hidden State Markov model is introduced as an alternative model for generating hydroclimatic time series with long-term persistence (Thyer & Kuczera, 2000). These approaches are comparatively complex in both calibration and implementation, and have been only modestly successful in comprehensively simulating the spatial coherency of observed precipitation data (e.g., Hay *et al.*, 1991; Hughes & Guttorp, 1994 a, b; Wilks, 1999c).

2.1.2 Precipitation intensity models

The next necessary element of a rainfall generator is a model for precipitation amounts on wet days. By combining the first-order Markov model for daily precipitation occurrence with a statistical model for daily non-zero precipitation amounts, Todorovic and Woolhiser (1975) were the first to produce a daily stochastic precipitation generator.

Most stochastic weather generators make the assumption that precipitation amounts on wet days are independent, and follow the same distribution. Several models have been proposed for the distribution of precipitation amounts given the occurrence of a wet day. These include the exponential (Todorovic & Woolhiser, 1975; Richardson, 1981); gamma (Ison *et al.,* 1971; Stern & Coe, 1984); two-parameter gamma (Buishand, 1978); three-parameter mixed exponential (Woolhiser & Pegram, 1979); kappa (Mielke, 1973); and Weibull distributions (Zucchini & Adamson, 1984). Woolhiser and Roldan (1982) tested the exponential, gamma and mixed exponential distribution, finding that the mixed exponential distribution fitted the model of precipitation amounts best for arid regions. The three most commonly used models are explained below.

a) Gamma distribution

The gamma distribution has been the most popular choice (e.g., Ison *et al.,* 1971; Edelsten, 1976; Buishand, 1978; Stern & Coe, 1984; Wilks, 1989, 1992), and has the following probability density function:

$$f(x) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp\left[\frac{-x}{\beta}\right] \qquad (2.3)$$

where α is the shape parameter, β is the scale parameter, $\Gamma(\alpha)$ is the gamma function evaluated at α and x is the precipitation amount

b) Exponential distribution

The exponential distribution requires specification of only one parameter, μ (the mean precipitation), which is easier to handle analytically (e.g., Todorovic & Woolhiser, 1975). This is probably the simplest reasonable model for daily precipitation amounts. The exponential distribution is a special case of the gamma with $\alpha=1$. Therefore its probability density function is

$$f(x) = \frac{1}{\mu}\exp\left[\frac{-x}{\mu}\right] \qquad (2.4)$$

where μ is the mean of the precipitation

c) Mixed exponential distribution

Another natural generalisation of the exponential distribution is the mixed exponential distribution, which is simply a probability mixture of two one-parameter exponential distributions. The mixed exponential distribution has rarely been used. However, it has been reported to provide substantially better overall fits to daily precipitation data than the gamma distribution (Smith & Schreiber, 1974; Woolhiser & Roldan, 1982; Foufoula-Georgion & Lettenmaier, 1987; Wilks, 1999a, Cahill, 2003). Its probability density function is

$$f(x) = \frac{\varepsilon}{\mu_1}\exp\left[\frac{-x}{\mu_1}\right] + \frac{(1-\alpha)}{\mu_2}\exp\left[\frac{-x}{\mu_2}\right] \qquad (2.5)$$

where ε is a weighting parameter with values between 0 and 1, μ_1 and μ_2 are the means of the smaller and the larger exponential distribution respectively.

2.1.3 Weather data generators

A range of weather generators have been developed such as WGEN (Richardson, 1981); TAMSIM (Larsen & Pense, 1982; Bristow & Campbell, 1984); SIMMETEO (Geng *et al.*, 1986); CLIGEN (Nicks & Gander, 1990); ClimGen (Campbell, 1990; Stöckle et al., 2003); LARS-WG (Racsko *et al.*, 1991); GEM (Hanson, *et al.*, 1994); PRISM (Daly *et al.*, 1994); SWELTER (Evans, 1996); Zucchini and Adamson model (1984), which can be used to generate weather data sets. The most popular WGEN weather data generator model and the Zucchini and Adamson model for South Africa are discussed below.

a) WGEN (The Richardson model)

Richardson's (1981) stochastic weather data generator simulates daily time series of maximum and minimum temperature, incident solar radiation, and precipitation. In WGEN the precipitation is generated independently for a given day. Maximum and minimum temperature and solar radiation are then generated depending on whether the generated day is wet or dry (Soltani *et al.*, 2000). The model is expressed in a conditional form, which is especially convenient for computer algorithms (Katz, 1996b). A lot of research has focused on the performance and improvement of such models (e.g., Wallis & Griffiths,

1995, Johnson *et al.*, 1996). Mearns *et al.* (1996) and Riha *et al.* (1996) adapted this generator to produce climate change scenarios as input to crop-climate models. WGEN models have found application in sensitivity studies of crop model responses to changes in climatic variability that can be controlled through judicious adjustment of the generator parameters (Mearns *et al.*, 1996; Riha *et al.*, 1996). Daily precipitation occurrence is represented by a two-state first-order Markov chain model. This model is completely characterised by the transition probabilities and it accounts for the stochastic dependence of the series of wet and dry days (Katz, 1996a).

b) Zucchini and Adamson model

Zucchini and Adamson (1984) described a daily rainfall model for South Africa. The model has been fitted at some 2550 sites throughout southern Africa and captures the probability properties of the daily rainfall process at those sites. The model comprises two components: the first component describes the occurrence of wet and dry days, and the second component describes the distribution of rainfall amounts on wet days. The model uses a first-order Markov chain (Zucchini *et al.*, 1992). As was previously discussed, Markov models work by randomly sampling a series of events where the probability of observing an event depends on the occurrence of previous events, i.e., wet or dry. A wet day is defined as a day with 0.25 mm or more precipitation. Lourens (1995) generated a 100 years of daily rainfall for South Africa using the Zucchini and Adamson (1984) algorithm and parameters. He obtained a high coefficient of determination (R^2) between the generated data and the measured rainfall.

2.2 Rainfall Predictors

The East Africa region contains some of the most varied topography in the world, including large lakes, rift valleys and snow-capped mountains on the equator. This heterogeneity gives rise to dramatic variations in climatological mean rainfall totals (Sun *et al.*, 1999 a, b). Interannual variability of rainfall in East Africa results from complex interactions of forced and free atmospheric variations. These include interactions between sea surface temperature (SST)

forcing, large-scale atmospheric patterns, and synoptic-scale weather disturbance including monsoon and trade winds, persistent mesoscale circulations, tropical storms, subtropical anticyclones, easterly waves, jet streams, the continental low level trough and extratropical weather systems (Ogallo, 1988; Beltrando, 1990; Hastenrath *et al.*, 1993; Mutai & Ward, 2000).

The seasonal patterns of rainfall in East Africa follow the seasonal migration patterns of the Inter-Tropical Convergence Zone (ITCZ). The convergence zone is determined by the convergence of the inter-hemispherical monsoonal wind systems. The monsoonal wind system is the major source of moisture flux into the region (Ogallo, 1988). The characteristics of these winds are generally controlled by the intensity, location and orientation of the major semi-permanent anticyclones of Africa (Ogallo, 1988). The rainfall is also associated with a westerly influx of moisture from the Atlantic Ocean and the moist Congo/Zaire basins. There have been several recent studies examining connections between observed rainfall and a number of large-scale climate signals (e.g., Makarau & Jury, 1997; Montecinos *et al.*, 2000; Wang *et al.*, 2000). Studies have looked for prediction links based on correlation with raw station data (e.g., Nicholls, 1981; Montecinos *et al.*, 2000) or area averages (e.g., Makarau & Jury, 1997).

Several studies have demonstrated that ENSO events have global impact on the atmospheric circulation (Latif *et al.*, 1999). Such studies were done for Indonesia (Hackert & Hastenrath, 1986), Australia (Allan, 1988; Kane, 1997; Joubert, 1998), East Africa (Ogallo, 1988; Camberlin, 1997; Ward, 1998), and the tropics as a whole (Ropelewski & Halpert, 1987, 1996; Kiladis & Diaz, 1989). Studies on crop yield were done by Legler *et al.* (1999).

ENSO is a recurring pattern of climate variability in the eastern equatorial Pacific that is characterised by anomalies in both sea-surface temperature (referred to as El Niño and La Niña for warming and cooling periods, respectively) and sea surface pressure deviations (Southern Oscillation) (Prasad & Singh, 1996; Naylor *et al.*, 2001). Statistical evidence shows that ENSO can

account for about 50% of the interannual rainfall variance in Eastern and Southern Africa (Ogallo, 1994). A number of studies have confirmed a relationship between rainfall and ENSO in parts of eastern and southern Africa (e.g. Lindesay *et al.*, 1986; Farmer, 1988; Janowiak, 1988; van Heerden *et al.*, 1988; Nicholson & Kim, 1997). Dilley (2000) found that elevated Nino3 (area in the Pacific Ocean) values are associated with reduced Southern Africa rainfall.

The magnitude of detectable ENSO signals varies significantly, not only from one location or region to another, but also from year to year. Ogallo (1988) for example, noted that over some parts of eastern Africa, ENSO signals shifted from a positive to a negative phase from the northern summer to the autumn season. During warm ENSO events, both the ITCZ and the SPCZ (South Pacific Convergence Zone) shift equator-ward, and appear to become merged over their western portions near the dateline (Diaz & Markgraf, 1992).

Lindesay (1988) found that rainfall for a NW-SE band through central South Africa was associated most closely with ENSO. The interdecadal variation of the Indian monsoon rainfall (IMR) is strongly correlated with the interdecadal variations of various indices of ENSO (Krishnamurthy & Goswami, 2000). Kripalani and Kulkarni (1997) showed that there are more El Niño-related droughts in the decades when the Indian monsoon is generally below normal than during the decades when the Indian monsoon is generally above normal. A detailed account of ENSO and other related phenomena can be obtained from Troup (1965), Rasmusson and Carpenter (1982), Philander (1990), Glantz (2001) and many others.

2.2.1 Southern Oscillation Index (SOI)

The SOI is calculated using the barometric pressure difference between Tahiti and Darwin (la Mer, 1994). Depending on SOI values in the current and immediately preceding month, the Southern Oscillation Index is categorised into five phases (Stone, 1996). The five SOI phases are: *phase 1*: strong negative (El Niño); *phase 2*: strong positive (La Niña); *phase 3*: rapidly falling; *phase 4*: rapidly rising; and *phase 5*: neutral. The negative phase of the SOI represents

below-normal air pressure at Tahiti and above-normal air pressure at Darwin. Prolonged periods of negative SOI values coincide with abnormally warm ocean water across the eastern tropical Pacific typical of El Niño episodes. Prolonged periods of positive SOI values coincide with abnormally cold ocean water across the eastern tropical Pacific typical of La Niña episodes.

The use of the SOI phase system has been found to lead to improved profitability and risk management in Australian wheat cropping systems (Stone, 1996). Rainfall anomalies in various parts of tropical and subtropical Africa correlate significantly with indices of the Southern Oscillation (Stoecjenius, 1981; Makarau, 2000). Ogallo (1988), Lindesay (1988), Stone (1996) and Haylock and McBride (2001) provided correlation studies between rainfall and SOI for Kenya (East Africa), South Africa, Australia and Indonesia respectively, while Pittock (1980) and Suppiah (1989) provided correlation results for Argentina and Sri Lanka rainfall respectively. Beltrando and Camberlin (1993) have shown some important linkage between the SOI and rainfall in Ethiopia.

Correlation between SOI and Zimbabwean seasonal rainfall, showed that positive values of the SOI coincided with rainfall 101-125 per cent of the normal. Negative values occurred generally with rainfall below the normal (Makarau & Jury, 1997). Bhalme *et al.*, (1983) found significant correlation between the Indian monsoon rainfall and the SOI. Rasmusson and Carpenter (1983) found less rainfall over the Indian region during the extreme negative phase1 of the Southern Oscillation i.e. El Niño years.

The SOI signal is most apparent in upper zonal wind over the equatorial Atlantic SST and in the central Indian Ocean. These climatic expressions of the Southern Oscillation modulate rainfall over southern Africa and explain about half of the inter-annual variance (Jury *et al.*, 1994). It has been found that the persistence of a low SOI leads to very low rainfall in Botswana (Nicholson *et al.*, 2001)

2.2.2 Sea surface temperatures (SSTs)

Many studies have addressed the predictability of rainfall anomalies in different regions of the African continent both on interannual and interdecadal time-scales and their relationships to Sea Surface Temperature (SST) anomalies (Latif *et al.*, 1999). Folland *et al.* (1986) for instance, show a remarkable relationship between Sahelian rainfall anomalies and global scale SST anomalies on interdecadal time-scales, while tropical Atlantic and Pacific SST anomalies appear to be more relevant at interannual time-scales (Palmer, 1986). The SST changes resulting from thermocline perturbations, upwelling and ocean currents apparently couple with atmospheric circulation and convection (Jury *et al.*, 1999).

Lamb and Peppler (1987) have described an apparent relationship between Moroccan rainfall variability and the North Atlantic oscillation. Cane *et al.* (1994) showed that Zimbabwean rainfall and maize yield correlates strongly on eastern equatorial Pacific SST anomalies. Anomalous SST in the Atlantic Ocean has been associated with rainfall fluctuations in eastern Africa (Reverdin *et al.*, 1986; Nicholson & Entekhabi, 1987; Nicholson *et al.*, 2001). Central Atlantic SSTs were positively correlated with summer rainfall over central South Africa (Makarau & Jury, 1997). Significant correlations between Atlantic wind stress and the SOI, hence Pacific SST, have been found (Servain *et al.*, 1998). Very high correlation, in excess of 0.8, between the eastern African "long rains" and the stratospheric quasi-biennial oscillation index for time lags of about 7 months have been found by different researchers (Ogallo *et al.*, 1994; Jury *et al.*, 1995).

The role of Indian Ocean SSTs in climate variability has also been discussed in a number of papers. Nicholls (1989), for instance, describes a relationship between Indian Ocean SST anomalies and Australian winter rainfall. Lanaman and Mason (1999b) describes the association between Indian Ocean SSTs and summer rainfall over South Africa and Namibia. Makarau and Jury (1997) find a relationship between summer rainfall in Zimbabwe and Indian Ocean SSTs. Analysis of observed rainfall and SST anomalies revealed that only about 20%

of the rainfall variability in winter (DJF) over eastern equatorial Africa can be attributed directly to the SST anomalies in the eastern equatorial Pacific, as measured by the Niño3 index. The larger part of the variability in eastern equatorial African rainfall is related to Indian Ocean SST anomalies (Latif *et al.*, 1999). An atmospheric general circulation model suggests that the Indian Ocean Sea Surface temperature exerts a greater influence over East Africa short rains than the Pacific SSTs (Goddard & Graham, 1999), specifically the western Indian Ocean (Cadet & Diehl, 1984; Mutai *et al.*, 1998). Smith (1994) showed that there are strong correlations between Indian Ocean SST and Australian winter rainfall. Hastenrath *et al.* (1993) showed a particularly close association between near-surface zonal wind anomalies in the equatorial Indian Ocean and East African rainfall, an association stronger than that with ENSO. This link between the Indian Ocean SST and the climate over eastern Africa provides some hope for improved seasonal climate forecasting in this region.

There are a number of articles that have compared the results of using different techniques in seasonal forecast development for a particular place. Ward and Folland (1991), for example, have compared the use of multiple regression and linear discriminant analysis in forecasting north Nordeste Brazil wet season rainfall.

2.3 The Use of the Geographical Information Systems (GIS)

An increasing world-wide population, coupled with misuse of land resources, requires the application of new technologies like a Geography Information Systems (GIS) to help maintain a sustainable food and water supply without degrading the environment. A GIS is a computer system designed to collect, store, retrieve, manipulate and display spatial data (Franklin, 1992). It is a powerful tool for meaningful combination and presentation of information on areas of the earth e.g., Lourens (1995) used the GIS/modelling system for both delimitation of drought stricken areas and indication of the intensity of the drought.

In rainfall studies ArcGIS software is helpful in developing a map of a continuous rainfall data set by interpolation techniques from measured point data. Spatial interpolation is a process of intelligent guesswork, in which the GIS attempts to make a reasonable estimate of the value of a field at places where the field has not actually been measured. It is used in various operations such as: the preparation of a contour map display (contour lines are used to portray surface relief as a set of lines that connect points of the same value), in estimating rainfall at places where there are no weather stations, in estimating the elevation of the surface in between the measured locations. The principle that underlies all spatial interpolation is the Tobler Law - "all places are related but nearby places are more related than distant places" (Longley *et al.*, 2001).

Different interpolation methods to interpolate point data measurements into a surface regular grid are employed by different researchers, e.g., distance weighting techniques (Ripley, 1980), multi-quadratic surfaces (Adamson, 1978) and Thiessen polygons (De Jong *et al.*, 1992).

As was mentioned previously, rainfall studies can contribute much towards the economy of a country. In this study a first-order Markov chain will be used. As stated by Zucchini & Adamson (1984) in using the first-order one assumes that, for the purposes of predicting whether day t will be wet or dry, knowing the state on day t-1 is equivalent to knowing the state on all days preceding t. This assumption could also hold true in semi-arid areas like Eritrea where there are relatively few runs of three or more consecutive rain days.

By using the software ArcGIS, three different interpolation techniques will be evaluated for developing a continuous rainfall data map. Finally, a statistical model for seasonal rainfall forecast will be developed using various predictors. Chapter 3 will deal with the preliminary annual and monthly data analyses of the central and southern zones of the highlands of Eritrea.

CHAPTER 3

PRELIMINARY DATA ANALYSIS

3.1 Introduction

Extreme weather or climate events can have major impacts on society, the economy and the environment. Rainfall in Eritrea is characterised by a high degree of variability and it is the element of climate most influential in determining the variety and abundance of flora and fauna, land use, economic development and practically all aspects of human activity. Hence a comprehensive analysis of rainfall data is a crucial component in water management and in agricultural production. Rainfall has the property that it is partly discrete and partly continuous, therefore it has a multivariate time series which is non-stationary and contains variables with special properties (Brandão & Zucchini, 1990). From analysis of the measured historical data it is possible to get some insight into problems related to the amount of plant available water. Thus, the main objective of this chapter is to analyse the characteristics of the general trend of the annual and monthly rainfall totals as a preliminary study.

When analysing rainfall series of a considerable number of years one must be aware that the data collected over a long period may not reflect uniform conditions. This could be due to changes in instrumentation, sensor calibration, maintenance procedure changes, change of site, observation method, personnel and codes. Prior to performing different analyses the homogeneity of observed data with respect to non-climatic influences must therefore be assessed. Without proper adjustments, erroneous conclusions may be inevitable. Thus, the first analysis of this chapter is dedicated to homogeneity tests.

3.2 Tests For Homogeneity

A homogeneous climate time series is defined as one where variations are caused only by variation in weather and climate (Conrad & Pollak, 1950). Inhomogeneities can bias a time series, i.e., slippage of the mean, trend or some oscillation that may lead to misinterpretations of the climate being studied (Buishand, 1977). It is important, therefore, to remove the inhomogeneities or at least identify them and determine the possible error they may cause. Buishand (1977) indicated that man-made departures from homogeneity usually consist of sudden obvious jumps in a data series. The probability of success in detecting jumps in a given rainfall time series depends on the situation of neighbouring rainfall stations. The assumption was made that rainfall observations at nearby stations are similarly influenced by the same general climatic trends and discontinuities. Numerous studies have used procedures such as neighbouring station checks and statistical tests to identify and adjust for inhomogeneities in seasonal or annual mean total rainfall (e.g. Lavery *et al.*, 1997; Peterson *et al.*, 1998). A detailed account of the application of homogeneity tests to annual rainfall data in Hong Kong is presented by Jayawardena and Lau (1990) where the Von Neumann Ratio, Cumulative Deviation and Bayesian Statistics have been used as test statistics.

Most homogeneity testing techniques are primarily used in comparisons with neighbouring stations. It often happens, however, that the homogeneity of rainfall series of neighbouring stations is also doubtful or, as in this particular study, that there is no independent close neighbouring rainfall station which has long-term rainfall data for comparison purposes. Then use can only be made of a single rainfall series and many possible jumps will pass unnoticed. On the other hand, using only data from an individual station is problematical because the change (or lack of change) one detects may be masked by real changes in climate. Zurbenko *et al.* (1996) describes a filter that has been applied to single station data to date a discontinuity.

The cumulative sum techniques (Buishand, 1977) as a homogeneity testing of Zwanenburg-Hoofdorp, Netherlands, was used for the detection and

quantification of jumps to determine homogeneity in the data source used in this data set. A first indication of departures from homogeneity due to different changes can be obtained by plotting the partial sum of the departures from the mean (Figures 3.1). Partial sum can be defined as follows: Let $\{a_n\}$ be a rainfall sequence. The n^{th} partial sum, S_n, is the sum $S_n = a_1 + a_2 + \ldots + a_n$ that is for $n = 1, 2, 3, \ldots$ the expression $S_n = \sum_{n=1}^{n} a_n$ is called the n^{th} partial sum of the series.

The other problem besides the availability of a near neighbouring stations is the history of the station. No historical record of metadata was available for either Asmara or Mendefera stations. It is difficult to know what had happened to the stations during the earlier days. For example, What was the area of the rain gauge? What type of rain gauge was used? What was the height of the rim? Was the site of the rainfall station changed? etc. The partial sum will assist in assessing the departures, if any, from the homogeneity, rather than assuming homogeneity and thus blindly splitting the data into three equal parts to do the other analyses.

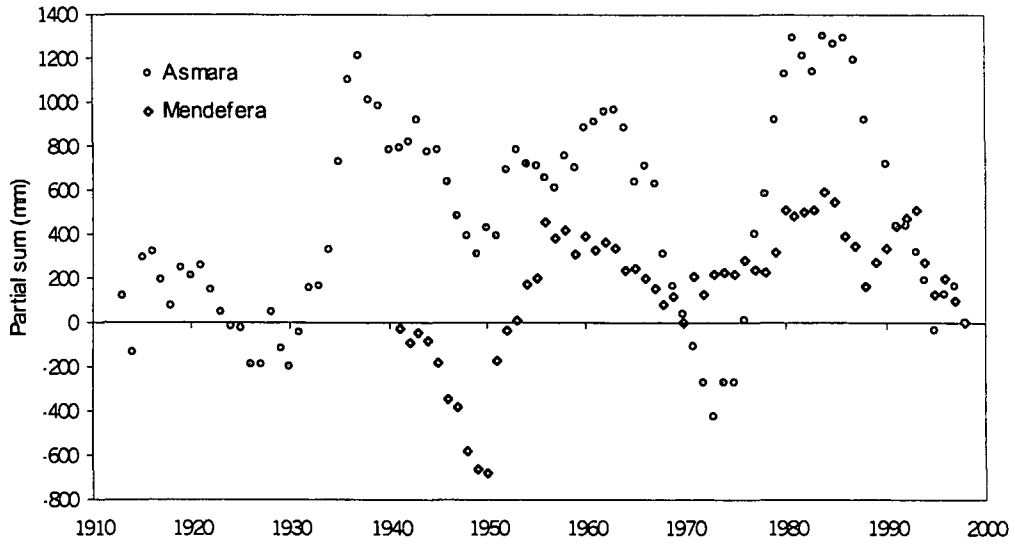


Figure 3.1 Partial sum of departures from the mean of annual totals for Asmara and Mendefera. (n = 86 and 58 respectively)

By using the partial sum patterns, the long term mean annual rainfall data for Asmara station representing the central zone was split into three sub-series, as follows: 1913-1949, 1950-1975 and 1976-2000. The same also applies to Mendefera station representing the southern zone with periods given by 1941-1956, 1957-1983 and 1984-2000. The main criterion for splitting into three sub-series is that each sub-series has to have a different pattern of rainfall. Estimates of the annual means, $\mu 1$, $\mu 2$, and $\mu 3$ for Asmara sub-series are 538.8, 497.4 and 537.7mm respectively, while for Mendefera sub-series the means are 553.1, 551.8 and 509.4mm. Homogeneity was tested under the assumption of equal variances and normality of the annual totals. The following Snedecor's F-statistics were done.

$Ho : \mu 1 = \mu 2 = \mu 3$

$Hi : \mu 1 \neq \mu 2 \neq \mu 3$

where: Ho = null hypothesis and Hi = alternative hypothesis

The result of the F-test for Asmara is 0.65 which is not significant at the 5% significance level (critical level is 0.43). The F-test for Mendefera was 0.49 which is also not significant at the 5% level (critical level is 0.72). This statistical analysis gives good evidence that each station is homogeneous.

3.3 Characteristics of Annual and Monthly Rainfall Totals

In terms of time resolution, rainfall data for Africa is most readily available in mean annual form, rather than as mean monthly values and least readily available as daily values. The mean annual total rainfall is of little use for most modelling applications. Mean monthly data have some use even though it is usually too coarse and daily rainfall records are often unavailable treasures. Chapter 5 deals with the generation of pseudo daily rainfall data from the measured monthly totals.

Figure 3.2 shows the monthly mean and standard deviation for some representative stations of the central and southern highland zones of Eritrea from observations over short periods (10 years). The monthly mean reaches its maximum in July or August and a minimum during the dry season from September to March. Stations that are further in the southern part are

characterised by high standard deviations for most of the rainy season. The monthly standard deviations are largest during the wet season and smallest during the dry season at all locations.

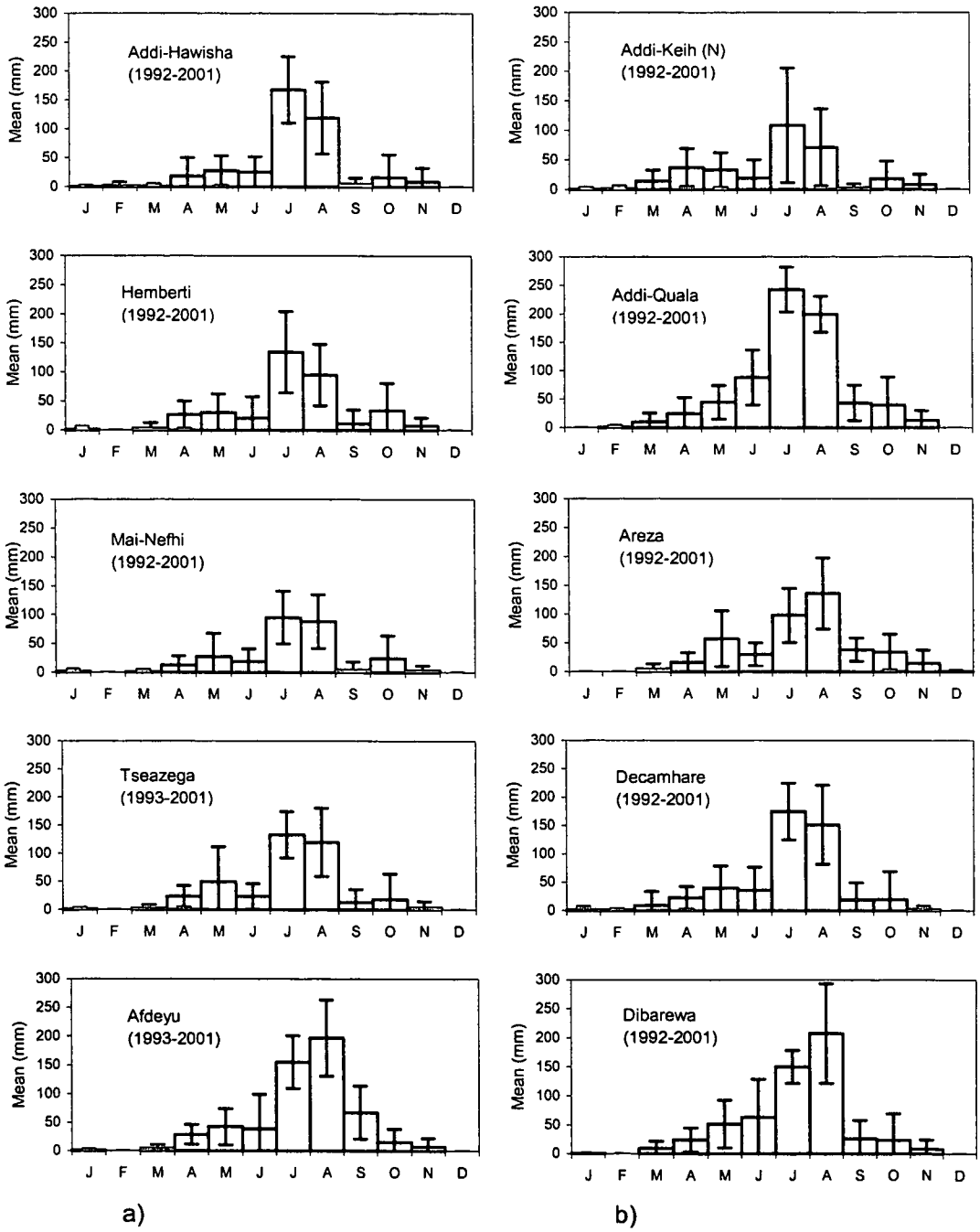


Figure 3.2 **Estimates of the means and standard deviations of monthly rainfall totals of some short-term rainfall series for (a) the Central zone and (b) the Southern zone stations.**

3.3.1 Time series analysis

A time series is nothing more than observed successive values of a variable or variables over regular intervals of time. It provides the basis in planning for future changes by identifying the nature of the phenomenon represented by the sequence of observations.

The review of historical data over time provides the decision-maker with a better understanding of what has happened in the past, and how estimates of future values may be obtained. Hoshmand (1998) states that the least squares method of fitting a straight line provides the best fit. The equations for fitting a least squares trend line are as follows:

$$a = \frac{\sum y}{n} \qquad (3.1a)$$

$$b = \frac{\sum xy}{\sum x^2} \qquad (3.1b)$$

where: a = y-intercept, b = slope, y = rainfall amount (mm), x = coded time.

Both equations are similar to those used in the simple regression model for estimating the value of a and b in practice. The value of x was substituted with a coded time scale, where $\sum x = 0$ (Hoshmand, 1998). The advantage of substitution by a coded time scale is avoiding dealing with the square of large numbers such as 1998, 1999, and so on. Furthermore, it simplifies the equation that is used in computing the intercept and the slope of the trend line. Simply subtracting the mean time from the value of each of the sample times results in a transformed time scale to the coded time. That is $x = 0$ in 1956 for Asmara and 1970 for Mendefera. (see Table 1.1).

Using the long-term annual rainfall for both representative stations of the highlands of Eritrea, it was shown that there has been a gradual decrease in annual rainfall for the periods 1913 to 2000 and 1941 to 2000 for Asmara and Mendefera respectively. Figure 3.3 shows that the least squares line has an annual trend decrease of 0.25mm and 0.58mm for Asmara and Mendefera respectively. But the statistical analysis of the deviation from zero is not

significant, that is, p = 0.28 and p = 0.78 for Asmara and Mendefera stations respectively. Thus it can be inferred that the rainfall distribution is fluctuating from time to time and the annual amount of rainfall has not significantly decreased over time during the last century.

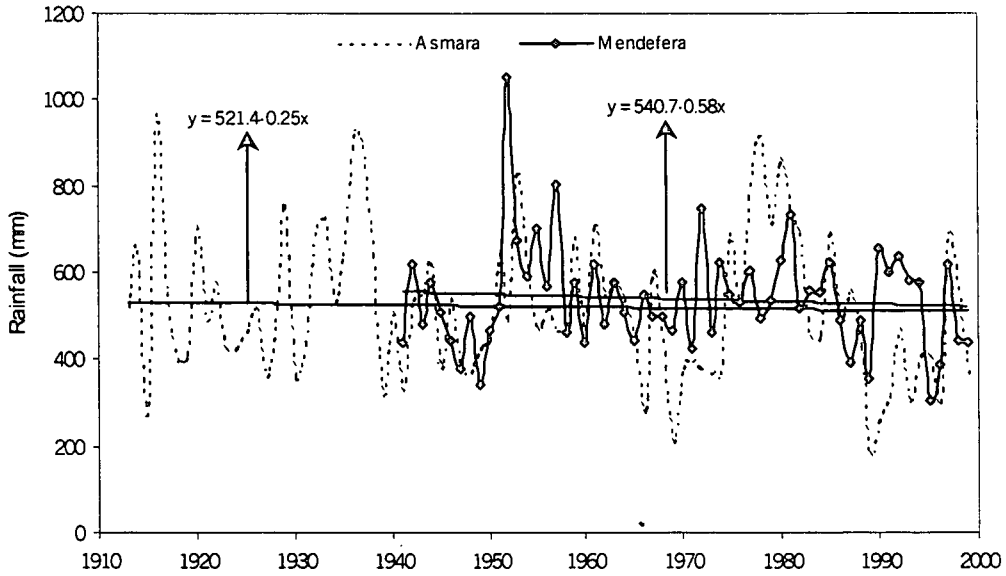


Figure 3.3 Trend line fitted by least squares method for the annual rainfall totals for Asmara and Mendefera.

Moving averages are calculated by taking a sequence of overlapping means over periods which are short compared to the total data period, and which work through the data chronologically. Trends and cycles in annual precipitation may be identified and isolated from the background year to year variations using the straightforward regression and moving averages respectively. The 5-year moving average amplitude (Figures 3.4 and 3.5) clearly show the year to year fluctuation to be much higher in Asmara than in Mendefera. It is also clear that the 5-year moving average floats above the actual rainfall during drier years.

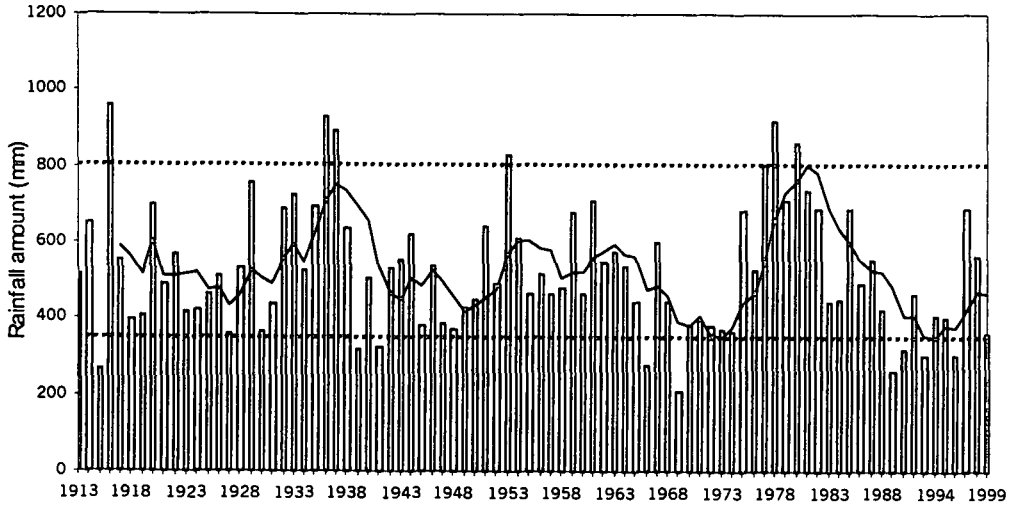


Figure 3.4 Rainfall for Asmara, with 5-year moving average shown by the solid line.

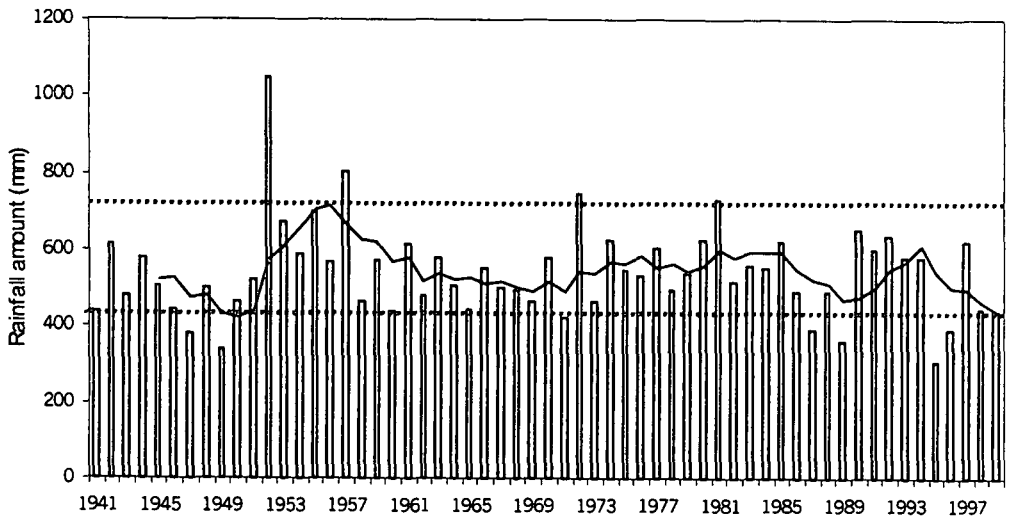


Figure 3.5 Rainfall for Mendefera with 5-year moving average shown by the solid line.

3.3.2 Percentiles

This is a useful method to determine, on the one hand, what the probabilities of exceeding specific limits are and, on the other hand, the position of a current event in terms of an historical dataset (White & O'Meagher, 1995). For a sample of n ungrouped data points which is arranged in ascending order, the k^{th} percentile is defined as the $k(n+1)/100^{th}$ value of the sample. The same rainfall amount can have different percentile values according to the climatic

characteristics of an area. For example, 50mm is above normal for the arid zone of the Kalahari but below normal for the humid zone of Eritrea. The percentile concept quantifies the values into categories for a specific site.

The 5-year return annual rainfall for Mendefera is 621mm that is the 80% point. Similarly the 20% point is 464mm. Therefore, one would expect more than 464mm in 4 years out of 5. Figure 3.6 shows percentiles for monthly rainfall.

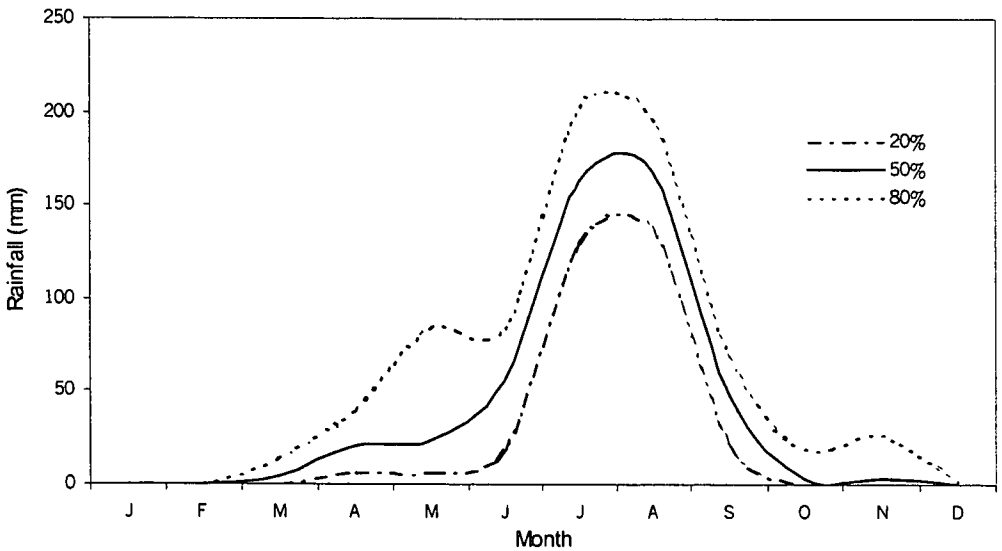


Figure 3.6 Percentiles 20, 50 and 80 for Mendefera.

These plots indicate that the distribution of the totals in each of these months is roughly symmetrical, with both months July and August having high values and May can be identified as short rains at the 80% level.

3.3.3 Using the Standardised Precipitation Index (SPI) to identify drought

The Standardised Precipitation Index (SPI) was designed by the Colorado Climate Center (McKee *et al.*, 1993) to quantify the precipitation deficit for multiple time scales. In technical terms, the SPI for a given historical precipitation record represents the number of standard deviations away from the mean for an equivalent normal distribution with a mean of zero and standard deviation of one (McKee *et al.*, 1993). Taking the difference between

the precipitation and the mean for a particular time scale, and then dividing by the standard deviation, gives the SPI. SPI is a dimensionless index where negative values indicate drought and positive values a wet condition (Heim, 2002). Hayes *et al.* (1999) stated that a drought event is defined as any time the SPI is continuously negative and reaches a magnitude where the SPI is –1.0 or lower, that is, less than the median precipitation. The drought event ends when the SPI again becomes positive. This indicates that SPI is helpful in monitoring the development and relief of a drought. Furthermore, SPI is capable of monitoring conditions important for both agricultural and hydrological applications. SPI was also used in the Great Plains in Nebraska for prediction of maize yield and risk probabilities (Hayes *et al.*, 2000). Hayes, *et al.*, 1999 suggests an SPI classification scale as given in Table 3.1.

Table 3.1 Classification scale for SPI values (from Hayes, *et al.*, 1999)

SPI values	Category
2.00 and above	Extremely wet
1.50 to 1.99	Very wet
1.00 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.00 to –1.49	Moderately dry
-1.50 to –1.99	Severely dry
-2.00 and less	Extremely dry

Based on Table 3.1, drought occurrence is analysed for Asmara and Mendefera stations of the study area. As illustrated in Figure 3.7, Asmara has experienced many years of moderately dry periods and the driest years on record are 1915, 1969 and 1990. In the same way for Mendefera, Figure 3.8 shows a near normal to moderate dry type with less variation, but there were no extremely dry conditions throughout this period. Likewise the more severe drought periods with the lowest SPI were recorded in 1949, 1989 and 1995.

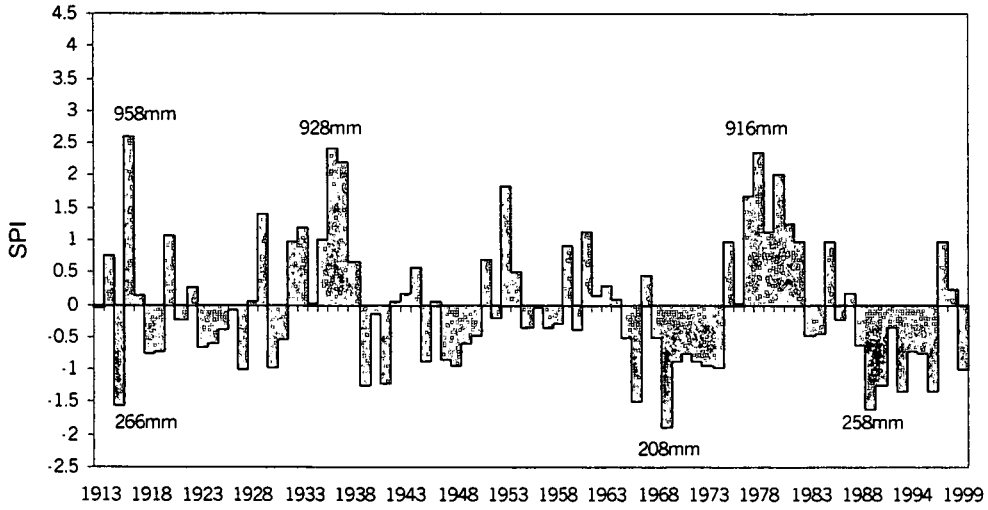


Figure 3.7 Annual standard precipitation indices for Asmara.

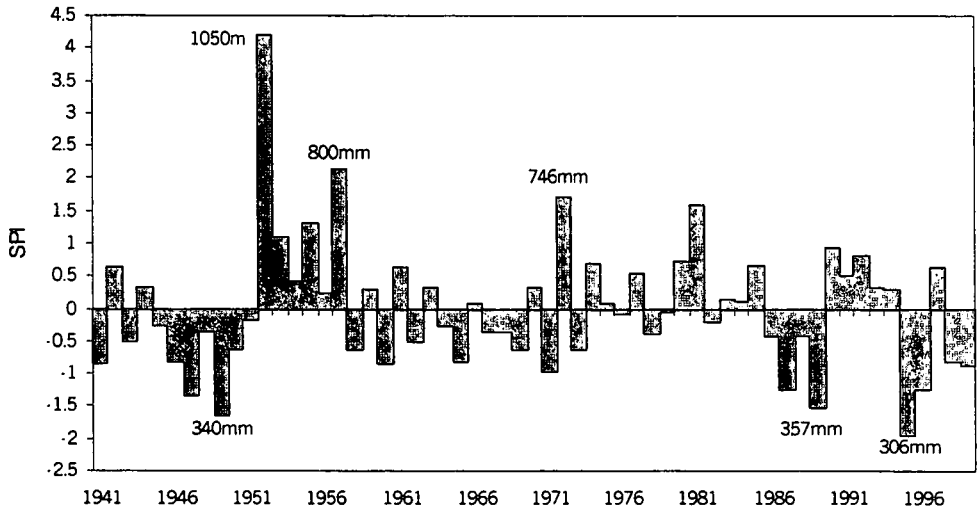


Figure 3.8 Annual standard precipitation indices for Mendefera.

3.4 Characteristics of Rainy Day, Spell Length and Water Balance

Determination of the number of rainfall days yielding specific amounts of rain, start of rainy season and the study of the events in general are some of the necessary steps towards an understanding of daily rainfall behaviour. The study of the wet and dry spells is necessary to complete the picture of the overall rainfall pattern of the highlands of Eritrea. It is of some importance in adapting a farming system to supplementary water resources to know how long a wet spell is likely to persist, how much rain the wet spell is likely to produce,

and what the probabilities are of experiencing dry spells of various duration at critical times during the growing season. The available data for the period 1943-1988 (Asmara station) was used for the analysis of rainy days, wet spells and dry spells in the central highland area. In this study, a rainy day is regarded as a day with measurable rain; i.e. a day yielding 0.8mm or more (Kottagoda & Horder, 1980). A rainless day has been defined as a day yielding less than 0.8 mm. This provides a cut-off to distinguish wet and dry days.

3.4.1 Start of the rainy season

A little rain could be risky by breaking a dry spell for some crops, thus awakening the vegetation from a state of dormancy induced by a prolonged dry period. A prolonged dry spell following the "false start" of the rains could affect seed germination and consequently lead to crop failure. Thus, a study of the start of the rainy season is very crucial for crop production.

Aspen and Douglas (2002) used a typed spatial view and defined the wet day as the day in the time series that had 75% or more stations reporting rain; dry days were those days with 35% or fewer stations reporting rain. Green (1966) used a different approach, whereby he defined the dry spell as a period of consecutive dry days broken by any day with measurable rain.

The start of the rainy season as defined by Raman (1974), which is adapted and modified by the software INSTAT+ (INSTAT+Climatic guide, 2001), has been used in this study. That is, the first occasion with more than 20mm within a 2day period after 1st June and with no dry spell of 10 days or more within the following 30 days. This definition adds the condition of a long dry spell immediately following the start of rain in the period, which is critical for crop production. The end of the season is defined as the first date the soil profile is empty after 1st September. Notice that (Figure 3.9) the earliest starting date of the rainy season is day 166 (June 14). The latest is day 218 (August 5). The mean starting date is late June (day 182) and the standard deviation is 12 days. For the end of season the earliest date is day 245 (September 1). And the latest

is day 272 (September 28). The mean end of season is day 254 (September 10) and the standard deviation is about eight days.

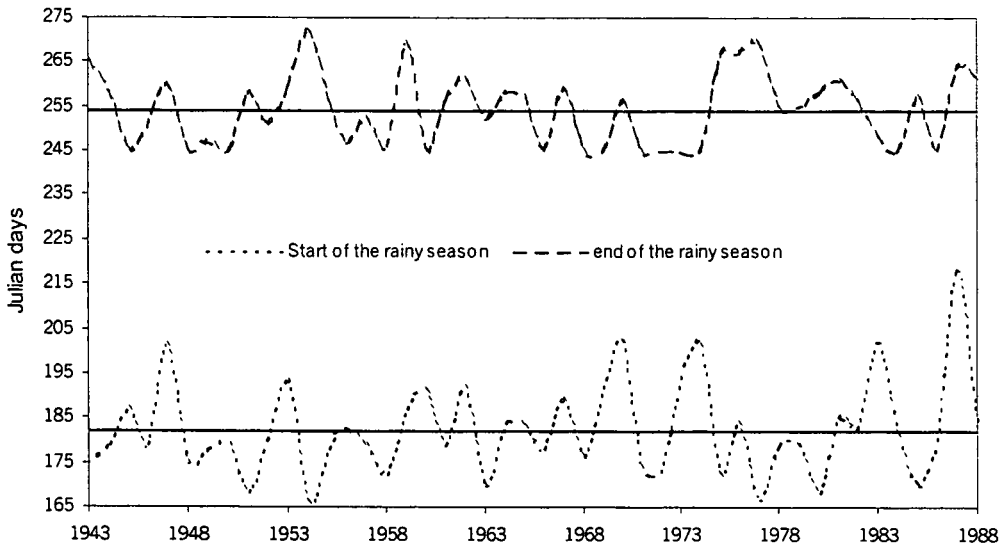


Figure 3.9 **Graph of the start of the rainy season (lower) of the first occasion with more than 20mm within a 2 day period after 1st June and end of the rainy season (upper) as the first date the soil profile is empty after 1st September for Asmara (1943-1988).**

3.4.2 Dry Spell length

In general, a wet spell may be defined as a run of rainy days, and a dry spell as a run of dry days. Thus, the definition of a wet day and a dry day are complementary. In many parts of the world the occurrence of long dry spells during the growing season of a crop is a major agricultural hazard (Stern & Coe, 1982; Shaw, 1987). But, it is also necessary to consider that a long dry spell is needed for harvesting some crops. The maximum dry spell lengths for June, July, August and September for 46 years for Asmara station was calculated (Figure 3.10). In comparison of the peak rainy months (JA) and the dry season (October through April) the number of dry spells fluctuates highly during the dry season.

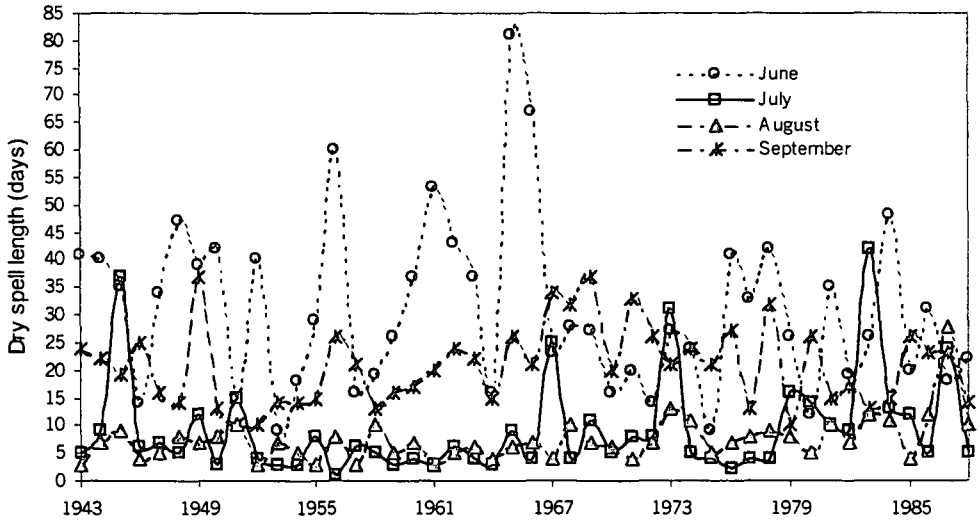


Figure 3.10 Maximum dry spell lengths for June, July, August and September for Asmara daily measured data (1943-1988).

3.4.3 Water Balance

Of vital importance for agriculture is the soil water balance, that is the balance between evapotranspiration, precipitation, drainage and runoff, and its variability within seasons, over the annual cycle, and from year to year. It is the water balance which determines plant available soil water, the frequency and severity of droughts or floods, and the availability of water for irrigation and natural river flows. The water balance is highly sensitive to small changes in climate with, in general, relatively large reductions in mean river flows and soil water as a result of high temperatures and hence evaporation.

A single value for evaporation is taken as 5.7mm per day, which was calculated by Stewart (1994), and the storage limit as 100mm for the central zone. The average water balance for the long-term record was tabulated and is given in Table 3.2. The results indicate that, without supplementing by irrigation for the dry period (5th June - 23rd June), planting on 6 May would not be recommended. In the earliest stage of crop growth the plant cover is too small to cover the soil, resulting in high soil surface evaporation. Planting on 23 June would be better because the profile is full by 8 July. If the end of the season is defined as the first date the soil profile is empty after 1 September, then Table 3.2 also shows the end day is 16 September. At the beginning of the season, the

average rainfall per day is less than the evaporation, so the resulting soil water remains near zero.

Table 3.2 **Average water balance (mm) for the long term record (1943-1988), for Asmara, (The code " -- " soil profile is empty and " ++ " when it is full)**

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	--	--	--	--	--	15	25	++	++	44	--	--
2	--	--	--	--	--	11	30	++	++	39	--	--
3	--	--	--	--	--	7	36	++	++	35	--	--
4	--	--	--	--	--	4	47	++	++	30	--	--
5	--	--	--	--	--	--	58	++	++	25	--	--
6	--	--	--	--	--	--	71	++	++	20	--	--
7	--	--	--	--	1	--	84	++	++	16	--	--
8	--	--	--	--	3	--	++	++	++	11	--	--
9	--	--	--	--	7	--	++	++	++	7	--	--
10	--	--	--	--	11	--	++	++	++	2	--	--
11	--	--	--	--	15	--	++	++	++	--	--	--
12	--	--	--	--	18	--	++	++	++	--	--	--
13	--	--	--	--	23	--	++	++	++	--	--	--
14	--	--	--	--	27	--	++	++	++	--	--	--
15	--	--	--	--	30	--	++	++	++	--	--	--
16	--	--	--	--	32	--	++	++	++	--	--	--
17	--	--	--	--	34	--	++	++	++	--	--	--
18	--	--	--	--	36	--	++	++	99	--	--	--
19	--	--	--	--	36	--	++	++	97	--	--	--
20	--	--	--	--	37	--	++	++	94	--	--	--
21	--	--	--	--	37	--	++	++	91	--	--	--
22	--	--	--	--	37	--	++	++	87	--	--	--
23	--	--	--	--	37	--	++	++	83	--	--	--
24	--	--	--	--	36	1	++	++	79	--	--	--
25	--	--	--	--	35	1	++	++	74	--	--	--
26	--	--	--	--	33	3	++	++	69	--	--	--
27	--	--	--	--	31	7	++	++	64	--	--	--
28	--	--	--	--	28	12	++	++	59	--	--	--
29	--	--	--	--	25	16	++	++	54	--	--	--
30	--		--	--	22	20	++	++	49	--	--	--
31	--		--		18		++	++		--		--

3.5 Fitting 1st Order Markov Chain Model for Daily Rainfall Data

It has already been indicated in the literature review that the 1st order Markov chain model has been found to be an adequate model for precipitation occurrence in many different regions. The Markov chain model was investigated for predicting the behaviour of dry spells for the Eritrea rainy season (June to September). In this fitting of Markov chains to the pattern of rainfall, the threshold level is 0.8 mm per day which defines all days with less than 0.8mm as dry. Figure 3.11 shows the curve fitted to probability data of rain. The curve was obtained from fitting two harmonics. The first-order Markov chain is given in Figure 3.12 for modelling of the probability of rain following a dry day (Pdw) and a rainy day (Pww). The software INSTAT+ (INSTAT+Climatic guide, 2001) was used for this analysis.

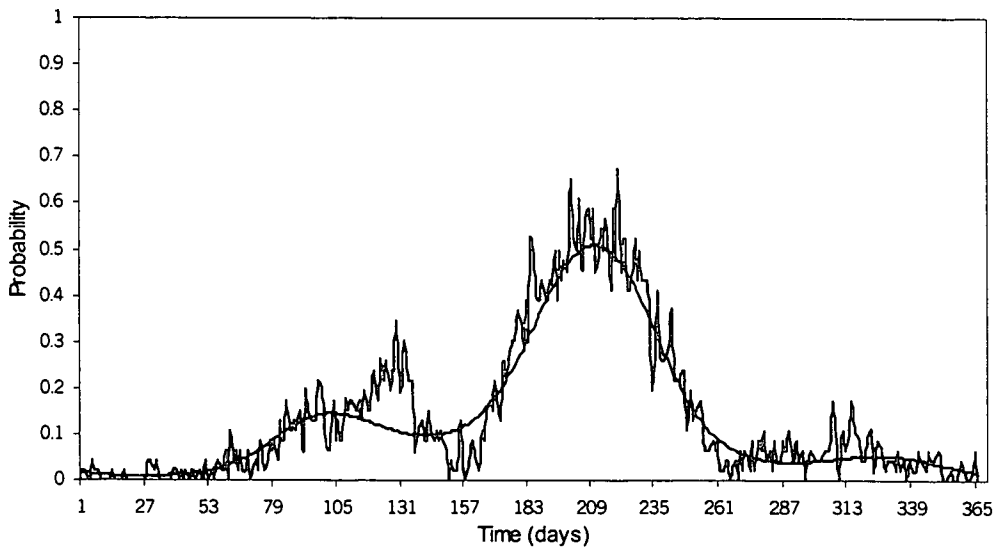


Figure 3.11 Curve fitted to rainfall probability data for Asmara (8th degree polynomial fit, $Y = a + bx + cx^2 + dx^3 + \ldots$) $R^2 = 0.84$.

The fitted model was used to estimate the risk of long dry spell lengths. The dry spell lengths for June and July were fitted and spell lengths of 8 and 10 days over periods of 30 days were calculated as shown in Figure 3.13. The results show that the probability of a dry spell of 8 days, within the 30 days following planting, has dropped to 0.5 by 15th June (day 167) and to 0.2 by 27th June (day 179). Figure 3.13 also tells that there is a 50 % probability of getting 10

days dry spell in day 154 but 8 days dry spell in day 167. This type of result can help determine risk involved in planting strategy.

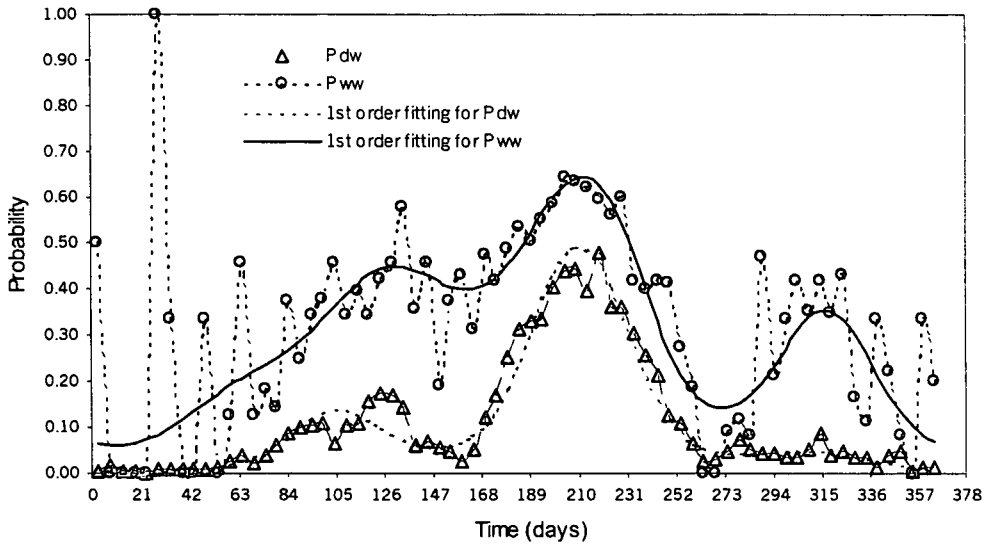


Figure 3.12 Fitting of the first order Markov chain for the probability of rain following a dry day (Pdw) and the probability of rain following a rainy day (Pww) for Asmara.

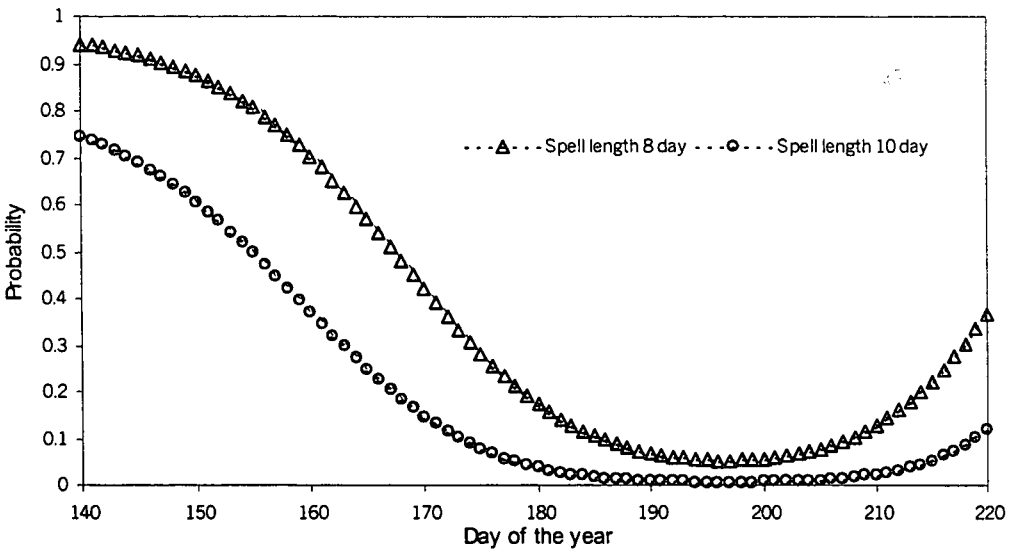


Figure 3.13 Modelling daily rainfall data with different spell lengths for Asmara.

3.6 Conclusion and Recommendations

With long series of observations it is common for the data collected to be inhomogeneous. This could be caused by changes of site, instrumental exposure or by changes in observation practices. Statistical tests were

conducted for both Asmara and Mendefera stations and revealed that both stations are homogeneous for the monthly rainfall dataset studied.

The least squares line of the time series analysis has shown the annual trend decrease of 0.25mm and 0.58mm for Asmara and Mendefera respectively. The rainfall distribution is fluctuating from time to time. It is incorrect, however, to think that the rainfall amount is decreasing as statistical analysis indicates that the deviation from zero slope is not significant.

The probability of wet and dry spells of different lengths were computed from historical sequences. Properties of rainy days, as well as of spells of wetness or dryness, are essential to the understanding of the complete rainfall pattern. The distribution of the daily rainfall and the wet and dry spells were investigated using a 1st order Markov model which was also fitted for estimation of the risk of long dry spells.

Strategic management is needed by analysing the time series for the prediction of a season to be a wet or dry season. During the dry season consideration should be given to the possibility of introducing drought resistant varieties, that is, crops more suited to the restricted water regime. Attention should also be given to the possibility of supplementary irrigation. The standard precipitation index (SPI) was used to identify the drought periods. In the central zone moderate drought occurs most frequently. In the southern zone the drought periods were mainly near normal to moderate dry type.

Analysis of rainfall data should not be based solely on the arithmetic mean. Means based on insufficiently long series are sensitive to distortion by runs of extreme values. Thus, chapter 5 of this document is dedicated to generation of artificial daily rainfall data. Spatial presentations of rainfall data may still be necessary to facilitate investigations designed to recognise and assess the various influences of rainfall factors on agricultural production. Therefore, spatial analysis for rainfall has been carried out in the next chapter.

CHAPTER 4

EVALUATION OF SPATIAL INTERPOLATION METHODS FOR ANNUAL RAINFALL

4.1 Introduction

The spatial estimation of monthly or annual rainfall from point measurements is an important prerequisite to the use of different areas of study like distributed modelling and agricultural land-use planning. It is therefore necessary to determine the best method of spatial interpolation. Rainfall, particularly at the daily time scale, typically displays complex spatial patterns. It is often related to topography, prevailing wind direction and usually displays a spatially varying dependence on elevation.

The most challenging task when creating a surface through interpolation, is selecting an appropriate technique. Different methods of interpolation will create a gridded surface, however, the result may not properly represent the way the data behave through space. Thus, the objective of this chapter is to find an appropriate method of spatial interpolation, in other words the best representation of reality for mean measured annual rainfall at a discrete set of locations (i.e., at points) in the highlands of Eritrea. Three spatial interpolation methods - Inverse Distance Weighting (IDW), Spline and Kriging were compared in constructing a continuous spatial map of rainfall. The errors from each of these techniques were evaluated using fundamental statistical parameters. Some work has been done with the errors of these interpolation methods. For example: Genton and Furrer (1998) used the root mean square error (RMSE) in predicting the performance of the highest and lowest values of rainfall measurements in Switzerland. Bancroft and Hobbs (1986), Siska and Maggio (1997), as well as Siska and Hung (2000, 2001), did some work with Kriging. Guan *et al.* (1999) used the determination coefficient (R^2) and RMSE to evaluate different interpolation methods for mapping submerged aquatic vegetation in

Florida, USA. RMSE is frequently used as an important parameter that indicates the accuracy of spatial analysis in GIS and remote sensing. RMSE is a measure of accuracy of prediction, defined as the mean square root of the sum of squared prediction errors.

4.2 Data and Methods

The data used in this assessment are the long and short-term annual rainfall amounts taken at twenty-two stations across the highlands of Eritrea. The altitude and the longitude of each station was converted into a decimal co-ordinate system and then changed to a raster format. The interpolated surfaces were created by estimating rainfall from 11 sampled points. The remaining 11 sites were used for the validation of the interpolated surfaces. MILA Grid Utilities in ArcGIS extension was used for extraction of interpolated point data. For display purposes all images are grouped into nine classes. The images show the average annual rainfall amount in the lightest colour (white). ArcGIS was employed as the software for this study. Evaluation of the interpolation methods follows the approach described by Willmott (1982) which calculates error statistics between measured and estimated values.

4.3 Spatial Interpolation

Spatial interpolation is a process of intelligent guesswork that estimates the values at locations where no measured values are available (Longley *et al.*, 2001). Spatial interpolation assumes two principles. First, the attribute data are continuous over space and this allows for the estimation of the attribute at any location within the data boundary. Second, the *Tobler law* (Longley *et al*, 2001) where the attribute is spatially dependent, indicating that the values closer together are more likely to be similar than the values farther apart. The quality of the interpolation results depends on the accuracy, number, and distribution of the known points used in the calculation and on how well the mathematical function models the phenomenon (Aronoff, 1989).

The analysis and interpretation of spatial data sets form an important part of geostatistics and is unfortunately highly subjective. For instance, it is well known that different individuals will take different approaches (Englund, 1990), yielding a large assortment of distinct solutions. This is partly due to the variety of available spatial interpolation methods, ranging from simple intuitive predictions to more sophisticated and complex procedures (Cressie, 1993). Myers (1994) clearly states the distinction between interpolating spatial data and contouring spatial data. Contouring might be best thought of as the algorithm that starts with data on a regular grid and produces the graphical contour plots, whereas interpolation (and smoothing) is used when the data are not available on a regular grid. Buys *et al.* (1979) used altitude contours as a guide in drawing maps of expectancy of monthly rainfall for southern Africa.

4.3.1 IDW ("Inverse Distance Weighting")

IDW employs the Tobler Law by estimating unknown measurements as weighted averages over the known measurements at neighbouring points, giving the greatest weight to the nearest point (Burrough & McDonnell, 1998). Thus, to predict a value for any unmeasured location, IDW uses the measured values surrounding the prediction location. Those measured values closest to the prediction location have more influence on the predicted value than those that are farther away (hence the name "inverse distance weighted"). The assumption is that each measured point has some local influence that diminishes with distance. Philip and Watson (1982) stated that IDW is not "ridge preserving", since the influence of input points on an interpolated value is distance related. The IDW method is computationally less intensive than Kriging (Guan *et al.*, 1999). Compared with other methods, most notably Kriging and Spline, the IDW method is simpler to program and does not require pre-modeling assumption in selecting a semi-variogram model. Seed (1992) examined a number of interpolation techniques and suggested that an IDW technique be used for interpolating daily rainfall. Yun (2003) used IDW for the spatial variation of the temperature for a rice simulation model in South Korea.

The IDW algorithm is a moving average interpolator that is usually applied to highly variable data, and the general equation for the IDW methods is (Burrough & McDonnell, 1998):

$$Z_j = \frac{\sum_{i=1}^{n} w_{ij} z_i}{\sum w_{ij}} \qquad (4.1)$$

where i= data point; j = interpolation point; n = total number of points; w_{ij} = weight; z_i = data point of mean annual rainfall amount. Z_j = the weighted average computed at j. The weight w_{ij} is:

$$w_{ij} = \frac{1}{(r_{ij} + c)^b} \exp(-\mu r_{ij}) \qquad (4.2)$$

where r_{ij} = distance between points i and j; μ = attenuation coefficient

b = exponent (divergence into k-dimensional space; b=k-1)

c = smoothing parameter

IDW has been used to analyse the annual rainfall values of the central and southern highlands of Eritrea as shown in Figure 4.1. The IDW parameters specified in ArcGIS are as follows: the power was set to the most commonly used value of two, that is the inverse distance squared weighted interpolation. The search radius was set to variable due to the irregularly spaced data sites.

IDW correctly shows a high rainfall area in the southern part of the study area, surrounded by a moderate type rainfall area where good agriculture is present. But, the disadvantage of the IDW interpolation technique is that it treats all sample points that fall within the search radius the same. Unlike Kriging the IDW creates small patches for surfaces which diminish with distance around the station used to create a smooth surface. Especially in the north-eastern part of the study area, this method of interpolation fails to give sufficient diversity.

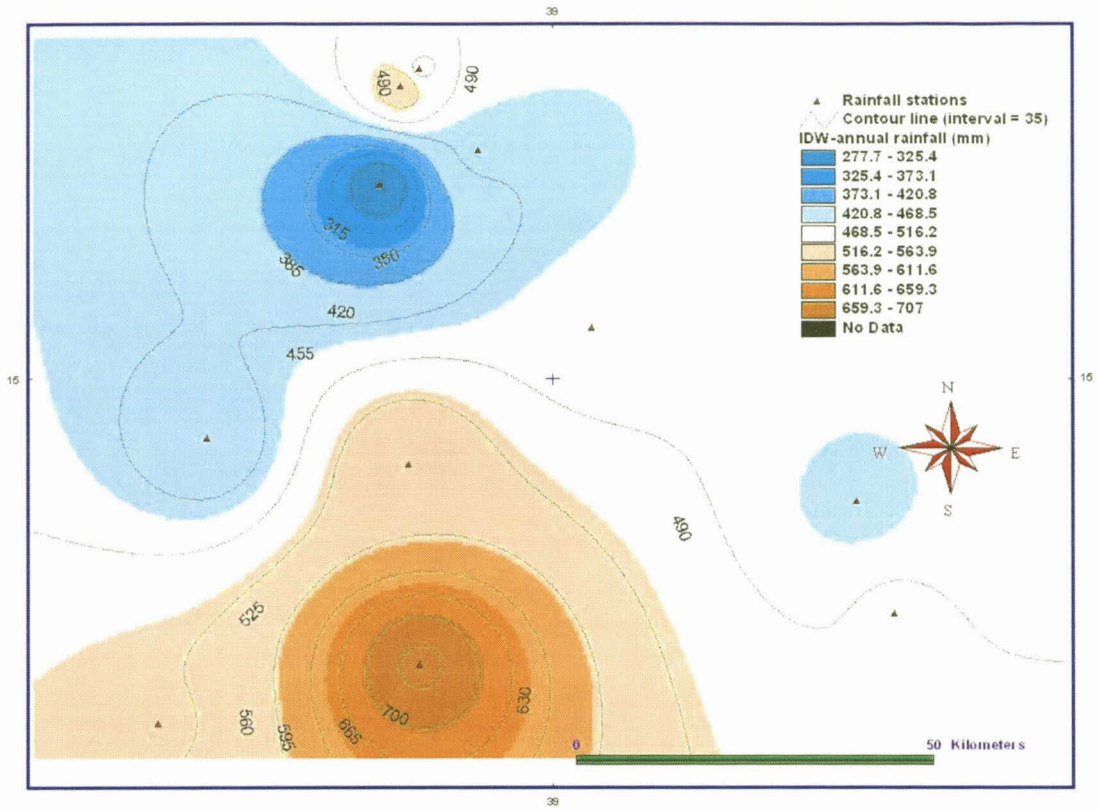


Figure 4.1 Interpolation results using IDW for the annual rainfall values of the central and southern highlands of Eritrea (11 number of stations).

4.3.2 Spline

A Spline is a polynomial curve or surface used to represent spatial variation smoothly. Spline interpolation fits a minimum-curvature surface, which passes through control points and has the least possible change in slope at all points. Conceptually, it is like bending a sheet of rubber to pass through the points, while minimising the total curvature of the surface results. Splines have the advantage of creating curves and contour lines which are visually appealing. A critical disadvantage of splines is perhaps that they provide a view of reality, which is unrealistically smooth and thus could generate misleading results (Burrough & McDonnell, 1998). Unlike IDW, spline is most suitable for those areas where there is little variation in the data within a short horizontal distance. In other words it is best for gently varying surfaces.

Despite their close formal connections, the actual formulations of smoothing Splines and Kriging are quite different. Splines are defined by minimising the

roughness of the interpolated surface, subject to having a prescribed residual from the data. Kriging surfaces are defined by minimising the variance of the error of estimation (Hutchinson & Gessler, 1994).

The parameters used in this interpolation method were tension type with the default weight of 0.1. Results are shown in Figure 4.2. Spline also correctly identifies the highest rainfall area in the southern part of the study area. But, it also incorrectly shows a high rainfall area towards the north-west where a large variability in rainfall occurs. It may be inappropriate to use the smoothing spline function for a highly heterogeneous area since it provides an unrealistic view of reality by overshooting estimated values or the underestimation by reducing variance such as in the north-eastern part of the study area. Splines tend to create values much greater than the highest known points and much lower than the lowest known points in the original data set for the study area.

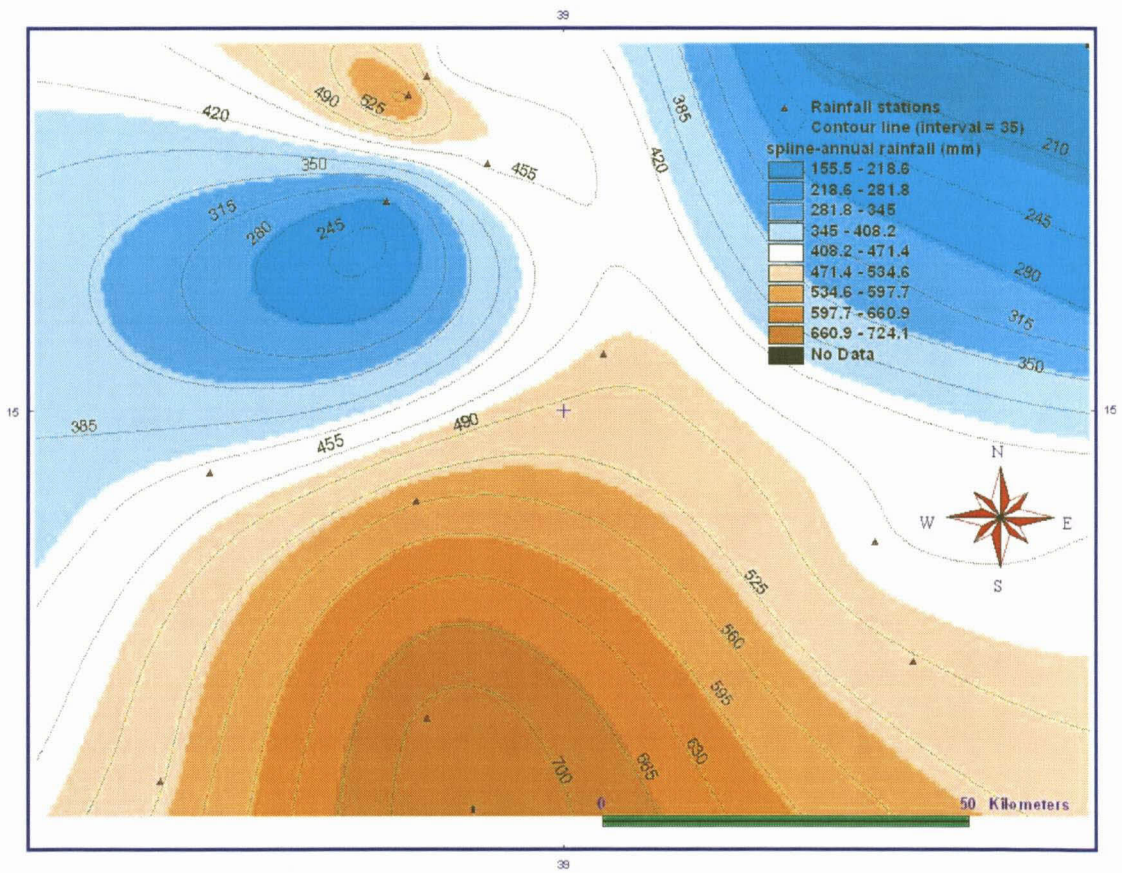


Figure 4.2 **Interpolation results by Spline method for the annual rainfall values of the central and southern highlands of Eritrea (using 11 stations).**

4.3.3 Kriging

Unlike IDW and spline which are deterministic models, Kriging is a stochastic model which is an advanced interpolation procedure that generates an estimated surface from a scattered set of points with z values (z value for this study is mean annual rainfall). Meneti *et al.* (1990) used ordinary Kriging to interpolate daily rainfall data points in Zambia. Kriging involves an interactive investigation of the spatial behaviour of the phenomenon represented by the z values before you select the best estimation method for generating the output surface. Kriging is based on the regionalized variable theory that assumes that the spatial variation in the phenomenon represented by the z values is statistically homogeneous throughout the surface; that is, the same pattern of variation can be observed at all locations on the surface. The spatial variation is quantified by the semi-variogram. Estimation and fitting of the semi-variogram, as well as variogram model selection, are crucial stages of spatial prediction, because the variogram determines the kriging weights of the predictor. Kriging uses the mathematical function specified with the method argument to fit a line or curve to the semi-variance data in the semi-variogram. Some advantages of this method are the incorporation of variable interdependence and the available error surface output. A disadvantage is that it requires more modelling time and more input.

The effectiveness of Kriging depends on the correct specification of several parameters that describe the semi-variogram. Thus, universal (U_2) Kriging with a linear drift semi-variogram model was selected to be used in this study. Results are shown in Figure 4.3. Kriging gave a reasonable explanation of the spatial distribution of rainfall in the study area. Unlike IDW, Kriging can use different weighting functions depending on the distance and orientation of sample points and the manner in which sample points are clustered.

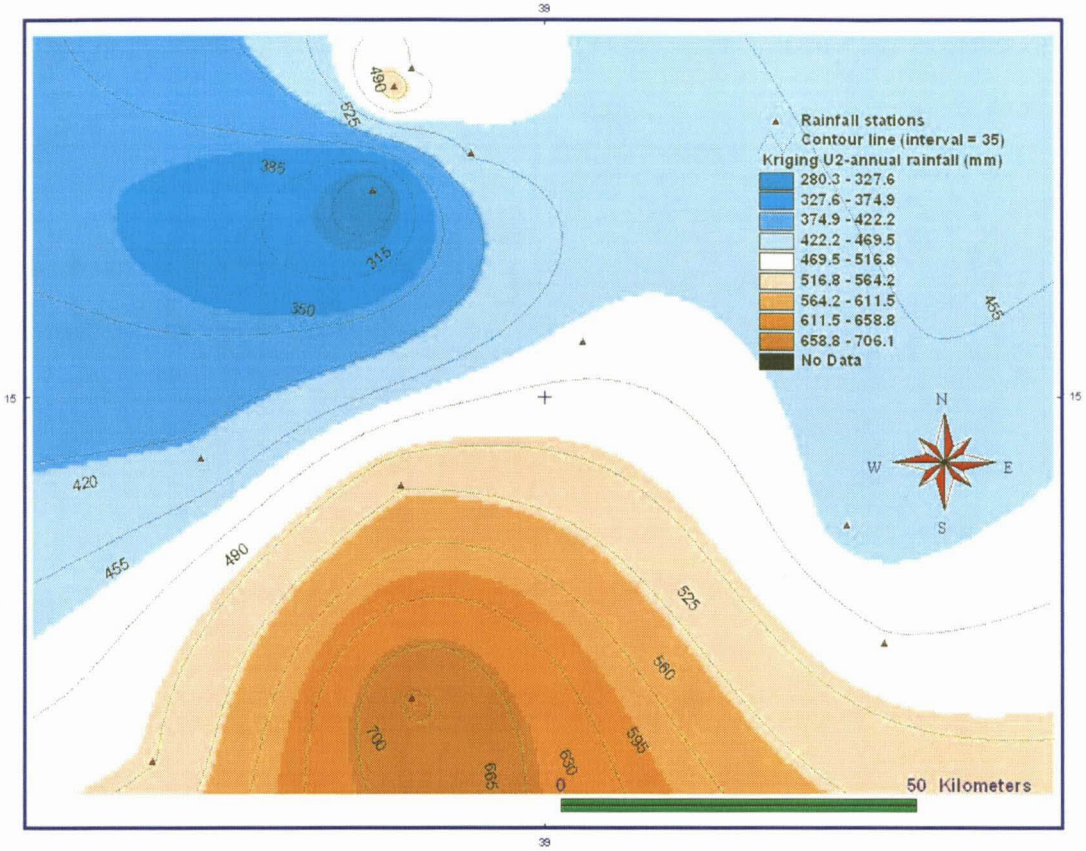


Figure 4.3 Interpolation results by Kriging for the annual rainfall values of the central and southern highlands of Eritrea (for 11 stations).

4.4 Statistical evaluation

For evaluation of the interpolation methods, correlation-based statistics were used with the deviation-based statistics (Willmott, 1981; 1982) using the 11 validation stations. The correlation-based analysis includes the coefficient of determination (R^2). The deviation-based analysis includes root mean square error (RMSE) and the index of agreement (D). Lower RMSE and higher D-index indicate an interpolator that is likely to give reliable estimates for the areas where the mean annual rainfall amount is not known. Mean absolute error (MAE) is sometimes preferred over RMSE as an evaluator because it is less sensitive to extreme values; however, RMSE is the error measure commonly computed for geographic applications. The MAE, RMSE and D-index are calculated using the following formulas:-

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (4.3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (4.4)$$

$$D = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(|y_i'| + |x_i'|)^2}\right] \quad (4.5)$$

where x_i and y_i are the measured and estimated values respectively, n is the number of the paired set data (errors), $x_i' = x_i - \bar{x}$ and $y_i' = y_i - \bar{x}$, and $\bar{x}$ is the measured mean (Willmott, 1981, 1982).

Systematic root mean square error (RMSEs, mm) and unsystematic root mean square error (RMSEu, mm) are also used for evaluating the performance of the interpolation methods (Willmott, 1982). Lourens (1995) states that the smaller the RMSEs is and the closer the RMSEu approaches the RMSE, the better the fit.

The systematic RMSEs can assess whether the model errors are predictable, whereas the unsystematic RMSEu identifies those errors that are not mathematically predictable. The D-index varies between 0.0 and 1.0, where the closer D is to 1.0 the better the agreement between the measured and estimated mean rainfall amount. These results are displayed in Tables 4.1 and 4.2.

From Table 4.1 one can see that the mean and standard error of the measured and estimated values are closest with the Kriging method. This gives some indication about the model potential for creating a continuous mean annual rainfall data set.

Table 4. 1 Summary univariate statistics of three interpolation methods (IDW, spline & Kriging) for annual rainfall totals (mm) for the Central and Southern zones, n = 11.

Methods	Summary univariate				
	Mean observed	Mean predicted	Standard error observed	Standard error predicted	Mean bias error
IDW	467.75	479.22	79.34	74.75	11.47
spline	467.75	462.81	79.34	108.69	-4.94
Kriging	467.75	464.93	79.34	83.97	-2.82

The average error, which is the difference of mean observed and predicted, is largest in IDW data (11.47) and smallest for Kriging data (2.83). The difference in the standard error suggests the IDW method is most similar, followed by the Kriging, with the Spline method resulting in the largest differences between the standard error observed and predicted. These suggest that the Spline method is slightly less able to reproduce the observed variance.

Table 4. 2 Difference measures between three interpolation methods (IDW, spline & Kriging) for annual rainfall totals (mm) for the Central and Southern zones, n = 11.

Methods	Difference measure				
	RMSE	MAE	RMSEs	RMSEu	D
Spline	98.49	69.46	31.18	93.43	0.68
IDW	71.93	52.18	39.28	60.25	0.75
Kriging	66.78	46.55	25.03	61.61	0.80

The RMSE between measured and estimated rainfall amounts of all 11 locations is calculated as 66.78mm for the Kriging method, which is a lower value than the IDW (71.93mm) and Spline (98.49mm) methods. The MAE and RMSEs measures have a similar response in that the Kriging method has the lowest value. The RMSEu for both IDW and Kriging is almost the same. The final error difference, D-index also indicates that Kriging (D=0.80) is better than both Spline (D=0.68) and IDW (D=0.75) for the mean annual rainfall amount. From this statistical analysis Kriging was identified as the best method for interpolating surfaces, followed by IDW and then Spline. Nevertheless, as illustrated in Figure 4.4, none of the interpolation methods gave a reasonable estimation of the measured mean annual rainfall amounts across the whole range of rainfall values from 350-650mm. This could be because the distance between the rainfall stations is really too small and the variation in the annual

rainfall amount with an 11 point data set is really a too small adequately to assess the accuracy of a method. Generally the results of the three interpolation methods for the western part of the study area are better than the eastern part. This could be due to the lack of observational stations in the eastern part, which is the mountainous area, and the spatial variability of mean annual rainfall could be very high.

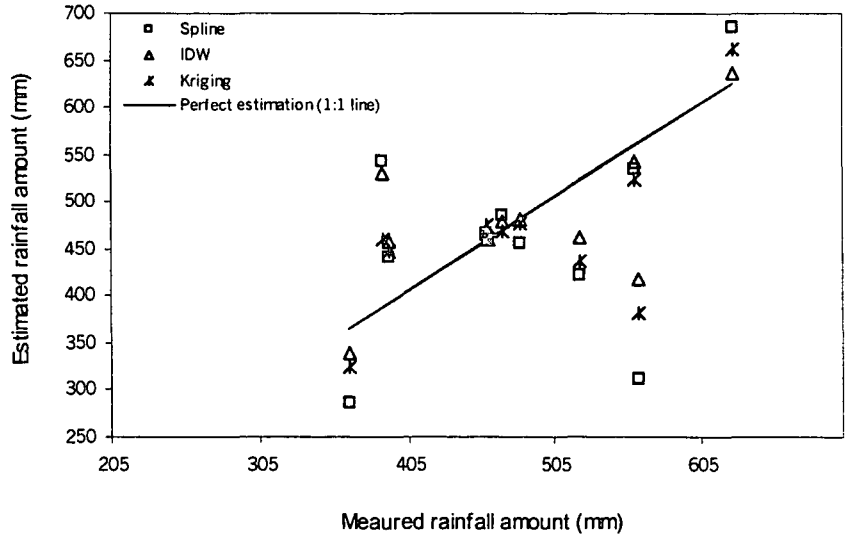


Figure 4.4 **Scatter plot of the measured annual rainfall amounts (mm) versus estimated amount of total annual rainfall by the IDW, Spline and Kriging methods for the central and southern zones of the highlands of Eritrea.**

4.5 Conclusions and Recommendations

The choice of an appropriate model is essential in order to obtain reasonable results. In this chapter the performance of three methods was evaluated based on the magnitude and distribution of errors. This study has shown that Kriging is most likely to produce the best estimation of a continuous surface of annual rainfall amount in the central and southern highlands of Eritrea. Kriging is recommended from this analysis to GIS practitioners as an efficient procedure to interpolate rainfall data and generate a map for practical applications. Nevertheless, these methods of interpolation do not adequately address the rainfall variability inherent in the central and southern highlands. Thus, further analysis is needed like Jackknife corrected estimates (Tomczak, 1998) to improve the performance of the Kriging interpolator by reducing the high value of RMSE (RMSE = 66.78mm).

CHAPTER 5

GENERATING PSEUDO DAILY RAINFALL DATA FROM MEASURED MONTHLY RAINFALL DATA

5.1 Introduction

Effective water resources management is essential in a country like Eritrea, which is particularly prone to the adverse effects of drought as well as heavy rains which cause floods and soil erosion. The representation of the rainfall climate by an average monthly or annual total (which are the only available and readily accessible data) is not sufficiently accurate for good water resources management when investigating problems concerned with the agricultural potential, hydrology or other fields of interest. Artificial rainfall sequences are useful as inputs to crop growth models (Riha *et al.*, 1996) which can then be used to determine the distribution of yield, the risk of crop failure due to adverse climate, optimal planting dates, the potential profitability of irrigation and so on. For such purposes artificial rainfall sequences generated by a good stochastic model are more useful than the original short historical record.

Monthly means provide little information on many properties of the rainfall that is relevant to the wide variety of rainfall-related activities. For example, the severity and duration of a drought and timing of rainfall within each month or year are all aspects of rainfall that are of extreme importance to agricultural decision making. The average rainfall figure can mask a lot of information. Tyson (1987) pointed out that the concept of mean monthly and annual rainfall must be treated with caution, as the rainfall pattern can be highly variable.

Developing a reliable method to generate artificial daily climate sequences will empower water resource and agricultural planners to assess the probable consequences of decisions of which the outcome depends on climatic factors. Thus, the objective of this chapter is to generate pseudo daily rainfall data from

available monthly measured data by developing a stochastic weather generator model for the site of interest.

5.2 Algorithms

The algorithm of the stochastic model for weather generator is given below. The basic idea of the model was taken from Zucchini and Adamson (1984) model for southern Africa. This model was coded by Dr. J. C. Venter (Promoter), whose assistance is gratefully acknowledged. The program was written in the BASIC language. The variables used in the program are described in Table 5.1. The model-coded program is given in appendix A.

The threshold level of the rainfall amount was taken as 0.1mm, being the smallest amount of rainfall usually recorded. This provides a convenient cut-off to distinguish wet and dry days. However, the model user can choose any amount of rain as a threshold. For example, 0.8mm was used in Chapter 3 for spell lengths and risk analysis.

The main assumption of the model is, when using monthly input data, that the probability of having a wet (or dry) day exhibits the same type of behaviour as the distribution of median monthly rainfall which is smoothed by cubic Spline interpolation (Figures 5.1 and 5.2). The Spline interpolation was done for the median monthly rainfall from the measured monthly historical data set. The precipitation depths on consecutive days are not independently distributed. It was assumed that the probability that a wet day will follow a wet day is higher than the probability that a wet day will follow a dry day. Consequently, the conditional distribution of precipitation depth on a given day depends on the state of precipitation on the previous day.

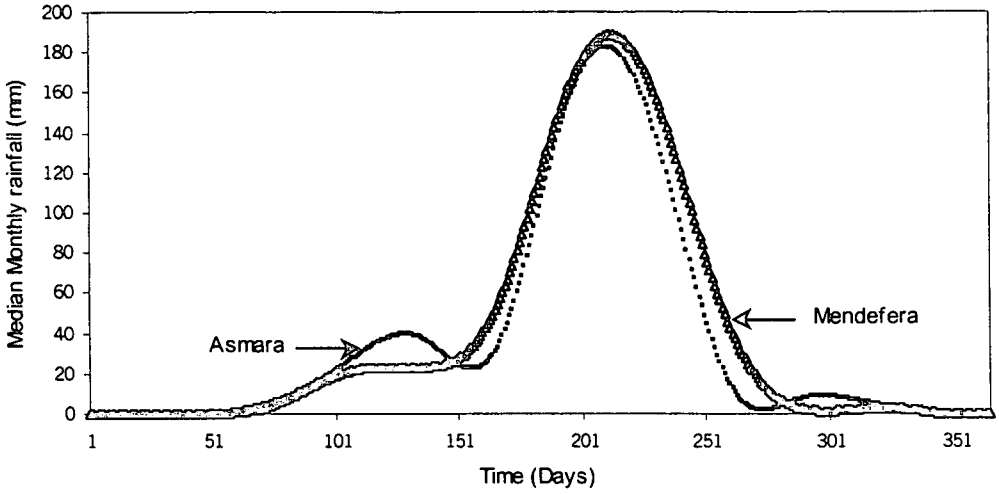


Figure 5.1 **Spline interpolation of median monthly rainfall for Asmara and Mendefera. The 24 data points were positioned at the middle of each month.**

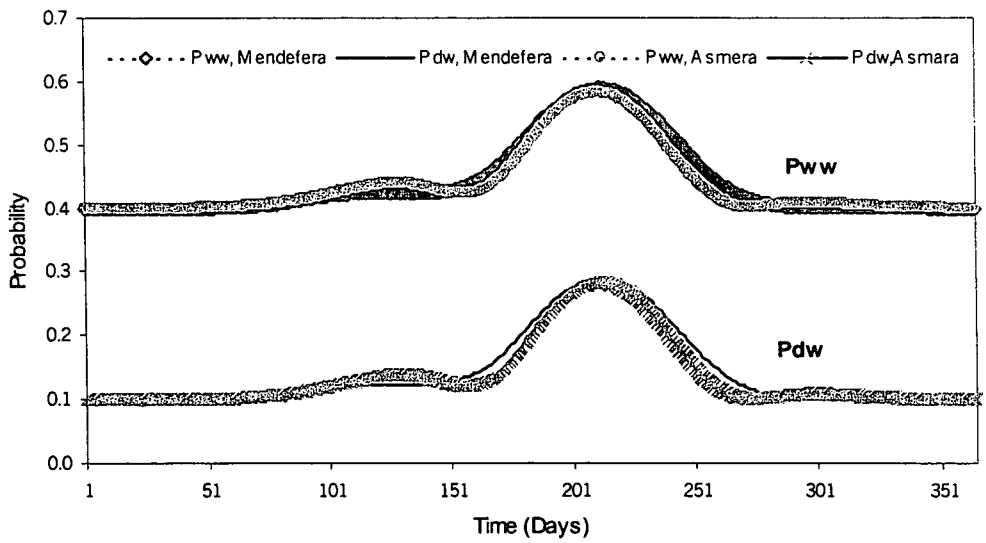


Figure 5.2 **The probability of a wet day given the previous day is wet (Pww, top curve) and the probability of a wet day given the previous day is dry (Pdw, bottom curve) for Asmara and Mendefera.**

The variables used in the program (Appendix A) are as follows:-

Table 5.1 The variable description used in the stochastic model.

Variable	Description
Pww (i)	The probability that day T is wet given that it is wet on day T-1, T = 1,2,...,365
Pdw (i)	The probability that day T is wet given that it is dry on day T-1, T = 1,2,...,365
r(i)	Generated daily rainfall (mm) for each day (i) of year (provisional)
gr(i)	Generated daily rainfall (mm) (final)
S_1-S_6	Site dependent "scale" parameters ≥ 0.
i_2; i_3	The days of year on which the rainy season starts and ends
m(i) >=0	Site rainfall dependent parameters, i = 1, 2, ..., 365
o$	Name of rainfall station (e.g. Asmara1)
n(i)	Julian day numbers
xr	The first seed (random number)
RND	Random number generator function
ng	Number of years to be generated

For discussion and easy understanding, the algorithm is divided into five main steps:

Step 1

Input both the n(i) and m(i) from the data files Asmara1 (central zone) or Mendef1 (southern zone). Cubic splines were used for interpolating the monthly median rainfall to obtain the site rainfall dependent parameters (m(i)).

Step2

Identify the start and the end of the rainy season (i_2, i_3). See Chapter 3 for a detailed explanation.

Step3

a) Input the site dependent "scale" parameters (S_1, S_2, S_3, S_4, S_5, and S_6) for both the dry and wet season. The scale parameters of the central and southern parts of the study area are given in (Table 5.2). Much effort has been expended in developing the S-values by trial and error. Pww and Pdw should always be in interval [0 , 1].

Table 5.2 S-parameters for the central and southern zones for wet and dry season.

Location	Season	S_1	S_2	S_3	S_4	S_5	S_6
Asmara	Wet	0.3	0.3	0.001	0.1	0.001	0.4
	Dry	0.009	0.0	0.001	0.1	0.001	0.4
Mendefera	Wet	0.2	0.6	0.001	0.1	0.001	0.4
	Dry	0.1	0.2	0.001	0.1	0.001	0.4

b) The variables $r(i) = S_1 * m(i) + S_2$ are also calculated in this step.

c) As was mentioned previously the critical assumption is that the probabilities (Pww, Pdw) follow the same kind of variation as the monthly rainfall. In particular, they depend linearly on the smoothed historical records of Figure 5.1.

$$Pdw(i) = S_3 * m(i) + S_4 \quad (5.1a)$$

$$Pww(i) = S_5 * m(i) + S_6 \quad (5.1b)$$

$(i = 1, 2, \ldots, 365)$

Step 4

Check that probabilities lie in the interval [0 , 1]. This check is only necessary during model calibration.

Step 5

a) Since the state (wet or dry) of the previous day has to be known, the model assumes that December 31 preceding the record to be generated, was dry.

b) To get different random numbers the model was programmed to loop over 195 seeds. This means that the number of seeds in file "seed" must not be less than 195.

c) By generating only with random number generation one can get a rectangular shape with the probabilities between one and zero, that is

$$y = RND\ (xr) \qquad (5.2)$$

To get a realistic representation of the rainfall pattern the negative logarithm, -log (y), is used as follows:

$$gr(i) = r(i) * (-\log (y)) \qquad (5.3)$$

$r(i) = S_1 * m(i) + S_2$ were calculated in step 3.

After generating a record of daily rainfall (which could stretch over many centuries depending on computer speed and memory storage), various statistical packages can be used to organise and analyse the generated data. In this study a record of 900 years of daily data were generated. Comparisons are made for different number of generated years (49, 86, 100, 500 and 900 years) and centuries (1st, 2nd, 3rd, 4th and 5th centuries) (Appendix B: B-i to B-iv). The resulting standard deviations for the generated precipitation values are lower than the standard deviation calculated from the measured data set. This is due to the fact that the model does not generate extreme values at the same frequency with which they actually occur in the observed dataset. The mean can be similar but one expects the generated standard deviation to be lower.
A detailed statistical analysis of the model performance is discussed below.

5.3 Evaluation of the Stochastic Model

Two sites were selected to evaluate the performance of the model. The choice of the sites was, of course, constrained by the availability of suitable historical records.

The aim of model evaluation is to establish that the modified model preserves the important properties of the historical records, at least to an appropriate degree, so that the generated sequences can be regarded as representative of the population of sequences which could occur. The statistics used were the cumulative distribution function, Kolmogorov-Smirnov test, Gamma distribution function, Chi-square test and Student's t-test.

5.3.1 Cumulative Distribution Function (CDF)

The CDF is a convenient method requiring little computation which provides reasonably good estimates of probabilities. The CDF was determined by ranking the generated data in ascending order and calculating their associated cumulative probability of non-exceedence as:

$$\frac{j}{(n+1)} * 100 \qquad (5.4)$$

where j = rank position; n = total number of generated data

The division by n+1 gives a better estimate of the population especially at the extremities of the distribution. The cumulative distribution function of total monthly rainfall for each month of the 100 years of generated data was analysed. For a particular station for instance, one hundred Januaries of generated data were analysed to determine the CDF for January, and so on. Twelve CDF's (Appendix C: C-i & C-ii) were therefore obtained for each station. Visual inspection of the graphs (Figures 5.3 & 5.4) gives an idea that the model appears to be a good representation of the measured values of the peak rainy months (JA).

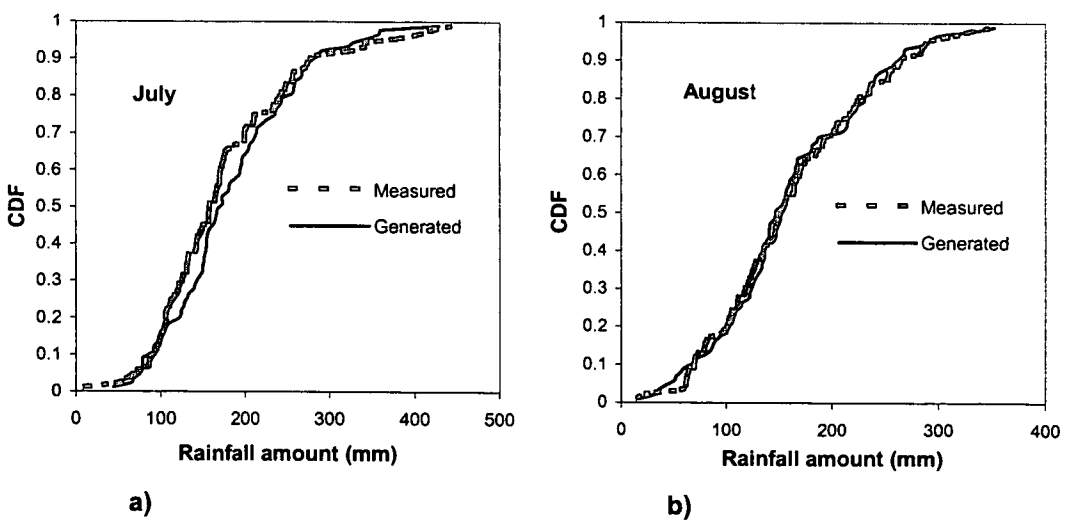


Figure 5.3 Probability of non-exceedence as a function of ranked precipitation for measured and generated rainfall for (a) July and (b) August months for Asmara.

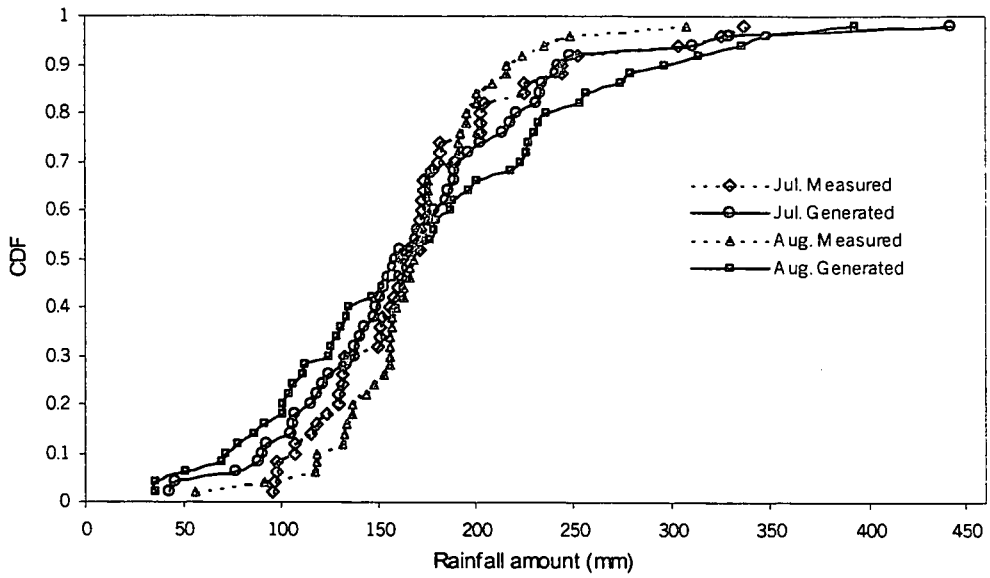


Figure 5.4 **Probability of non-exceedence as a function of ranked precipitation for measured and generated rainfall for July and August months for Mendefera.**

For a statistical proof the Kolmogorov-Smirnov (KS) test was conducted. The KS-test was used to determine if two distribution functions (measured and generated) are significantly different in a continuous variable like rainfall (Conover, 1971; Steele *et al.*, 1997). The KS-test involves a pairwise comparison, whereby the test statistics is applied to the maximum vertical distance between distributions. The two-distribution functions are different if the maximum vertical distance between them exceeds the critical level for a given level of significance chosen as 0.05. The KS-test has the advantage of making no assumption about the distribution of data i.e., it is non-parametric and distribution free. The classical KS-test is a widely used goodness of fit test which is based on differences between the empirical and the theoretical distribution function, and the statistic D is defined by: D = max (D^+, D^-)

$$D^+ = \underset{-\infty<x<\infty}{Sup}\,(F_N(x) - F_0(x)) = \max_{1\le i\le N}\left\{\frac{i}{N} - F_0(x_{(i)})\right\} \qquad (5.5a)$$

$$D^- = \underset{-\infty<x<\infty}{Sup}\,(F_0(x) - F_N(x)) = \max_{1\le i\le N}\left\{F_0(x_{(i)}) - \frac{i-1}{N}\right\} \qquad (5.5b)$$

where: $F_N(x)$ is the empirical distribution and $F_0(x)$ the cumulative distribution function.

For a specified significance level, the fitting is rejected if the computed value D is equal to or greater than the critical value. In other words, high D statistics and associated low ρ-values would offer evidence to reject the null hypothesis. From the KS-test (Table 5.3) it was proved that the CDFs of the peak rainy months (JA) are not statistically different.

Table 5.3 Kolmogorov-Smirnov (KS) test statistics, D, for the comparison of the CDFs of the measured and generated daily rainfall data by month.

Location		Asmara		Mendefera	
Factor		D-value	ρ-value	D-value	ρ-value
month	January	0.857	0.000	0.686	0.000
	February	0.735	0.000	0.674	0.000
	March	0.265	0.052	0.361	0.000
	April	0.347	0.004	0.233	0.016
	May	0.225	0.147	0.302	0.001
	June	0.184	0.346	0.174	0.131
	July	0.143	0.665	0.151	0.257
	August	0.265	0.502	0.058	0.998
	September	0.122	0.832	0.291	0.001
	October	0.408	0.000	0.465	0.000
	November	0.429	0.000	0.407	0.000
	December	0.632	0.000	0.674	0.000
	Annual	0.102	0.950	0.116	0.580

5.3.2 Gamma Distribution Function

The rainfall at all the stations was observed to have a skewed distribution. In this case, the gamma distribution function could be employed for describing the data. From Appendix D: D-i and D-ii it can be seen that there is a high degree of agreement between the statistics obtained for the measured rainfall data and those from the generated data. The poorest agreement occurred between skewness values especially for the dry months (October, November, December, January, February, March and April). Differences in skewness can be attributed to an inability in the rainfall data generator to simulate unusually high values. Also, the distribution of precipitation depth when rain does occur is positively skewed, that is, a smaller amount occurs more frequently than the larger amounts. The results also reveal that the generated sequence in the wet months

(June, July and August) has an almost similar distribution as that of the measured data. Consequently, the monthly mean, standard deviation, skewness coefficient, etc., of the measured data are adequately reproduced by the model.

To concisely summarise and quantitatively describe the annual cycle of histogram shape, a flexible probability density function (PDF) was sought. The two parameters of the gamma density function was selected for its flexibility in shape, where α the shape parameter and β the scale parameter can be estimated from the mean, E(x), and variance, V(x):

$$E(x) = \alpha\beta \quad (5.6)$$

$$V(x) = \alpha\beta^2 \quad (5.7)$$

The general formula of the gamma probability density function has been described in the literature review. For $\alpha = 1.0$, the gamma density function is the exponential density, while for $\alpha > 1.0$, the function is a non-symmetrical bell shaped distribution (Schmidt *et al.*, 1997).

As shown in Table 5.4, the parameters α and β for the measured and generated data were estimated for each month. At both stations the parameter α exhibited a better correspondence between generated and measured data than the parameter β. Nevertheless, no clear explanation can be given for the wet and dry seasons from the results.

The Chi-square test was used to evaluate the gamma distribution function goodness of fit. The Chi-square test makes a comparison between the actual number of observations and the theoretical or expected number of observations that fall in the class intervals employed. The fitting is rejected if the computed value is greater than the critical value at the selected significance level. The Chi-square goodness-of-fit test was carried out to compare the measured values with the generated values.

Table 5.4 The shape (α) and scale (β) parameters of the gamma probability density function for Asmara and Mendefera for each month.

Location		Asmara				Mendefera			
factor		shape (α)		scale (β)		shape (α)		scale (β)	
		Measured	Generated	Measured	Generated	Measured	Generated	Measured	Generated
month	January	0.108	1.534	15.291	1.266	0.056	2.003	2.245	0.491
	February	0.197	1.651	12.789	1.442	0.056	1.689	39.089	0.966
	March	0.686	1.698	17.227	5.614	0.423	1.193	26.312	6.815
	April	1.179	2.291	27.599	13.018	0.925	3.367	29.125	10.380
	May	1.145	2.043	37.869	25.179	0.869	2.603	45.541	14.859
	June	1.543	2.958	26.234	14.069	1.922	3.097	29.156	19.854
	July	4.281	5.694	40.908	32.142	9.556	5.669	17.927	30.150
	August	4.915	4.975	32.825	32.154	17.457	4.335	9.745	40.213
	September	1.033	1.751	23.969	15.897	1.614	2.325	32.979	22.279
	October	0.496	1.453	26.239	5.889	0.54	0.78	18.12	10.52
	November	0.620	2.089	25.397	3.428	0.552	1.843	27.083	2.578
	December	0.263	2.350	11.400	1.158	0.172	2.128	14.348	0.906
	Total	9.949	18.251	52.809	28.819	21.930	22.431	25.435	24.869

In both stations the difference between the two-tailed P-value (Table 5.5) were not statistically significant for the peak rainfall months (JA). That is, the measured and generated values show a similar behaviour for these two months (Figures 5.5 & 5.6) for both stations. The same was found for the P-value of the annual total rainfall. This means that the gamma distribution describes the distribution for the peak rainy months. Fitting the gamma distribution with scale and shape parameters by maximum likelihood to the measured and generated annual totals, showed the gamma model to be appropriate (Appendix E: E-i & E-ii). The histogram of the annual total for both locations also showed a similar cycle.

Table 5.5 Chi-squared goodness of fit test for the gamma probability density function for the peak rainfall amount (July and August) for Asmara and Mendefera.

Location	Asmara		Mendefera	
Factors	Chi-squared	P-value	Chi-squared	P-value
July	20.799	0.054	5.487	0.601
August	15.968	0.104	29.569	0.051

The fitted gamma probability density function of the measured and generated monthly totals for both stations was also further analysed for each month (Appendix F: F-i & F-ii). During the driest months, the distribution was exponential with a very high frequency of small monthly totals and a low frequency of moderate to large totals. During the wet season the monthly histogram gradually changed from an exponential type distribution to a broad, normal distribution. Progressing from wet to dry season, the histograms gradually returned to the exponential shape.

5.3.3 Student's t-Test

A t-test was also conducted to evaluate the amounts of the measured and the generated rainfall for both stations with 0.05 levels of significance. The null hypothesis was two-sample assuming unequal variances. The degrees of freedom were 170 and 96 respectively for Asmara and Mendefera. As indicated in Table 5.6 the null hypothesis is rejected in most cases, which means that the two parameters (measured and predicted) have similar variances.

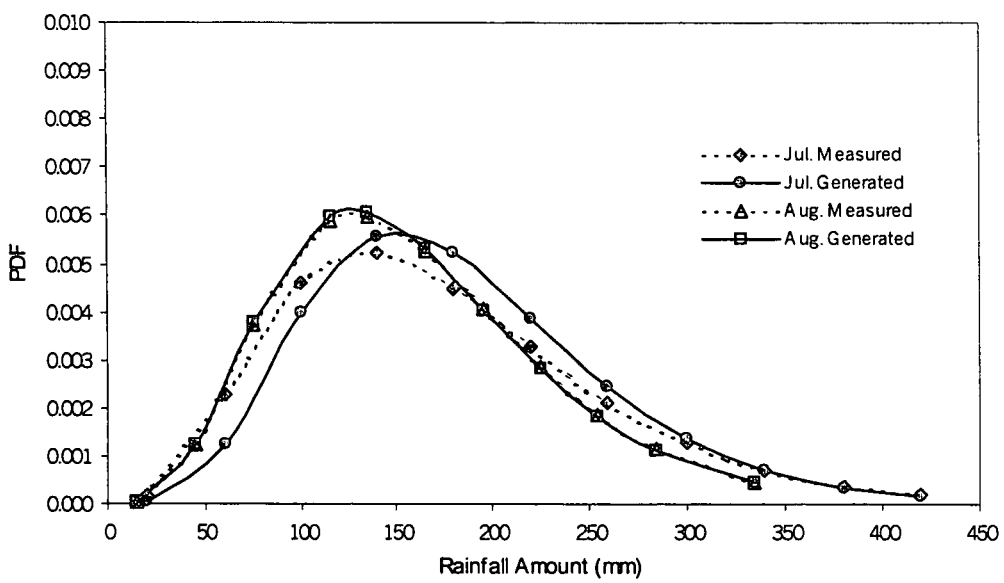


Figure 5.5 Probability density function (PDF) of the measured and generated July and August rainfall for Asmara.

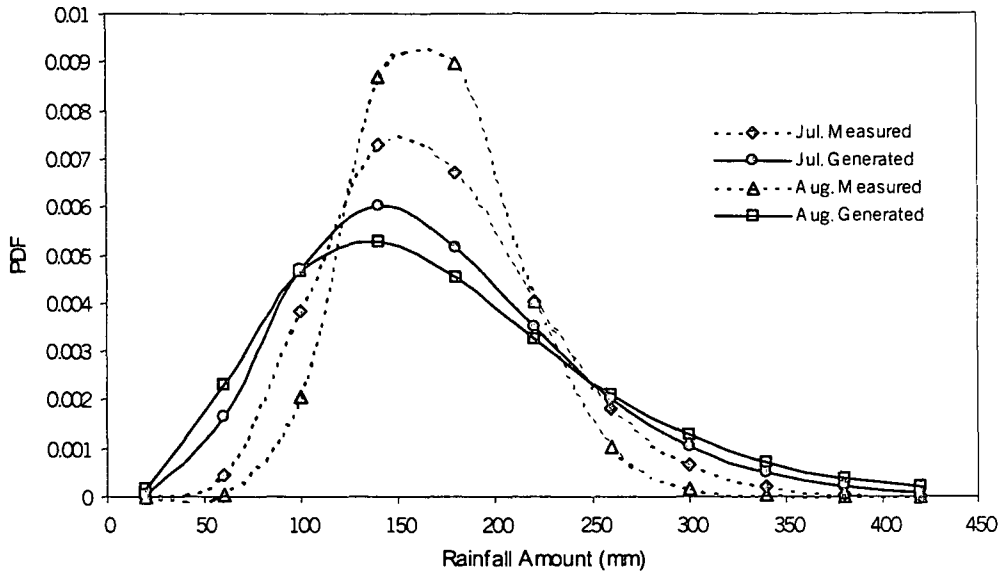


Figure 5.6 **Probability density function (PDF) of the measured and generated July and August for Mendefera.**

Table 5.6 **A two-tailed student's t-test (two-sample assuming unequal variances) for Asmara and Mendefera: 0.05 confidence level, (* significant)**

Location		Asmara	Mendefera
factor		P-value	P-value
month	January	8.84E-10	0.617*
	February	0.684*	0.826*
	March	0.265*	0.189*
	April	0.101*	0.478*
	May	0.897*	0.169*
	June	0.477*	0.795*
	July	0.977*	0.525*
	August	0.752*	0.903*
	September	0.852*	0.379*
	October	0.490*	0.039
	November	0.0009	0.0002
	December	0.539*	0.679*
	Annual	0.998*	0.979*

5.3.4 Comparison of the spell lengths

The generated and measured daily sequences were also compared to evaluate the model performance concerning the start of the rainy season and frequency of dry spells. 46 years of generated daily data were compared with the 46 years

of measured historical data on a daily basis. The model somehow has a failure in preserving the frequency of the dry and wet days (Figure 5.7). The mean number of dry days for each month has been preserved by the model. For example, the mean number of dry days for July is 10 per year historically and 9 per year for generated data. For August the corresponding figures are 10 days and 7 days.

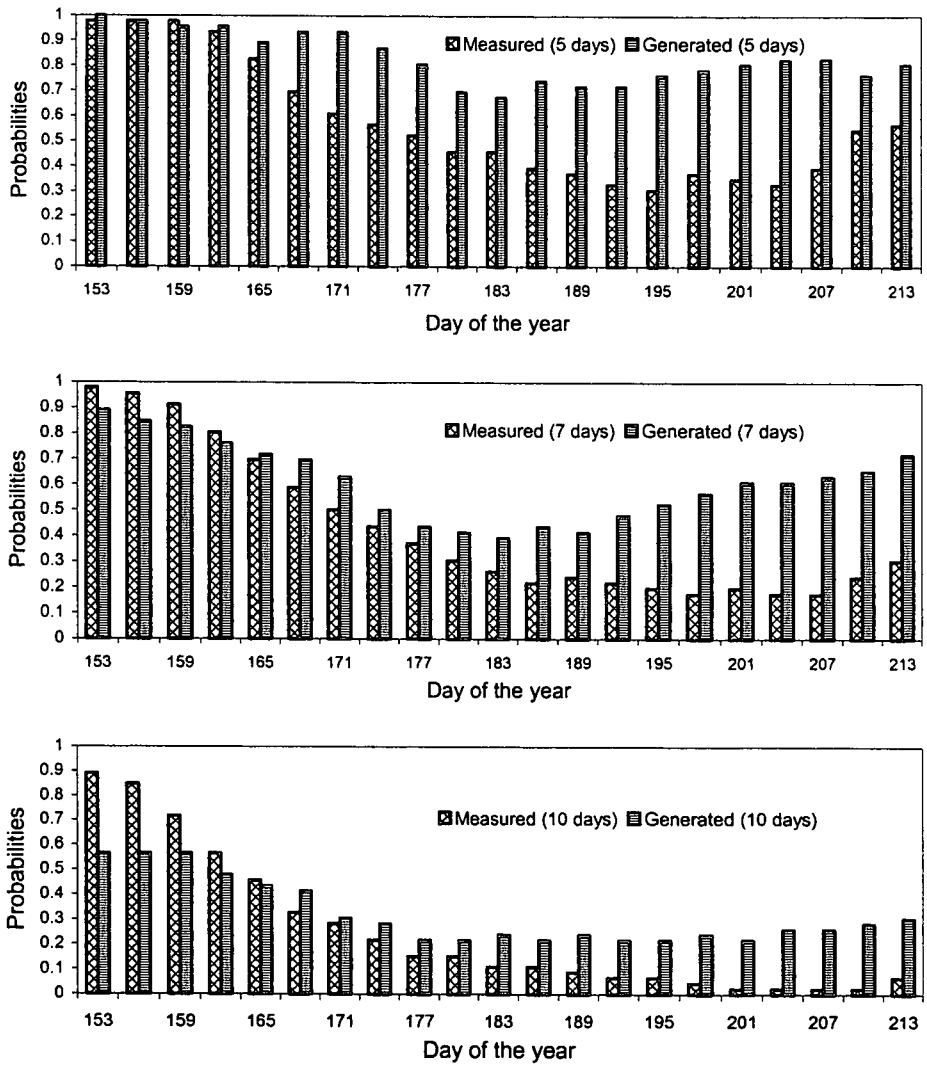


Figure 5.7 **Comparison of the probability of commencing a dry spell for different length of days in the rainy season at Asmara for 46 years daily data.**

5.4 Possible Model Application

In many agricultural regions, such as the highlands of Eritrea, the density of meteorological stations is low, and reliable long-term continuous data are very scarce. Thus the use of a generated dataset is essential. Some of the potential model applications include: assessment of the rainfall climate of a country; input to crop models such as Putu (De Jager *et al.*, 2001); input to hydrological models to obtain an understanding of the system for agrohydrology performance under a specific rainfall regime (e.g. Richardson, 1982).

The point of being able to generate sequences of artificial rainfall is that it enables one to estimate statistics relating to rainfall events. For example, the model could be helpful for risk analysis and analysis of drought occurrence. In effect one estimates probabilities of this type by simply regarding the artificial rainfall sequence generated as a real rainfall record. One can do this because the model used to generate the sequences has been shown to preserve the properties of real rainfall sequences. The yield of a crop is obviously related to the amount of precipitation that falls during some critical period of growth and the occurrence of long dry spells during the growing season of a crop is a major agricultural hazard. The risk of a long dry spell in the 30 days after sowing for any potential sowing date was estimated using the generated 100 years daily rainfall data for Asmara station. Knowledge of these risks may help the user in his decision of whether to sow and the planner in defining a sowing strategy.

Assuming day $\mathbf{x}$ was a sowing date and therefore had rain, each event therefore gives the lengths of the longest dry spell in the 30 days. Supposing that in the next year, late June (day 175) is a potential sowing date, one can see the probability of a dry spell of more than 10 days within the next 30 days is only about 0.2 (i.e. 1 year in 5) (Figure 5.8). About half of the years are estimated to have a dry spell of more than 5 days within the next 30 days. The longest dry spells endured for 180 days, from (29 April to 16 October). June is the most likely month for moderately wet spells, and July for extremely wet spells.

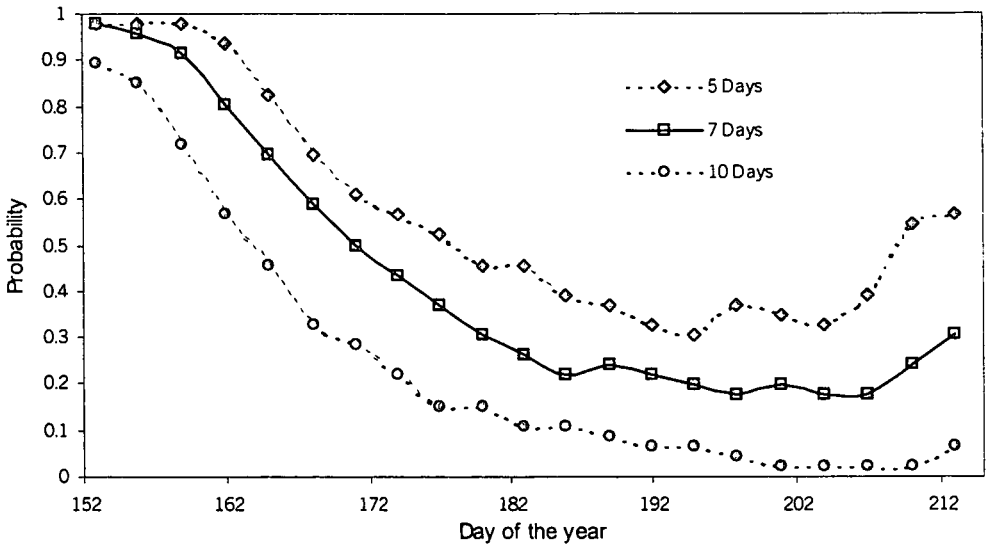


Figure 5.8 The probability of dry spells of 5, 7 and 10 days length for a 100 years of generated daily rainfall data for Asmara.

5.5 Conclusion and Recommendations

As rainfall events are intermittent, a long data series is required to allow accurate and representative statistics to be calculated. This requirement is seldom satisfied in practice. But, one can get a wealth of information from the analysis of long data series by using a weather data generator. A continuous daily rainfall record provides the information for the agriculturist to solve operational problems. Researchers and decision-makers can routinely use the daily rainfall generator models in a wide variety of applications.

Nine hundred years of daily rainfall data were generated for the Asmara and Mendefera stations in the highlands of Eritrea. A number of statistics were determined for both the generated rainfall and measured historical data. The comparison between generated and measured rainfall data is fairly good but not excellent.

The first order Markov chain was adopted in the stochastic model because it simplifies the determination of the number of states to be used and the estimation of the corresponding transition probabilities. A trial and error procedure was used to get a representative state. The model incorporates the seasonal effect by interpolation of monthly data. The developed model could be

used with confidence in various fields of interest like drought monitoring, crop modelling and hydrology. The model has a weakness in that the model failed to account for extreme values. This problem arises because of the monthly median rainfall, smoothed by spline interpolation being used, instead of daily data. The monthly histogram for both actual and generated values was observed to vary in a very regular manner throughout the year. A gamma probability density model was successfully fitted to the monthly histogram.

Further research is required to develop methodology for generating simulated sequences of daily rainfall for an area rather than a single point, in such a way as to preserve the appropriate spatial correlation of individual rainfall occurrences.

CHAPTER 6

STATISTICAL MODEL FOR SEASONAL FORECASTS OF RAINFALL

6.1 Introduction

Climate variations have an important impact on agriculture and water supply. With an ever-increasing population, effective water management has become essential. Seasonal forecasting has good prospects for early warning of low rainfall totals to help prepare for, and mitigate the effect of, the famine, which so often results from drought in Eritrea. The need for providing accurate forecasts for a coming rainfall season is becoming more and more necessary. Farmers could make more efficient management decisions if they had a better assessment of the forthcoming season in adequate preparation time.

Climate variability is caused by complex interactions between the atmosphere and the water, ice, and land surfaces beneath it. The extreme pressure anomalies in the Tropical Pacific and Indian Oceans involve dislocations of the rainfall distribution in the tropics, bringing drought to some regions and torrential rains to other areas. These anomalies can last for many months (Philander, 1990; Nicholls, 1991).

Approaches to seasonal forecasting can broadly be divided into two categories, statistical (empirical) techniques and numerical, dynamical modelling. Statistical modelling assumes that historical climate will be repeated in the future, that information contained in historical data can be extracted, analysed and reduced to one or more equations that can be used theoretically to replicate historical patterns. The statistical methods used in current statistical forecasting include regression analysis, principal component analysis, discriminant analysis, cluster analysis, analogue methods, time-series analysis and period analysis. A wide variety of statistical techniques have been used in

the development of empirical seasonal forecasts to identify lagged statistical associations between SST indices and predictand climate variables. A basic starting point is simply correlation analysis between predictand and predictor variables. Different techniques can be used for seasonal forecasting models, for example, multiple regression (Colman *et al.*, 1996), quadratic discriminant analysis (Mason, 1998), canonical correlation analysis (CCA) (Landman & Mason, 1999; Thiaw & Barnston, 1999), or neural networks (Hastenrath *et al.*, 1995; Greischar & Hastenrath, 1999).

In East Africa, Mutai *et al.*, (1998) have found promising seasonal forecast skill for the Oct-Nov-Dec "short" rains using multiple regression techniques and predictors based on eigenvectors of global SST. Hastenrath *et al.*, (1998) has attempted to forecast February-March discharge of the Caroni River in Venezuela using multiple linear regression and the previous Jul-Aug Southern Oscillation and river discharge as predictors. Martin *et al.*, (1999) used correlation and multiple regression methods to create a maize yield forecast, by relating the water stress index and SSTs and sea level pressures at several months lead time. Fraedrich and Smith (1989) discuss a linear regression method of combining two statistical forecast schemes, and apply this method to long-term forecasting of the monthly mean tropical Pacific sea surface temperatures.

The objective of this chapter is to develop a simple statistical model forecast for the peak rainy months (JA). A simple model is needed because simple models are easier to understand, easier to test, and are less costly to put into practice in predicting and controlling the outcome in the future also, sometimes the effort for minimal improved reliability and sophisticated is hardly justified.

6.2 Data

The data used in this assessment are the long-term monthly rainfall amounts for two representative stations of the highlands of Eritrea and 11 predictors from the Indian, Atlantic and Pacific Oceans as well as the sunspots from 1950 to 2000. Eleven predictors used are indicated as follows:

Niño1+2 (0-10°S, 90-80°W);

Niño3 (5°N-5°S, 150-90°W);

Niño4 (5°N-5°S, 160°E-150°W);

Niño3.4 (5°N- 5°S, 170-120°W);

North Atlantic SST (5-20°N, 60-30°W);

South Atlantic SST (0-20°S, 30°W-10°E);

Global Tropics SST (10°S-10°N, 0-360);

South Indian SST (0-15°S, 45-60°E);

Southern Oscillation Index (SOI);

Pacific Decadal Oscillation (PDO) and Sunspot index.

These parameters were selected and tested as described below to determine which have a significant influence on Eritrea rainfall and can be used in a prediction model. It would be interesting to include more predictors like ITCZ, QBO, winds at different hPa and the Red Sea convergence zone into the analysis, but the availability of the data is the main stumbling block, so they were not included at this stage.

6.3 Model Development

6.3.1 Correlation Matrix

A correlation matrix is used to show all possible correlation coefficients between all variables. The matrix is useful in showing how strong each independent variable is related to the dependent variable at different lag times. The method used here is a row-wise deletion.

A correlation matrix was set up for all 12 candidates including rainfall itself as one of the candidates. For each candidate monthly lags (1-12) were calculated, such that lag 1 is a 1-month difference between the predicted rainfall and the candidates. From the results it was found that lag11 and lag12 between the previous year's rainfall and the coming rainy season are significant, which can be explained by the autocorrelation property of the long-term rainfall pattern. From the resultant correlation matrix further analysis was carried out to see how the different lags of the rainfall correlated with the different lags of the other parameters.

6.3.2 Stepwise Regression

As was mentioned before, this chapter will focus on JA (July-August) which form the peak rainy months in the area of interest. These two months contribute 65% of the total annual rainfall in the highlands of Eritrea. All the predictors data were categorised into 6 sets consisting of 2 month sums of JF, MA, MJ, JA, SO and ND. From the results of the correlation matrix, stepwise regression was employed for further analysis. Stepwise regression is a technique for choosing the variables to be included in a multiple regression model. The basic procedures involved in stepwise regression are:

a) Identifying an initial model,

b) Repeatedly changing the model by adding or removing a predictor variable in agreement with the stepping criteria,

c) Terminating the search when a specified maximum number of steps has been reached.

A stepwise selection procedure was carried out to select the most significant candidate from the 66 variables. These 66 variables represent from lag 1 to 6 for each of the 11 predictors for the 6 categories of the 2-month sums. The NCSS statistical software (Hintze, 2000) was used for all regressions and multiple regressions.

6.3.3 Multiple Regression Analysis

Regression analysis is a statistical tool that allows us to create a model for predicting the behaviour of the dependent variable. Multiple regression models of the dependent and independent variable are of the form:

$$R_i = \beta_0 + \beta_1 I_{1i\,(1\text{-}6lag)} + \beta_2 I_{2i\,(1\text{-}6lag)} + \ldots \beta_j I_{ji\,(1\text{-}6lag)} + E_i \qquad (6.1)$$

(i = 1, 2, ..., n)

where the dependent variable (R) and the independent variables (I) are monthly total data. β_0, β_1... β_j are the regression coefficients to be estimated. n is the number of years, and E_i is the i^{th} year model error.

Using matrix notation, the model in Equation 6.1 can be written:

$$R_i = \beta I + E \qquad (6.2)$$

$$\begin{bmatrix} R_{i1} \\ R_{i2} \\ . \\ . \\ . \\ R_{ij} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_j \end{bmatrix} \begin{bmatrix} 1 & I_{11} & I_{21} & . & . & . & I_{i1} \\ 1 & I_{12} & I_{22} & . & . & . & I_{i2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & I_{1i} & I_{2i} & . & . & . & I_{ij} \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \\ . \\ . \\ . \\ E_j \end{bmatrix}$$

where R_i = Monthly rainfall; β = coefficients; I = input variable (eleven predictors) and E = the error.

It was found that (lag1 to 6) of the South Indian SST, South Atlantic (lag6) (R^2 = 0.63), North Atlantic (lag3) (R^2 = 0.61) and SOI (R^2 = 0.6) have an influence on the amount of rain received during the peak rainfall months (JA). For the sake of simplicity it was decided to focus on the highly influential region that was identified as the South Indian SST, but ignoring lag1 from an operational point of view as it is hardly useful as it is only available at the beginning of July month.

From multiple regression estimates the regression coefficients were calculated as indicated in Table 6.1. Multi-collinearity is also not a problem in this case as indicated in the Table. The absence of multi-collinearity is essential to a multiple regression model. In regressions when several predictors are highly correlated, they are called multi-collinear. If this is the case, the presence of one input variable in the model may mask the effect of another input. Variation inflation factor (VIF) and tolerance of a variable are the two methods to identify collinearity. VIF = $1/(1-R^2)$. When R^2 and VIF values are high for any of the predictors the fit is affected by multi-collinearity. A general rule is that the VIF should not exceed 10. If any variable has a VIF greater than 10, collinearity could be a problem (Belsley, *et al.*, 1980; Hintze, 2000). A value of near one for tolerance of variables indicates independence. If the tolerance value is close to zero, the variables are multi-collinear.

The normal probability plot of residuals of 95% confidence interval for the rainfall fits well as shown in Figure 6.1. Most of the residuals fall within the confidence bands for the normal probability plot. But the model is not sensitive

to the extreme residuals. Attempts were also made to force the extreme residuals into the confidence bands via the transformation of the logarithmic model.

Table 6.1 Multiple regression statistics for rainfall and different lags of South Indian SST.

months	variable	Regression coefficient	Standardised coefficient	Standard error	t-value	Probability level	Power (5%)	R-squared	multicollinearity	
									Variance inflation	tolerance
	Intercept	129.232	0.000	540.26	0.239	0.811	0.056	0.79		
	Indian lag2	4.049	0.031	4.791	0.845	0.398	0.134		1.904	0.525
JA	Indian lag3	33.461	0.253	4.996	6.697	0.000	0.999		2.081	0.481
Asmara	Indian lag4	-10.50	-0.079	5.947	-1.766	0.078	0.421		2.923	0.342
	Indian lag5	85.348	0.642	5.002	17.064	0.000	1.000		2.063	0.485
	Indian (lag6	-114.67	-0.862	4.790	-23.939	0.000	1.000		1.894	0.528
	Intercept	-141.41	0.000	540.26	-0.285	0.776	0.059	0.84		
	Indian lag2	1.421	0.010	4.791	0.323	0.747	0.062		1.904	0.525
JA	Indian lag3	43.849	0.323	4.996	9.555	0.000	1.000		2.081	0.481
Mendefera	Indian lag4	-6.95	-0.051	5.947	-1.273	0.204	0.245		2.924	0.342
	Indian lag5	80.546	0.590	5.002	17.533	0.000	1.000		2.063	0.485
	Indian lag6	-116.31	-0.852	4.790	-26.436	0.000	1.000		1.894	0.528

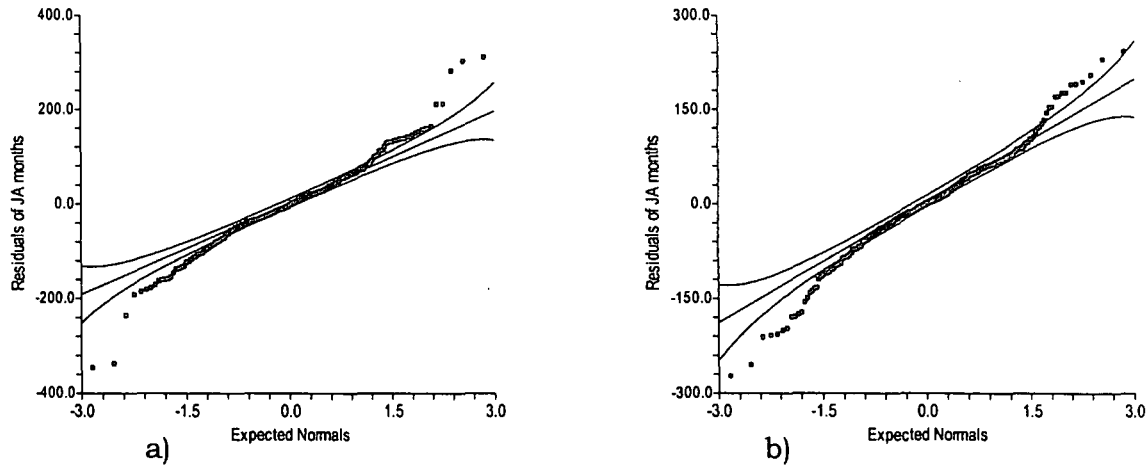


Figure 6.1 Normal probability plot of 95% confidence for the residuals of the peak rainfall months (JA) for (a) Asmara and for (b) Mendefera.

6.3.4 Statistical Model Development

The model is developed using the lags of the South Indian SSTs and the lags of rainfall amount. As was mentioned before the rainfall amount of the previous year has an influence on the current rainy season. This idea is used in model development.

The input data required by the model: The difference in rainfall between JA (July-August) and ND (November-December) (lag4) can be explained as follows:

$$Ri_{,j} - Ri_{,j-4} = f(SST_{\text{S. Indian}}) + \text{Error} \qquad (6.3)$$

($R_{i,j} = 0$ if $R_{i,j} < 0$)

The formula for the peak rainy months (JA) is given as:

a) Asmara

$$\mathbf{Ri_{,j} = Ri_{,j-4} + (+0.03(SIn_{t-2})+0.25(SIn_{t-3})-0.08(SIn_{t-4})+0.64(SIn_{t-5})-0.86(SIn_{t-6}))}$$

b) Mendefera

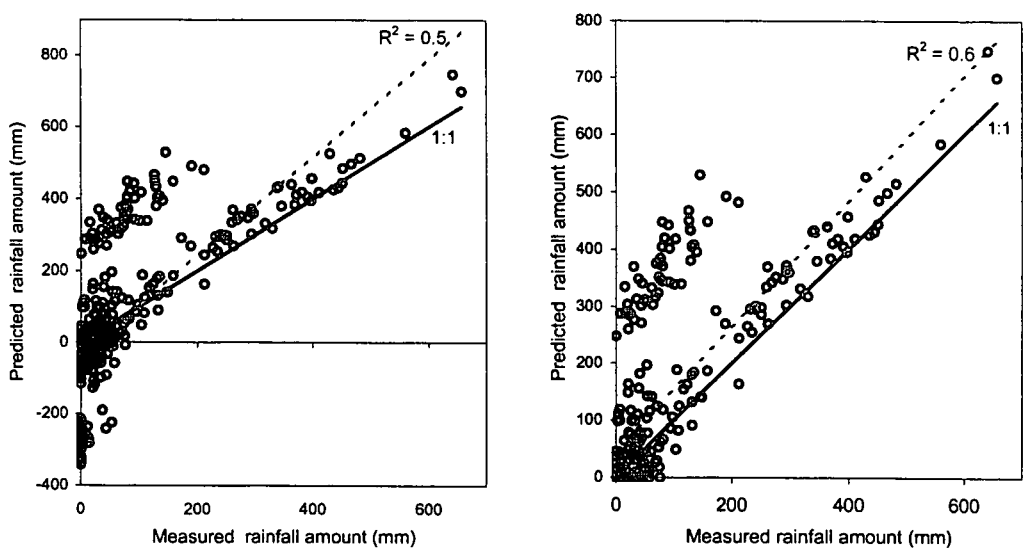
$$\mathbf{Ri_{,j} = Ri_{,j-4} + (+0.01(SIn_{t-2})+0.32(SIn_{t-3})-0.05(SIn_{t-4})+0.59(SIn_{t-5})-0.85(SIn_{t-6}))}$$

where: $Ri_{,j-4}$ = Total rainfall amount of the previous year for ND

$Ri_{,j}$ = Current rainfall amount for JA

SIn = South Indian Ocean SST.

As is indicated in Figure 6.2a, the model gives negative values for precipitation in dry months. Negative values amount to 28% for the peak rainfall months (JA). To improve the model a correction of the 28% was added to the model, but this modification did not give the desired result. When the negative values were changed to zero (Figure 6.2b) the R^2 increased from 0.5 to 0.6. Only the peak rainy months (JA) were chosen for further analysis.



a) b)

Figure 6.2 Comparison of the measured and predicted values of the 2 month totals for (a) 6 month sequence for Asmara and (b) Truncated model.

6.4 Model Validation

Once a model has been identified and the parameters estimated, it remains to decide whether the model is adequate for its purpose. Model validation is performed with the objective of assessing the performance of the model and to uncover any possible lack of fit.

Different methods were used in order to determine the model accuracy, in other words how near the predicted value is to reality. These are: Jack-knife skill test; hit, over and under estimation type of analysis and Chi-squared test.

6.4.1 Jack-knife Skill Test

Before the model can be used to forecast in real-time, validation of the model should be made on independent time-periods with independent data. Trail forecasts (hindcasts) were made using the Jack-knife method. Jack-knife forecasts were made for every year in the data set period. The forecast year is always excluded from the regression equation. The process is repeated by removing the next year and so on until 28 forecasts have been made for the data set from 1950 to 2000. Thus the coefficients of the predictors in the equations change from year to year. The 22 years were selected by assuming that the rainfall cycle could be linked in some way with sunspot cycles which have an approximate 11-year period. Another important point is that the model needs to be trained on the most recent information (Jury, *et al.*, 1997).

Jack-knife multiple regression forecasts are plotted against measured rainfall amounts in Figures 6.3 and 6.4. The correlation of 0.89 and 0.85 for Asmara and Mendefera is very high for a climate prediction. The anomaly departures are calculated by subtracting the mean and dividing by the standard deviation (Jury, *et al.*, 1997).

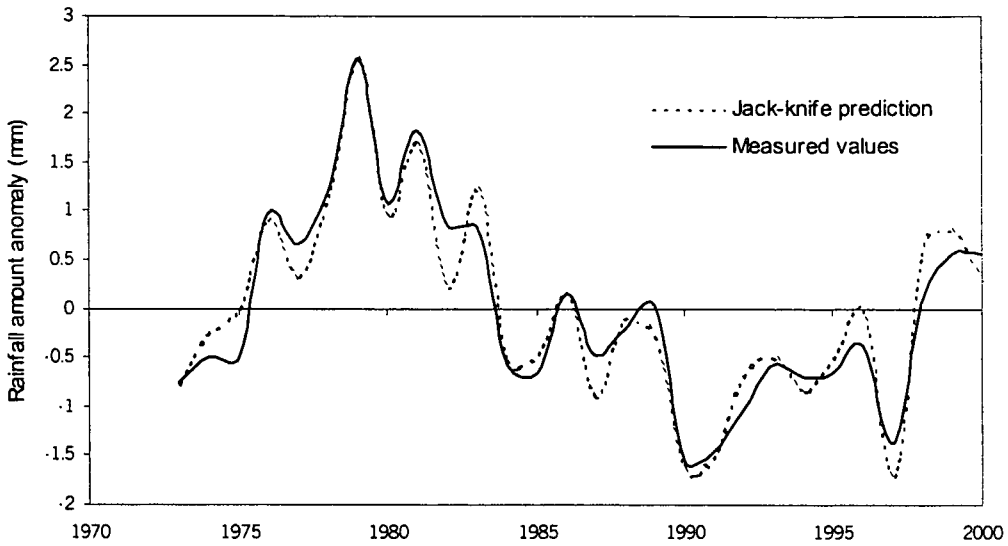


Figure 6.3 **Jack-knife skill tests for S. Indian SSTs (lag_{2-6}) 28-year models, showing rainfall anomalies versus Jack-Knife predictions for Asmara for peak rainy months (JA). [R^2 = 0.89; D = 0.91 and RMSE = 50.62]**

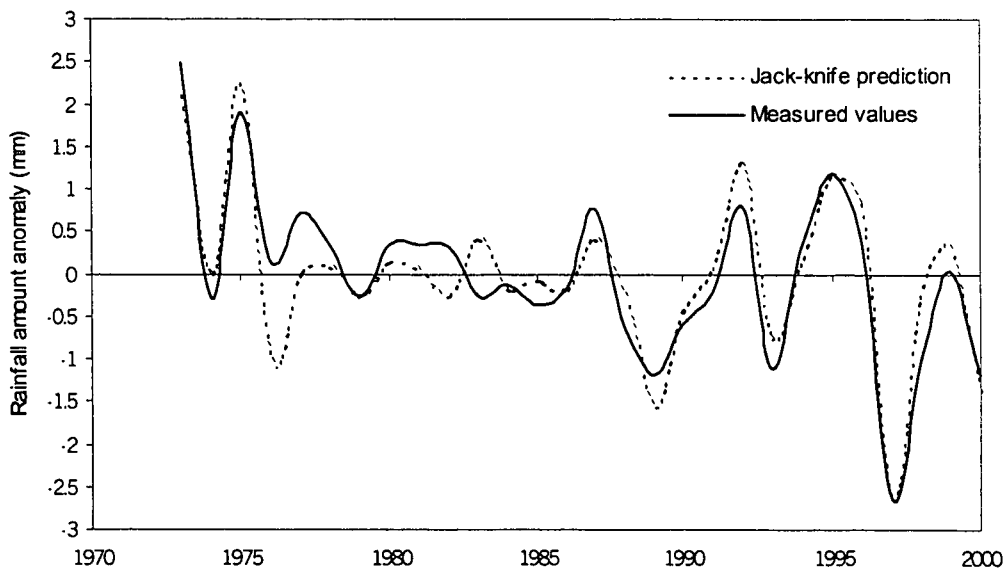


Figure 6.4 **Jack-knife skill tests for S. Indian SSTs (lag_{2-6}) 28-year models, showing rainfall anomalies versus Jack-Knife predictions for Mendefera for peak rainy months (JA). [R2 = 0.81; D = 0.79 and RMSE = 85.89]**

6.4.2 Hit Rate

The hit rate is the simplest method of evaluation (Ward & Folland, 1991). For this method of evaluation the data were separated into two continuous discrete segments of time. Half of the data (1950-75) was used for training and another half (1976-2000) for validation of the model. For the hit and miss kind of analysis a probability <33% is taken as below normal (BN); 33% - 66% probability as near normal (NN) and the probability> 66% used as above normal (AN). From the cumulative probability distribution the 33% and 66% were calculated as is shown in Table 6.2.

Table 6.2 The 33% and 66% probability level of the measured rainfall amount for the 2 month of peak rainy months for Asmara and Mendefera (1950-2000)

Month	Asmara		Mendefera	
	33%	66%	33%	66%
JA	261.4mm	381.7mm	297.0mm	347.5mm

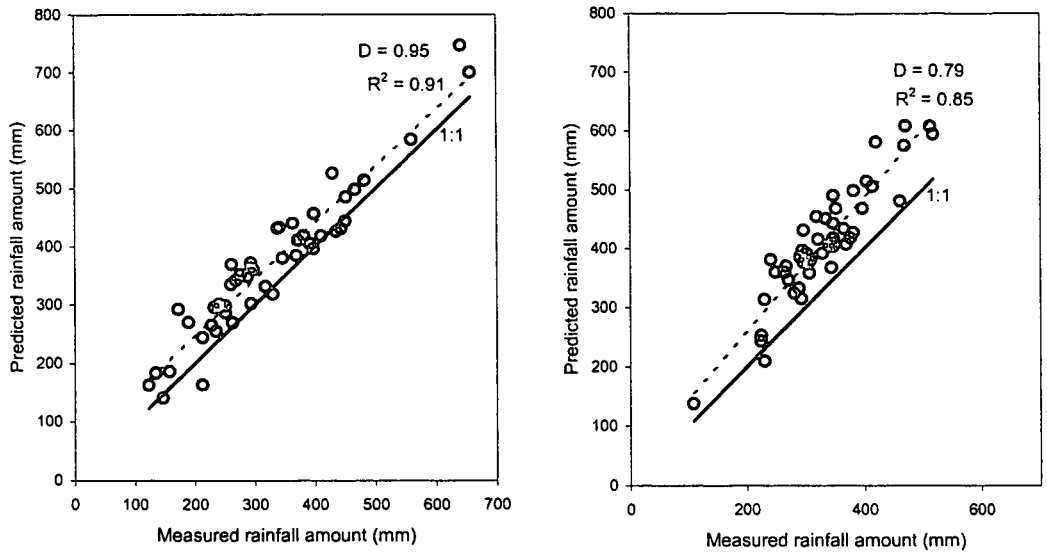
The following assumptions were made for Asmara, for example (Table 6.3):

BN < 261.4mm ≤ NN ≤ 381.7mm <AN.

Table 6.3 The percentage of hit, over and under estimate of the predicted values for Asmara and Mendefera.

Location	Asmara	Mendefera
Month	July-August	
Over-estimated	5/25 = 20%	8/25 = 32%
Hit	20/25 = 80%	17/25 = 68%
Under-estimated	0/25 = 0%	0/25 = 0%

In general, the model overestimates the amount of rainfall received during the peak rainy months (20 and 32%) (Figures 6.5a, 6.5b). The hit rate for the model is very good for both stations.



a) b)

Figure 6.5 Regression results of measured and predicted values for peak rainy months (JA) of Asmara (a) and Mendefera (b).

6.4.3 Chi-Squared Test

For the chi-square test, the data were divided into 10 classes. The classes were categorised as follows:

$$(X-1)(\text{Max. } R_i/10) \leq \text{ and } < (X)(\text{Max. } R_i/10)$$

where: X = 1,2,..., 10.

Max. R_i = Maximum rainfall amount (mm) recorded in 25years.

The Chi-square test results indicated that the model predicted values are not significantly different from the measured values (Table 6.4).

Table 6.4 The Chi-squared two-tailed p-value goodness of fit test for the predicted and measured amount of rainfall for both stations.

Month	Asmara		Mendefera	
	Chi-square	p-value	Chi-square	p-value
JA	3.300	0.654	9.292	0.167

An inspection of the results of model validation using different statistics clearly indicated that the model predicted the peak rainy months amount of rainfall very well. This is very critical for crop production. Generally, the results

indicated that the influence of the Indian Ocean SST on the rainfall in the Eritrean highland is very high.

Indian Ocean SSTs are used in numerous statistical long-range rainfall forecasts for Africa (Jury, 1993; 1996). Oceanic and atmospheric circulation over the Indian Ocean have a marked impact on convection over Eritrea. Indian Ocean SST forms the basis for long-range climate forecasts in many parts of Africa (Barnston *et al.*, 1996). A warmer tropical Indian Ocean is frequently associated with wet conditions over eastern Africa (Mason, 1995). Unlike the decadal oscillations of SST in the tropical Atlantic, the tropical Indian Ocean has exhibited ENSO oscillations at a period of about 4 years since 1950 (Mason, 1995). Walker (1990) suggests that lower SSTs in the central Indian Ocean suppress uptake of cross-equatorial monsoon flow by tropical cyclones, and coincide with wetter conditions over Africa. Pathack (1993) demonstrated a close coupling between equatorial convection over the Indian Ocean during the onset of the NE monsoon and African rainfall at lead times of 3 to 6 months. Most recently Clark *et al.* (2003) demonstrated that there are strong correlations between Indian Ocean SST and coastal East African short rains (OND). They also identified a significant change in this relationship between 1983 and 1993.

6.5 Conclusion and Recommendations

Initially, of the 11 candidate predictors, southern Indian SST has shown very high correlations. Therefore a simple most successful model was developed to predict the rainfall amount for the peak rainfall months (JA). The model uses rainfall from previous November-December and SST from South Indian Ocean for a 2 month sum for the whole of the previous year of 5 values of South Indian Ocean SST i.e., MJ, MA, JF, ND, SO. The model could provide a measure that can be used in developing a seasonal forecast system, which would allow farmers to make better management decisions.

The model was validated using three different methods: Jack-knife skill test, hit rate and Chi-squared test. From the validation it can be concluded that the

model is reproducing and describing the pattern of the rainfall for both Asmara (R^2 = 0.89; D = 0.91) and Mendefera (R2 = 0.81; D = 0.79) stations of the Eritrean highlands.

Further research is needed to explore the ways in which the Indian Ocean can influence the Eritrean peak rainy months in conjunction with other rainfall predictors.

All the information of the seasonal forecast is useless unless it can be converted into an organised and comprehensive strategy to minimise risk and vulnerability on the ground. Good communication and feedback (Klopper, 1999) between the forecaster and the end-user will be needed. Training of end users will also be essential.

A further study should be done for a more specific seasonal forecast that relates to crop yields, rather than just rainfall prediction, which is very critical for drought prone countries like Eritrea. Improved prediction of expected rainfall behaviour in the approaching crop season enables improved decision making at the field level and outlooks need to be developed for crop specific information (Walker *et al.*, 2001).

CHAPTER 7

CONCLUSION AND SUMMARY

The conflict between wanting to use daily rainfall data to achieve unbiased predictions and having to make do with the more coarse available data (actual monthly data) can be resolved by generating pseudo daily data. The main stumbling block in the analysis was the lack of suitable detailed historical records. Nevertheless, for the purpose of modification of the weather data generator, they are adequate. Considerable time was dedicated to investigating these techniques.

The stochastic model used for generating an artificial daily rainfall dataset was developed and calibrated at two stations, namely one in the central zone, Asmara and one in the southern zone, Mendefera. The developed model used a first-order Markov chain to describe the occurrence of wet and dry days. The median rainfall was used to develop the interpolated daily values from monthly totals via a spline interpolation. This is because the median has certain advantages over the mean as a measure of average rainfall. However, it is not so sensitive to occasional extremes. Extensive tests were carried out on the model. The tests were CDF test, KS-test, Gamma density function and student's t test. The results show that, on the whole, the model performs remarkably well. Two stations were used to evaluate the performance of the model.

Researchers and decision-makers routinely use the daily rainfall data in a wide variety of applications. It is hoped, now that this model is applicable for the central and southern Eritrea, that it will find even wider application.

In order to improve the spatial coverage of the rainfall stations, spatial interpolation techniques were employed. From evaluation of three spatial interpolation methods, the Kriging (Universal U_2) method of interpolation was selected as the best model to interpolate rainfall data in Eritrea. The statistics

used for evaluation were the root mean square error (RMSE) and the Willmott index of agreement (D).

Given an improved understanding of Eritrean climate, numerous benefits could be expected in many related activities: better management of agriculture and water resources stemming from more reliable seasonal predictions. In this study the Indian Ocean Sea Surface temperature was identified out of 11 predictors to be the most influential predictor for July August of the rainfall in the highlands of Eritrea. A statistical model was developed for peak rainy months (July-August) of the study area. The model jack-knife skill test gave the correlation of 0.89 and 0.85 for Asmara and Mendefera stations, which is very high for rainfall prediction. Thus, validation of the model shows that the model can reproduce the measured monthly sum for JA rainfall totals with confidence.

Further research is needed for an in depth study on the processes governing climate variability over Eritrea. The value of the forecasts will depend not just on their accuracy, but also on the management options available to the user to take advantage of applying the forecasts. Successful mitigation of the impact of climate hazards cannot purely be concerned with seasonal forecast development. It requires an interactive approach between the physical and social sciences. A communication (two-way) process should be strengthened in order for the envisioned development to have the desired outcome.

7.1 Future Priorities

Intensive research will be required to develop methodology for generating simulated sequences of daily rainfall for an area rather than a single point, in such a way as to preserve the appropriate spatial correlation of individual rainfall occurrence.

Research is needed to combine the spatial interpolated results as well as the most effective rainfall predictors with the weather generator in order to produce generated weather data at a high level of accuracy throughout Eritrea , particularly areas where there are little or no rainfall data available.

Scientific collaboration between international and Eritrean scientists should be encouraged. Links between advanced research institutes need to be established and nurtured by gaining funding for joint projects from international funding agencies.

REFERENCES

Adamson, P. T. (1978). The analysis of areal rainfall using multiquadratic surfaces. Department of Water Affairs and Forestry, Technical Report No TR82, Pretoria.

Allan, R. J. (1988). El Niño-Southern Oscillation influences on the Australian Region. *Progress in Physical Geography* 12: 4-40.

Aronoff, S. (1989). Geographic Information System: A Management Perspective. WDL Publications, Ottawa, 294 pp.

Aspen, M. & Douglas, M. W. (2002). Characteristics of wet and dry spells over the Pacific side of Central America during the rainy season. *Monthly Weather Review* 130: 3054-3073.

Bancroft, B. A. & Hobbs, G. R. (1986). Distribution of kriging error and stationarity of the variogram in a coal property. *Mathematical Geology* 8(7): 635-651.

Barnston, A. G., Thiao, W. & Kumar, V. (1996). Long-lead forecasts of seasonal precipitation in Africa using CCA. *Weather and Forecasting* 11: 506-520.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and sons, New York, 292 pp.

Beltrando, G. (1990). Space-time variability of rainfall in April-November over East Africa during the period 1932-1983. *International Journal of Climatology* 10: 691-702.

Beltrando, G. & Camberlin, P. (1993). Interannual variability of rainfall in the eastern horn of Africa and indicators of atmospheric circulation. *International Journal of Climatology* 13: 533-546.

Bhalme, H. N., Moodley, D. A. & Jadhav, S. K. (1983). Fluctuations in the drought / flood area over India and relationship with the Southern Oscillation. *Monthly Weather Review* 111: 86-94.

Brandão, A. de G. & Zucchini, W. (1990). A Stochastic Daily Climate Model for South African Conditions. WRC Report No 200/1/90. Water Research Commission, Pretoria, 284 pp.

Brately, P., Fox, B. L. & Schrage, L. E. (1987). A Guide to Simulation. Springer-Verlag, New York, 397 pp.

Bristow, K. L. & Campbell, G. S. (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology* 31: 159-166.

Buishand, T. A. (1977). Stochastic Modeling of Daily Rainfall Sequences. Mededelingen Landbouwhogeschool Wageningen, 211 pp.

Buishand, T. A. (1978). Some remarks on the use of daily rainfall models. *Journal of Hydrology* 36: 295-308.

Burrough, P. A. & McDonnell, R. (1998). Principles of Geographical Information Systems. Oxford University Press, Oxford, 346 pp.

Buys, M. E. L., Fabricius, A. F., Van Den Bergh, P. & Klopper, A. P. J. (1979). Analysis of rainfall in South Africa: Expectancy of monthly rainfall. Department of Agricultural Technical Services, Technical Communication No 148: 33, Pretoria.

Cadet, D. L. & Diehl, B. (1984). Interannual variability of the surface field over the Indian Ocean during the recent decades. *Monthly Weather Review* 112: 21-25.

Cahill, A. T. (2003). Significance of AIC differences for precipitation intensity distributions. *Advances in Water Resources* 26(4): 457-464.

Camberlin, P. (1997). Rainfall anomalies in the source region of the Nile and their connection with the Indian summer monsoon. *Journal of Climate* 10: 1380-1392.

Campbell, G. S. (1990). ClimGen: Climatic data generator. Scientific report, department of Crop and Soil Sciences. Washington State University, Washington.

Cane, M. A., Eshel, G. & Buckland, R. W. (1994). Forecasting maize yield in Zimbabwe with eastern equatorial Pacific sea surface temperature. *Nature* 370: 204-205.

Chin, E. H. (1977). Modeling daily precipitation occurrence process with Markov chain. *Water Resources Research* 13 (6): 949-952.

Clark, C. O., Webster, P. J. & Cole, J. E. (2003). Interdecadal variability of the relationship between the Indian Ocean zonal model and east African coastal rainfall anomalies. *Journal of Climate* 16: 548-553.

Coe, R. & Stern, R. D. (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology* 21: 1024-1030.

Cole, J. A. & Sherrif, J. D. F. (1972). Some single- and multi-site models of rainfall within discrete time increments. *Journal of Hydrology* 17: 97-113.

Colman, A., Davey, M., Harrison, M., Evans, T. & Evans, R. (1996). Multiple regression and discriminant analysis predictors of Jul-Aug-Sep 1996 rainfall in the Sahel and other tropical North African regions. NOAA *Experimental Long-Lead forecast Bulletin 5:2.*

URL:http://nic.fb4.noaa.gov/products/predictions/experimental/bulletin/ *(data viewed-Dec, 2002).*

Conover, W. J. (1971). Practical Nonparametric Statistics. John Wiley & Sons Inc., New York, 462 pp.

Conrad, V. & Pollak, L. W. (1950). Methods in Climatology. Harvard University Press. Massachusetts, Cambridge, 459 pp.

Cpc, (2002). Monthly Atmospheric and SST Indices.
URL: http://www.cpc.ncep.noaa.gov/data/indices *(data viewed-Jan, 2002).*

Cressie, N. (1993). Statistics for Spatial Data. Wiley, New York, 900 pp.

Daly, C., Neilson, R. & Phillips, D. (1994). Statistical topographical model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33: 140-158.

De Jager, J. M., Mottram, R. & Kennedy, J. A. (2001). Research on a Computerised Weather-Based Irrigation Water Management System. WRC report No 581/1/01. Water Research Commission, Pretoria, 227 pp.

De Jong, R., Dumanski, J. & Bootsma, A. (1992). Implications of spatial averaging of weather and soil moisture data for broad scale modeling activities. *Soil Use and Management* 8 (2): 74-79.

Dennett, M. D. J., Rodgers, J. A. & Keating, J. D. H. (1983). Simulation of a rainfall record for a new site of a new agricultural development: an example from northern Syria. *Agricultural Meteorology* 29: 247-258.

Diaz, H. F. & Markgraf, V. (1992). El Niño Historical and Paleoclimatic Aspects of the Southern Oscillation. Cambridge University Press, Cambridge, 476 pp.

Dilley, M. (2000). Reducing vulnerability to climate variability in Southern Africa: Growing role of climate information. *Climate Change* 45: 63-73.

Edelsten, P. R. (1976). A stochastic model of the weather at Hurley in S.E. England. *Meteorological Magazine* 105: 206-214.

Englund, E. J. (1990). A variance of geostatisticians, *Mathematical Geology* 22: 417-455.

Evans, S. P. (1996). SWELTER-Synthetic Weather Estimator for Land Use and Terrestrial Ecosystem Research. Cranfield University, Soil Survey and Land Research Centre, Silsoe Campus, Silsoe, Bedford.

FAO (1994). Agricultural Sector Review and Project Identification, Food and Agricultural Organization of the United Nations, Rome.

Farmer, G. (1988). Seasonal forecasting of the Kenya coast short rains, 1901-1984. *Journal of Climate* 8: 489-497.

Flohn, H. (1987). Rainfall teleconnections in northern and north-eastern Africa. *Theoretical and Applied Climatology* 38: 191-197.

Folland, C. K., Palmer, T. N. & Parker, D. E. (1986). Sahel rainfall and world-wide sea surface temperature. *Nature* 320: 602-607.

Foufoula-Georgion, E. & Lettenmaier, D. P. (1987). A Markov renewal model for rainfall occurrences. *Water Resources Research* 23: 875-884.

Fraedrich, K. & Smith, N. R. (1989). Combining predictive schemes in long-range forecasting. *Journal of Climatology* 2: 291-294.

Franklin, C. (1992). An introduction to geographic information systems: Linking maps to database. *Database* 4: 12-21.

Gabriel, K. R. & Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society* 88: 90-95.

Geen, J. R. (1964). A model for rainfall occurrence. *Journal of the Royal Statistical Society* B 26: 345-353.

Geng, S., Penning De Vries, F. W. T. & Supit, I. (1986). A simple method for generating daily rainfall data. *Agricultural and Forest Meteorology* 36: 363-376.

Genton, M. G. & Furrer, R. (1998). Analysis of rainfall data by simple good sense: Is spatial statistics worth the trouble? *Journal of Geographic Information and Decision Analysis* 2: 12-17.

Glantz, M. H. (2001). Currents of Change: El Niño and La Niña impacts on Climate and Society, 2nd Edition. Cambridge University, New York, 252 pp.

Goddard, L. & Graham, N. E. (1999). The importance of the Indian Ocean for GCM-based climate forecasts over eastern and Southern Africa. *Journal of Geophysical Research* 104 (D16): 19099-19116.

Green, G. C. (1966). The Evaluation of Methods of Rainfall Analysis and the Application to the Rainfall Series of Nelspruit. M.Sc. thesis, University of the Orange Free State, Bloemfontein, 147 pp.

Guan, W., Chamberlan, R. H., Sabol, B. M. & Doering, P. H. (1999). Mapping submerged aquatic vegetation with GIS in the Caloosahatchee Estuary: Evaluation of different interpolation methods. Ecosystem Restoration Department, South Florida Water Management District, West Palm Beach, Florida.

Hackert, E. C. & Hastenrath, S. (1986). Mechanisms of Java rainfall anomalies. *Monthly Weather Review* 114: 745-757.

Hanson, C. L., Cumming, K. A., Woolhiser, D. A. & Richardson, C. W. (1994). Microcomputer Program for Daily Weather Simulation in the Contiguous United States. U.S. Department of Agriculture, Agricultural Research ARS-114, 38 pp.

Hantel, M. & Acs, F. (1998). Physical aspects of the weather generator. *Journal of Hydrology* 212: 393-411.

Hartkamp, A. D., White, J. W. & Hoogenboom, G. (2003). Comparison of three weather generators for crop modeling: A case study for subtropical environments. *Agricultural Systems* 76: 539-560.

Hastenrath, S., Greischar, L., Colon, E. & Gil, A. (1998). A forecast of the February-March 1999 anomalous discharge of the Caroni River, Venezuela. NOAA Experimental Long-Lead Forecast Bulletin 7:4.

URL: http://grads.iges.org/ellfb/ *(data viewed-Jan, 2003).*

Hastenrath, S., Greischar, L. & Van Heerden, J. (1995). Prediction of summer rainfall over South Africa. *Journal of Climate* 8: 1511-1518.

Hastenrath, S., Nicklis, A. & Greischar, L. (1993). Atmospheric-hydrosphere mechanisms of climate anomalies in the western equatorial Indian Ocean. *Journal of Geophysical Research* 98(C11): 20219-20235.

Hay, L. E., McCabe, J. R, Wolock, G. J. & Ayres, M. A. (1991). Simulation of precipitation by weather type analysis. *Water Resources Research* 27: 493-501.

Hayes, M. J., Shapiro, C. A., Walters, D. T., Tamoah, C. F & Francis, C. A. (2000). Standardized Precipitation Index and nitrogen rate effects on crop yields and risk distribution in maize. *Agricultural Ecosystems and Environment* 80: 113-120.

Hayes, M. J., Svoboda, M. D., Wilhite, D. A. & Vanyarkho, O. V. (1999). Monitoring the 1996 drought using the Standardized Precipitation Index. *Bulletin of the American Meteorological Society* 8(3): 429-438.

Hayhoe, H. N. (1998). Relationship between weather variables in observed and WXGEN generated data series. *Agricultural and Forest Meteorology* 90: 203-214.

Haylock, M. & McBride, J. (2001). Notes and correspondence. Spatial coherence and predictability of Indonesian wet season rainfall. *Journal of Climate* 14: 3882-3887.

Heim, R. R. (2002). A review of twentieth century drought indices used in the United States. *American Meteorological Society* 8: 1149-1163.

Heneker, T. M., Lambert, M. F. & Kuczera, G. (2001). A point rainfall model for risk-based design. *Journal of Hydrology* 247: 54-71.

Hintze, J. L. (2000). NCSS-Statistical System for Windows. Number Cruncher Statistical Systems. User's guide. Kaysville, Utah, 961 pp.

Hoshmand, A. R. (1998). Statistical Methods for Environmental and Agricultural Sciences. 2nd Edition. CRC press, Boca Raton, 439 pp.

Hughes, J. P. & Guttorp, P. (1994a). A class of stochastic models for relating synoptic scale atmospheric patterns to regional hydrologic phenomena. *Water Resources Research* 30 (5): 1535-1546.

Hughes, J. P. & Guttorp, P. (1994b). Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology* 33: 1503-1515.

Hutchinson, M. F. & Gessler, P. E. (1994). Splines more than just a smooth interpolator. *Geoderma* 62: 45-67.

INSTAT+ Climatic Guide (2001). The Analysis of Climatic Data. Statistical Services Center, the University of Reading, Whiteknights, Berkshire.

Ison, N. T., Feyerherm, A. M. & Bark, L.D. (1971). Wet period precipitation and the gamma distribution. *Journal of Applied Meteorology* 10: 658-665.

Janowiak, J. (1988). An investigation of interannual variability in Africa. *Journal of Climate* 1: 240-255.

Jayawardena, A. W. & Lau, W. H. (1990). Homogeneity tests for rainfall data. *Hong Kong Engineer* September: 22- 25.

Johnson, G. L., Hanson, C. L., Hardegree, S. P. & Ballard, E. B. (1996). Stochastic weather simulation: Overview and analysis of two commonly used models. *Journal of Applied Meteorology* 35: 1878-1896.

Jones, P. G. & Thornton, P. K. (1993). A rainfall generator for agricultural applications in the tropics. *Agricultural and Forest Meteorology* 63: 1-19.

Jones, P. G. & Thornton, P. K. (1997). Spatial and temporal variability of rainfall related to a third-order Markov model. *Agricultural and Forest Meteorology* 86: 127-138.

Jones, P. G. & Thornton, P. K. (2000). Mark Sim: Software to generate daily weather data for Latin America and Africa. *Agronomy Journal* 92: 445-453.

Joubert, A. (1998). El Niño, La Niña and summer rainfall variability over South Africa. *South African Journal of Science* 94: 1.

Jury, M. R. (1993). A preliminary study of climatological associations and characteristics of tropical cyclones in the SW Indian Ocean. *Meteorology and Atmospheric Physics* 51: 101-115.

Jury, M. R. (1996). Regional teleconnection patterns associated with summer rainfall over South Africa, Namibia and Zimbabwe. *International Journal of Climatology* 16: 135-153.

Jury, M. R., Mc Queen, C. & Levey, K. (1994). SOI and QBO signals in the African region. *Theoretical and Applied Climatology* 50: 103-115.

Jury, M. R., Mulenga, H. M. & Mason, S. J. (1999). Exploratory long-range models to estimate summer climate variability over southern Africa. *Journal of Climate* 12: 1892-1899.

Jury, M. R., Mulenga, H. M., Mason, S. J. & Brandao, A. (1997). Development of an Objective Statistical System to Forecast Summer Rainfall over Southern Africa. WRC Report No 672/1/97. Water Research Commission, Pretoria, 45 pp.

Jury, M. R., Parker, B. A., Raholijao, N. & Nassor, A. (1995). Variability of summer rainfall over Madagascar: Climatic determinants at interannual scales. *International Journal of Climatology* 15: 1323-1332.

Kane, R. P. (1997). On the relationship of ENSO with rainfall over different parts of Australia. *Australian Meteorological Magazine* 46(1): 39-49.

Katz, R. W. (1974). Computing probabilities associated with the Markov chain model for precipitation. *Journal of Applied Meteorology* 13: 953-954.

Katz, R. W. (1977). Precipitation as a chain dependent process. *Journal of Applied Meteorology* 16: 671-676.

Katz, R. W. (1996a). Use of conditional stochastic models to generate climate change scenarios. *Climatic Change* 32: 237-255.

Katz, R. W. (1996b). Mixing of stochastic processes: Application to statistical downscaling. *Climate Research* 7: 185-193.

Kiladis, G. N. & Diaz, H. F. (1989). Global climatic anomalies associated with extremes in the Southern Oscillation. *Journal of Climatology* 2: 1069-1090.

Kittle, T. G. F., Rosenbloom, N. A., Painter, T. H., Schimel, D. S. & Vemap Modeling Participants (1995). The VEMAP States ecosystem / vegetation sensitivity to climate change. *Journal of Biogeography* 22: 857-862.

Klopper, E. (1999). The use of seasonal forecasts in South Africa during the 1997 / 98 rainfall season. *Water SA* 25: 311-316.

Kottagoda, N. T. & Horder, M. A. (1980). Daily flow model based on rainfall occurrences using pulses and a transfer function. *Journal of Hydrology* 47: 215-234.

Kripalani, R. H. & Kulkarni, A. (1997). Climatic impact of El Niño / La Niña on the Indian monsoon: A new perspective. *Weather* 52: 39-46.

Krishnamurthy, V. & Goswami, B. N. (2000). Indian-Monsoon-ENSO relationship on interdecadal timescale. *Journal of Climate* 13: 579-595.

La Mer, D. (1994). El Niño and Climate Prediction. University Corporation for Atmospheric Research pursuant to National Ocean and Atmospheric Administration (NOAA), No 3.

Lamb, P. J. & Peppler, R. A. (1987). North Atlantic oscillation: Concept and an application. *Bulletin of American Meteorological Society* 68: 1218-1225.

Landman, W. A. & Mason, S. J. (1999a). Operational long-lead prediction of South African rainfall using canonical correlation analysis. *International Journal of Climatology* 19: 1073-1090.

Landman, W. A. & Mason, S. J. (1999b). Change in the association between Indian Ocean SSTs and summer rainfall over South Africa and Namibia. *International Journal of Climatology* 19: 1477-1492.

Larsen, G. A. & Pense, R. B. (1982). Stochastic simulation of daily climatic data for agronomic models. *Agronomy Journal* 75: 510-514.

Latif, M., Dommenget, D., Dima, M. & Grotzner, A. (1999). The role of Indian Ocean Sea Surface Temperature in forcing East African rainfall anomalies during December-January 1997/98. *Journal of Climate* 12: 3497-3504.

Lavery, B., Joung, G. & Nicholls, N. (1997). An extended high-quality historical rainfall dataset for Australia. *Australian Meteorological Magazine* 46: 27–38.

Legler, D. M., Bryant, K. J. & O'Brien, J. J. (1999). Impact of ENSO-related climate anomalies on crop yields in the U.S. *Climate Change* 42: 351-375.

Lindesay, J. A. (1988). South African rainfall, the Southern Oscillation and a Southern hemisphere semi-annual cycle. *Journal of Climatology* 8: 17-30.

Lindesay, J. A., Harrison, M. S. J. & Haffner, M. P. (1986). The Southern Oscillation and South African rainfall. *South African Journal of Science* 82: 196-198.

Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2001). Geographic Information System and Science. John Wiley & Sons, New York, 454 pp.

Lourens, U. W. (1995). A System for Drought Monitoring and Severity Assessment. PhD Thesis, University of the Orange Free State, Bloemfontien, 161 pp.

Lowry, W. P. & Guthrie, D. (1968). Markov chains of order greater than one. *Monthly Weather Review* 96: 798-801.

Makarau, A. (2000). Rainfall predictors used in Southern Africa. *Southern Africa Regional Climate Outlook Forum.* Report presented at SASAS Conference, Gaborone, Botswana.

Makarau, A. & Jury, M. R. (1997). Predictability of Zimbabwe summer rainfall. *International Journal of Climatology* 17: 1421-1432.

Martin, R., Washington, R. & Downing, T. E. (1999). Seasonal maize forecasting for South Africa and Zimbabwe derived from an Agroclimatological model. A report by Environmental Change Unit and School of Geography, University of Oxford, Oxford.

Mason, S. J. (1995). Sea-surface temperature - South African rainfall associations, 1910-1989. *International Journal of Climatology* 15: 119-135.

Mason, S. J. (1998). Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *International Journal of Climatology* 18: 147-164.

McCaskill, M. R. (1990a). TAMSIM – a program for preparing meteorological records for weather-driven models. CSIRO, Division of Tropical Crops and Pastures, *Tropical Agronomy Technical Memorandum* No. 65.

McCaskill, M. R. (1990b). An efficient method for generation of full climatological records from daily rainfall. *Australian Journal of Agricultural Research* 41: 595-602.

McKee, T. B., Doesken, N. J. & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In: R. E. Hallgren (Eds), *8th Conference on Applied Climatology. American Meteorological Society*, Anaheim California. pp 179-186.

McNeill, L., Brandao, A., Zucchini, W. & Joubert, A. (1993). Interpolation of the Daily Rainfall Model. WRC Report No 305/1/94. Water Research Commission, Pretoria, 164 pp.

Mearns, L. O., Rosenzweig, C. & Goldberg, R. (1996). The effect of change in daily and interannual climatic variability on CERES-wheat yield: Sensitivity study. *Climatic Change* 32: 257-292.

Mearns, L. O., Rosenzweig, C. & Goldberg, R. (1997). Mean and variance changes in climate scenarios: Methods, agricultural applications and measures of uncertainty. *Climatic Change* 35: 367-396.

Meneti, M., Huygen, J., Azzali, S. & Berkhout, J. A. A. (1990). Early warning on agricultural production with satellite data and simulation model in Zambia. In: J. P. Gastellu-Etchegorry (Eds), *Satellite Remote Sensing for Agricultural Project.* World Bank Technical Paper No. 128. World Bank Washington DC. pp 187-215.

Mielke, P. W. (1973). Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology* 10: 275-280.

Mohammedberhan, N. (1998). Satellite Based Rainfall Estimation over Eritrea. M.Sc. thesis, Department of Meteorology, University of Reading, Whiteknights, Berkshire, 56 pp.

Montecinos, A., Diaz, A. & Aceituno, P. (2000). Seasonal diagnostic and predictability of rainfall in subtropical South America based on tropical Pacific SST. *Journal of Climatology* 13: 746-758.

Mutai, C. C. & Ward, M. N. (2000). East African rainfall and the tropical circulation / convection on interseasonal to interannual timescales. *Journal of Climate* 13: 3915-3939.

Mutai, C. C., Ward, M. N. & Colman, A. W. (1998). Towards the prediction of the East Africa short rains based on the Sea Surface Temperature - Atmospheric Coupling. *International Journal of Climatology* 18: 975-997.

Myers, E. D. (1994). Spatial interpolation: An overview. *Geoderma* 62: 17-28

Naylor, R. L., Falcon, W. P., Rochberg, D. & Wada, N. (2001). Using El Niño/ Southern Oscillation climate data to predict rice production in Indonesia. *Climatic Change* 50: 225-265.

Nicholls, N. (1981). Air-sea interaction and the possibility of long-range weather prediction in the Indonesian archipelago. *Monthly Weather Review* 109: 2435-2443.

Nicholls, N. (1989). Sea surface temperature and Australian winter rainfall. *Journal of Climate* 2: 965-973.

Nicholls, N. (1991). Advances in long-term weather forecasting, In: R. C. Muchow & J. A. Bellamy (Eds), *Climatic Risk in Crop Production: Models and Management for the Semi-Arid Tropics and Sub-Tropics*. CAB International, Wallingford. pp 427-444.

Nicholson, S. E. & Entekhabi, D. (1987). Rainfall variability in equatorial and Southern Africa: Relationship with sea surface temperature along the southwestern coast of Africa. *Journal of Climate and Applied Meteorology* 26: 561-578.

Nicholson, S. E. & Kim, J. (1997). The relationship of the El Niño-Southern Oscillation to African rainfall. *International Journal of Climatology* 17: 117-135.

Nicholson, S. E., Leposo, D. & Grist, J. (2001). The relationship between El Niño and drought over Botswana. *Journal of Climate* 14: 323-335.

Nicks, A. & Gander, G. (1990). CLIGEN. USDA Agricultural Research Services (ARS) lab in Durant, Oklahoma.

NMSA (National Meteorological Services Agency of Ethiopia), 1996. Assessment of Drought in Ethiopia, a scientific Report. Addis Ababa, 259 pp.

Ogallo, L. A. (1988). Relationship between seasonal rainfall in East Africa and Southern Oscillation. *Journal of Climatology* 8: 34-43.

Ogallo, L. A. (1994). Validity of the ENSO-related impact in Eastern and Southern Africa. In: M. H. Glanz (Eds), *Usable Science: Food Security, Early Warning and El Niño*. UNEP (Nairobi) and NCAR (Boulder, CO). pp 179-184.

Palmer, T. N. (1986). Influence of the Atlantic, Pacific, and Indian Oceans on Sahel rainfall. *Nature* 322: 251-253.

Pathack, B. M. (1993). Modulation of South African Summer Rainfall by Global Climatic Processes. PhD Thesis, Oceanography department, University of Cape Town, Cape Town, 214 pp.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mester, O., Szentimrey, T., Salinger, J., Forland, E., Hanssen-Bauer, I., Alexandersson, H., Jones P, & Parker, D. (1998). Homogeneity adjustments of *in situ* atmospheric climate data: A review. *International Journal of Climatology* 18: 1493–1517.

Phien, H. N. & Warakittimalee, S. (1981). Simulation of daily rainfall sequences using Markov chain. *Water SA* 7(4): 193-202.

Philander, S. G. (1990). El Niño, La Niña and the Southern Oscillation. Academic Press, Inc. San Diego, California, 293 pp.

Philip, G. M. & Watson, D. F. (1982). A precise method for determining contour surfaces. *Australian Petroleum Exploration Association Journal* 22: 205-212.

Pittock, A. B. (1980). Patterns of climatic variation in Argentina and Chile. *Monthly Weather Review* 108: 1347-1361.

Prasad, K. D. & Singh S. V. (1996). Seasonal variations of the relationship between some ENSO parameters and Indian rainfall. *International Journal of Climatology* 16: 923-933.

Racsko, P., Szeidl, L. & Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological Modeling* 57: 27-41.

Raman, C. V. R. (1974). Analysis of commencement of Monsoon rains over Maharashtra State for agricultural planning. Scientific Report 216, India Meteorological Department, Poona.

Rasmusson, E. M. & Carpenter, T. H. (1982). Variations in the tropical sea surface temperature and surface wind fields associated with the Southern Oscillation El Niño. *Monthly Weather Review* 110: 354-384.

Rasmusson, E. N. & Carpenter, T. H. (1983). The relationship between eastern Pacific Sea surface temperature and rainfall over India and Sri Lanka. *Monthly Weather Review* 111: 517-528.

Reverdin, G., Cadet, D. L. & Gutzler, D. (1986). International displacements of convection and surface circulation over the equatorial Indian Ocean. *Quarterly Journal of the Royal Meteorological Society* 112: 46-67.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research* 17: 182-190.

Richardson, C. W. (1982). A comparison of three distributions for the generation of daily rainfall amounts. In: V. P. Singh (Eds), *Statistical Analysis of Rainfall and Runoff*. Water Resources Publications, Littleton, Colorado. pp 67-78.

Riha, S. J., Wilks, D. S. & Simoens, P. (1996). Impact of temperature and precipitation variability on crop model predictions. *Climatic Change* 32: 293-311.

Ripley, B. (1980). Spatial Statistics. John Wiley & Sons, London, 252 pp.

Roldan, J., & Woolhiser, D. A. (1982). Stochastic daily precipitation models. 1. A comparison of occurrence processes. *Water Resources Research* 18: 1451-1459.

Ropelewski, C. F. & Halpert, M. S. (1987). Global and regional scale precipitation pattern associated with El Niño/Southern Oscillation. *Monthly Weather Review* 115: 1606-1626.

Ropelewski, C. F. & Halpert, M. S. (1996). Quantifying Southern Oscillation-precipitation relationships. *Journal of Climate* 9: 1043-1059.

SG-2000 (Sasakawa Global), 1997. Annual Report Crop Season 1996 (unpublished report). Ministry of Agriculture, Asmara

Schmidt, G. M., Smajstrla, A. G. & Zazueta, F. S. (1997). Long-term variability of monthly total precipitation. *American Society of Agricultural Engineers - Transactions* 40(4): 1029-1039.

Seed, A. W. (1992). The Generation of a Spatially Distributed Daily Rainfall Database for Various Weather Modification Scenarios. WRC Report No 373/1/92. Water Research Commission Pretoria, 43 pp.

Semenov, M. A. & Barrow, E. M. (1997). Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change* 35: 397-414.

Servain, J., Busalacchi A., Moura A., Mcphaden, M. J., Reverdin G., Vanna M. & Zebiak, S. (1998). A Pilot Research Moored Array in the Tropical Atlantic (PIRATA). *Bulletin of the American Meteorological Society* 79: 2019-2031.

Shaw, A. B. (1987). An analysis of the rainfall regimes on the coastal region of Guyana. *Journal of Climatology* 7: 291-302.

Singh, S. V. & Kripalani, R. H. (1986). Analysis of persistence in daily monsoon rainfall over India. *Journal of Climatology* 6: 625-639.

Siska, P. P. & Hung, I. K. (2000). Data quality on applied spatial analysis. In: F. A. Schoolmaster (Eds), *Papers and Proceedings of the Applied Geography Conferences* No 23 PP 199-205.

Siska, P. P. & Hung, I. K. (2001). Assessment of kriging accuracy in the GIS environment. *Proceedings of the ESRI users conference 2001.* **URL:**http://www.esri.com/library/usersconf/proc01/professional/*(data viewed May, 2002).*

Siska, P. P. & Maggio, R. C. (1997). The role of relief dissectivity in distribution of Kriging errors from digital elevation modeling. In: F. A. Schoolmaster (Eds), *Papers and Proceedings of the Applied Geography Conferences* No 20 PP 186-194.

Smith, I. (1994). Indian Ocean sea surface temperature patterns and Australian winter rainfall. *International Journal of Climatology* 14: 287-305.

Smith, R. E. & Schreiber, H. A. (1973). Point processes of seasonal thunderstorm rainfall, I. Distributions of rainfall events. *Water Resources Research* 9: 871-884.

Smith, R. E. & Schreiber, H. A. (1974). Point processes of seasonal thunderstorm rainfall, II. Rainfall depth probabilities. *Water Resources Research* 10: 418-423.

Soltani, A., Latifi, N. & Nasiri, M. (2000). Evaluation of WGEN for generating long term weather data for crop simulations. *Agricultural and Forest Meteorology* 102: 1-12.

Srikanthan, R. & McMahon, T. A. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences* 5(4): 653-670.

Steele, R., Torrie, J. & Dickey, D. (1997). Principles and Procedures of Statistics: A Biometrical Approach. 3rd Edition, McGraw-Hill, New York, 666 pp.

Stern, R. D. & Coe, R. (1982). The use of rainfall models in agricultural planning. *Agricultural Meteorology* 26: 35-50.

Stern, R. D. & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society Series* A 147: 1-34.

Stewart, J. I. (1994). Utilising rainfall probabilities referenced to crop seed germination dates to assess risks and develop response farming strategies for Asmara, Eritrea. Workshops for the famine mitigation activities support activity of USAID, U.S. Foreign Disaster Assistance, Asmara.

Stöckle, C. O., Donatelli, M. & Nelson, R. (2003). CropSyst: A cropping systems simulation model. *European Journal of Agronomy* 18: 289-307.

Stoecjenius, T. (1981). Interannual variations of tropical precipitation patterns. *Monthly Weather Review* 109: 1233-1247.

Stone, R. (1996). SOI phase relationships with rainfall in eastern Australia. *International Journal of Climatology* 12: 625-636.

Sun, L., Semazzi, F. H. M., Giorgi, F. & Ogallo, L. (1999a). Application of the NCAR regional climate model to eastern Africa. Part I: Simulation of the short rains of 1988. *Journal of Geophysical Research* 104(D6): 6529-6548.

Sun, L., Semazzi, F. H. M., Giorgi, F. & Ogallo, L. (1999b). Application of the NCAR regional climate model to eastern Africa. Part II: Simulation of interannual variability of short rains. *Journal of Geophysical Research* 104(D6): 6549-6562.

Suppiah, R. (1989). Relationship between the Southern Oscillation and the rainfall of Sri Lanka. *International Journal of Climatology* 9: 601-618.

Thiaw, W. & Barnston, A. (1999). CCA forecast for Sahel rainfall in Jul-Aug-Sep 1999. NOAA *Experimental Long-Lead forecast Bulletin 8:1.*

URL: http://grads.iges.org/ellfb/ *(data viewed-Jan, 2003).*

Thyer, M. & Kuczera, G. (2000). Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model. *Water Resources Research* 36: 3301-3310.

Todorovic, P. & Woolhiser, D. A. (1975). A stochastic model of n-day precipitation. *Journal of Applied Meteorology* 14: 17-24.

Tomczak, M. (1998). Spatial interpolation and its uncertainty using automated anisotropic Inverse Distance Weighting (IDW)-cross-validation / Jackknife approach. *Journal of Geographic Information and Decision Analysis* 2(2): 18-30.

Troup, A. J. (1965). The Southern Oscillation. *Quarterly Journal of the Royal Meteorological Society* 91: 490-506.

Tyson, P. D. (1987). Climatic Change and Variability in Southern Africa. Oxford University Press, Cape Town, 220 pp.

Van Buskirk, R. & Araia, A. (1994). Rainfall Statistics of Eritrea. Environmental Information Collection and Analysis Project - Department of Physics, University of Asmara, Asmara.

Van Heerden, J., Terblanche, D. E. & Schulze, G. C. (1988). The Southern Oscillation and South African summer rainfall. *Journal of Climatology* 8: 577-597.

Walker, N. D. (1990). Links between South African summer rainfall and temperature variability of the Agulhas and Benguela Current Systems. *Journal of Geophysical Research* 95: 3297-3319.

Walker, S., Mukhala, E., Van Den Berg, W. J. & Manley, C. R., 2001. Assessment of communication and use of climate Outlooks and development of scenarios to promote food security in the Free State province of South Africa. Final Report, DMCH / WB / NOAA / OGP 6.02/31/1. 162 pp.

Wallis, T. W. R. & Griffiths, J. F. (1995). An assessment of the Weather Generator (WXGEN) used in the Erosion Productivity Impact Calculator (EPIC). *Agricultural and Forest Meteorology* 73: 115-133.

Wang, B., Wu, R. & Fu, W. (2000). Pacific-East Asia teleconnection: How does ENSO affect east Asian climate? *Journal of Climate* 13: 1517-1536.

Ward, M. N. (1998). Diagnosis and short-lead time prediction of summer rainfall in tropical North Africa at interannual and multidecadal timescales. *Journal of Climate* 11: 3167-3191.

Ward, M. N. & Folland, C. K. (1991). Prediction of seasonal rainfall on North Nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology* 11: 711-744.

White, D. H. & O'Meagher, B. (1995). Coping with exceptional droughts in *Australia. Drought Network News* 7(2): 13-17.

Wilks, D. S. (1989). Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research* 25: 1429-1439.

Wilks, D. S. (1992). Adapting stochastic weather generation algorithms for climate change studies. *Climatic Research* 22: 67-84.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology* 210: 178-191.

Wilks, D. S. (1999a). Interannual variability and extreme value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology* 93: 153-169.

Wilks, D. S. (1999b). Multi-site downscaling of daily precipitation with a stochastic weather generator. *Climate Research* 11: 125-136.

Wilks, D. S. (1999c). Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agricultural and Forest Meteorology* 96: 85-101.

Wilks, D. S. (2002). Realization of daily weather in forecasting seasonal climate. *Journal of Hydrometeorology* 3: 195-207.

Wilks, D. S. & Wilby, R. L. (1999). The weather generation game: A review of stochastic weather models. *Progress in Physical Geography* 23 (3): 329-357.

Willmott, C. J. (1981). On the validation of models. *Physical Geography* 2: 184-194.

Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin American Meteorological Society* 63: 1309-1313.

Woolhiser, D. A. & Pegram, G. G. S. (1979). Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *Journal of Applied Meteorology* 18: 34-42.

Woolhiser, D. A. & Roldan, J. (1982). Stochastic daily precipitation models: II. A comparison of distributions of amounts. *Water Resources Research* 18: 1461-1468.

Yun, J. I. (2003). Predicting regional rice production in South Korea using spatial data and crop-growth modeling. *Agricultural Systems* 77 (1): 23-38.

Zucchini, W. & Adamson, P. T. (1984). The Occurrence and Severity of Droughts in South Africa. WRC Report No 91/1/84. Water Research Commission, Pretoria, 237 pp.

Zucchini, W., Adamson, P. & McNeill, L. (1992). A model of Southern African rainfall. *South African Journal of Science* 88: 103-107.

Zurbenko, I., Porter, P. S., Rao, S. T., Ku, J. Y., Gui, R. & Eskridge, R. E. (1996). Detecting discontinuities in time series of upper air data: Development and demonstration of an adaptive filter technique. *Journal of Climate* 9: 3548-3560.

APPENDICES

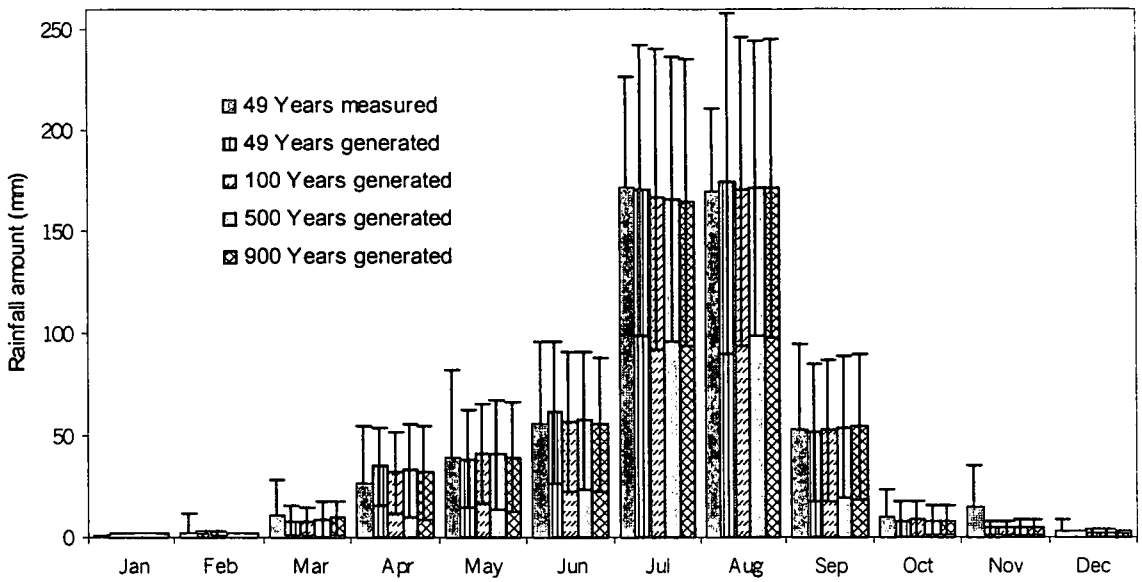
Appendix A Listing of BASIC code computer programme for the developed stochastic model.

```
10    DIM n (365), pww (365), pdw (365), m (365), r (365), gr (365)
20    PRINT
30    REM File "seeds" contains random seeds for the
40    REM random number generator
50    OPEN "seeds" FOR INPUT AS #1
60    OPEN "asmara1" FOR INPUT AS #2
70    OPEN "rainfall" FOR OUTPUT AS #3
80    REM s1 to s6 are site dependent "scale" parameters >= 0.
90    REM i2 and i3 are the days of year on which the rainy
100   REM season starts and ends.
110   REM m(i) >= 0 are site rainfall dependent parameters
120   INPUT #2, o$
130   PRINT o$
140   INPUT #2, i1, i3
150   FOR i = 1 TO 365
160   INPUT #2, n(i), m(i)
170   NEXT i
180   INPUT #2, s1, s2, s3, s4, s5, s6
190   i2 = i1 + 1: i4 = i3 + 1
200   FOR i = 1 TO i1
210   r(i) = s1 * m(i) + s2
220   pdw (i) = s3 * m(i) + s4
230   pww (i) = s5 * m(i) + s6
240   NEXT i
250   FOR i = i4 TO 365
260   r(i) = s1 * m(i) + s2
270   pdw (i) = s3 * m(i) + s4
280   pww (i) = s5 * m(i) + s6
290   NEXT i
300   INPUT #2, s1, s2, s3, s4, s5, s6
310   FOR i = i2 TO i3
```

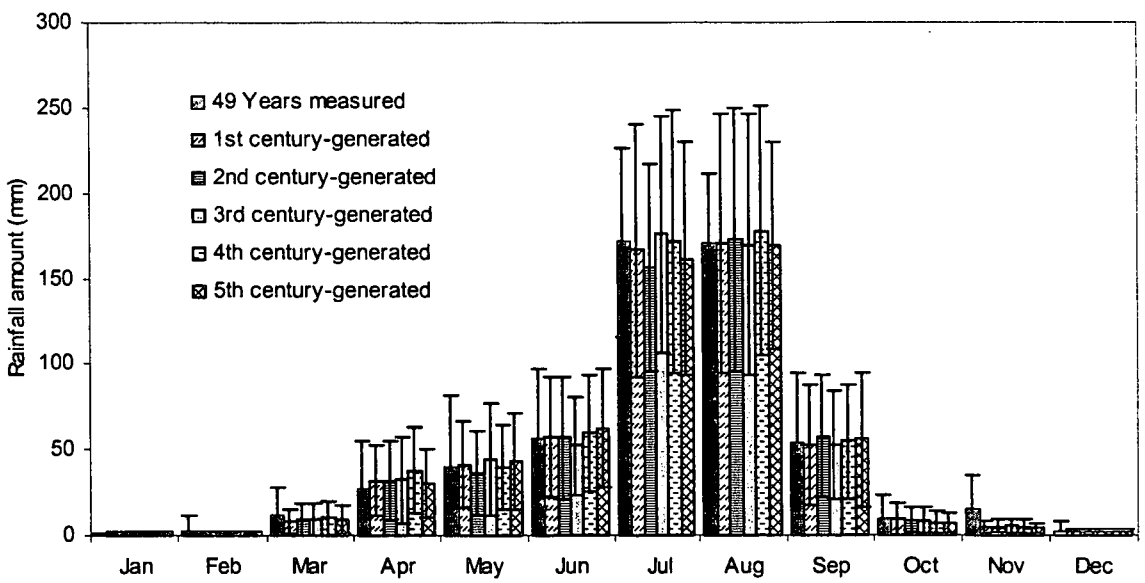
```
320   r(i) = s1 * m(i) + s2
330   pdw (i) = s3 * m(i) + s4
340   pww (i) = s5 * m(i) + s6
350   NEXT i
360   FOR i = 1 TO 365
370   IF (pdw (i) < 1) THEN GOTO 390
380   PRINT "pdw > 1"
390   IF (pdw i) > 0) THEN GOTO 410
400   PRINT "pdw < 0"
410   IF (pww (i) < 1) THEN GOTO 430
420   PRINT "pww > 1"
430   IF (pww (i) > 0) THEN GOTO 450
440   PRINT "pww < 0"
450   NEXT i
460   p = pdw (365)
470   REM The above implies that day 365 which precedes our
480   REM generated record is assumed to be dry
490   REM PRINT p
500   ne = 0
510   INPUT #1, xr
520   y = RND (xr)
530   ind = 1
540   IF (y < p) THEN GOTO 560
550   ind = 0
560   INPUT "required no of years in generated record"; ng
570   FOR ny = 1 TO ng
580   FOR t = 1 TO 365
590   ne = ne + 1
600   IF (ne < 195) THEN GOTO 640
610   ne = 1
620   CLOSE #1
630   OPEN "seeds" FOR INPUT AS #1
640   IF (ind = 0) THEN GOTO 720
650   INPUT #1, xr
660   y = RND (xr)
670   IF (y < pww(t)) THEN GOTO 700
680   ind = 0
```

```
690   GOTO 830
700   ind = 1
710   GOTO 830
720   ne = ne + 1
730   IF (ne < 195) THEN GOTO 770
740   ne = 1
750   CLOSE #1
760   OPEN "seeds" FOR INPUT AS #1
770   INPUT #1, xr
780   y = RND (xr)
790   IF (y < pdw(t)) THEN GOTO 820
800   ind = 0
810   GOTO 830
820   ind = 1
830   IF (ind = 1) THEN GOTO 860
840   PRINT #3, ny; t; 0
850   GOTO 950
860   ne = ne + 1
870   IF (ne < 195) THEN GOTO 900
880   CLOSE #1
890   OPEN "seeds" FOR INPUT AS #1
900   INPUT #1, xr
910   y = RND (xr)
920   gr(i) = r(i) * (-LOG(y))
930   IF (gr(i) < .1) THEN GOTO 950
940   PRINT #3, ny; i; USING "####.#"; gr(i)
950   NEXT t
960   NEXT ny
970   PRINT #3, 999; 0; 0
980   STOP
```

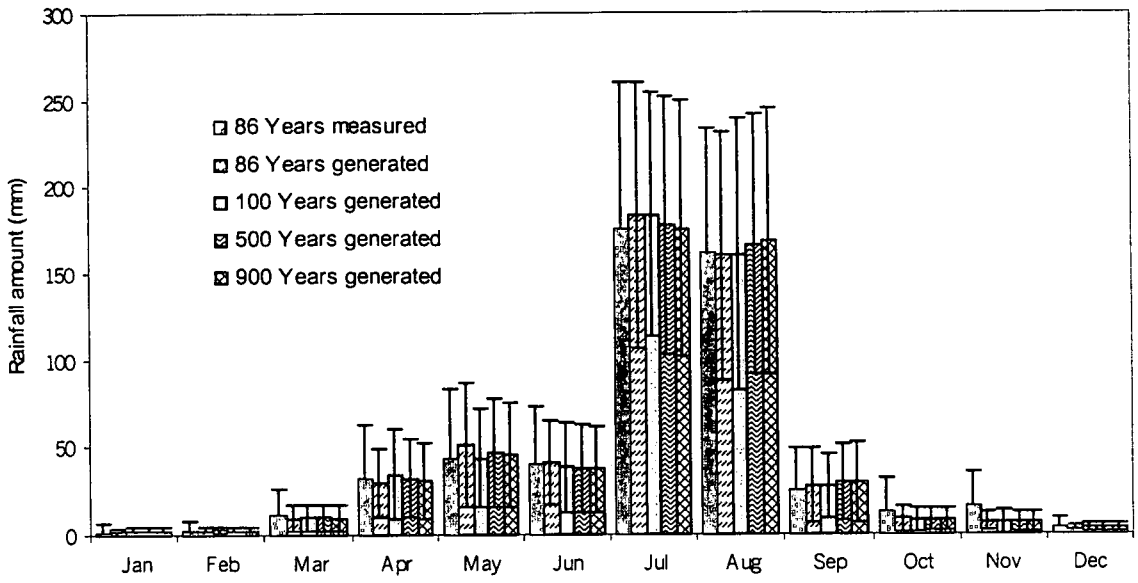
Appendix B Histogram comparison of the mean annual rainfall for measured and generated data for different generated years and centuries.



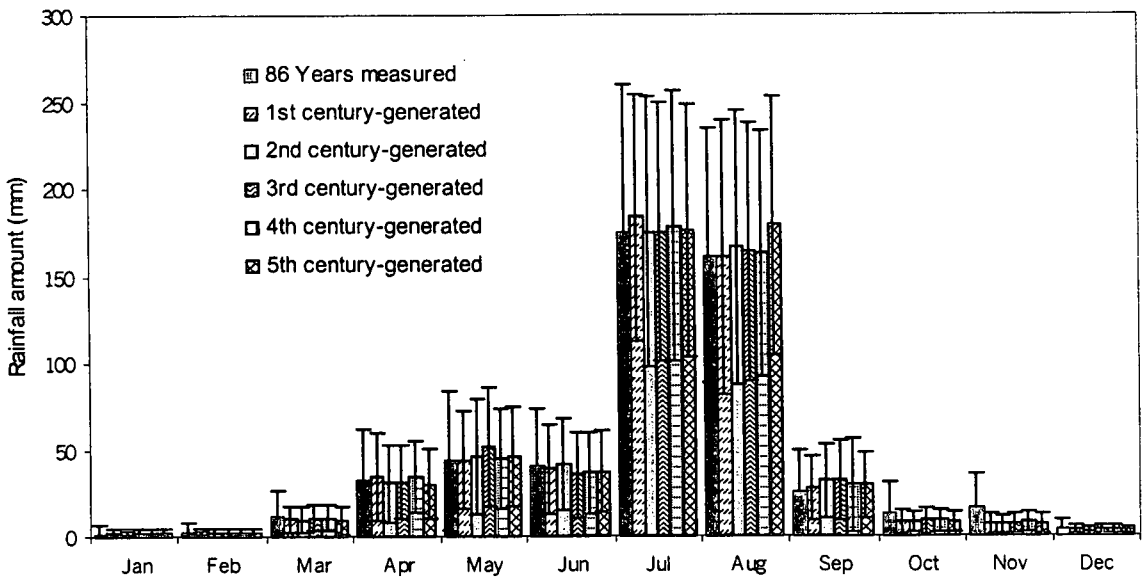
B-i) Histogram comparison of the mean annual rainfall for measured and generated data for Mendefera for different generated centuries.



B-ii) Histogram comparison of the mean annual rainfall for measured and generated data for Mendefera for different generated centuries.



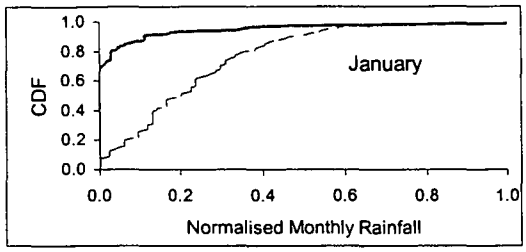
B-iii) Histogram comparison of the mean annual rainfall for measured and generated data for Asmara for different generated years.

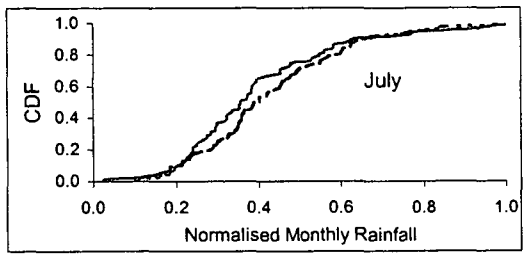


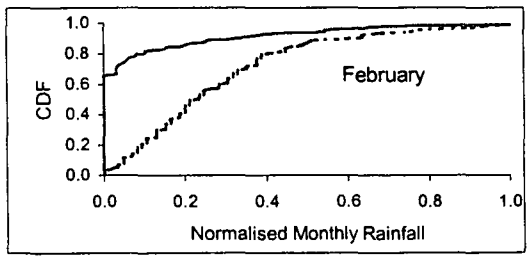
B-iv) Histogram comparison of the mean annual rainfall for measured and generated data for Asmara for different generated centuries.

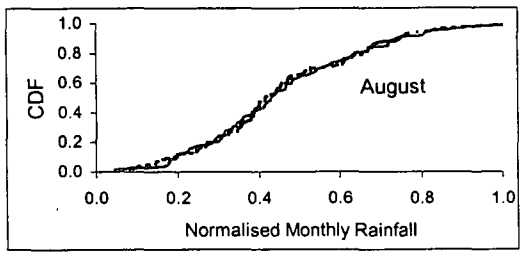
Appendix C **Probability of non-exceedence as a function of ranked precipitation for measured (___) and generated (....) rainfall amount for each month.**

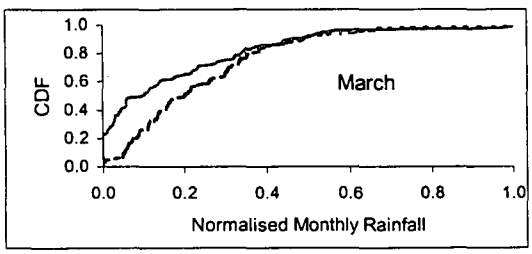
C-i) **Asmara station.**

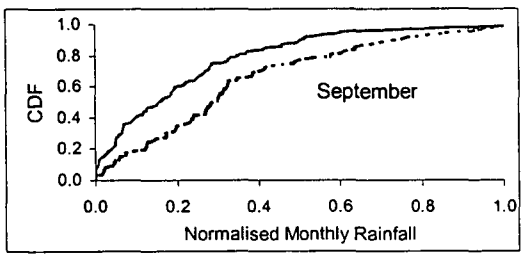


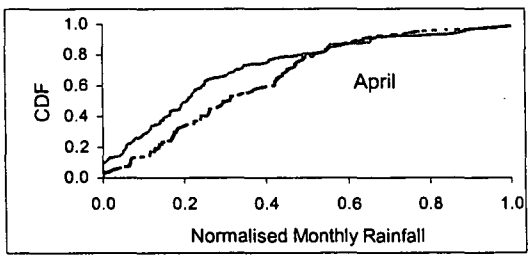


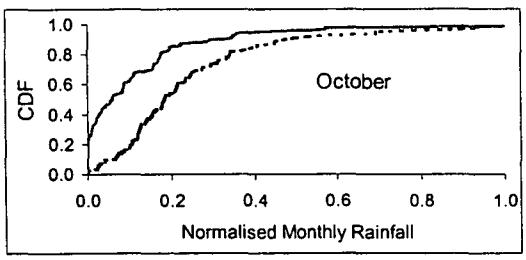


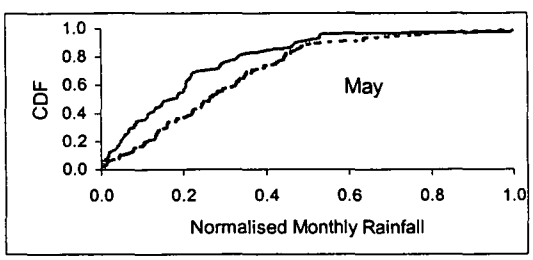


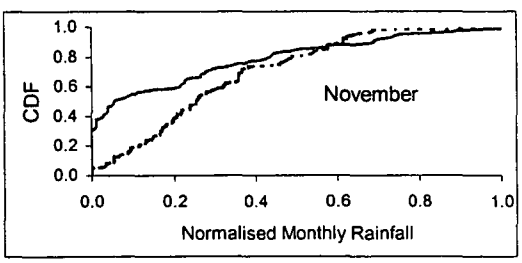


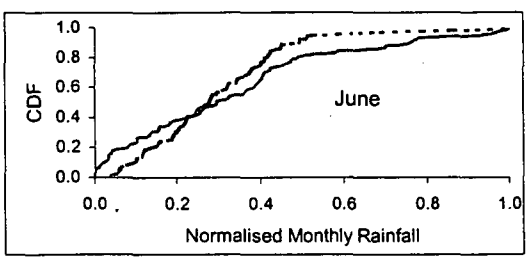


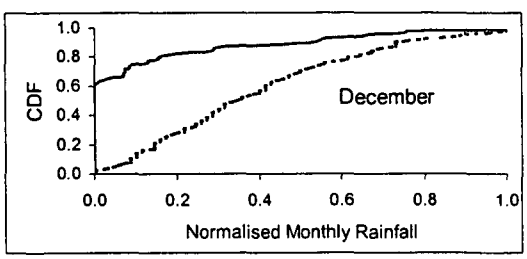




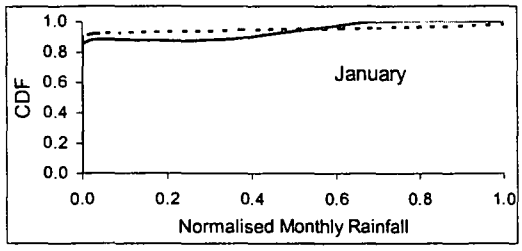


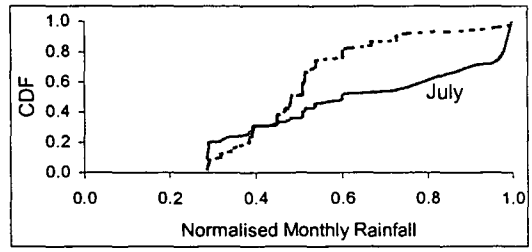


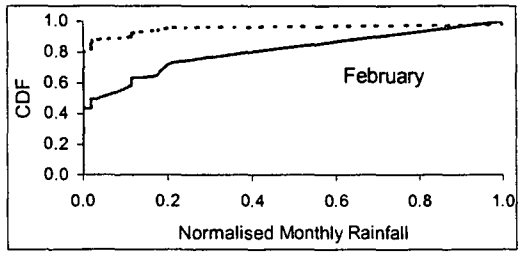


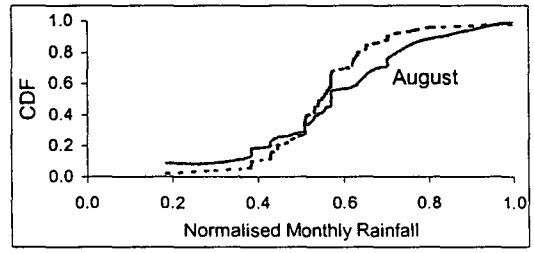


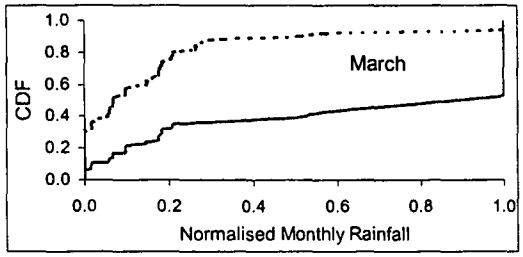
C-ii) Mendefera station.

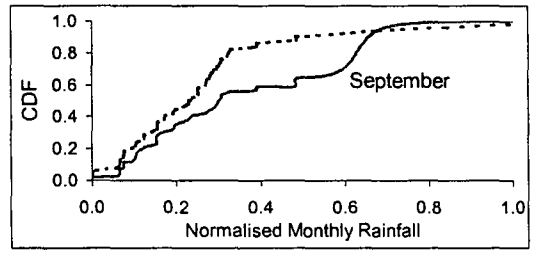


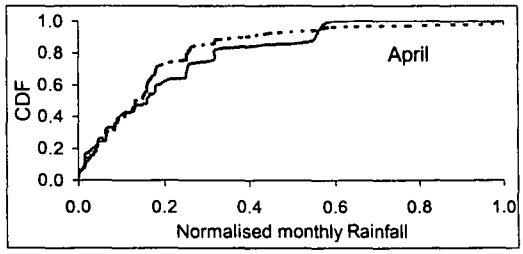


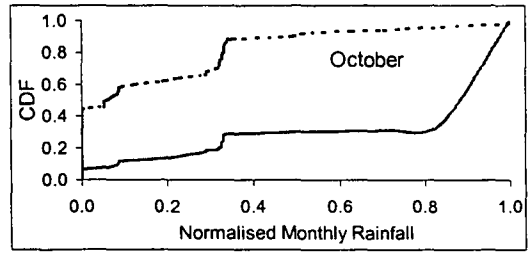


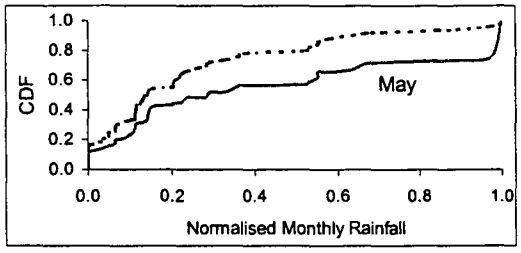


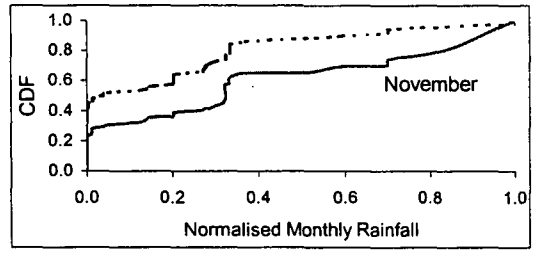


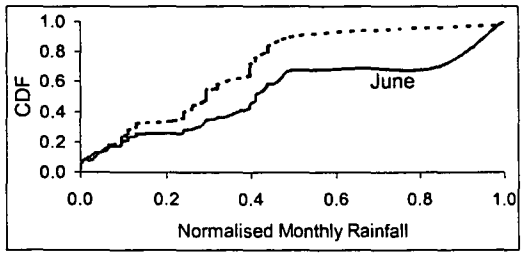


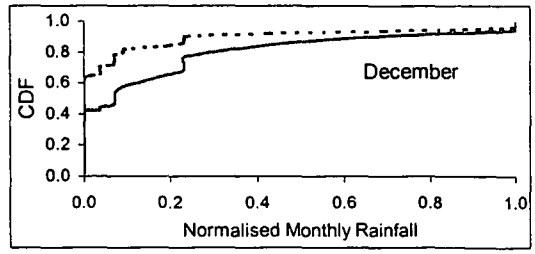












Appendix D Median, mean, skewness and coefficient of variation of the historical and generated data.

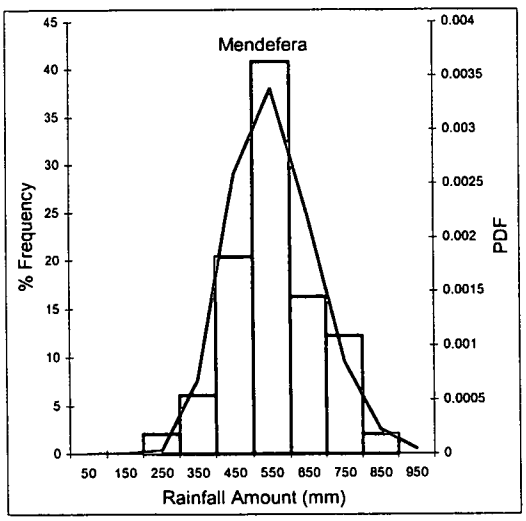
D-i) Asmara station (n = 88).

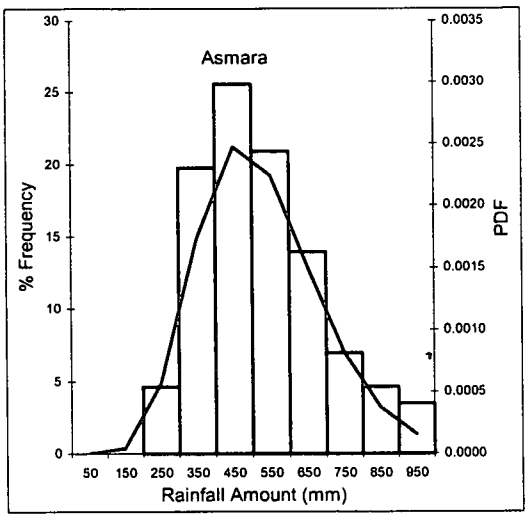
Parameters		Mean		Median		Coefficient of Variation		Skewness	
		Actual	Generated	Actual	Generated	Actual	Generated	Actual	Generated
Month	January	1.66	1.94	0.00	1.70	303.80	80.75	4.645	1.333
	February	2.52	2.38	0.00	1.95	225.20	77.83	2.864	1.261
	March	11.82	9.53	6.65	8.00	120.80	76.74	1.650	1.286
	April	32.57	29.82	24.70	26.80	92.06	66.08	1.138	0.690
	May	43.35	51.43	36.30	46.25	93.46	69.97	1.736	0.995
	June	40.48	41.62	36.20	39.40	80.50	58.14	0.787	1.244
	July	175.20	183.00	157.00	169.40	48.33	41.91	1.074	0.670
	August	161.30	159.90	151.40	146.60	45.11	44.83	0.425	0.390
	September	24.77	27.83	18.70	24.10	98.38	75.57	1.423	0.841
	October	13.03	8.56	7.00	6.60	141.90	82.96	2.818	1.822
	November	15.75	7.16	4.35	6.20	126.90	69.17	1.206	0.725
	December	2.99	2.72	0.00	2.60	195.10	65.23	2.189	0.565
	Annual	525.40	525.90	497.20	517.70	31.70	23.41	0.604	0.079

D-ii) Mendefera station (n = 60).

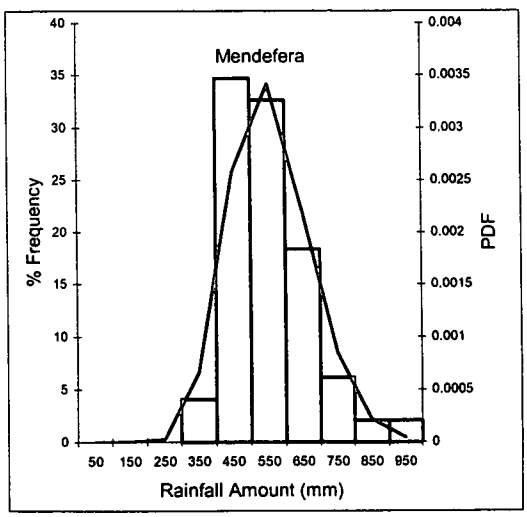
Parameters		Mean		Median		Coefficient of Variation		Skewness	
		Actual	Generated	Actual	Generated	Actual	Generated	Actual	Generated
Month	January	0.13	0.98	0.00	0.80	421.20	70.66	4.312	0.719
	February	2.18	1.63	0.00	1.50	423.90	76.96	5.762	1.119
	March	11.13	8.13	4.60	6.30	153.80	91.57	2.327	1.757
	April	26.93	34.95	20.30	35.60	103.90	54.49	2.373	0.189
	May	39.58	38.68	23.10	34.30	107.30	61.98	1.349	0.398
	June	56.03	61.49	56.30	58.50	72.13	56.83	0.899	0.670
	July	171.30	170.90	162.90	159.10	32.35	41.99	1.131	1.175
	August	170.10	174.30	168.90	162.00	23.93	48.03	0.407	0.524
	September	53.24	51.79	48.30	54.10	78.70	65.59	1.673	0.614
	October	9.79	8.19	3.30	4.70	135.90	113.36	1.596	3.162
	November	14.95	4.75	3.10	4.20	134.60	73.66	1.446	0.629
	December	2.47	1.93	0.00	1.90	241.30	68.56	3.005	0.697
	Total	557.80	557.84	538.80	551.30	21.35	21.11	0.368	1.587

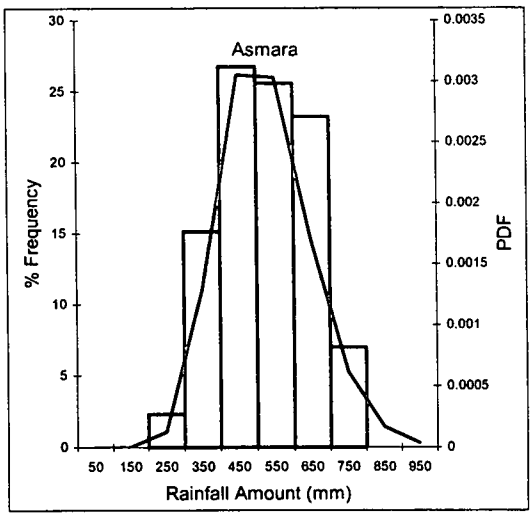
Appendix E Fitting gamma probability density functions (PDF) for measured and generated annual histograms.





E-i) Historical measured histograms and fitted gamma probability density functions (PDF) for annual total amount of rainfall (mm) for Mendefera and Asmara.

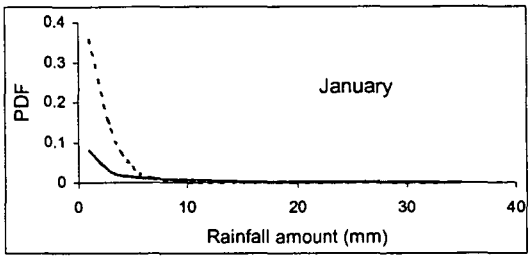


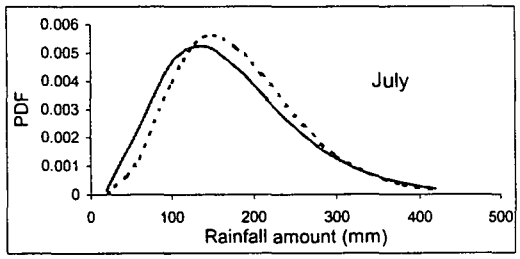


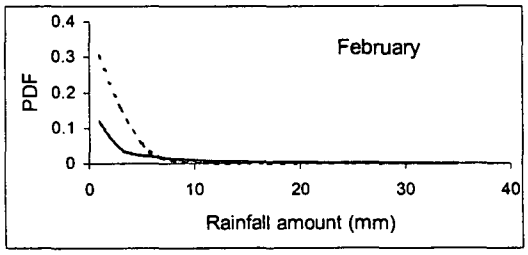
E-ii) Generated histograms and fitted gamma probability density functions (PDF) for annual total amount of rainfall (mm) for Mendefera and Asmara

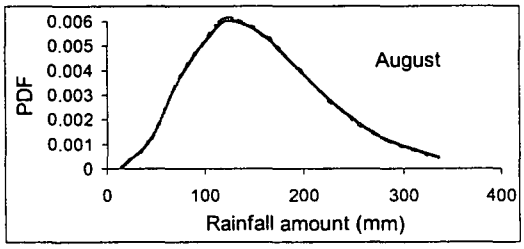
Appendix F Probability density function (PDF) of the measured (___) and generated (....) rainfall for each month.

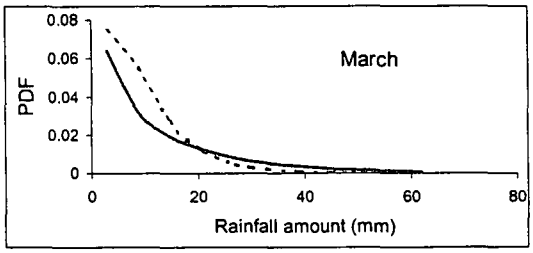
F-i) Asmara station.

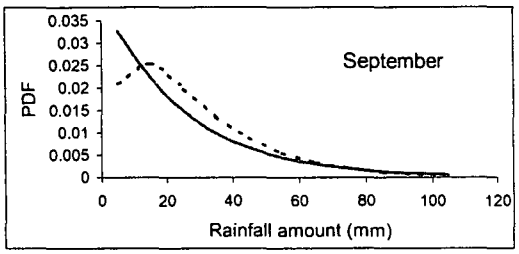


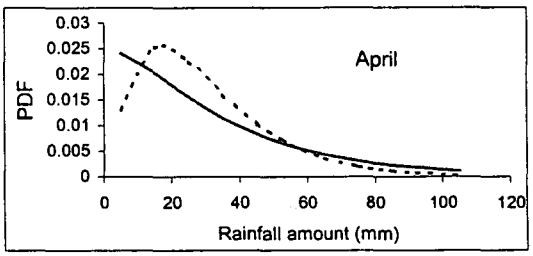


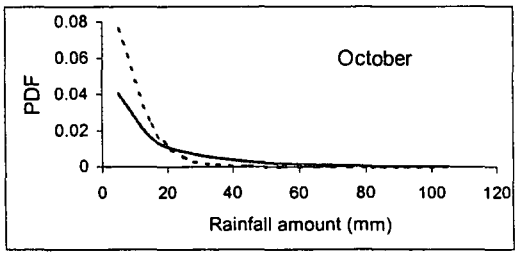


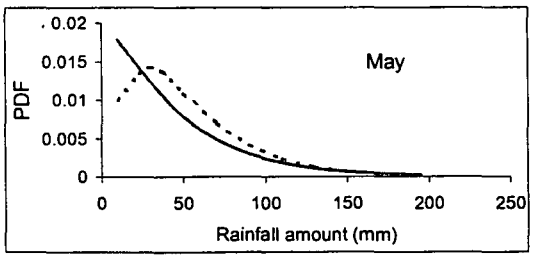


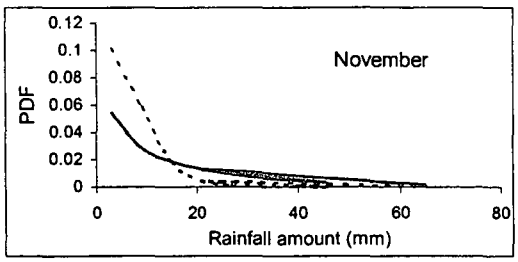


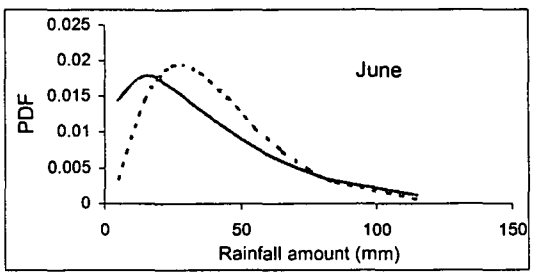


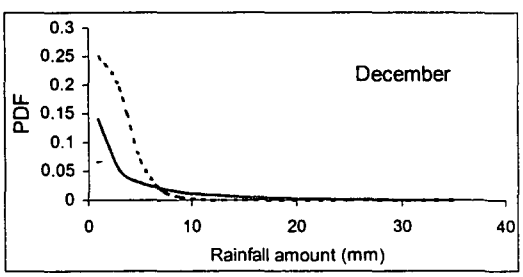












F-ii) Mendefera station.