THE ANALYSIS OF DAILY RAINFALL DATA

(A CASE STUDY OF NYANKPALA)

by

Emmanuel Bright Owusu BSc. Mathematics (Hons.)

A thesis submitted to the Department of Mathematics,

Kwame Nkrumah University of Science and Technology

in partial fulfilment of the requirements for the degree

of

MASTER OF SCIENCE

Faculty of Physical Sciences,

College of Science

DECLARATION

I hereby declare that this submission is my own work towards the Master of Science degree and that, to the best of my knowledge it contains neither material previously published by another person nor material which has been accepted for award of any other degree of the University, except for references to other people's work, which have been duly acknowledged.

Emmanuel Bright Owusu, PG20040453		22/09/08.
Student's Name& ID	Signature	Date

Certified by:

Dr. S.Ohemeng -Dapaah		28/09/09
Supervisor's Name	Signature	Date

Certified by:

Dr. S.K. Amponsah		28/09/09
Head of Department's Name	Signature	Date

DEDICATION

This dissertation is dedicated to my parents, Mr. and Mrs. E.B. Owusu.

TABLE OF CONTENTS

	PAGE
Title Page	i
Declaration	ii
Dedication	iii
Table of contents	iv
Acknowledgement	viii
Abstract	ix
CHAPTER ONE	
INTRODUCTION (TO THE GENERAL CLIMATIC AND AGRICULTURAL CHARACTERISTICS OF GHANA)	1
1.1 Introduction	1
1.2 The boundary and the population of Ghana	3
1.3 Vegetation, Soils, and Climate	5
1.4 Temperature	6
1.5 Agriculture economy of Ghana	8
1.6 Role of Agriculture	9
1.7 Methods of the analysis	14
1.8 Organization of study	15
CHAPTER TWO: LITERATURE REVIEW	17
2.1 Introduction	17
2.2 Markov chains	17

2.3 Gamma distribution 20

2.4 Fourier series and periodic functions 23

2.5 Other methods of analysis 24

2.5.1 Alternating renewal process 24

2.5.2 Resampling models 27

2.5.3 Time series models of the Arma type 29

2.6 Conclusion 30

CHAPTER THREE : THEORY AND CONCEPT 31

3.1 Introduction 31

3.2 Generalized linear model 32

3.2.1 Definition 32

3.2.2 Estimation of model parameters 33

3.2.3 Method of maximum livelihood 33

3.2.4 Method of least squares 35

3.3 Extension of Markov chains models 37

3.3.1 Introduction 37

3.3.2 Properties of Markov chains 39

3.3.2.1 Reducibility 39

3.3.2.2 Periodicity 39

3.3.2.3 Recurrence 40

3.3.2.4 Ergodicity 41

3.3.2.5 Steady-state analysis and limiting distribution 41

3.3.2.6 Markov chains with a finite state space 43

3.4 First order Markov chain 43

3.5 The deviance 46

CHAPTER FOUR: SIMPLE METHODS FOR ANALYZING DAILY RAINFALL DATA 48

4.1 Introduction 48

4.2 Data Availability 49

4.3 Daily Rainfall Data (1953-2003) 49

4.3.1 Average Rainfall 51

4.4 Dry Spells 52

4.5 Start of the growing season 58

4.5.1 Introduction 58

4.5.2 The definition for the start of the growing season 59

4.5.3 Rainfall amount summarized over days 60

4.5.4 Application of definition to Nyankpala data 63

4.6 The end of the growing season 66

4.7 The length of the season 68

4.8 Conclusion 69

CHAPTER FIVE: MODELING THE DAILY RAINFALL SEQUENCE BY MARKOV CHAIN 71

5.1 Introduction 71

5.2 Modeling the occurrence of rain 71

5.3 Modeling the transition probabilities 73

5.4 Application of Markov chain model to Nyankpala data 74

5.4.1 Structure of the data 74

5.4.2 The order of the Markov chain 75

5.4.3 Conclusion 79

5.5 Modeling rainfall amounts 79

5.5.1 Introduction 79

5.5.2 Fitting the model 80

5.5.3 Selection of harmonics 81

5.5.4 Estimating the shape parameter K 82

5.6 Application of the gamma distribution to the Nyankpala data 83

5.6.1 Selection of harmonics 84

5.6.2 Estimating the shape parameter K 86

5.7 Conclusion 87

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES 89

REFERNCES 92

APPENDIX A 103

A.1 How the daily rainfall data collected was arranged for the analysis 103

A.2 The results of the length and the end of the growing season 104

ACKNOWLEDGEMENT

The author wishes to express his sincere gratitude to his supervisor, Dr. S. Ohemeng-Dapaah of the Crop Research Institute (CRI) of the Council for Scientific and Industrial Research (CSIR), for his comments, tolerance, methodological guidance and suggestions that enhanced the quality of this thesis.

Special thanks also goes to all the Lecturers of the Department of Mathematics KNUST, especially Mr. Kwaku Darkwah, whose concern and directions helped shaped the structure of this thesis. I am also grateful to Mr. Simon Kojo Appiah and Mr Edmond Agyemang for their support and concern in the preparation of this thesis.

To all members of the Msc. Mathematics 2005-year group especially Francis Kwadwo Awuah and Derick Folson, I also say a big thank you.

Finally, I am very grateful to the Most High God for his help and protection at all stages of this thesis preparation.

ABSTRACT

The Northern Savanna Zone of Ghana is mostly semiarid. The irregular onset and distribution of the rains in this part of Ghana makes interpretation of agronomic experiments difficult. This work seeks to determine the optimum time of planting of maize in relationship with the length of the growing season in the Northern Savanna zone of Ghana.

Several statistical methods have been proposed for the analysis of rainfall data. In recent years one of the most popular methodologies are the use of Markov chains and Gamma distribution.

Markov chains are fitted to the occurrence of rain, and gamma distribution with parameters which vary with-the time of year, used to fit the rainfall amounts.

By relating climatic data analysis which incorporate soil factors for over 51 years (1953-2003) daily rainfall data to a maize cropping system, the planting time for maize in this area were determined. Results indicated that the growing season varied between 122 to 223 days. The best planting time for maize was found to be during the last two weeks of May in order to meet the moisture requirements during flowering and growing period lengths.

CHAPTER ONE

AN INTRODUCTION TO THE GENERAL CLIMATIC AND AGRICULTURAL CHARACTERISTICS OF GHANA

1.1 INTRODUCTION

Ghana being an agricultural economy, crop yields play a vital role in its national income. A major determinant for better yields is timely rainfall as the growth responses of plants are often a compromise between photosynthesis and transpiration with an optimization of water use efficiency being the prime aim. The dry periods at the early stage of a crop can damage the crop and on the other hand are usually useful at the ripening stage. The lack of rainfall in a certain area is the factor for declaring it a desert. Rainfall data is generally available for most areas on daily basis. There is a need to know the probability of having a dry period or having a consecutive dry period of 7 or 10 days during the growing season of a crop. This information would improve decisions about crops or varieties and the timings of the plantings of crops.

According to Gangopadhya and Sarkar (1965), the rainfall has a direct relationship with the yield of crops. Robertson (1970) discussed the rainfall and water variability and again Robertson (1976) and Gabriel and Neumann (1962) used the method of Markov chains to calculate dry and wet spells.

Stern and Coe (1982) gave a way to analyse and simulate the daily rainfall data; Stern et al. (1984) combine the Fourier analysis, Markov chains and distribution function of daily rainfall amount. Salimi et al. (1988) and Muhammad et al. (1988) have calculated the rainfall probabilities and the probabilities of the number of rainy days.

Rainfall in Ghana is characterised by a high degree of variability and it is the element of climate most influential in determining the variety and abundance of land use, economic development and practically all aspects of human activity. Hence a comprehensive analysis of rainfall data is a crucial component in agricultural production. From analysis of the daily rainfall data it is possible to get some insight into problems related to the plant water requirement.

This chapter is designed to highlight on the introduction to the general climatic and agricultural characteristics of Ghana. Section 1.2 gives the boundary and the population of Ghana whilst section 1.3 gives a brief discussion of the vegetation, soils and climate of Ghana.

Section 1.4 discusses temperature after which sections 1.5 and 1.6 presents a brief discussion of the Agricultural economy and the role of agriculture in Ghana. The methods of the analysis of the daily rainfall data are considered in section 1.7 whilst section 1.8 gives the organisation of study.

1.2 THE BOUNDARY AND THE POPULATION OF GHANA.

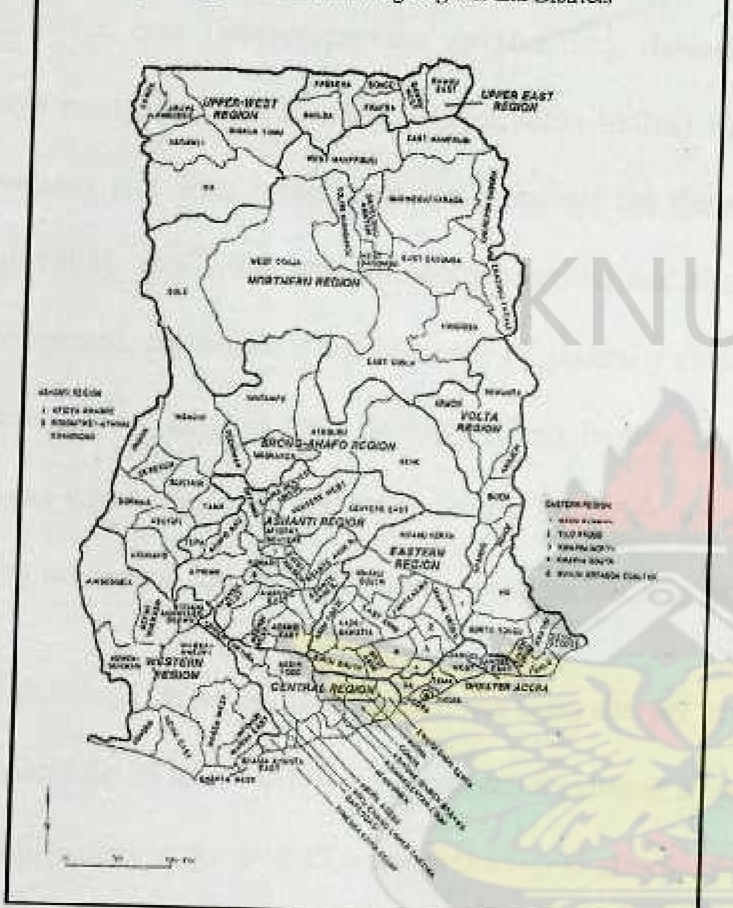
Ghana lies between latitudes 4° 44'N and 11°15'N and longitudes 3°15'W and 1° 12'E with a land area of 238,539 km^2. It is bordered on the east by the Republic of Togo, in the west by Côte d'Ivoire and in the north by Burkina Faso. Administratively, the country is divided into ten regions and one hundred and thirty districts with Accra as the capital.

Most of the country's surface is flat and the altitude varies between 500m and 200m above sea level. The Volta River basin dominates the county's river system, including the 8 480 km^2 covered by Lake Volta. The south has an extensive rain forest while the north is mostly savannah.

The population as at March 2007 was estimated to be approximately 22.93 million, of which more than 45 percent is below 15 years. Ghana's population growth is estimated at about 2.6 percent per year (CIA World Factbook, 2007). The average population density is around 52 persons per km^2. Most of the population is concentrated in the southern part of the country, with the highest densities in the urban and cocoa-producing areas.

Ghana ranks 119^{th} in the Human Development Index of UNDP (1991). It is a low-income food-deficit country with a per capita income of less than US $340 per Year. The World Bank has estimated that 31 percent of the population lives below the poverty line. Northern, Upper East and Upper West Regions are the poorest areas of the country, with levels of malnutrition estimated at more than double the national average.

Map 1 :Map of Ghana Showing Regions and Districts



50

Figure1.1 The map of Ghana.

1.3 VEGETATION, SOILS AND CLIMATE

Climate is the dominant factor in Ghana's physical environment. The natural vegetation of the study area is typical of Guinea savannah woodland, which is composed of trees of varying size and densely dispersed in a ground cover of tall perennial bunch-grasses and associated herbs. Shea butter tree (Butyrospernum paradoxum), dawadawa (Parkia clapertiniana), mahogany (Khaya senegalensis) and neem (Azadirachra indica) trees are now the dominant tree species. Dawadawa and shea butter tree are protected for their economic value (Runge-Metzger and Diehl, 1993). Soils are predominantly lateritic, and the texture is mainly silt or sandy loam. Their main characteristic is the presence of generally shallow depths below the surface of a more cemented layer of iron pan, through which rainwater does not penetrate easily. It therefore becomes water-logged in the rainy season but dries out completely during the dry season. Soil fertility is a major constraint for agricultural production (Runge-Metzger and Diehl, 1993).

The public interest on climatic change has risen sharply in recent years. Changes in weather have been also related to a worldwide increase of extreme events. A recent analysis by Diehl (1993) shows increases in the overall areas of the world affected by either drought or excessive wetness.

Climate is defined as the average weather conditions experienced in a given area over a considerable or long period of time usually not less than thirty-five years.

The annual distribution of rainfall is bi-modal in the south of Ghana and other high rainfall areas, and uni-modal in northern Ghana. Annual variability is quite high resulting in considerable drought risks.

The study area is characterised by distinct climate conditions such as one rainy season per year. Mean annual rainfall is approximately 1100 mm and occurs over 95 rainy days. It builds up gradually from small rains in March/April to a maximum in August in the north, and then declines sharply, coming to a complete stop in mid-November when the dry Saharan winds usher in the harmattan season. From December till February, the Northern Region is characterised by very distinct climate with relative humidity dropping to 15-26%, which enables the farmers to dry harvested cassava roots naturally. (Runge-Metzger and Diehl, 1993).

1.4 TEMPERATURE

Temperature refers to the degree of hotness or coldness of a place. Temperatures are uniformly high all the year, except for few places where temperatures are low due to high altitudes. Typical values are around 37^0C maximum and 26^0C minimum. The pattern of temperature during the year is represented graphically in figures 1.2 and 1.3 demonstrating the extreme conditions found in the region as represented by Kumasi and Tamale.

For most part of the country, the highest daily or monthly mean maximum occurs in February or March (suitable for land preparation for farmers) while the least occurs in August. The annual mean maximum temperatures occur in January or along the Coast in August.

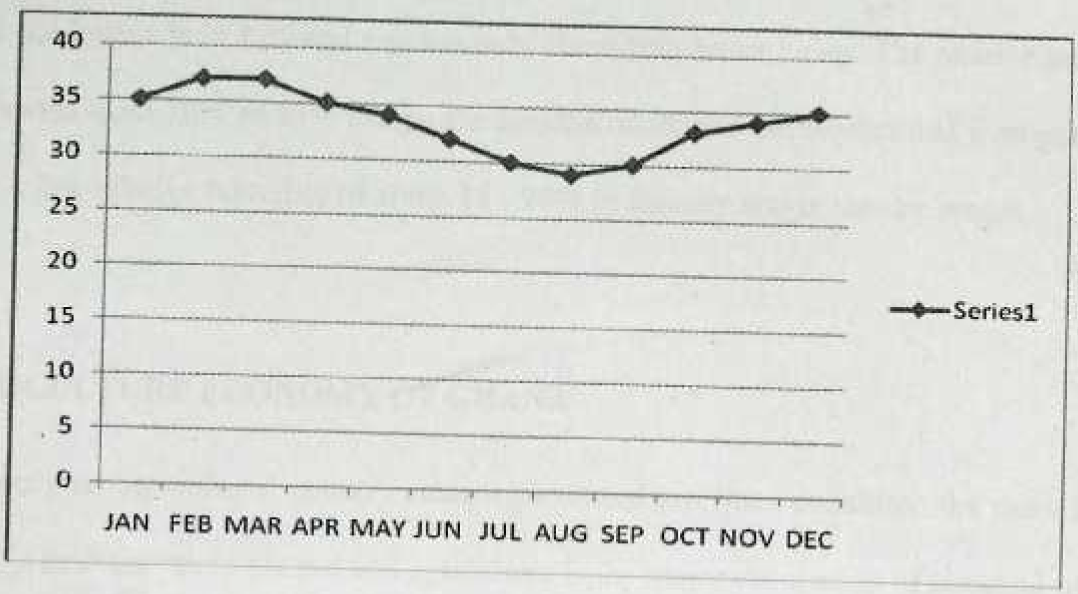


Figure 1. 2 Temperature distribution in Tamale

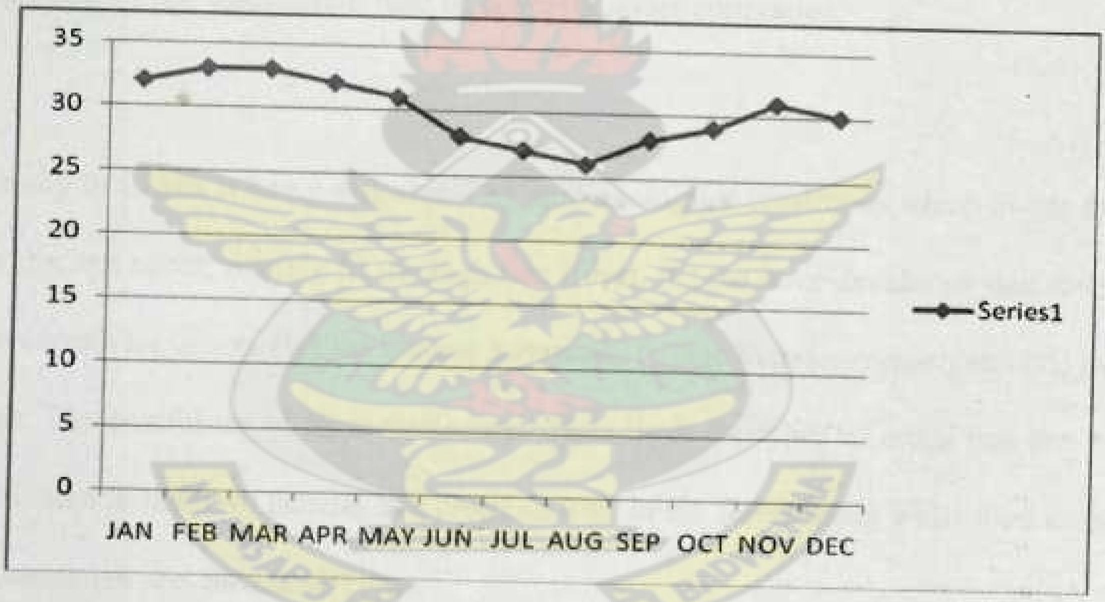


Figure 1.3 Temperature distribution in Kumasi

The study area is characterised by high temperatures throughout the year but exhibit wide variations between day and night. Between November and February, maximum day temperatures can reach 33° C to 37° C and minimum night temperatures vary between 20° C and 22° C. The greatest amount of sunshine occurs in the period from November to February for about 8.5 hours a day, while the lowest is in July and reaches only about four hours a day. The relative humidity of the area varies from 78% to 83% during the months of June to September and then gradually decreases to a low relative humidity of about 15 - 26% in January and in the dry season.

1.5 AGRICULTURE ECONOMY OF GHANA

Ghana is mainly an agricultural country since agricultural activities constitute the main use to which Ghana's land resources are put and agriculture is the major occupation of about 47 percent of the economically active age group (15 to 64). The country covers an area of approximately 239 million kilometers of which agricultural land forms about 57 percent of the total land area. Only about 20 percent of this agricultural land is, however under cultivation.

Agricultural activity in Ghana is being influenced by agro-ecological conditions which divide the country into six distinct zones, namely (i) the high rain forest, (ii) the semi-deciduous rain forest (iii) the forest-savannah transition, (iv) the Guinea Savannah, (v) the Sudan savannah and (vi) the coastal savannah. The conditions of these ecological zones limit the types of crops that can be successfully cultivated in them. In general tree crops do well in the forest zones while food crops do well in the transitional and Savannah zones.

Crop production in Ghana is for three main purposes; namely, food production for consumption, raw materials for industry and production for export. The major staple food crop includes cereals mainly rice and maize and starchy staples which include yams, cassava and plantain. Industrial raw materials include cotton, oil palm, and tobacco and bast fiber. The main export crop of

Ghana is cocoa for which Ghana was for a long time the leading world producer. Ghana however lost its place as the highest producer of cocoa, with it recorded share of world cocoa exports declining from 35 percent in 1961-1965 to only 15 percent in 1981.

The traditional crop farming system still prevails in Ghana, particularly in food production where small-scale farming predominates. Under this system, land preparation is accomplished by slashing and burning the vegetation. The seed is obtained from the previous harvest. Usually several crops are intercropped in a haphazard fashion, perhaps to avoid risk of total crop failure, particularly on small-scale subsistence farms. The field is cultivated for a few seasons and abandoned for several years when yields are observed to be too low. The cultivation is shifted to a "new" land or previously abandoned field thus earning the name "shifting cultivation". The abandoned land regenerates the fertility through natural means.

Thus, the traditional farming system depends mainly on natural soil fertility and very little on chemical fertilizers. The system works better in regenerating soil fertility in the forest zones which have higher vegetative cover than in the savannah zones with lower vegetative cover. The longer the land is allowed to rest, the higher the level of fertility generated. However, due to increasing population pressure, the fallow period is being progressively shortened, resulting in lower crop yields where fertilizers are not used.

1.6 ROLE OF AGRICULTURE

The importance of agriculture in the economy of Ghana cannot be overemphasized. Agriculture contributes immensely to the Gross Domestic Product.

During the first half of the 1980s, the sector's contribution averaged about 55 percent and declined to about 42 percent during the first half of the 1990s. The main reason for the decline in agriculture's contribution to GDP is increasing influence of the services sector in Ghana's economy. In fact, the services sector has, since 1992, taken over from the agriculture sector as the highest contributor to GDP with its contribution averaging about 44 percent in the first half of the 1990s.

The agricultural sector is a major source of government revenue, mainly through duties paid on exports of agricultural commodities, particularly cocoa. The contribution of agriculture to government revenue has, however, declined steadily from about 26 percent in 1987 to an average of about 20 percent in the first half of the 1990s. The decline has been a deliberate government strategy in order to boost cocoa production in the country through exchange rate devaluations which have the effect of raising prices in the domestic currency.

Agriculture's contribution to foreign exchange earnings averaged about 30 percent during the second half of the 1980s and declined to about 26 percent during the first half of the 1990s. The contribution has traditionally come mainly from the export of cocoa and timber. And, until gold took over in 1994, cocoa accounted for the highest proportion of the foreign exchange earned by the country each year. Since 1986, the government has been promoting the export of non-traditional commodities of which agricultural commodities such as raw food crops, seafood and processed commodities feature prominently.

From 1986 to 1989, the agricultural commodities in the non-traditional exports fetched the country an average of about 67 percent of the foreign exchange from this source. During the first half of the 1990s, however, the average proportion declined to about 34.3 percent.

The agricultural sector has continued to offer job avenues to highest proportion of the economically active population in the country as farmers, farm labourers and other workers in agriculture related activities such as processing and marketing. However, as consistent with economic development everywhere in the world, the proportion of the economically active population in agriculture has been declining gradually over the years from over 60 percent in the 1970s to an estimated 47 percent in 1994. Economic transformation is often accompanied by structural changes in subsistence agriculture, which often leads to agricultural diversification and specialization. The structural transformation manifests itself in changes in labour force participation in agriculture as shown in figure 1.4. As economic development proceeds, the agricultural sector plays an important role of supplying the labour force needed by the other emerging sectors.

When agriculture is in the subsistence stage, where production is mainly for home consumption, available capital and skills are insufficient to generate non-agricultural employment at a rate that is concomitant with the increase in the number of job seekers. The percentage of agricultural workers in the total labour force is decreasing, but their absolute number is still increasing. Planned generation of employment in agriculture is an absolute necessity and economically beneficial (Weitz, 1979).

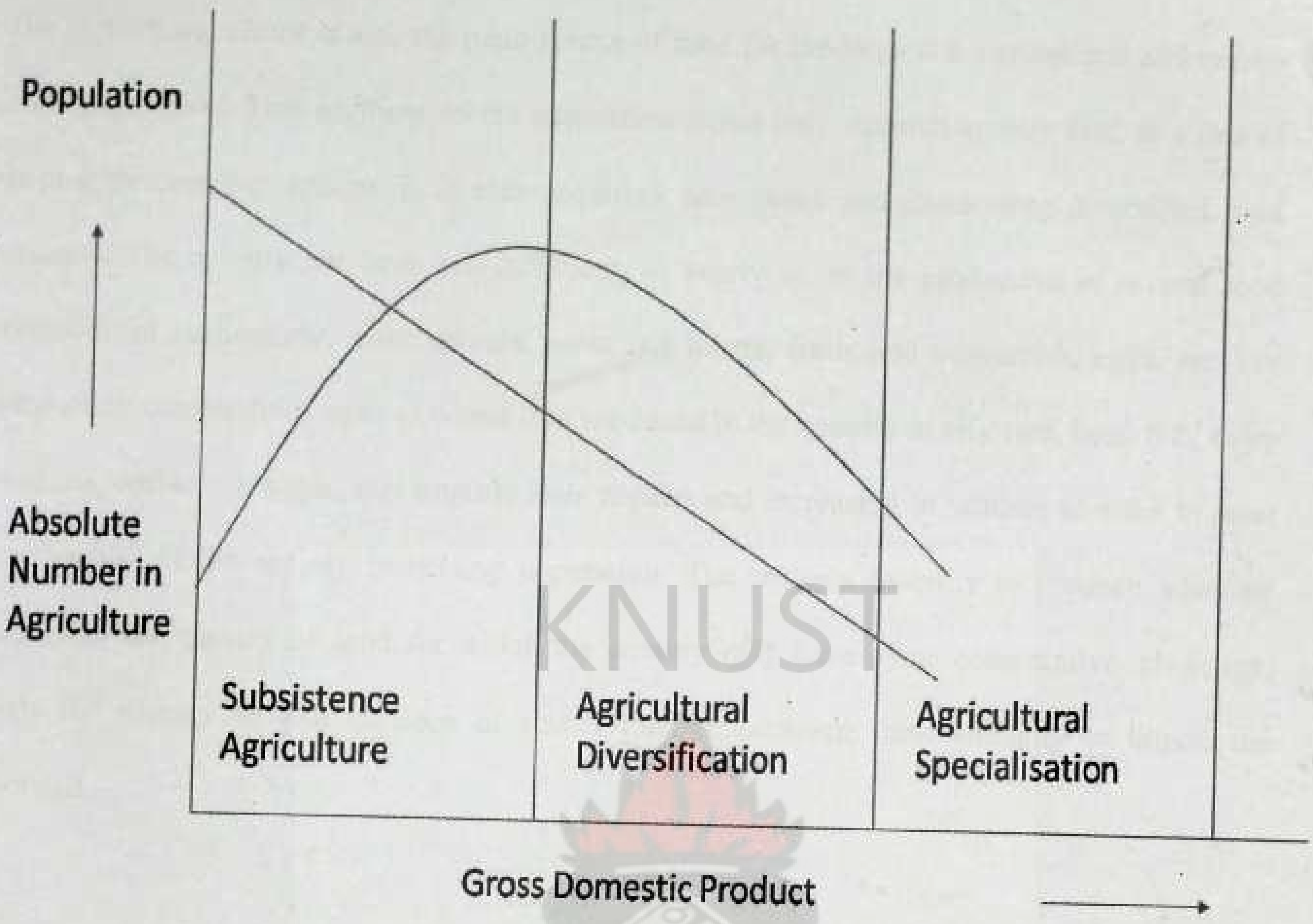


Figure 1.4 Labour Force participation in Agriculture in the process of Economic transformation.

Agricultural diversification represents a transitions period where the increase in demand for employment is almost of the same order of magnitude as the increase in non-agricultural job opportunities. The percentage of those engaged in agriculture is decreasing, but their absolute number remains more or less constant. Agricultural specialization is the desired stage in economic transformation.

It represents a period in which both the percentage and absolute number employed in agriculture are decreasing thus releasing labour to other sectors of the economy.

The agricultural sector is also the main source of food for the large non-agricultural and mainly urban population. This segment of the population is not only expanding very fast, at a rate of about 6 percent per annum, it is also acquiring new tastes and demanding diversified food products. The country has been self-sufficient, or nearly so, in the production of several food commodities particularly, some cereals, roots and tubers, fruits and vegetables, eggs, etc. For some other commodities such as wheat (not produced in the country at all), rice, beef, fish, dairy products, edible oil, sugar, etc. imports have regular and increasing in volume in order to meet the demand of the rapidly increasing population. The sector's inability to produce adequate quantities and variety of food for which the country may have some comparative advantage, costs the country several millions of scarce foreign exchange used annually to import the shortfall.

Agriculture also supplies the bulk of the raw materials needed for processing by the agro-based industries. The sector's failure to play this role effectively is partly the cause of the low capacity utilization of many of the agro-based industries (jebuni, etal, 1990, Seini, 1987; Ministry of Agriculture, 1990, 1991). The large agricultural population in the country is capable of providing a substantial market for the output of the industrial and service sectors.

However, as many of the predominantly small scale farmers are compelled to produce mainly for subsistence due to limited access to markets and therefore low income generation, the actual contribution of the agricultural sector to the size of market for the output of the other sectors has been relatively low and therefore a constraint to the development of those sectors and economy as a whole. Despite the decline in some areas of the agricultural sector's contribution, the rate of the country's economic development is still heavily dependent on the performance of the sector. To be able to play it role effectively in the medium-term, the agricultural sector is envisaged to grow area rate of about 4 percent per annum (Ministry of Agriculture, 1986).

Long sequences of daily rainfall are required increasingly, not only for hydrological purposes but also to provide inputs for models of crop growth, landfills, tailing dams, land disposal of liquid waste and other environmentally-sensitive projects. Rainfall is generally measured at the daily time scale and this forms the basis for monthly and annual rainfall series. Because daily data form this basic data set, modelling of the daily rainfall process has attracted a lot of interest in the past.

This work is therefore concerned with analysing daily rainfall data to look at its implications on agriculture and the techniques used for analysing such a data.

1.7 METHODS OF THE ANALYSIS

In this section, we discussed the various methods that were employed in the analysis of the rainfall data. Exploratory data analysis would be conducted in order to get insight of the data. Descriptive statistics for the mean, mode, and standard deviation in the start, end and the length of the rains of the daily rainfall data would be investigated by the use of *INSTAT CLIMATIC SOFTWARE* and *MICROSOFT EXCEL*.

Non-stationary Markov chains would be used to model the occurrence of rain, and gamma distributions, with parameters which vary with the time of year, would be fitted to the rainfall amounts. Fourier series would be used to model the mean amount of rainfall. The process of fitting and using these models provides a straight-forward and flexible analysis for rainfall records. All the models mentioned are examples of a Generalised Linear Models (GLMs). This thesis is therefore concerned with the analysis and interpretation of daily rainfall data. The contents of each chapter in the thesis are described in the rest of this chapter.

1.8 ORGANIZATION OF STUDY

In chapter 1, introduction to the general climatic and agricultural characteristics of Ghana have been briefly discussed whist in chapter 2, a review of the literature on some of the major methods for analysing the daily rainfall data is presented. The Generalized Linear Models, in this context of analysing daily rainfall data involves fitting Non-stationary Markov chains to model the occurrence of rain, and gamma distributions, with parameters which vary with the time of year, to the rainfall amounts. Fourier series would be use to model the mean amount of rainfall. Chapter 3 considers the theoretical and conceptual framework of the Generalized Linear Models.

In chapter 4, simple methods for analysing rainfall data were conducted using the data set collected from Nyankpala. We look at dry spells, the various starts, end and the length of the growing season. Descriptive statistics such as the mean, standard deviation of the start, end and the length of the season were also looked at. Rainfall data summarised over seven days was also considered.

In chapter 5, Statistical analysis using the Generalized Linear Models was conducted. This includes, fitting a non-stationery Markov chains to the occurrence of rainfall and gamma distribution to the rainfall amount on a rainy days. The fitting of Markov chains would help us to know the probability of rain or dry on any particular day whilst the fitting of gamma distribution will help us to know the amount of rainfall on any rainy day. A brief discussion of the results obtained in the data analysis also given in chapter 5.

Finally, concluding remarks of this research and possibility for further study are presented in Chapter 6. Appendix A gives additional results which were not given in the main chapters. These includes, how the daily rainfall data was arranged in Instat for the analysis, the results of the start, length, and the end of the growing season and the calculated risk from planting using the definition of the start of the growing season. In appendix B references of literature used are provided.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

In this chapter we present a literature review on some of the most important statistical models which have been used for the analysis of the daily rainfall data. Section 2.2 discusses Markov chains models after which section 2.3 presents a brief discussion of Gamma distribution. Fourier series and periodic functions are also considered in sections 2.4 and 2.5 respectively. Section 2.6 presents other methods used in modelling rainfall data. The chapter ends with a conclusion on the methods discussed and this is presented in section 2.7.

2.2 MARKOV CHAINS

Most stochastic models of daily rainfall consist of two parts: a model for the occurrence of dry and wet days and a model for the generation of rainfall amount on wet days. Modelling of daily rainfall occurrences are usually done by Markov chains introduced by Andrei Markov (1856-1922). Markov chains specify the state of each day as 'wet' or 'dry' and develop a relation between the state of the current day and the states of the preceding days. A Markov chain is a process that consists of a finite number of states and some known probabilities p_{ij}, where p_{ij} is the probability of moving from state i to state j. We may have more than two states. For example, political affiliation: Democrat, Republican, and Independent. For example, p_{ij} represents the probability of a son belonging to party i if his father belonged to party j.

The order of the Markov chain is the number of preceding days taken into account. Most Markov chain models referred in the literature are first order (lag one). Gabriel and Neumann (1962) used a first-order stationary Markov chain to model the daily rainfall occurrences. The models have since been extended to allow for non-stationarity, both by fitting separate chains to different periods of the year (Caskey, 1963; Dumont and Boyce, 1974; Heerman *et al.,* 1968; Jackson, 1981) and by fitting continuous curves to the transition probabilities (Feyerherm and Bark, 1965; Woolhiser and Pegram, 1979).

The order of Markov chain required has been discussed extensively (Lowry and Guthrie, 1968; Gates and Tong, 1976; Chin, 1977), the obvious conclusion being that different sites require different orders.

It is this flexibility of the Markov chain models, as well as the ease with which parameters are estimated, that leads us to use them. Another advantage which Markov chains have over other models of rainfall occurrence is the ease, with which results can be obtained from the fitted model without resorting to simulation (Caskey, 1963; Weiss, 1964; Hopkins and Robillard, 1964; Feyerherm and Bark, 1965, 1967; Lowry and Guthrie, 1968; Selvalingam and Miura, 1978; Stern, Richardson, 1981; Stern and Coe, 1984).

Models of second or higher orders have been studied by Chin (1977), Gates and Tong (1976), Eidsvik (1980), Pegram (1980) and Singh *et al.* (1981). The results varied with the climate characteristics of the rainfall stations investigated, with the statistical tests used and with the length of record. The Akaike information criterion, introduced by Akaike (1974), was widely used to determine the order of the Markov chains. Katz (1981) derived the asymptotic distribution of the Akaike information criterion (AIC) estimator but found that the estimator is inconsistent. The Bayesian information criterion (BIC) proposed by Schwarz (1978) was shown to be consistent and asymptotically optimal. However, Hurvich and Tsai (1989) provided a correction for AIC for model selection in small samples and the corrected AIC does not over fit the models as the AIC tends to do.

Jimoh and Webster (1996) determined the optimum order of a Markov chain model for daily rainfall occurrences at five locations in Nigeria using AIC and BIC. The AIC consistently gave a higher order for the Markov chain than the BIC.

The optimum order was also investigated by the generation of synthetic sequences of wet and dry days using zero-, first- and second-order Markov chains. They found that the first-order model was superior to the zero-order model in representing the frequency distribution of wet and stochastic generation of annual, monthly and daily rainfall data. It was concluded that caution is needed with the use of AIC and BIC for determining the optimum order of the Markov model and the use of frequency duration curves can provide a robust alternative method of model identification.

Jimoh and Webster (1999) investigated the intra-annual variation of the Markov chain parameters for seven sites in Nigeria. They found that there was a systematic variation in P_{01} (probability of a wet day following a dry day) as one moves northwards and a limited regional variation in P_{11} . A general conclusion is that a first-order model is adequate for many locations but a second- or higher order model may be required at other locations or during some times of the year.

2.3 GAMMA DISTRIBUTION

Some authors have attempted to describe rainfall amounts by fitting Markov chains with many states each representing a range of amounts (Khanal and Hamrick, 1974; Haan *et al.*, 1976). One unsatisfactory element of these models has been the large number of parameters to be estimated. One form of analysis which has received much attention in the modelling of the daily rainfall amount is the gamma distribution model.

A random variable is said to have a gamma distribution with parameters (α, λ), $\lambda, \alpha > 0$, if its density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq o \\ 0 \;, & x < 0 \end{cases}$$

where $\Gamma(\alpha)$, called the gamma function, is defined as

$$\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} dy .$$

The integration by parts of $\Gamma(\alpha)$ ~~yields that~~

$$\Gamma(\alpha) = -e^{-y} y^{\alpha-1} \Big|_0^{\infty} + \int_0^{\infty} e^{-y}(\alpha-1) y^{\alpha-2} dy$$

$$= (\alpha - 1)\int_0^\infty e^{-y} y^{\alpha - 2} dy = (\alpha - 1)\Gamma(\alpha - 1)$$

$$\begin{aligned}
\Gamma(n) &= (n-1)\Gamma(n-1) \\
&= (n-1)(n-2)\Gamma(n-2 \\
&= (n-1)(n-2)....3.2\Gamma(1) \\
&= (n-1)(n-2)....(4)(3)(2)(1).
\end{aligned}$$

Since $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$.

It follows that for integral values of n, $\Gamma(n) = (n-1)!$

When α is a positive integer, say $\alpha = n$, the gamma distribution with parameters (α, λ) often arises, in practice, as the distribution of the amount of time one has to wait until a total of n event has occurred. The gamma distribution with $\lambda = 1/2$ and $n/2$ (n being positive integer) is called the χ_n^2 (read ''Chi-squared'') distribution with n degrees of freedom.

Models used for modelling daily rainfall amounts include the two parameter Gamma distribution (Jones *et al.,* 1972; Goodspeed and Pierrehumbert, 1975; Stern and Coe, 1982; Richardson, 1981; Woolhiser and Roldan, 1982), mixed Exponential distribution (Woolhiser and Pegram, 1979; Woolhiser and Roldan, 1982, 1986), a skewed Normal distribution (Nicks and Lane, 1989) and a truncated power of Normal distribution (Bardossy and Plate, 1992; Hutchinson *et al.,* 1993).

Wang and Nathan (2000) developed a daily and monthly mixed (DMM) algorithm for the generation of daily rainfall. Daily rainfall data are generated month by month using the usual two-part model with two sets of parameters for the Gamma distribution, one estimated from the daily rainfall data and the other from monthly rainfall data. The monthly total is obtained by summing the daily values generated from the monthly Gamma parameters and adjusted for serial correlation. The generated daily rainfalls from the daily Gamma parameters are linearly scaled to match the serially correlated monthly rainfalls. Results for the Lake Eppalock catchment rainfall and for six other sites around Australia showed that the DMM algorithm reproduced the mean, coefficient of variation and skewness of daily, monthly and annual rainfall.

The results were examined in detail for the Lake Eppalock catchment (in southern Australia); the algorithm worked well in reproducing the mean, coefficient of variation and skewness of monthly maximum daily rainfall, but not as well for the annual maximum rainfall.

For the other six sites, the algorithm worked well in reproducing the mean and coefficient of variation but not as well as in reproducing the skewness of the annual maximum daily rainfall. Chapman (1998) investigated the impact of adjoining wet days on the distribution of rainfall amounts and found that the models which take this into account resulted generally in a better fit than the models which lump the data together. Chin and Miller (1980) examined the possible conditional dependence of the using 25 years of daily rainfall data at 30 stations in the continental United States. They concluded that, except for the winter season in the Pacific Northwest, the distribution of daily rainfall did not depend on whether the preceding day was wet or dry.

Menabde and Sivapalan (2000) ~~used~~ Levystable distributions to fit the storm duration and rainfall totals. They showed that this distribution having a fat tail fits the storm duration and amounts better than the Exponential or Gamma in the tail. Chapman (1994, 1998) compared the following five models for rainfall amounts, the Exponential (one parameter), the mixed Exponential (three parameters), the Gamma (two parameters), a skewed Normal (three parameters) and the Kappa distribution (two parameters).

Based on the AIC, the ranking of the models was consistent, the best being the Gamma followed by skewed Normal distribution, followed by the mixed Exponential, the Kappa, and last the Exponential. There was also consistency in the model selected for different groups of data (solitary wet days, first day of a wet spell etc.). He observed little variation in the coefficient of variation between different groups and relatively little between months.

Yevjevich and Dyer (1983) suggested that the latter feature may be a general characteristic of daily rainfall series and this could lead to a significant parsimony in the number of parameters to model seasonal variations.

2.4 FOURIER SERIES AND PERIODIC FUNCTIONS

Fourier series is an expansion of periodic functions in terms of infinite sum of sines and cosines. The computational and study of Fourier series is called harmonic analysis. Fourier series is a mathematical tool used for analysing periodic functions by decomposing such a function into weighted sum of much simpler sinusoidal component function called Normal Fourier Modes. The weights or coefficients of the modes are a one-to-one mapping of the original function.

Fourier series serve many useful purposes, as manipulation and conceptualization of the modal coefficients are often easier than with the original function. Suppose that f is a periodic, with period $2L$, and is such that the integrals

$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\left(\frac{\pi nx}{L}\right)dx$, and $b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\left(\frac{nx}{L}\right)dx$, $n = 1,2,3\ldots$ are defined.

Then the Fourier series of f is the series: $\frac{a_0}{2} + \sum_{n=1}^{\infty}\left\{a_n \cos\left(\frac{\pi nx}{L}\right) + b_n \sin\left(\frac{\pi nx}{L}\right)\right\}$.

The coefficients a_n and b_n are called the Fourier coefficients of f. We write

$$f(x) \approx \frac{a_0}{2} + \sum_{n=1}^{\infty}\left\{a_n \cos\left(\frac{\pi nx}{L}\right) + b_n \sin\left(\frac{\pi nx}{L}\right)\right\}.$$

The transition probabilities are likely to change smoothly through the year and can thus be modelled by continuous functions of time. Fourier series is considered because rainfall distribution function is periodic function. These have the desirable properties of modelling complex bimodal, rainfall patterns with few parameters and, where the whole year is modelled (T=366), of being continuous from day 366 to day 1.

2.5 OTHER METHODS OF ANALYSIS.

2.5.1 Alternating Renewal Process

In the alternating renewal process, the daily rainfall data is considered as a sequence of alternating wet and dry spells of varying length.

The wet and dry spells are assumed to be independent and the distributions may be different for wet and dry spells.

Distributions investigated include the logarithmic series (Williams, 1947), a modified logarithmic series (Green, 1964), truncated negative binomial distribution (Buishand, 1977), and the truncated geometric distribution (Roldan and Woolhiser, 1982).

Roldan and Woolhiser (1982) compared the alternating renewal process with truncated geometric distribution of wet sequences and truncated negative binomial distribution of dry sequences with a first-order Markov chain. For five US stations with 20-25 years of record lengths, the first-order Markov chain was superior to the alternating renewal process according to the Akaike information criterion (Akaike, 1974). The parameters of the distributions were assumed to be either constant within seasons or to vary according to Fourier series.

One of the disadvantages of the alternating renewal process is that the seasonality is difficult to handle. The starting day of the sequence is usually used to determine the season to which the sequence belongs.

Small and Morgan (1986) derived a relationship between a continuous wet-dry renewal model with Gamma distributed dry intervals and a Markov chain model for daily rainfall occurrence. The Markov process model was shown to provide a good representation in certain parts of the United States while in other areas, where the Markov model is inappropriate due to event clustering or other phenomena, the Gamma model provides an improved characterization of the relationship between continuous and discrete rainfall occurrence.

Foufoula-Georgiou and Lettenmaier (1987) developed a Markov renewal model for rainfall occurrences in which the time between rainfalls occurrences were sampled from two different geometric distributions. The transition from one distribution to the other was governed by a Markov chain.

Smith (1987) introduced a family of models termed Markov-Bernoulli processes that might be used for rainfall occurrences. The process consists of a sequence of Bernoulli trials with randomised success probabilities described by a first-order, two-state Markov chain. At one extreme the model is a Bernoulli process, at the other a Markov chain.

A binary discrete autoregressive moving average (DARMA) process was first used by Buishand (1977). He found that an alternating renewal process was superior to the DARMA model for the data from The Netherlands but the DARMA model looked more promising in tropical and monsoonal areas. Chang *et al.* (1984) and Delleur *et al.* (1989) used four seasons for two stations in Indiana (USA) and found that either the first-order autoregressive or the second-order moving average model was appropriate for different seasons.

Chapman (1994) compared five models, namely, Markov chains of orders 1, 2 and 3 (MC1, MC2 and MC3), truncated negative Binomial distribution (TNBD) and the truncated Geometric distribution (TGD) with separate parameter values for each month using data from 17 Australian rainfall stations.

Three of the above models (MC1, TNBD and TGD) were also compared with parameters varying smoothly throughout the year according to a Fourier series having 0, 1 and 2 harmonics.

The Fourier series representation with one harmonic for parameter variation throughout the year using the MC1 or TGD model was successful for five stations with high rainfall in southern Australia. The monthly MC2 model or monthly TNBD model fitted the remaining stations best. Different record lengths appear to affect the selection of the best model, particularly when wet and dry spells are considered separately. For combined results, different models were selected for the 20-50 years records in four out of ten cases, for the 20-100 years records in two out of five cases, and for the 50-100 years records in one case out of five. He concluded that the prospects for regionalisation of parameters are poor unless there is a good sample of long records.

In a later study, Chapman (1997) compared the above distributions and Markov chain for the rainfall stations from 22 islands in the Western Pacific. He concluded that a first-order Markov chain or truncated geometric distribution with a Fourier series representation for parameter variation over months. R. Srikanthan and T.A. McMahon (1985) were successful for stations with latitude greater than 14°.

For the stations close to the Equator, the seasonal regularity was less important and the models with individual monthly values or constant value throughout the year performed well.

2.5.2 Resampling Models

Lall *et al.* (1996) developed a non-parametric, wet-dry spell model for re-sampling daily rainfall at a site. All marginal, joint and conditional probability densities of interest (dry spell length, wet spell length, precipitation amount and wet spell length given prior to dry spell length) are estimated non-parametrically using at-site data and kernel probability density estimators.

The model was applied to daily rainfall data from Silver Lake station in Utah (USA) and the performance of the model was evaluated using a number of performance measures. The model reproduced satisfactorily the wet day precipitation, wet spell length and dry spell length.

Rajagopalan *et al.* (1996) presented a non-homogeneous Markov model for generating daily rainfall at a site. The first-order transition probability matrix was assumed to vary smoothly day by day over the year. A kernel estimator was used to estimate the transition probabilities through a weighted average of transition counts over a symmetric time interval centered at the day of interest. The rainfall amounts on each wet day were simulated from the kernel probability density estimated from all wet days that fall within a time interval centred on the calendar day of interest over all the years of available data. Application of the model to daily rainfall data from Salt Lake City, Utah, showed that the wet and dry-spell attributes and the rainfall statistics were reproduced well at the seasonal and annual time scales. Sharma and Lall (1997, 1999) used a nearest-neighbour conditional bootstrap for re-sampling daily rainfall for Sydney. The dry spell lengths were conditioned on the number of days in the previous wet spell and the wet spell lengths were conditioned on the number of days in the previous dry spell. The rainfall amounts were conditioned on two variables, the rainfall amount on the previous day and the number of days from the start of the current spell. Results from the model showed its ability to simulate sequences that are representative of the historical record. A limitation of the non-parametric density estimation approach is the rather limited extrapolation of daily rainfall values beyond the largest value recorded.

The simulations from the k-nearest-neighbour method do not produce values that have not been observed in the historical data and this is a major limitation if extreme values outside the available record are of interest (Rajagopalan and Lall, 1999). Sample sizes needed for estimating the probability density function of interest are likely to be larger than for parametric estimation.

The non-parametric methods have been tested on a limited range of sites and testing over a greater range of climates is needed for broader applicability.

2.5.3 Time Series Models of the ARMA Type

In this approach, time series models similar to stream flow data generation are used to generate daily rainfall data.

Adamowski and Smith (1972) used a first-order Markov model to generate standardised daily rainfall data. The major problem with this procedure is the cyclical standardization which occurs if there are numerous zero daily values.

A truncated power of Normal distribution has been suggested to model daily rainfall (Hutchinson *et al.,* 1993; Hutchinson, 1995). The underlying Normal distribution can be put into a simple first-order autoregressive scheme to account for the day-to-day persistence of wet and dry days. The correlations in the amounts of rainfall on successive wet days from this model were found to be much larger than the observed correlations in the rainfall and, to a first approximation, could be ignored (Hutchinson, 1995).

Such systematic differences between correlations based on occurrence and intensity have not been recognised in the applications of such models as described in Bardossy and Plate (1992).

2.6 CONCLUSION

Recent advances in statistical methods have dramatically improved the range of techniques available for analysing data that are not from normal distribution. These new techniques, which are used in this study, parallel those used in the analysis of variance and regression for normally distributed data. This development is of considerable importance, since daily rainfalls are clearly not normally distributed (Stern *et al.*, 1982). This chapter has reviewed the literature of a large range of statistical methods applicable to the analysis of daily rainfall data. The big question is that which model or method is preferable and for which data and which objective.

More often than not it is necessary to perform several analyses and compare the results. A large number of these statistical methods including the Generalized Linear Models (Markovs chains, Gamma distribution and Fourier series) are very useful in daily rainfall data analysis. We proceed to the next chapter where the conceptual and the theoretical framework of the Generalised Linear Models are been discussed.

CHAPTER THREE

THEORY AND CONCEPT

3.1 INTRODUCTION

In the advent of Modem Powerful Computers, one can easily analyse daily rainfall data without knowing the mathematical concepts behind it. However, to be able to effectively analyse a data by a statistical tool (model) it is very important to understand the theoretical and conceptual framework of the model.

In chapter two we presented a review literature on some of the most important statistical models that is the Generalized Linear Models (Markov chains, Gamma distribution and Fourier series) which have been used in this work.

This chapter is to highlight on the theoretical and conceptual framework of the Generalized Linear Models (GLMs) used in this work in a systematic fashion coupled with the mathematical theory behind the model. In section 3.2 of this chapter, a brief description of the Generalized Linear model is also presented whilst section 3.3 considers the extension of Markov chain models. In section 3.4 first order Markov chains is also discussed.

3.2 GENERALISED LINEAR MODEL

3.2.1 Definition

The unity in many statistical methods involving linear combinations of parameters was demonstrated by Nelder and Wedderburn (1972) using the idea of generalized linear model. This is defined in terms of a set of independent random variables $Y_1, Y_2, ..., Y_n$ each with a distribution from the exponential family with the following properties:

(i) The distribution for each Y_i is of the canonical form and depends on a single parameter θ_i (the θ_i 's do not all have to be the same), thus

$$f(y_i;\theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]$$

(ii) The distribution of all Y_i 's are of the same form (e.g. all normal or all binomial) so that the subscripts b, c and d are not needed. Thus the joint probability density function of $Y_1, Y_2, ..., Y_n$ is

$$f(y_1, y_2, ... y_n; \theta_1, \theta_2, ..., \theta_n) = \exp\left[\sum_{i=1}^{n} y_i b_i(\theta_i) + \sum_{i=1}^{n} c_i(\theta_i) + \sum_{i=1}^{n} d_i(y_i)\right]$$

For model specification the parameter θ_i are usually not of direct interest (since there may be one for each observation).

For a generalised linear models we consider a smaller set of parameters, $\beta_1, \beta_2, ..., \beta_p$ (where $p < N$) such that a linear combinations of the β_i's is equal to some function of the expected value μ_i of Y_i, i.e. $g(\mu_i) = x_i^T \beta$, where g is a monotone, differentiable function called the link function;

x_i is a px1 vector of explanatory variables (covariates and dummy variables for levels of factors); β is the px1 vector of parameters.

3.2.2 Estimation of Model Parameters.

Two of the most commonly used approaches to the statistical estimation of parameters are the method of maximum likelihood and the method of least squares. The method of maximum likelihood is usually used for generalized linear models. The estimates have to be obtained numerically by an iterative procedure which turns out to be closely related to weighted least squares estimation.

3.2.3 Method of Maximum Likelihood

The method of maximum likelihood would be used to estimate the shape parameter (k) of the Gamma distribution function in modelling the rainfall amounts of Nyankpala.

Let $Y_1, Y_2, ..., Y_N$ be N random variables with the joint probability density function $f(y_1, y_2, ..., y_n; \theta_1, \theta_2, ...\theta_p$ which depends on parameters $\theta_1, \theta_2, ...\theta_p$.

We let $y = [y_1, y_2, ...y_N]'$ and $\theta = [\theta_1, \theta_2, ...\theta_N]'$, So that the probability density function is denoted by $f(y; \theta)$.

The likelihood function L $(\theta; y)$ is algebraically the same as $f(y;\theta)$ but the change in notation reflects a shift of emphasis from the random variable y, with θ fixed, to the parameter of θ with y fixed (where y represents the observations).

Let Ω denote the set of all possible values of the parameter vector θ (Ω is called the parameter space). The maximum likelihood estimator of θ is the value $\hat{\theta}$ which maximizes the likelihood function, that is $L(\hat{\theta}; y) \geq L(\theta; y)$ for all θ in Ω.

Equivalently, $\hat{\theta}$ is the value which maximizes the log-likelihood function $l(\hat{\theta}; y) = \log L(\theta; y)$ (since the logarithmic function is monotonic). Thus $l(\hat{\theta}; y) \geq l(\theta; y)$ for all θ in Ω.

Often it is easier to work with the log-likelihood function than with the likelihood function itself. Usually the estimator $\hat{\theta}$ is obtained by differentiating the log-likelihood function with respect to each element θ_j of θ and solving the simultaneous equations.

$$\frac{\partial l(\theta; y)}{\partial \theta_j} = 0 \text{ for all } j = 1,2,\ldots p.$$

It is necessary to check that the solution do correspond to maxima of $l(\theta; y)$ by verifying that the matrix of second derivatives $\frac{\partial^2 l(\theta; y)}{\partial \theta_j \partial \theta_k}$, evaluated at $\theta = \hat{\theta}$ is negative definite. Also, it is necessary to check if there are any values of θ at the edges of the parameter space Ω which gives local maxima of $l(\theta; y)$.

When all local maxima have been identified, the value of $\hat{\theta}$ corresponding to the largest one is the maximum likelihood estimator.

3.2.4 Method of Least Squares

This method would be used to estimate the parameters $a_1, a_2, \ldots, a_5$ and $b_1, b_2, \ldots, b_5$ in the harmonic analysis in the modelling of rainfall occurrences and rainfall amounts. Let $Y_1, Y_2, \ldots, Y_N$ be N random variables with expected values $E(Y_i) = \mu_i$ for $i = 1, 2, 3, \ldots, N$ and let the $\mu_i's$ be functions of parameters $\beta_1, \beta_2, \beta_3, \ldots, \beta_p$ (where $p < N$) which are to be estimated.

Let $\beta = [\beta_1, \beta_2 \ldots \beta_p]'$

Consider the formulation $Y_i = \mu_i + e_i$ for $i = 1, 2, 3, \ldots, N$ in which for μ_i represents the 'signal' component of Y_i and e_i represent the 'noise' component.

The method of least squares consist of finding estimators β, also denoted by b, which minimize the sum of squares of the error term e_i; that is , it involves minimizing the function

$$S = \sum e_i^2 = \sum [Y_i - \mu_i(\beta)]^2 .$$

In matrix notation this is

$$S = (y - \mu)^T (y - \mu) \text{ where}$$

$y = [Y_1, Y_2, \ldots Y_N]'$ and $\mu = [\mu_1, \mu_2, \ldots \mu_N]'$.

Usually, the estimator $\hat{\beta}$ is obtained by differentiating S by each element β_1 of β and solving the simultaneous equation $\frac{\partial S}{\partial \beta_i} = 0 \ \ i = 1,2,3,\ldots,p$

It is necessary to check that the solutions corresponds to minima (i.e. the matrix of second order derivatives is positive definite) and to identify the global minimum from among these solutions and any local minima at the boundaries of the parameter space. An important distinction between the methods of least squares and maximum likelihood is that the least squares can be used without making assumptions about the distributions of the response variables Y_i beyond specifying their expectations and possibly their variance-covariance structure. In contrast, to obtain maximum likelihood estimators we need to specify the joint probability distribution Y_i's.

However, to obtain the sampling distribution of the least squares estimators b, additional assumptions about the Y_i's are generally required. Thus in practice there is a little advantage in using the method of least squares unless the estimation equations are computationally simpler.

3.3 EXTENSION OF MARKOV CHAINS MODELS

3.3.1 Introduction

Consider a sequence of random variables $X_0, X_1, X_2, ..., X_n$ and suppose that the set of possible values of these random variables is $\{0,1,2,3,...,M\}$. It will be helpful to interpret X_n as being the state of some system at time n, and, in accordance with this interpretation, we say that the system is in state i at time n if $X_n = i$. The sequence of random variables is said to form a Markov chain if each time the system is in state i there is some fixed probability P_{ij} that it will next be in state j. That is, for all the $i_0, i_1, i_2, ..., i_n, i, j$,

$$\Pr\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0\} = P_{ij}.$$

The values P_{ij}, $0 \le i \le M$, $0 \le j \le N$, are called the transition probability of the Markov chain and have the following features,

1. $P_{ij} \ge 0, \sum_{j=0}^{M} P_{ij} = 1$, for all $i = 0,1,2,...,M$.
2. It is a square matrix, since all possible states must be used both as rows and as columns.
3. All entries are between 0 and 1, inclusive; this is because all entries represent probabilities.

Knowledge of the transition probability matrix and the distribution of X_0 enable us in theory, to compute all probabilities of interest.

A transition matrix is a constant square matrix P of order n such that the entry in the ith row and jth column indicates the probability of the system moving from the ith state to the jth state on the next observation or trial. An example of a transition matrix is

$$p = \begin{bmatrix} 0.8 & 0.2 \\ 0.35 & 0.65 \end{bmatrix}$$

For a first order Markov chain, the transition probability matrix used to calculate the occurrence of rain is given by

$$p = \begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix}$$

where

$$p_{11} = p\{wet\ day /\ \ previous\ day\ wet\}$$
$$p_{01} = p\{wet\ day /\ \ previous\ day\ dry\}$$
$$p_{00} = p\{dry\ day /\ \ previous\ day\ dry\}$$
$$p_{10} = p\{dry\ day /\ \ previous\ day\ wet\}$$

A Markov chain of order m (or a Markov chain with memory m) where m is finite, is where $\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots)$

$\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m})$, for all n. This would be used to determine whether the rainfall occurrences depend on the event of the previous day (first order) or two previous day's event (second order).

3.3.2 Properties of Markov Chains

Define the probability of going from state i to state j in n time steps as

$p_{ij}^{(n)} = \Pr(X_n = j | X_0 = i)$ and the single-step transition as $p_{ij} = \Pr(X_1 = j | X_0 = i)$.

3.3.2.1 Reducibility

A state j is said to be accessible from state i (written $i \to j$) if, given that we are in state i, there is a non-zero probability that at some time in the future, we will be in state j. That is, that there exists an n such that $\Pr(X_n = j | X_0 = i) > 0$

A state i is said to communicate with state j (written $i \leftrightarrow j$) if it is true that both i is accessible from j and that j is accessible from i. A set of states C is a communicating class if every pair of states in C communicates with each other. (It can be shown that communication in this sense is an equivalence relation). A communicating class is closed if the probability of leaving the class is zero, namely that if i is in C but j is not, then j is not accessible from i. Finally, a Markov chain is said to be irreducible if its state space is a communicating class; this means that, in an irreducible Markov chain, it is possible to get to any state from any state.

3.3.2.2 Periodicity

A state i has period k if any return to state i must occur in some multiple of k time steps and k is the ~~largest number~~ with this property. For example, if it is only possible to return to state i in an even number of steps, then i is periodic with period 2. Formally, the period of a state is defined as

$$k = \gcd\{n : \Pr(X_n = i | X_0 = i\ \} > 0$$

(where "gcd" is the greatest common divisor). Note that even though a state has period k, it may not be possible to reach the state in k steps. For example, suppose it is possible to return to the state in {6, 8, 10, 12...} time steps; then k would be 2, even though 2 does not appear in this list.

If $k = 1$, then the state is said to be aperiodic; otherwise ($k>1$), the state is said to be periodic with period k. It can be shown that every state in a communicating class must have the same period. A finite state irreducible Markov chain is said to be ergodic if its states are aperiodic.

3.3.2.3 Recurrence

A state i is said to be transient if, given that we start in state i, there is a non-zero probability that we will never return to i. Formally, let the random variable T_i be the next return time to state i (the "hitting time"):

$$T_i = \min\{n : X_n = i | X_0 = i\ \}$$

Then, state i is transient if and only if there exists a finite T_i such that:

$$\Pr\{T_i < \infty\ \} = 1$$

If a state i is not transient (it has finite hitting time with probability1), then it is said to be recurrent or persistent. Although the hitting time is finite, it need not have a finite average. Let M_i be the expected (average) return time,

$$M_i = E[T_i].$$

Then, state i is positive recurrent if M_i is finite; otherwise, state i is null recurrent (the terms non-null persistent and null persistent are also used, respectively).

$$\sum_{n=0}^{\infty} P_{ij}^{(n)} = \infty$$

A state i is called absorbing if it is impossible to leave this state. Therefore, the state i is absorbing if and only if

$p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$

3.3.2.4 Ergodicity

A state i is said to be ergodic if it is aperiodic and positive recurrent. If all states in a Markov chain are ergodic, then the chain is said to be ergodic.

3.3.2.5 Steady-state Analysis and Limiting Distributions

If the Markov chain is a time-homogeneous Markov chain, so that the process is described by a single, time-independent matrix p_{ij}, then the vector π is a stationary distribution (also called an equilibrium distribution or invariant measure) if its entries π_j sum to 1 and satisfy

$$\pi_j = \sum_{i \in s} \pi_i P_{ij} .$$

An irreducible chain has a stationary distribution if and only if all of its states are positive-recurrent. In that case, π is unique and is related to the expected return time:

$$\pi_j = \frac{1}{M_j} .$$

Further, if the chain is both irreducible and aperiodic, then for any i and j,

$$\lim_{n \to \infty} P_{ij}^{(n)} = \frac{1}{M_j}.$$

Note that there is no assumption on the starting distribution; the chain converges to the stationary distribution regardless of where it begins. If a chain is not irreducible, its stationary distributions will not be unique (consider any closed communicating class in the chain; each one will have its own unique stationary distribution). Any of these will extend to a stationary distribution for the overall chain, where the probability outside the class is set to zero).

However, if a state j is aperiodic, then

$$\lim_{n \to \infty} P_{jj}^{(n)} = \frac{1}{M_j}.$$

and for any other state i, let f_{ij} be the probability that the chain ever visits state j if it starts at i,

$$\lim_{n \to \infty} P_{ij}^{(n)} = \frac{f_{ij}}{M_j}.$$

3.3.2.6 Markov Chains with a Finite State Space

If P is the transition matrix and S_0 is an initial-state matrix for a Markov chain, then the Kth-state matrix is given by

$$S_k = S_0 P^k$$

The entries in the i th row and j th column of P^k indicates the probability of the system moving from the i th state to the j th state in K observations or trails. The sum of the entries in each row of P^k is 1.

The state matrix $S = [s_1 \quad s_2 \quad . \quad . \quad . \quad s_n]$ is a stationary matrix for a Markov chain with transition matrix P if

$S = S_i P$ where $S_i \geq 0$, $i = 1,2,3,....,n$ and $s_1 + s_2 + ... + s_n = 1$.

3.4 FIRST ORDER MARKOV CHAIN

If we are interested in the rainfall between day t_1 and t_2 of the year and we have N years record of daily rainfall data for a station, then this forms an $N \times T$ matrix, where $T = t_2 - t_1 + 1$, the entries being the amount of rain falling each day in each year. If the whole year is to be analysed then $T = 366$. Day 60 (29 February) only contains data in leap years.

To fit a two state Markov chain, each day is first classified into one of the two states, wet or dry.

It will be sometimes appropriate to define a threshold value for rain and define only the days with amount above this as rain days. If a first- order, two- state Markov chain is to be fitted then the data matrix may be condensed into a $4 \times T$ matrix. 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 the number of dry days following dry days. The new matrix then becomes a 2x2xT table with entries on day $t[\eta_{ij}(t)]$ giving the number of years when day t is in state $j(1 = rain,\ 0 = dry)$ and day $t-1$ is in state i. If the full year is to be analysed then day 366 (31 December) for previous year is used to give $\eta_{ij}(1)$. Otherwise day $t-1$ gives. In non-leap years, when day 60 (29 February) does not exist, the state i of day 59 is used to calculate $n_{ij}(61)$.

In a first-order Markov chain, the probability of rain falling on any day depends only on the state (wet or dry) of the previous day. The parameters to be estimated are therefore the transition probabilities $p_i(t)$, that is, the probability of rain on day t conditional on day $t-1$ being in state i *for* $i =$ 0, 1. The obvious estimates of $p_i(t)$, which are also the maximum likelihood estimates (Anderson and Good-man, 1957), are the observed proportions of years with day $t-1$ in state i that had rain on day t, that is

$$r_i(t) = \eta_{i1}(t)[\eta_{i1}(t) + \eta_{i0}(t)]^{-1}.$$

The next step in the analysis is to fit a function to model the time dependence of the probabilities of rain ~~$p_i(t)$~~ through the year.

This involves fitting a regression-type equation to the $r_i(t)$, using the correct form of the distribution for the data. The $\eta_{i1}(t)$ may be considered as observation from independent binomial distribution with number of trials $[\eta_{i0}(t)+\eta_{i1}(t)]$ and the probability of success $p_i(t)$. If the equations for the time dependence of probabilities are given by $p_i(t) = h(g_i(t))$, $i =$ 0, 1,

Where h is a known function and $g_i(t)$ is any function linear in unknown parameters, then the model is a Generalised Linear model. Therefore the unknown parameters in $g_i(t)$ can be found using the maximum likelihood estimation. The link function h used here is the logistic

$$p_i(t) = \exp[g_i(t)]\{1+\exp[g_i(t)]\}^{-1}, \; i = 0, 1.$$

Where the link function makes the values of $g_i(t)$ lies between 0 and 1. There are many types of functions of $g_i(t)$ that may be used for modelling the time dependence throughout the year. Which of these is appropriate depends on the use that is to be made of the fitted model. The simplest model assumes a constant probability of rain, that is

$$g_i(t) = a_i \quad i = 0, 1.$$

This method was used by Gabriel and Neumann (1962) in modelling a daily data from Tel Aviv.

An alternative way is also by making the function $g_i(t)$ to be a step function which makes the probability of rain to vary.

The function $g_i(t)$ then becomes:

$$g_i(t) = a_{1i}, \quad for \; t_1 \leq t \leq s_1,$$
$$= a_{2i}, \quad for \; s_1 \leq t \leq s_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$= a_{mi}, \quad for \; s_{m-1} \leq t \leq t_2$$

Which takes a different value for each of the m periods (s_{k-1}, s_k). This model was used by Heermann et al. (1968) and Jackson (1981).

Probabilities that vary continuously with time are more attractive. The simplest are polynomials

$$g_i(t) = \sum_{k=0}^{m} a_{ik} t^k.$$

Fourier series may also be used in which case

$$g_i(t) = a_{i0} + \sum_{k=1}^{m} \; [a_{ik} \sin(kt') + b_{ik}(kt')], \text{ where } t' = \pi(t-183)/183.$$

3.5 THE DEVIANCE

The fitting of any of the functions of $g_i(t)$ is analogous to fitting regression equations for normally distributed data. For normal data a residual sum of squares is calculated for each model. Different models are then compared by considering the magnitude of the difference between the residual sums of squares. With non-normal data, the analogous statistic is the deviance, which is given for a binomial data as

$$G^2 = 2\sum_{t=t_1}^{t_2} \{r_i(t) \ln[r_i(t) / \hat{p}_i(t) + [1 - r_i(t)] \ln[(1 - r_i(t))/(1 - \hat{p}_i(t))]\},$$

where $\hat{p}_i(t)$ is the fitted value of $p_i(t)$, and $\hat{g}_i(t)$ is the fitted value of $g_i(t)$ with the unknown parameters replaced by their estimates. If m parameters have been estimated and the model is the correct one then G^2 has approximately a χ^2 distribution with $(T - m)$ degrees of freedom.

CHAPTER FOUR

SIMPLE ANALYSIS OF DAILY RAINFALL DATA

4.1 INTRODUCTION

Throughout the world considerable effort is devoted to the collection of rainfall data. Many long records exist and most countries now have a reasonably dense network of rainfall stations. The work on data collection is not, at present, matched by a corresponding effort on analysis. This chapter of the thesis introduces the 'direct' methods of analysing daily rainfall data. Nyankpala, a suburb of Tamale notable for its maize production is used as a case study. The need for rainfall analysis has already been emphasized in chapter two and in this chapter simple methods are used to describe the analysis. The methods used are applied to agronomic questions on dry spells and on the start of the growing season for maize.

Section 4.2 discusses the data availability and software for the analysis, whilst section 4.3 considers how daily rainfall data is captured in Instat Climatic Software and graphical display of average rainfall. Section 4.4 discusses dry spells in the daily rainfall data. The start of the growing season and the definition of the growing season are also considered in section 4.5. In section 4.6 the end of the growing season has also been looked at whilst section 4.7 also considers the length of the growing season. The chapter ends with a conclusion on the results obtained and this is presented in section 4.8.

4.2 DATA AVAILABILITY

The data for Nyankpala covering 51 years records (1953-2003) was made available by the Ghana Meteorological Department in Accra. Unlike the advanced countries where climatic data are in computerised form, in Ghana the data extracted from the archives at the Meteorological Department were on pieces of paper. One of the major problems faced in the initial analysis was the time spent for entering all the daily data for the 51 years. The data was first entered into Excel and then transferred to the Instat software. This was arranged to forms a data matrix of 366×51 and a section is presented in Table A.1 in Appendix A. To ensure that any analysis performed draw valid conclusions, the quality of the data was investigated at the initial stage of the analysis. Monthly totals were calculated and compared to monthly totals supplied by the Meteological Department and where there was any disagreement, the necessary corrections were made. The data for 1965 was incomplete and was not used for the analysis.

Instat Software provides a climatic data base structure together with tailored programs within the software. These programs are partly to provide data entry, quality control and dense number of programs for climatic analysis.

4.3 DAILY RAINFALL DATA (1953-2003)

Table 4.1 shows a display of the daily rainfall data (mm) Nyankpala for the year 2002. This shows how daily rainfall data is displayed by day and month in Instat.

The codes "--" shows days of no rainfall in the data. It is clear from Table 4.1 that the first rainy day in 2002 occurred in March 9th with 6.4mm of rain while October 19th was the last rainy day in 2002 with 8.9mm of rain.

Table 4.1 DAILY RAINFALL DATA FOR 2002 AT NYANKPALA

Mon Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	--	--	--	--	--	--	19.2	1.6	0.5	2.7	--	--
2	--	--	--	--	--	6.1	--	--	--	--	--	--
3	--	--	--	--	--	--	6.1	--	20.7	--	--	--
4	--	--	--	--	--	2.0	--	8.8	--	--	--	--
5	--	--	--	40.2	19.1	--	3.3	32.3	9.1	0.5	--	--
6	--	--	--	--	--	9.6	--	--	--	--	--	--
7	--	--	--	--	--	--	--	1.7	5.5	--	--	--
8	--	--	--	--	5.0	--	--	--	--	--	--	--
9	--	--	6.4	--	--	20.2	--	--	--	--	--	--
10	--	--	--	--	--	--	--	--	5.4	12.4	--	--
11	--	--	--	--	--	--	8.2	--	--	--	--	--
12	--	--	--	--	--	--	--	--	--	16.5	--	--
13	--	--	--	22.0	8.8	0.7	--	15.3	--	7.8	--	--
14	--	--	--	--	--	--	--	8.0	19.1	7.9	--	--
15	--	--	--	--	--	--	--	--	--	--	--	--
16	--	--	--	--	--	--	29.2	--	--	--	--	--
17	--	--	--	0.2	--	2.3	2.9	--	--	--	--	--
18	--	--	--	--	4.7	18.2	9.3	--	34.6	--	--	--
19	--	--	--	--	7.9	--	--	--	--	8.9	--	--
20	--	--	--	--	--	43.0	--	25.6	--	--	--	--

21	--	--	--	--	--	--	24.2	85.6	--	--	--	--
22	--	--	-- --	33.8	--	--	--	--	--	--	--	--
23	--	--	--	--	--	--	15.8	--	0.3	--	--	--
24	--	--	--	--	--	34.9	--	--	2.5	--	--	--
25	--	--	--	--	--	--	--	--	--	--	--	--
26	--	--	--	--	--	23.6	--	--	--	--	--	--
27	--	--	--	--	--	--	--	23.0	--	--	--	--
28	--	--	--	--	--	--	--	3.3	--	--	--	--
29	--	--	--	--	--	0.9	--	4.7	--	--	--	--
30	--		--	8.6	--	--	15.0	--	--	--	--	--
31	--		--	--	15.5		10.3	3.5		--		--

From Table 4.1 we noticed that, there were 62 days of rainfall in 2002 with a maximum daily rainfall of 85.6mm which occurred in August 25th. August has the highest monthly rainfall in 2002 with an average of 97.7mm while the annual rainfall was below 1000mm of rain.

4.3.1 AVERAGE RAIN FALL

Rainfall variability constitutes the dominant character of the climate of Ghana. Figure 4.1 shows a graph of the average monthly rainfall in Nyankpala in 2002. It is clear from figure 4.1 that Nyankpala has a single rainy season with monthly totals increasing slowly from March and reaching its peak in August. Although monthly average rainfall or monthly rainfall totals are useful in defining the pattern of rainfall, it does not reveal the periods without rain which are useful for agricultural planning.

In section 4.4, we shall look in detailed the distribution of dry spells for different periods within the 51 years of daily rainfall data from Nyankpala.

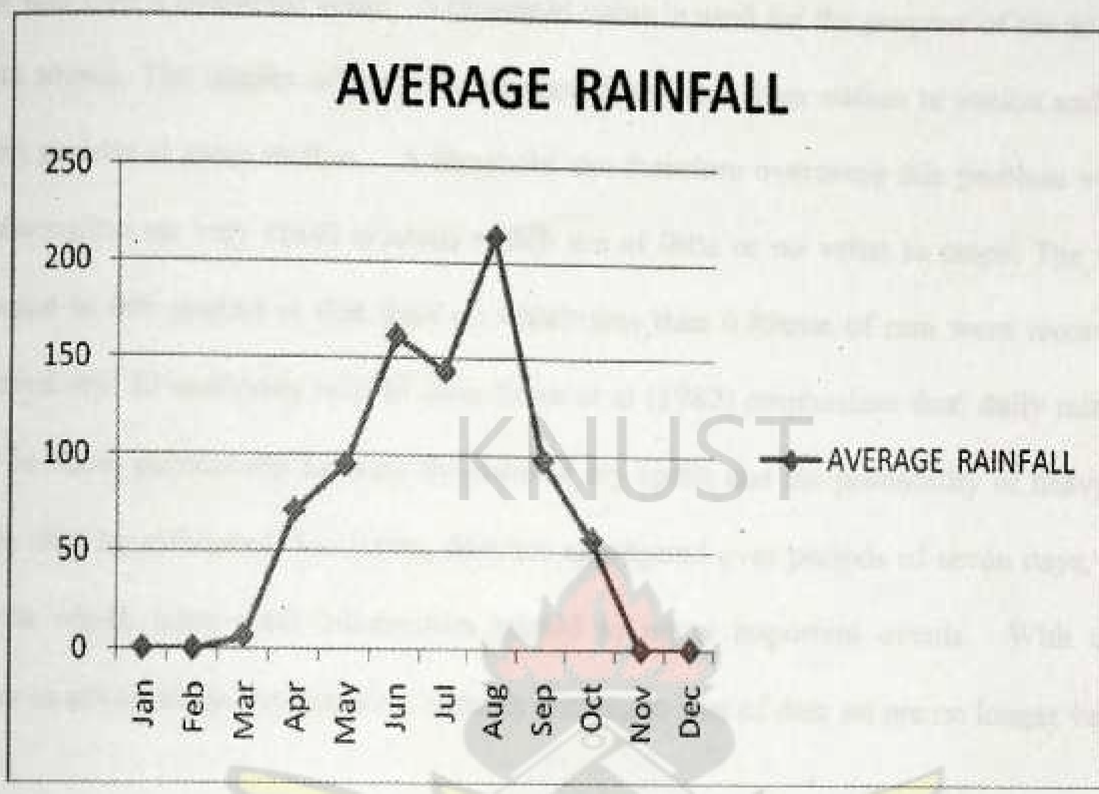


Figure 4.1 Average monthly rainfalls in Nyankpala in 2002

4.4 DRY SPELLS

Maize is known to be particularly sensitive to water shortage immediately after germination and flowering. Maize stomata do not recover after a severe drought of 7 or more days although leaves may recover turgidity. The knowledge of the relationship between the numbers of dry days in the period of maize growth could be of great economic and agronomic value especially in West Africa where lack of adequate and assured water from rains is a major setback in the yield of a crop.

Crops should therefore not be sown during periods of high risks of long dry spells as this will affect germination. A dry spell is defined as a sequence of consecutive dry days where rainfall is zero or less than a threshold value. A threshold value is used for the purpose of the accuracy of the data source. The quality of data records sometimes vary from station to station and possibly between periods at same station. A threshold can therefore overcome this problem while only lost information on very small rainfalls which are of little or no value to crops. The threshold value used in this project is that days on which less than 0.85mm of rain were recorded were considered dry. In analyzing rainfall data, Stern et al (1982) emphasizes that, daily rainfall data should be used, particularly because the risks of dry spells and the probability of heavy erosive rainfalls may be estimated. Too often, data are aggregated over periods of seven days, ten days or month which loses vital information related to many important events. With computer facilities as advanced as day are now, excuses relating to size of data set are no longer valid.

In this study, daily observations were used and each day was coded as dry or wet based on the threshold value. The sequence of wet and dry days in 2002 at Nyankpala is presented at Table 4.2. In the year 2002 it was clear that there was no dry spell of more than 8 days from May to September. During the whole year the longest dry spell was 73 days which began after October 19 and ended on December 31. There were no rains in the months of January and February. The early part of the year is coded as missing (m). In the absence of information at the end of 2001 it is not possible to give the spell lengths for 2002 until the first wet day. As shown in Table 4.2 the first rainfall was on 9th March. Table 4.1 earlier showed that the rainfall on 9th March was just over 6mm.

This is not enough for sowing a cereal crop and the results in Figure 4.2 show that there was a dry spell of 26 days ending on April 5th, so sowing early would be unlikely to be a success. Table 4.3 gives the maximum dry spell for each month.

Table 4.2 DRY SPELLS AND WET DAYS FOR 2002 AT NYANKPALA

Mon Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	m	-	m	23	1	1	--	--	1	--	13	43
2	m	-	m	24	2	--	1	1	2	1	14	44
3	m	-	m	25	3	1	--	2	--	2	15	45
4	m	-	m	26	4	--	1	--	1	3	16	46
5	m	-	m	--	--	1	--	--	--	4	17	47
6	m	-	m	1	1	--	1	1	1	5	18	48
7	m	-	m	2	2	1	2	--	--	6	19	49
8	m	-	m	3	--	2	3	1	1	7	20	50
9	m	-	--	4	1	--	4	2	2	8	21	51
10	m	-	1	5	2	1	5	3	--	--	22	52
11	m	-	2	6	3	2	--	4	1	1	23	53
12	m	-	3	7	4	3	1	5	2	--	24	54
13	m	-	4	--	--	4	2	--	3	--	25	55
14	m	-	5	1	1	5	3	--	--	--	26	56
15	m	-	6	2	2	6	4	1	1	1	27	57
16	m	-	7	3	3	7	--	2	2	2	28	58
17	m	-	8	4	4	--	--	3	3	3	29	59
18	m	-	9	5	--	--	--	4	--	4	30	60
19	m	-	10	6	--	1	1	5	1	--	31	61
20	m	-	11	7	1	--	2	--	2	1	32	62

21	m	-	12	8	2	1	--	--	3	2	33	63
22	m	-	13	9	--	2	1	1	4	3	34	64
23	m	-	14	10	1	3	--	2	5	4	35	65
24	m	-	15	11	2	--	1	3	--	5	36	66
25	m	-	16	12	3	1	2	4	1	6	37	67
26	m	-	17	13	4	--	3	5	2	7	38	68
27	m	-	18	14	5	1	4	--	3	8	39	69
28	m	-	19	15	6	2	5	--	4	9	40	70
29	m	-	20	16	7	--	6	--	5	10	41	71
30	m	-	21	--	8	1	--	1	6	11	42	72
31	m	-	22		--		--	--		12		73
Maximum		-								(Overall: 73)		
	31	29	22	26	8	7	6	5	6	12	42	73

In discussing dry spells care should be taken to differentiate between the ways by which spells lengths are calculated, since their interpretation can be misleading. Archer (1981) emphasizes that to show the change in probabilities more clearly, analysis of dry spells should be done on overlapping periods. Figure 4.2 presents the probabilities of a dry spell length of 7, 10, and 15 days in overlapping 30 day period from January 1st to December 31st.

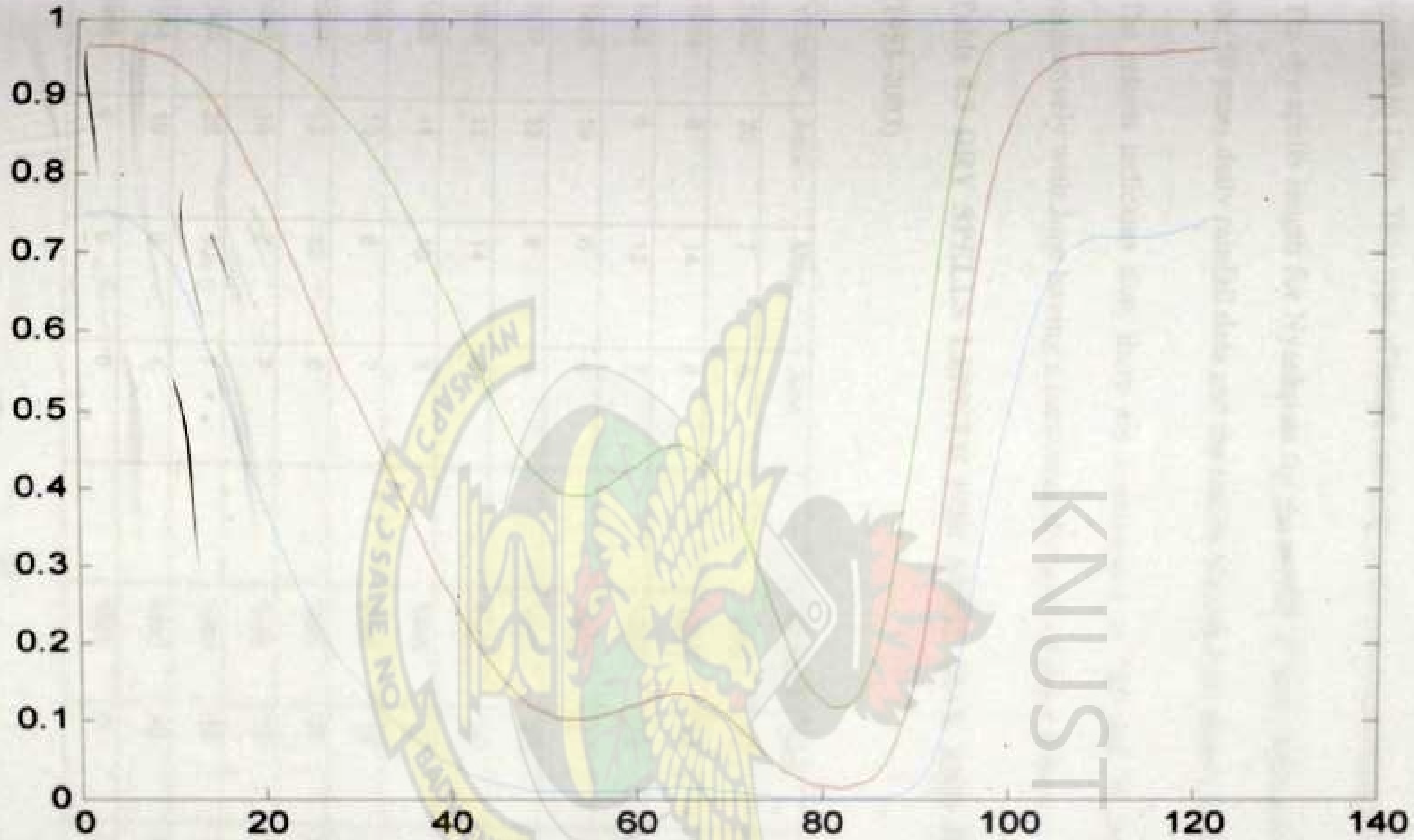


Figure 4.2 Nyankpala; Probability of dry spells of length greater than or equal to 7, 10 and 15 in the next 30 days.

The results show that the graphs of the chance of a dry spell of more than 5, 7 10, and 15 days, in overlapping 30 days period from January 1st to December 31st shows similar pattern.

The chance of a dry spell of 10 and 15 days have drop below 0.2 from March 20 to June 14th (day 90 to 126). This type of result can help to determine planting strategy.

The dry spells length for Nyankpala for the month of May, June and July was calculated for all the 50 years daily rainfall data and the results obtained are shown in Table 4.3.

The results indicates that, there are a maximum of 166 and 36 dry spells in April and May respectively with June having a maximum dry spell of about 8 days.

Table 4.3 DRY SPELLS LENGTH FOR APRIL MAY AND JUNE AT NYANKPALA. (1953-2003)

YR/MON	April	May	June		YR/MON	April	May	June
1953	20	7	6		1979	23	12	6
1954	8	14	6		1980	36	8	7
1955	6	12	7		1981	19	9	5
1956	19	6	8		1982	7	9	4
1957	12	8	9		1983	29	36	5
1958	11	14	6		1984	16	9	6
1959	11	10	5		1985	13	16	7
1960	10	6	7		1986	13	12	7
1961	12	12	8		1987	23	34	7
1962	16	7	7		1988	15	7	6
1963	26	10	7		1989	12	11	3
1964	19	9	4		1990	42	10	9
1966	8	9	6		1991	0	0	0

1967	8	6	5		1992	33	8	5
1968	9	10	6		1993	13	11	5
1969	7	5	12		1994	19	14	5
1970	30	6	9		1995	21	6	6
1971	9	9	8		1996	13	11	6
1972	15	3	7		1997	16	4	7
1973	8	9	6		1998	158	10	7
1974	18	9	10		1999	6	7	6
1975	7	15	6		2000	14	6	6
1976	166	7	4		2001	10	7	8
1977	30	8	11		2002	26	8	7
1978	22	8	10		2003	14	9	8

4.5 START OF THE GROWING SEASON

4.5.1 Introduction

In Ghana where rainfall is highly variable, there is a risk of crops failure due to water shortage whenever they are planted. This means that the planting date is crucial and by choosing it correctly the risk of crop failure due to water shortage may be minimised. In defining an event to mark the start of the growing season, Stern et al (1982), emphasized that daily rainfall values should be used as opposed to studies which have used rainfall amounts summarized over years making it difficult to interpret the resulting average dates.

4.5.2 Definition for the start of the growing season

Many possible criteria exist to define the start of the growing season because the start of the growing season for a particular crop may require more or less rain than other crops. Any definition used in rainfall analysis in West Africa due to the peculiar nature of the distribution of rainfall pattern should reflect on the farmers experience and the benefit of early harvest. Most authors have used rainfall amounts summarized over years in their studies of the start of the rains. Woodhead et al., (1970), Panabokke, (1974) considered percentage points of the total rainfall in successive 7 or 10-day periods in the analysis of daily rainfall data. While these can be a useful guide to which crops are viable, there are many other aspects of the rainfall pattern that are also important.

For example, in many areas of the seasonally arid tropics, crops must be planted early and the date of the start of the growing season may coincide with the first heavy rainfall. Davy (1976) observed that millet was often planted in Nigeria after occurrence of at least 20mm of rain over a 2-day period. The distribution of the date of this event is therefore of interest. Crops will be at risk from dry spells occurring during the growing season. The level of the risk can sometimes be assessed by evaluating the probability that a long dry spell occurs when plant is particularly sensitive, such as just after germination, or at flowering. A calculation of the chance of long dry spells through the season is therefore often useful. The criteria used by Stern et al (1982) for the start of the rains consisted of three components;

1. The start of the season is considered until after a stated date, D.

2. An event, E, then indicates a potential start date, defined as the first occurrence of at least (x) mm totaled over 't' consecutive days.

3. The potential start could be false start if an event, F, occurs afterwards, when F is defined as a dry spell of 'n' or more days in the next 'm' days.

This definition as stated has not been restricted to any particular crop neither is it based on any particular soil water capacity. The 'x'mm,'t' days, 'n' dry spell and 'm' should be chosen for different crops on different soils.

4.5.3 Rainfall amount summarized over days

The dangers in using rainfall amounts summarized over days as mentioned above is discussed in detail in this section. An example is used to illustrate the problem.

In this example rainfall amounts have been grouped over a period of seven days to determine when there is sufficient rainfall amount to mark the start of the growing season.

The example has been chosen with no reference to any particular place and the rainfall amounts are carefully chosen so that the real problem is revealed. The start of the rains is assumed to coincide with the first occurrence of 20mm of rain totalled over seven- day's period.

Table 4.4 RAINFALL AMOUNTS (MM) SUMMARISED OVER SEVEN-DAY TOTALS.

Day	Rainfall Amount (mm)	Totals
1	0	
2	10	
3	0	
4	5	
5	2	
6	2	
7	0	19
8	10	
9	2	
10	1	
11	2	
12	2	
13	1	
14	0	18
15	17	
16	4	
17	2	
18	0	
19	3	
20	2	
21	5	33

Table 4.5 SEVEN-DAY RUNNING TOTALS OF RAINFALL AMOUNTS (MM)

Day	Rainfall Amount (mm)	Totals
1	0	
2	10	
3	0	
4	5	
5	2	
6	2	
7	0	19
8	10	29
9	2	21
10	1	22
11	2	19
12	2	19

13	1	18
14	0	18
15	17	25
16	4	27
17	2	28
18	0	26
19	3	29
20	2	28
21	5	33

The calculations in Table 4.4 and Table 4.5 indicate two different methods in choosing a suitable planting date. Table 4.4 indicates when seven-day totals (no overlapping) are used and Table 4.5 when seven –day running totals are used. It is clear from Table 4.4 that week three (day 21) is the first occurrence of at least 20mm of rain totalled over seven consecutive days. A suitable planting time using this method is from 15th -21st. In table 4.5, the groupings of the running seven-day totals are as follows ; 1-7, 2-8,3-9, 4-10, 5-11,.... The first occurrence of at least 20mm is on day 8. By using the running totals we make efficient and detailed use of the daily data which leads to suitable time for the start of the growing season. In Table 4.4 the choice of the suitable time was limited to only three numbers while in Table 4.5 the choice came from 15 numbers. It should be noted that for most part of the years the start of the growing season will occur earlier if running totals are used instead of data summarised over days. In the proceeding sections we shall restrict ourselves to the use of running totals of rainfall amounts in the growing season.

4.5.4 Application of Definition to Nyankpala Data

The definition used in this project is stated as follows; the earliest date, D, when rainfall is sufficient to provide water equivalent to or greater than one half of reference crop evapotranspiration (ET_0) and remain greater when ET_0 for the remainder of the growing season provided that a dry spell of five days or more did not begin in the week after this date. The earliest date, D, defined here is chosen to be 1st May beyond which farmers would normally not do any planting. Studies that have used this definition (rainfall $\geq 1/2\mathrm{ET}_0$) include Benoit (1977) in the studies for the start of the growing season in Northern Nigeria. The reference crop evapotranspiration which is the effect of climate on crop water requirement is defined as " rate of evaporation plus transpiration from plants on extensive surface of 8 to 15 cm tall, green grass cover of uniform height, actively growing, completely shade the ground and not short " (Doorenbos and Pruitt, 1977). There are different methods used for calculating ET_0 values. The Priestly Taylor method is used in this project. The minimum requirements for the Priestly Taylor equation are the daily net radiation and the mean daily air temperatures. The Priestly Taylor equation predicts potential evaporation from a horizontally uniform surface as

$$\text{Potential Evaporation} = \alpha \frac{\Delta}{\Delta + \gamma}(R_N - G)$$

where α = Priestly-Taylor coefficient (taken 1.26)

Δ = rate of change of saturated water vapour pressure with temperature

γ = psychometric constant (here taken as $0.0006\text{mbar}/c^0$)

$R_N =$ net radiation(mm water /day)

$G =$ Soil heat flux (here considered to be zero over 24 hour period)

In order to calculate the rate of change of saturated water vapour with temperature (Δ), an equation was used in which saturated water vapour (e_s) was related to temperature (T) (°C) as

$e_s(T) = 6.108.10^{(7.5T/(273.3+T))}(mb)$, and

$$\Delta = \frac{de_s}{dT} = \frac{4098.e_s(T)^2}{(237.3+T)^2} \text{ (mb /°C)}$$

In this thesis the method of seven day running totals was used to get the rainfall amounts for each day and the definition for the start of the growing season defined above was used. The suitable planting date for each year was determined and the results of the analysis are presented in Table 4.6 and figure 4.3. Reference to table 4.6 indicates that for a sample of 50 years, early planting had occurred in 7 years.

The earliest date, D, is day 122 (May 1st) and the latest date is day 223 (August 10th) beyond which farmers would normally not do any planting. The mean starting date is day 142 (May 21st) and the standard deviation is 19 days.

Figure 4.3, shows a time series plot of the results of the start of the rains for the definition used. We noticed that there is great variability in the start days, ranging from 122nd day to 223rd day.

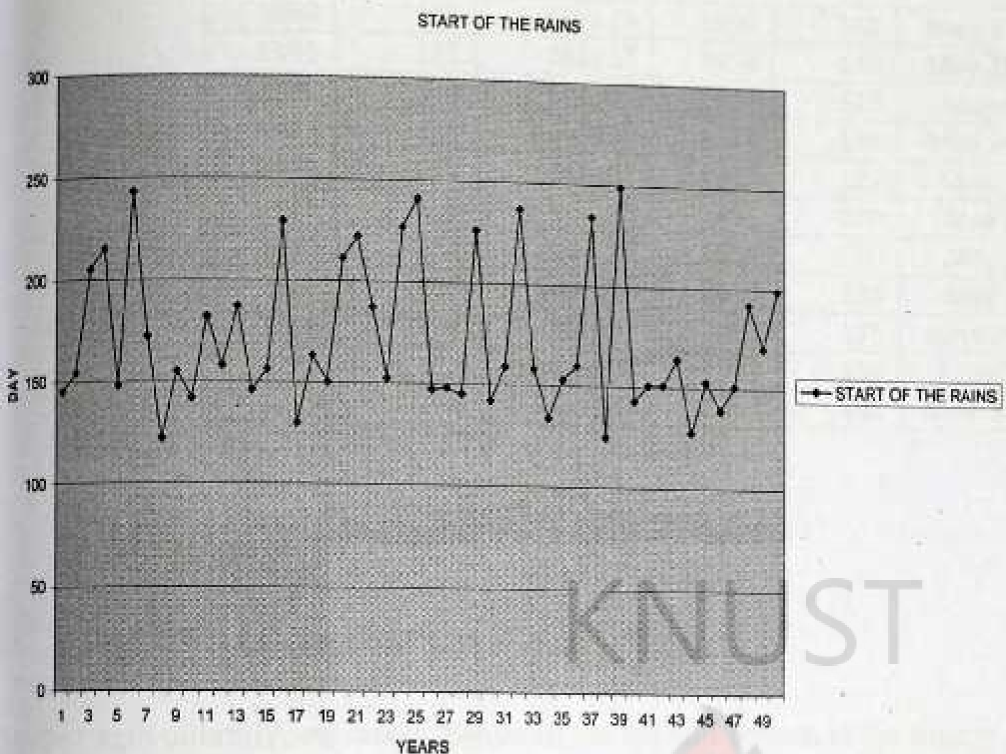


Figure 4.3 The start of the rains for the various years (1953-2003)

Table 4.6 THE START OF THE GROWING SEASON

YEAR	DAY	DATE	YEAR	DAYS	DATE
1953	129	May 8th	1979	147	May 26th
1954	154	June 2nd	1980	147	May 26th
1955	147	May 26th	1981	145	May 24th
1956	122	May 1st	1982	137	May 16th
1957	139	May 18th	1983	129	May 8th
1958	149	May 28th	1984	131	May 10th
1959	135	May 14th	1985	160	June 8th
1960	122	May 1st	1986	144	May 23rd
1961	154	June 2nd	1987	133	May 12th
1962	133	May 12th	1988	122	May 1st
1963	139	May 18th	1989	150	May 29th
1964	141	May 20th	1990	189	July 7th
1966	147	May 26th	1991	122	May 1st
1967	124	May 3rd	1992	154	June 2nd

1968	136	May 24	1993	142	May 21st
1969	122	May 1st	1994	146	May 25th
1970	130	May 9th	1995	125	May 4th
1971	163	June 11th	1996	135	May 14th
1972	122	May 1st	1997	122	May 1st
1973	128	May 7th	1998	126	May 5th
1974	154	June 2nd	1999	130	May 9th
1975	150	May 29th	2000	124	May 3rd
1976	124	May 3rd	2001	175	June 23rd
1977	223	August 10th	2002	154	June 2nd
1978	176	June 24th	2003	132	May 11th

4.6 THE END OF THE GROWING SEASON.

Another agriculturally important feature of the rainfall pattern is the timing of the end of the wet season. If this occurs too soon the crop may not have sufficient water to reach maturity. However, excessive wet weather may prevent ripening or harvesting. In general, crops will use stored soil moisture for growth beyond the end of the rains, and so the end of the growing season is the date when the soil profile is too dry for growth to continue.

This date can be evaluated by considering a water balance model with rainfall as input to the soil and evaporation (plus possibly runoff and drainage) as output.

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. 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, and what probabilities are of experiencing dry spells of various duration at critical times during the growing season.

In this study, a rainy day is regarded as a day with measurable rain; i.e. a day yielding 0.85mm or more. A rainless day (a dry day) has been defined as a day yielding less than 0.85 mm. This provides a cut-off to distinguish wet and dry days. The end of the season was defined as the first date the soil profile is empty after 1st September. This was defined by Raman (1974), which is adapted and modified by the + (INSTAT+Climatic software, 2001). Considering Table 4.7 below, the Water balance table of Nyankpala.

Table 4.7 WATER BALANCE OF NYANKPALA

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	--	--	--	--	2	37	31	36	41	50	--	--
2	--	--	--	--	--	32	26	31	36	46	--	--
3	--	--	--	--	--	27	32	26	31	41	--	--
4	--	--	--	--	--	34	27	25	43	36	--	--
5	--	--	--	--	--	++	22	20	47	31	--	--
6	--	--	--	35	--	55	17	15	42	30	--	--
7	--	--	--	30	--	++	15	10	37	33	--	--
8	--	--	--	25	--	55	10	5	32	28	--	--
9	--	--	--	20	--	55	5	--	57	23	--	--
10	--	--	--	15	--	53	--	24	++	18	--	--
11	--	--	--	10	22	48	--	30	++	13	--	--
12	--	--	--	18	17	53	--	25	++	8	--	--
13	--	--	--	13	12	48	2	33	55	22	6	--
14	--	--	--	18	7	43	5	28	51	17	1	--

15	--	--	--	13	2	57	--	24	49	54	--	--
16	--	--	--	8	13	52	13	19	44	49	--	--
17	--	--	--	3	8	47	++	15	39	53	--	--
18	--	--	--	--	3	42	56	15	53	54	--	--
19	--	--	--	8	--	42	51	31	++	49	--	--
20	--	--	--	3	28	37	51	26	++	44	--	--
21	--	--	--	--	23	32	46	21	++	39	--	--
22	--	--	23	--	18	27	41	16	55	34	--	--
23	--	--	18	3	13	22	36	43	50	29	--	--
24	--	--	13	--	8	17	31	38	54	24	--	--
25	--	--	8	1	3	12	26	33	49	19	--	--
26	--	--	3	--	--	7	21	28	44	14	--	--
27	--	--	--	--	6	2	43	23	45	10	--	--
28	--	--	--	--	4	46	38	18	40	5	--	--
29	--	--	--	1	--	41	33	26	++	--	--	--
30	--		--	--	--	36	33	32	55	--	--	--
31	--		--		--		28	46	--	--	--	--

The earliest date for the end of the season is day 245 (September 1) and the latest date is day 324 (November 20). The mean end of season date is day 285 (October11), and the standard deviation is about 22 days.

4.7 THE LENGTH OF THE SEASON

The length of the growing season refers to the time between the start of the rains and the end of the growing season (the end of the rains).

Table A.2 Appendix A shows the values of the various ends and the length of the growing season of Nyankpala from 1953 to 2003.

Looking at the table, the mean length of the season is about 144 days and the maximum length of the season is 192 days.

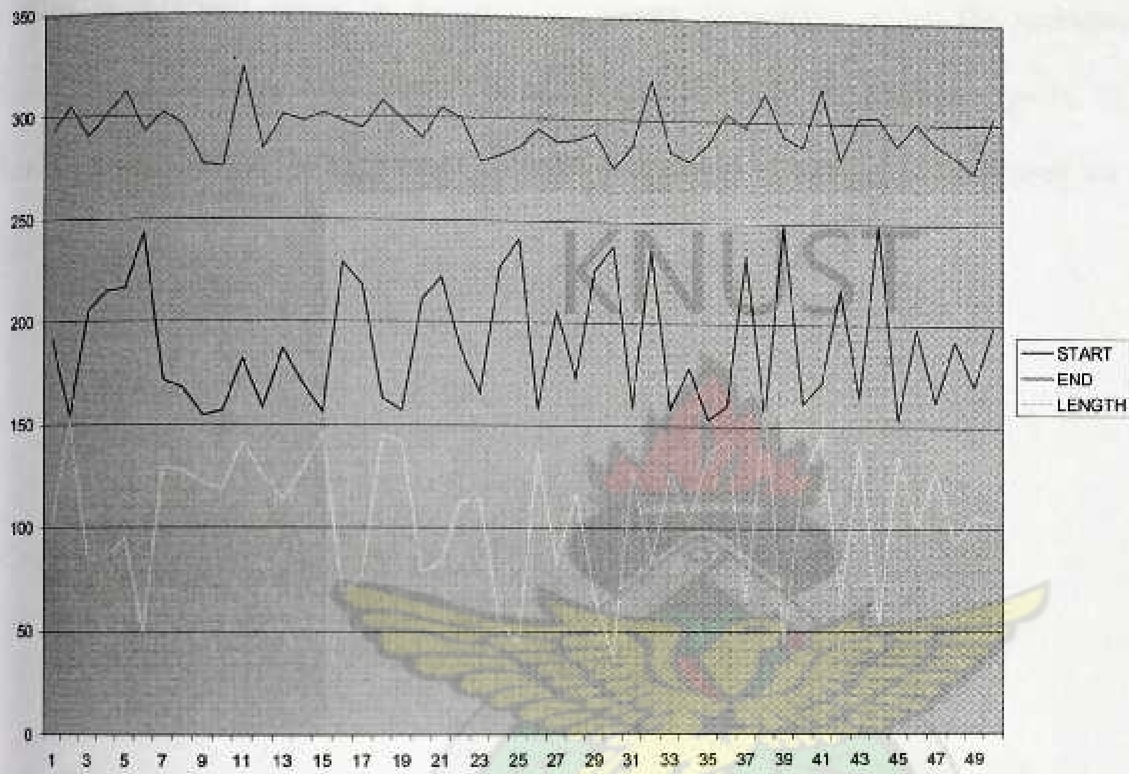


Figure 4.4 The start, end and the length of the growing season.

4.8 CONCLUSION

From the exploratory analysis discussed in this chapter, there was no significant change in climate so far as the rainfall patterns are concerned. The start days appear to be randomly scattered over the years, showing no particular increase or decrease over the years, taking dry spells into account. In terms of change, this would imply that there does not seem to be much evidence of increased risks for farmers over the years.

We proceed to the next chapter where the analysis went further to quantify the probability of rain, the amount of rain on any particular day and the probability of long dry spells of different lengths at different times in the growing season particularly round the sensitive period of flowering to consider how these findings might relate to the farmers' needs. In this work, however, we confined ourselves to consider changes in climate as it influenced the start of the rainy season.

CHAPTER FIVE

MODELING THE DAILY RAINFALL SEQUENCE BY MARKOV CHAIN

5.1 INTRODUCTION

Agricultural scientists and statisticians have shown considerable interest in the modelling and simulation of daily rainfall data. Many methods for anaylsing rainfall data has been used but among the proposed methods, a combination of Markov Chain model for analyzing the probability of rainfall occurrence and the gamma distribution function for the amount of rain are particular popular.

5.2 MODELING THE OCCURRENCE OF RAIN

To use a Markov chain model for the probability for the occurrence of rain each day is classified as a wet day (w) if the amount of rain is greater than or equal to a threshold value δ mm or a dry day (D) if the amount of rainfall is less than δ mm. δ mm is consider to be of doubtful value both to the crop and in terms of the reliability with which the data are recorded. The classification of a sequence W's and D's can be regarded as a two - state Markov Chain with wet and dry day as the two states.

It is assumed that the probability of rainfall on any day depends only on whether the previous day was wet or dry, i.e. whether rainfall did or did not occur. Given the event on a previous day, the probability of rainfall is assumed independent of events of further preceding days. Such a probability model is referred to as a first order Markov chain model whose parameters are the two conditional probabilities:

$P_1 = pr\,\{wet\ day\,/\ \ previous\ day\ wet\}$
$P_0 = pr\,\{wet\ day\,/\ \ previous\ day\ dry\}$

The model is fitted to the T days of the year from day t_1 to day t_T.

$$\begin{aligned} Let\ R(t) &= 1 \quad if\ x(t) \geq \delta mm \ \ imply\ wet\ day \\ &= 0 \quad if\ x(t) < \delta mm \ \ imply\ dry\ day \end{aligned}$$

Where $x(t_i)$ is the amount of rain on day t and $t = t_1 t_2 t_3 \ldots t_T$

The first order Markov chain is

$$P[R(t) = 1 / R(t-1), R(t-2) \ldots\ldots] = 1 / R(t-1) = i] \quad i = 0,1$$

Similarly in a second order Markov chain it is assumed that the probability of rainfall occurrence on day t depends on the state of the two previous days. Thus

$$\begin{aligned} P[R(t) &= 1 / R(t-1), R(t-2), \ldots\ldots] \\ &= 1 / R(t-1) = i, R(t-2) = h] \\ &= p_{hi}(t),\ where\ h = 0,1,\ i = 0,1]. \end{aligned}$$

The usual analysis of Markov chains assumes stationarity that is

$$p_{hi}(t) = p_{hi},\quad t = t_1, t_2, \ldots\ldots t_T.$$

In general $p_{hi}(t)$ are estimated from observed values of $\eta_{hij}(t)$, the number of times day t is in state $x(t) = j[(j = 1,\ wet\ and\ j = 0\ dry)]$ with $x(t-1) = i,\ x(t-2) = h$, over the years.

The conditional probabilities are thus estimated by the appropriate relative frequencies. These are the maximum likelihood estimates $\hat{r}_i(t) = \dfrac{\eta_{i1}(t)}{\eta_{i1}(t) + \eta_{i0}(t)}$

where $\hat{r}_i(t)$ is the proportion of rain days wet for day t. For a second order model

$$\hat{r}_{hi}(t) = \frac{\eta_{hi1}(t)}{\eta_{hi1}(t) + \eta_{hi0}(t)}.$$

Higher order chains (i.e. greater than second order) have been used in few studies. The strength of the Markov chain model lies in its simplicity. All the properties of rainfall occurrences are derived from the model and require only the calculation of the transition probabilities.

5.3 MODELING THE TRANSITION PROBABILITIES

Modelling the transitional probabilities of the first order chain is indicated here. Higher order models can be applied accordingly. The individual number of rainy days $n_{i1}(t)$ are binomially distributed with probability of rain $p_i(t)$ and the number of trials $n_{i0}(t) + n_{i1}(t)$. In the fitting process ordinary regression cannot be applied since the $\hat{r}_i(t)$, the estimates of $p_i(t)$, are proportions.

We let $p_i(t) = f(g_i(t))$ $i = 0, 1$ where f is known function and $g_i(t)$ is any function linear in unknown parameters. The logit transformation is applied hence $g_i(t)$ gives a value of $p_i(t)$ as [0, 1].

The function $g_i(t)$ can thus be modelled in a number of ways:

1. $g_i(t) = a_i$ $i = 0,1$. The model assumes constant probability of rain.

2. $g_i(t) = \sum_{h=0}^{m} a_{ih} t^h$. Polynomials with time of the year as the independent variable.

3. $g_i(t) = a_{i0} + \sum_{h}^{m} a_{ih} \sin\frac{2\pi ht}{366} + b_{ih} \cos\frac{2\pi ht}{366}$ which is a Fourier series where m is the number of harmonics fitted. The Fourier series have the advantage that the function is continuous at the ends of the year. However, the selection of harmonics can be quite challenging. This major question in fitting data to the function using Fourier series is which harmonics (terms of the form $\left(a_{ih} \sin(\frac{2\pi ht}{366}) + b_{ih} \cos(\frac{2\pi ht}{366})\right)$ should be retained in estimating $g_i(t)$? If a_0, a_1 *and* b_1 are the only constants considered to be different from zero, then the fitted function is unimodal and symmetrical. Retention of any further harmonics may indicate bimodality in the fitted curves.

5.4 APPLICATION OF MARKOV CHAIN MODEL TO NYANKPALA DATA

5.4.1 Structure of the Data

The Nyankpala data are analysed in this section. Stern et al. (1980) used first order Markov chains in a comparison of the climate of eleven places in West Africa.

The motivation then is to see how adequately Markov chains fit the Nyankpala data and to find the number of harmonics that also fits the data. In each year, each day is classified as wet day if $X(t) \geq 0.85mm$, and dry day if otherwise. $X(t)$ is defined as the amount of rain recorded on day t. Four vectors are thus required summarised over the years as follows:

- number of dry days with previous day dry
- number of dry days with previous day wet
- number of wet days with previous day dry
- number of wet days with previous day wet

After summarizing the daily data into these categories the data were put into groups of 5 days to produce shorter vectors. The advantage is that the analysis with Instat is quicker and requires less storage. The assumption here is that there is not much variability within each group. Instat was used to fit the model using Fourier series.

5.4.2 The Order of the Markov chain

The simplest model is zero order Markov chain. Here the probability of rain on any given day is estimated by the proportion of rain days on that day. The first order model has been found to be adequate for many practical purposes in most countries. A second order model has also been found necessary for some data especially in stations in the Indian sub-continent. Zero, first and second order Markov chain models were therefore fitted to the data. The results of the comparison of the models are presented in Table 5.1. These results follow the fitting of seven alternative models to the daily rainfall data.

Table 5.1 Results from fitting Markov chains of various orders to the data from Nyankpala, Ghana with time dependence modelled by Fourier series to the data in 5-day group

			Deviances of Models fitted			
Harmonics(m)	Degree of freedom	$p(r)$	$p(r\|r)$	$P(r\|d)$	$p(r\|rr)$	$P(r\|dd)$
1	70	9.17	952.44	7718.5	1207.87	6276.29
2	68	4.02	379.14	2675.38	761.88	2820.67
3	66	2.22	254.05	2513.1	644.16	2662.26
4	64	0.93	247.43	1672.76	624.92	1862.04
5	62	0.87	240.844	1543.44	569.32	1714.75

From Table 5.1, the largest deviances occurred by fitting a second order Markov chain models. This makes it obvious that the second order Markov chain curves gives a poor fit.

The largest deviance for fitting the more complex model of the second order models indicates that they are not required. It can then be conclude that a first order Markov chain is adequate for modelling the data.

The model fitted is presented in figure 5.1 whiles the parameter estimates and the results of fitting the first order Markov chain with the time dependence modelled by Fourier series to 5-day group data from Nyankpalan is also presented in table 5.2 and 5.3 respectively.

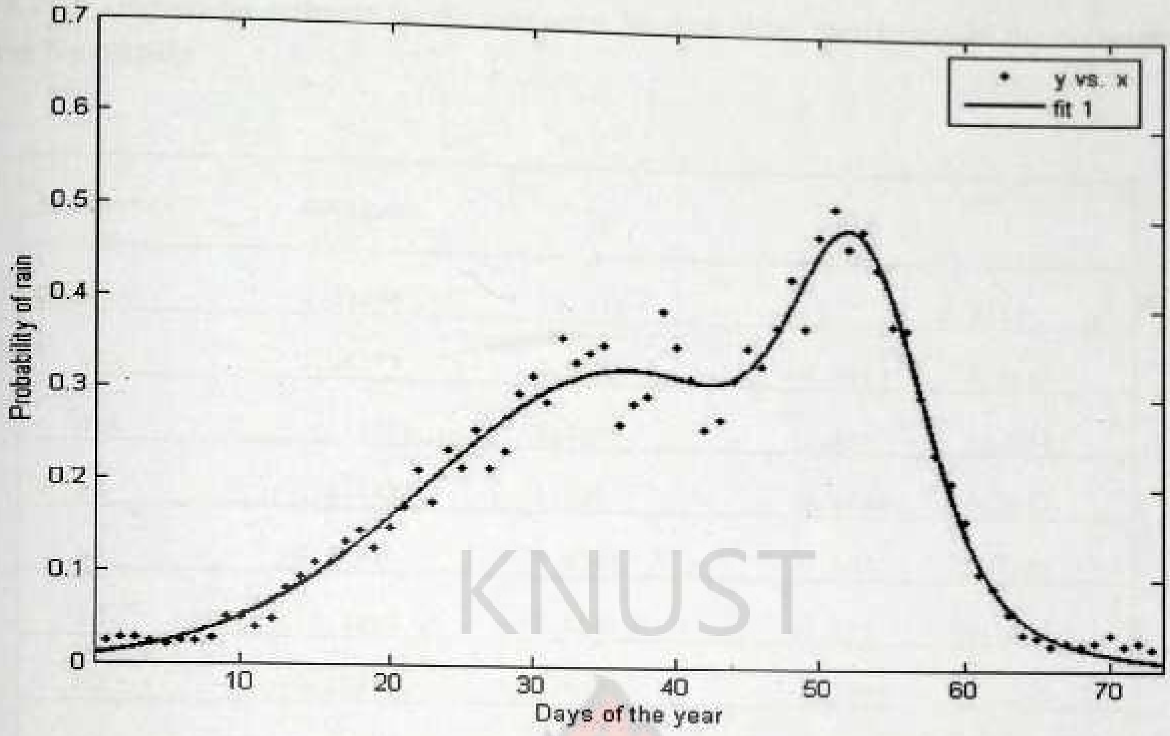


Figure 5.1 Fitted Fourier curve for first order probability of rain with five harmonic.

The notable features of the graph are as follows:

- The first order probability of rain curve has a unimodal pattern. The diagram in figure 5.1 shows clearly that the probability of rain increase gradually from January to July where it falls slightly till the end of July.
- The curve increase till second week in September where it attains its peak.
- The curve falls drastically from second week in September for the rest of the year.

Table 5.2 Parameter estimate for the first order Markov chain used to model the occurrence of rain in Nyankpala

Parameter	Estimate	SE	95% CI	
Const	0.37595	0.038	0.3	0.4519
a1	0.05377	0.0481	-0.0423	0.1499
dc1	-1.1814	0.0598	-1.299	-1.064
a2	0.27443	0.055	0.1646	0.3843
b2	0.60787	0.0524	0.503	0.7127
a3	-0.1208	0.0541	-0.229	-0.0126
b3	-0.18365	0.0531	-0.2898	-0.0775
a4	-0.03945	0.0519	-0.1432	0.0643
b4	-0.07873	0.0526	-0.184	0.0265
a5	0.01957	0.0469	-0.0742	0.1134
b5	0.05798	0.0471	-0.0361	0.1521

Table 5.3 Results of fitting the first order Markov chain with the time dependence modelled by Fourier series to 5-day group data from Nyankpala

Harmonic(s)	Degrees of freedom	Deviance
1	70	952.44
2	68	379.14
3	66	254.05
4	64	247.43
5	62	240.84

5.4.3 Conclusion

The probability of rainfall occurrence presented with the first order Markov chain model, enabled us to use the rainfall data sufficiently. The first order Markov chain model is found to be fitting well and the results obtained is presented in figure 5.1. The results obtained from fitting the first order Markov chain model are very useful for agricultural planning which indicates the potential of the Markov chain models for modelling the occurrence of rain. The ways such models can be used are discussed in Chapter Six.

5.5 MODELING RAINFALL AMOUNTS

5.5.1 Introduction

The daily rainfall amounts have a J-shaped distribution with a large proportion of small values and a few large values. The distribution is therefore skewed to the left. The Gamma distribution has been widely used to model the distribution of the daily amounts of rain falling on days when rain occurs and for rainfall amounts for longer periods.

The density function of the gamma distribution is given by

$$f(x,k,\mu)=\begin{cases}\dfrac{(k/\mu)^k x^{k-1}\exp[-kx/\mu]}{\Gamma(k)} & \mu,k,x>0\\ 0, \quad elsewhere & \end{cases}$$

Where $\Gamma(k)$ is the gamma function defined by

$$\Gamma(k)=\int_0^\infty x^{k-1}\exp(-x)dx$$

The Gamma distribution has two parameters, the mean amount of rain on a rainy day (μ) and k the shape parameter. $1/k$ is the square of the coefficient of variation. When $k = 1$, the distribution of the amounts is exponential and is a special case of the Gamma distribution.

It is found that generally there is a seasonal variation in the mean rain per rainy day (μ) but k may be assumed to be constant through the year. Since μ is related to time of the year, t, by defining a function $\log \mu(t) = g(t)$, $\mu(t)$ is always positive $(\mu(t) > 0)$.

The function $g(t)$ can modelled by any of the methods described in section 5.2. This study uses the Fourier series approach and the reason is the same as discussed in section 5.2.

5.5.2 Fitting the model

Given that the amounts of rain recorded on day t are $x_i(t)$, $i = 1,\ldots\ldots,n(t)$, $t = t_1,\ldots t_T$, the distribution of $x_i(t)$, depends on the time of the year, t , and the events of previous days.

The dependence of $\mu(t)$ on $R(t-1), R(t-2),\ldots.$ then incorporates the Markov chain modeling in the fitting process. The fitting of the first order Markov chain is discussed in this section.

The total number of rainy days on day t, $n(t)$ calculated by $n(t) = n_{01}(t) + n_{11}(t)$ where $n_{01}(t)$ is the number of rainy days given that it rained on day t and dry on day $t-1$ and $n_{11}(t)$ is the number of rainy days given that it rained on day t and day $t-1$. In fitting the model, eight vectors are required summarized over the years as:

- Number of dry days with the previous day dry

- Number of dry days with the previous day wet
- Number of wet days with the previous day dry
- Number of wet days with the previous day wet
- Sum of the amounts on wet days with previous day dry
- Sum of the amounts on wet days with previous day wet
- Sum of the logs of the amounts on wet days with previous day dry
- Sum of the logs of the amounts on wet days with previous day wet

The analysis proceeds by fitting separate gamma distribution to the mean amount of rain per rain day for rainy days following rainy days $(\mu_1(t))$ and for rainy days following dry days $(\mu_0(t))$.

5.5.3 Selection of harmonics

In fitting data to the function $g(t)$ using Fourier series, the right number of harmonics should be selected so that the model does not either over or under fit the data.

If we fail to choose the appropriate number of harmonics we may end up by presenting misleading results.

For gamma model, the natural parameter is $-1/\mu$ and the scale parameter is $1/k$. Thus k^* the deviance is approximately χ^2. k is unknown and by calculating the mean deviance from the analysis of deviances, the ratio of the mean deviances where the denominator is the within – day deviance has an approximate F distribution (Baker and Nelder, 1978). The within – day deviance D^2 is the deviance obtained when a separate mean is fitted to each value of t.

$$D^2 = 2\sum_{t=t_1}^{t_2} n(t)[\log \bar{x}(t) - \overline{\log x(t)}]$$

where

$$\overline{\log x(t)} = \sum_{j=1}^{n(t)} \log x_j(t) / n(t)$$

$$D^2 = 2\sum_{t=t_1}^{t_2} [n(t)\log \bar{x}(t) - \sum_{j=1}^{n(t)} \log x_j(t)]$$

5.5.4. Estimating the shape parameter k

There are two methods of estimating k. These are the crude and the maximum likelihood estimates. The method of maximum likelihood was used in this project. The maximum likelihood estimate of K is obtained by the solution of $\log k - \psi(k) = D^2 / 2n$

(Shenton and Bowman, 1977) where n is the total number of rainy days, $\sum_{t=t_1}^{t_2} n(t)$ and $\psi(K)$ is a digamma function defined by $\frac{d}{dk}\Gamma(k)$.

$$\frac{D^2}{2n} = \sum_{t=1}^{t_2} [\frac{n(t)\log \bar{x}(t)}{n} - \frac{n(t)\overline{\log x(t)}}{n}]$$

Using $\overline{\log x(t)} = [\sum_{j=1}^{n(t)} \log x_j(t)] / n(t)$

$$\frac{D^2}{2n} = \sum_{t=1}^{t_2} [\frac{n(t)\log \bar{x}(t)}{n} - \sum_{j=1}^{n(t)} \frac{\log x_j(t)}{n}]$$

$$\log K - \psi(K) = \sum_{t=1}^{t_2} [\frac{n(t) \log \bar{x}(t)}{n} - \sum_{j=1}^{n(t)} \frac{\log x_j(t)}{n}] \qquad (1)$$

$$\log K - \psi(K) = \log A - \log G$$

There are solutions to equation (1) presented by Person and Hentley (1970) in the form of tables.

5.6 APPLICATION OF THE GAMMA DISTRIBUTION TO THE NYANKPALA DATA

This section applies the methodology discussed above to the Nyankpala data. After data organization, the first analysis was to find which model to fit the data. This involved model selection which is simply a comparison of regression problem. The models were fitted using a threshold value of 0.85mm. Table 5.4 shows the analysis of deviance produced by fitting Fourier series to the data. These results follow the fitting two alternative models to the data.

Table 5.4 The analysis of deviance for selecting the order of the Markov chain using 5 harmonics on the data.

Source	Deviance (D)	Degrees of freedom	Model
1 Curve	689.47	365	I
2 curves	566.35	364	II

To select a suitable model, the deviances were compared. The reduction in deviance for fitting the second model (two curves) indicates that model II was better than model I and II was then chosen otherwise model. This shows that two curves (first order) is adequate to model the Nyankpala data, that is, the mean rain per rainy day depends mostly on the conditions of the previous day.

5.6.1 SELECTION OF HARMONICS

The analysis of deviances for selecting the number of harmonics to be used in the Fourier series is presented in Tables 5.5 with the parameter estimates at Table 5.6.

The total variation is broken down into the within-day deviance and between-day deviance. The contributions from the harmonics and residual terms constitute the between-day deviance.

Table 5.5 Analysis of deviance for selecting harmonics, 5 day groups used

Source	Degrees of freedom	Deviance	Mean Deviance	F
Between day	72	1394.5		
1 harmonic	2	364.90	182.45	1.56
2 harmonic	2	233.81	116.905	1.11
3 harmonic	2	211.69	105.845	1.02
4 harmonic	2	207.36	103.68	1.02
5 harmonic	2	203.84	101.92	36
Residual	62	172.91	2.79	93
Within day	18228	566.35	0.03	

Total	18300	1960.85		

Significant at 5%

Table 5.6 The parameter estimates for the first order Markov chain used to model the mean rain per the mean rain per rain day Nyankpala.

Parameter.	Estimate	SE	Prob>|t|	95% CI	
Const	2.2213	0.048	0.0000	2.125	2.317
a1	0.01252	0.0555	0.8221	-0.0984	0.1234
b1	0.76854	0.0785	0.0000	0.6116	0.9254
a2	-0.03635	0.0691	0.6010	-0.1745	0.1019
b2	-0.44726	0.0669	0.0000	-0.581	-0.3135
a3	0.07366	0.0691	0.2907	-0.0645	0.2118
b3	0.17613	0.0654	0.0091	0.0454	0.3069
a4	-0.055	0.063	0.3857	-0.1808	0.0708
b4	-0.05962	0.0628	0.3459	-0.1851	0.0659
a5	0.01042	0.0519	0.8415	-0.0933	0.1142
b5	0.05285	0.052	0.3136	-0.0511	0.1568

The results indicate that five harmonics are required and the plot of the observed and fitted values of the mean rain per rain day is presented at Fig. 5.5.

The plot shows that the model ~~implies a~~ steady rising amount of rain from early January and reaches a constant level (15 mm) between May and early June and falls slightly in July until August when reaches a constant level (15 mm).

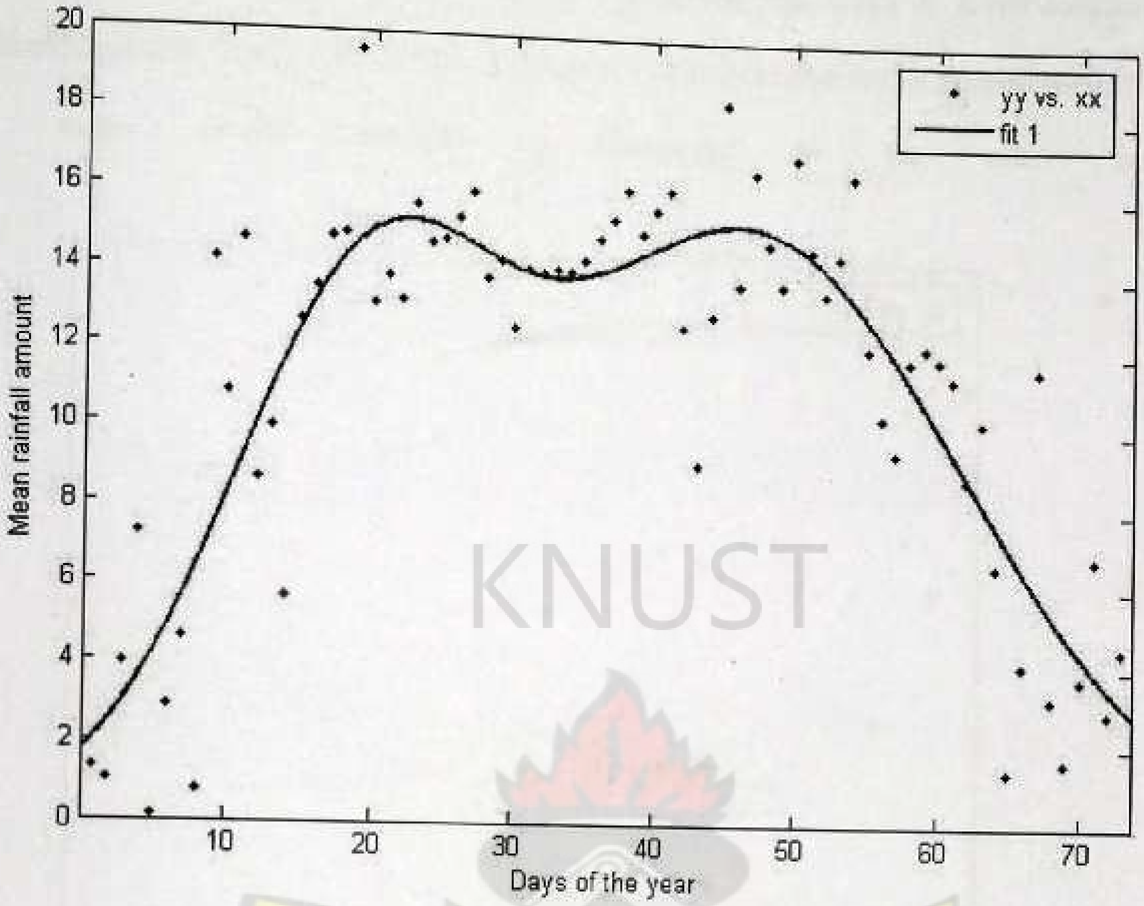


Figure 5.5 Mean rain per rain day at Nyankpala. Fitted Fourier curve with five harmonics and observed values.

5.6.2 ESTIMATING THE SHAPE PARAMETER K

The maximum likelihood estimate of K is calculated as $K = 0.6760$. By using tables which gives solution to $\log(k) - \psi(k) = D^2/2n$. The analyses have assumed that K is constant throughout the year. As a test of this assumption the within day deviance for each (5-day group) mean was plotted against t in fig. 5.6. The plot indicates that K may be slightly smaller

($1/K$ slightly greater) during August than at any time of the year. Thus K is not constant. It is not coincidence that K has a lower value in August. The implication is that August tends to have lower mean rainfall amount as compared to the other months.

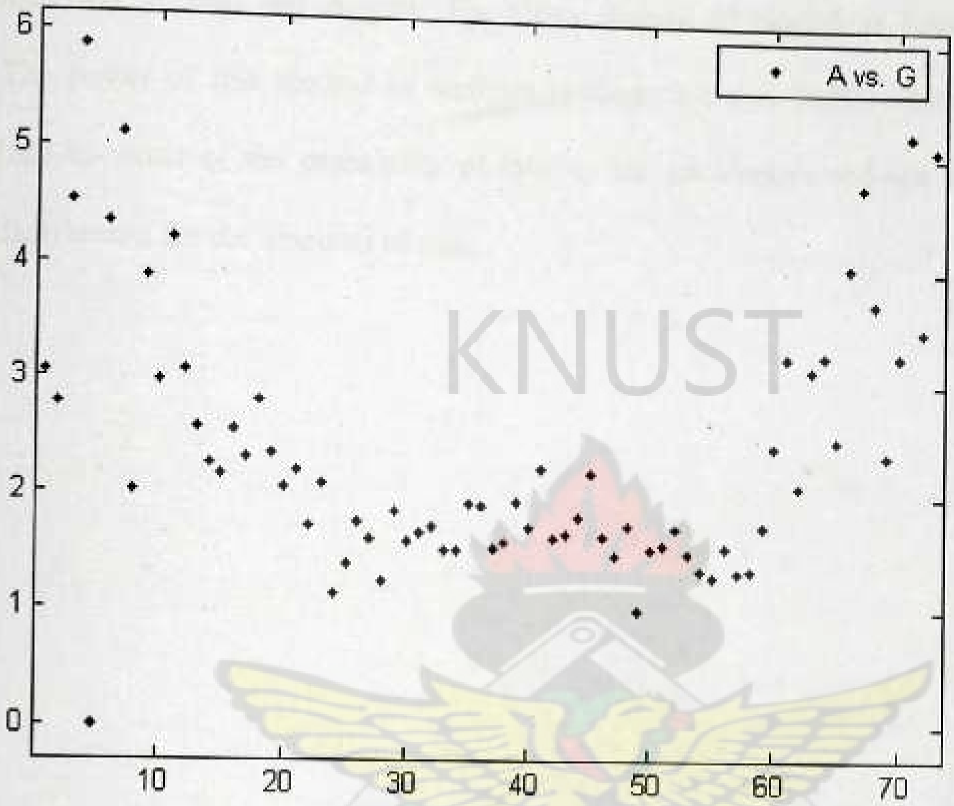


Figure 5.6 Gamma model for the amount of rain at Nyankpala, a plot of the deviance for each 5-day group by mean time.

5.7 CONCLUSION

The results of the analysis of the modelling approach indicated that for the probability of rain a first order Markov Chain was adequate to model data. From the Markov Chain analysis, the results further indicated that Nyankpala, a site in the Savannah zone has a unimodal pattern of rainfall distribution. The unimodality is less marked for probability of rain given dry than probability of rain given rain.

The mean rainfall per rainy day varies throughout the whole year with a maximum rainfall of 15mm occurring in May and August. The mean amount of rainfall is lower during the dry season. The power of this method of analysis is clear; a set of 18300 observations has been summarized in terms of the probability of rain by ten parameters and ten parameters by the gamma distribution for the amounts of rain.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

This thesis shows how a class of models may be fitted to daily rainfall data and then used to provide the type of information that agricultural planners commonly require from the data. The main objective of the study is to analyse daily rainfall data to solve the agricultural problems in Nyankpala and provide an account of the need for rainfall data analysis which is crucial to crop planners and agronomists in planning various crop operations. We have examined the most efficient methods currently being used for analysing daily rainfall data. These methods for analysing daily rainfall data have been reviewed in chapter two. The case for using these models is made in two stages. The first is the claim that a comprehensive analysis of rainfall data should use daily records and not based on 7-, 10-day or monthly totals. There is then a choice between a direct method (preliminary) analysis of the characteristics of interest and an analysis via a model of the pattern of rainfall on daily basis. The preliminary analysis and the indirect methods were both used in this work.

The conceptual and theoretical framework of the Generalised Linear Models and its examples has been presented in chapter three. In chapter four, a preliminary data analysis was conducted to look at dry spells, start, end and the length of the rains using Instat Climatic Software. Chapter five deals with the statistical modelling of the daily rainfall data using Markov chains and Gamma distribution. In conclusion, the analysis done with the use of Markov chain modelling techniques enabled us to use the rainfall data sufficiently. The Markov chain models are found to be fitting well and the results obtained from these techniques are also very useful for agricultural planning. More importantly, the Markov chain modelling techniques were successfully used in the study of the probability of long dry spells in the growing season of the study area. An important feature of the preliminary analysis is its simplicity, making it attractive to non statisticians. We have however found out that the types of models fitted are also quite easy for non-statisticians to understand. The model is not difficult to explain in terms of ''chances of rain'' and the ''amount of rain when it occurs''. Analysis of rainfall data between 1953 and 2003 indicated that the earliest date, D, for the start of the rains defined here is chosen to be day 122 (May 1) and the latest date is day 223 (August 10th) beyond which farmers would normally not do any planting. The mean starting date is day 142(May 21st) and the standard deviation is 19 days. We noticed that there is great variability in the start days, ranging from day 126 to day 200. The earliest date for the end of the season is day 245 (September 1) and the latest date is day 324 (November 20). The mean end of the season date is day 285 (October11), and the standard deviation is about 22 days. We also noticed that, the mean length of the season is about 113 days and the maximum length of the season is 189days. The minimum length of the season is 19 days with a standard deviation of about 41 days. The aanalysis of the daily rainfall data between 1953 and 2003 also indicated that the best time for planting maize at Nyankpala is during the last two weeks in May in order to meet the moisture requirements during flowering and the growing season length. The May planting favours the maize crop since a substantial portion of the rains for the season will fall during the growing period until maturity.

In the forecasting of the daily rainfall occurrences, the first order Markov model was accurate up to 86% and no improvements were observed by developing higher order models. Hence, considering the complexity involved, developing higher order Markov models to forecast rainfall occurrence in the northern region of Ghana cannot be justified. Markov chain modelling indicated that the mean amount of rainfall is 15mm per day occurring in May.

The following future studies are recommended so that the results obtained from this work can be used in the operational works of the National Meteorological Services Department of Ghana.

As it was mentioned in the previous section, daily rainfall data analysis is found to be very useful for agricultural planning. Hence, this work should be extended to other drought prone areas and to all over the country at large. Looking at the general characteristics of rainfall distribution in every region of Ghana and identifying risks in time and space with respect to agricultural activity which are very useful for agricultural planning.

REFERENCES

1. Abramowitz, M. and Stegun, I A. (1968) *Handbook of Mathematical Functions.* New York: Dover.
2. Adamowsky, K. and Smith, A. F., 1972. Stochastic generation of rainfall. *J. Hydraul. Eng., ASCE*, **98**, 1935–1945.
3. Adedokun, J. A. (1979) Towards achieving an in-season forecast of West African precipitation. *Arch. Met. Geoph. and Biokl.* Ser A, 28, 19-38.
4. Akaike, H., 1974. A new look at statistical model identification. *IEEE Trans. Auto. Control*, AC **19**, 716–722.
5. Archer D. R. (1981) Rainfall sequences in Northern Malawi. *Weather,* 36, 2-9.
6. Andrie. A Markov (1856-1922) Finite Mathematics, Margret Lial, Raymond N Greenwell and Nathan P. Ritchey.
7. Baker, R. J. and Nelder, J. A. (1978) *The GLIM System.* Oxford: NAG.
8. Bardossy, A. and Plate, E.J. 1992. Space time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res.*, **28**, 1247–1259
9. Benoit, P. (1977) The start of the growing season in Northern Nigeria. *Agric. Meteor.,* 18, 91-99.
10. Buishand, T. A. (1977) *Stochastic Modelling of Daily Rainfall Sequences.* Mededlingen Landbouwhoge School, Wageningen.
11. Caskey, J. E. (1963) A Markov chain model for the probability of precipitation occurrence in intervals of various lengths. *Mon. Weather Rev.*, 91, 298-301.

12. Chang, T.J., Kavvas, M.L. and Delleur, J.W., 1984. Daily precipitation modelling by discrete autoregressive moving average processes. *Water Resour. Res.*, **20**, 565–580. R. Srikanthan and T.A. McMahon **668**

13. Chapman, T., 1994. *Stochastic models for daily rainfall.* Hydrology and Water Resources Symposium, The Institution of Engineers, Australia, National Conference Publication 94/ 15, 7–12.

14. Chapman, T., 1997. Stochastic modelling of daily rainfall in the Western Pacific. *Math. Comput. Simulat.*, **43**, 351–358.

15. Chapman,T., 1998. Stochastic modelling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts. *Environ. Model. Software*, **13**, 317–324.

16. Chin, E. H. (1977) Modelling daily precipitation process with Markov chain. *Water Resources Res.*, 13,949-95.6.

17. Chin, E.H. and Miller, J.F., 1980. On the conditional distribution of precipitation amounts. *Mon. Weather Rev.*, **108**, 1462–1464.

18. CIA, World Fact Book 2007.

19. Cocheme, J. and Franquin, P. (1967) A study of the agroclimatology of the semi-arid area south of the Sahara.FAO/UNESCO/WMO Tech. Bull. 86.

20. Cooke, D. S. (1953) The duration of wet and dry spells at Moncton, New Brunswick. *Quart.* J.*Roy. Met.Soc.,* 79,536-538.

21. Davy, E. G., Mattei, F. and Solomon, S. I. (1976) An evaluation of climate and water resources for development of agriculture in the Sudano-Sahelian zone of West Africa. WMO Special Environmental Report No. 9. WMO, Geneva, Switzerland.

22. Dennett, M. A., Rodgers, J. A. and Stern, R. D. (1981) Rainfall at Kampi-ya-Mawe and Katumani, Kenya. Report No. 3, Tropical Agric. Meteor.

23. Dick, K. 1988. A review of insect infestation of maize in farm storage in Africa. With special reference to the ecology and control of *Prostephanus truncatus.* ODNRI, Bulletin No. 18.

24. Doorenbos, J. and Pruitt, W. 0. (1977) Crop water requirements. FA0 Irrigation and Drainage Paper No. 24. FAO, Rome, Italy.

25. Dumont, A. G. (1974) The probabilistic simulation of weather variables. J.*Agric. Eng. Res.,*19.131-145.

26. Eidsvik, K.J., 1980. Identification of models for some time series of atmospheric origin with Akaike's information criterion. *J. Appl. Meteorol.*, **19**, 357–369.

27. Feyerherm, A.M. and Bark, L.D., 1965. Statistical methods for persistent precipitation pattern. *J. Appl. Meteorol.*, 4, 320–328.

28. Feyerherm, A.M. and Bark, L.D., 1967. Goodness of fit of Markov chain model for sequences of wet and dry days. *J. Appl. Meteorol.*, **6**, 770–773.

29. Fienberg, S. E. (1980) *The Analysis of Boss-classified Categorical Data,* 2nd ed. Cambridge, Mass.: MIT Press.

30. Foufoula-Georgiou, E. and Lettenmeier, D., 1987. Markov renewal model for rainfall occurrence. *Water Resour. Res.*, **23**, 875–884.

31. Gabriel, K. R. and Neumann, J. (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Royal Met. Soc.*, 88, 90-95,1962

32. Gangopadhya M, Sarker RP : Influence of rainfall distribution on the yield of wheat crop. Agri Met J, 2:331-350, 1965.

33. Gates, P. and Tong, H. (1976) On Markov chain modelling to some weather data. J. *Appl. Meteor.*, 15, 1145-1151.

34. Garbutt, D. J., Stern, R. D., Dennett, M. D. and Elston, J. (1981) A comparison of the rainfall climate of eleven places in West Africa using a two-part model for daily rainfall. *Arch. Met. Geoph. Biokl.* Ser B, 29, 137-155.

35. Goodspeed, M. J. and Pierrehumbert, C. L., 1975. Synthetic inputdata time series for catchment model testing. In: *Prediction in Catchment Hydrology,* T.G.Chapman and F.X.Dunin(Eds.), Aust. Acad. of Sci., Canberra, 359–370.

36. Green, J.R., 1964. A model for rainfall occurrence. *J. Roy. Statist. Soc.*, **B26**, 345–353.

37. Green, J. R. (1970) A generalised probability model for sequences of wet and dry days. *Monthly Weather Rev.*, 98,238-241.

38. Greenwood, J. A. and Durand, D. (1960) Aids for fitting gamma distribution by maximum likelihood. *Technometrics,* 2, 55-65.

39. Haan, C. T., Allen, D. M. and Street, J. 0. (1976) A Markov chain model of daily rainfall. *Water Resources Res.,*12,443-449.

40. Heerman, D. F1, Finkner, M. D. and Hiler, E. A. (1968) Probability of sequences of wet and dry days for eleven western states and Texas. *Colorado Agric. Exp. Station Tech. Bull.* 117.

41. Hills, R. C. and Morgan, J. H. T. (1981) An interactive approach to the analysis of rainfall records for agricultural purposes. *Expl. Agric.,* 17, 1-16.

42. Hutchinson, M.F., Richardson, C.W. and Dykes, P.T., 1993. *Normalisation of rainfall across different time steps*. Management of Irrigation and Drainage Systems, 21-23 July

1993, Park City, UT. Irrigation and Drainage Division, ASCE, US Department of Agriculture, 432–439.

43. Hutchinson, M.F., 1995. Stochastic space time weather models from ground based data. *Agric. Forest Meteorol.*, **73**, 237–264.

44. Jackson, I. J. (1981) Dependence of wet and dry days in the tropics. *Arch. Met. Geoph. and Biokl.* Ser B, 29,167-179.

45. Jones, J. W., Colwick, R. F. and Threadgill, E. D. (1970) A simulated environment model of temperature, evaporation, rainfall and soil temperature. ASAE Paper No. 70-404, American Soc. Agric. Engrs.

46. Jebuni, C.D. and Seini, Wayo. 1992. Agricultural Input Policies Under Structural Adjustment: their Distributional Implications. Cornell Food and Nutrition policy.

47. Jimoh, O.D. and Webster, P., 1996. Optimum order of Markov chain for daily rainfall in Nigeria. *J. Hydrol.* **185**, 45–69.

48. Jimoh, O.D. and Webster, P., 1999. Stochastic modelling daily rainfall in Nigeria, intra-annual variation of model parameters *.J. Hydrol.,* **222**,1–17.

49. Katz, R. W. (1974) Computing probabilities associated with the Markov chain model for precipitation. J.*Appl. Meteor.*, 43,953-954.

50. (1977) Precipitation as a chain dependent process. J.*Appl. Meteor.*, 16, 671-676.

51. Kawas, M. L. and Delleur, J. W. (1981) A stochastic cluster model of daily rainfall sequences. *Water Resources Res.,* 17, 1151-1160.

52. Khanal, N. N. and Hamrick, R. L. (1974) Misc. Publ. No. 1275, U.S. Dept Agric., 197-210.

53. Kowal, J. M. and Knabe, D. T. (1 972) *An Agroclimatological Atlas of the Northern States of Nigeria.* Zaria, Nigeria: Ahmadu Bello University Press.

54. Lall, U. and Sharma, A., 1996. A nearest neighbour bootstrap for resampling hydrologic time series. *Water Resour. Res.*, **32**, 679–693.

55. Lawrence, E. N. (1954) Application of mathematical series to the frequency of weather spells. *Meteor. Mag.*, 83,195-200.

56. Le Cam, L. (1961) A stochastic description of precipitation. *Proc. 4th Berkeley Symp.*, pp. 165-186.

57. (Lloyd E. H., O'Donnell, T. and Wilkinson, J. C. eds.) pp. 73-85. London: Academic Press.

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

59. Manning, H. L. (1955) Calculation of confidence limits of monthly rainfall. J.*Agric. Sci.*, 45, 154-156.

60. Menabde, M. and Sivapalan, M., 2000. Modelling of rainfall time series and extremes using bounded random cascades and Levystable distributions. *Water Resour. Res.*, **36,** 3293–3300.

61. Ministry of Food and Agriculture (1986) Ghana Agricultural Policy: Action Plan and Strategies (1986-1988)

62. Muhammad F, Haider F, Nabi G : Statistical probabilities of the number of rainy days. Abst 1st Isl Count Con, p 58, 27-31 August, Lahore, Pakistan, 1988.

63. Mooley, D. A. (1971) Independence of monthly and bimonthly rainfall over Southeast Asia during the summer monsoon season. *Mon. Wea. Rev.*, 99,532-536.

Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalised linear models. J.*R.Statist. Soc.* A, 135,370-384.

64. Nicks, A.D. and Lane, L.J., 1989. Weather generator. In: *USDA –Water Erosion Prediction Project*, L.J. Lane and M.A. Nearing,(Eds.), National Soil Erosion Research Lab., Report No. 2, West Lafayette. Over, T.M. and Gupta, V.K., 1994. Statistical analysis of mesoscale rainfall: Dependence of a random cascade generator on largescale forcing. *J. Appl. Meteorol.*, **101**, 319–326, 331. Over, T.M. and Gupta, V.K., 1996. A space-time theory of mesoscale rainfall using random cascades. *J. Geophys. Res.*, **101**, 26319–26331.

65. O'Brien, G. L. (1974) Limit theorems for sums of chain-dependent processes. J. *Appl. Prob.,* 11, 583-587.

66. Panabokke, C. R. and Walgama, A. (1974) The application of rainfall confidence limits to crop water requirements in dry zone agriculture in Sri Lanka. J. *National Sci. Council Sri Lanka,* 2,95-113.

67. Rajagopalan, B., Lall, U. and Tarboton, D.G., 1996. A nonhomogeneous Markov model for daily precipitation simulation. *J. Hydrol. Eng.-ASCE,* **1**, 33–40.

68. Rajagopalan, B., Lall, U., Tarboton, D.G. and Bowles, D.S., 1997a. Multivariate non-parametric resampling scheme for generation of daily weather variables. *Stoch. Hydrol. Hydraul.*, **11**, 65–93.

69. Rajagopalan, B. and Lall, U., 1999. A k-nearest-neighbour simulator for daily precipitation and other weather variables. *Water Resour. Res.*, **35**, 3089–3101.

70. Richardson, C.W., 1981. Stochastic simulation of dailyprecipitation, temperature and solar radiation. *Water Resour. Res.*, **17**, 182–190.

71. Robertson GW: Rainfall and soil water variability with reference to land use planning. Tech Ser No 1, WMO/UNDP Dept. of Meteorology Univ., Philippines, 1970.

72. Robertson GW: Daily global energy estimates from observation of bright sunshine and rainfall. Project Field Report No A-3, UNDP/FAO and FELDA Malaysia, 1976.

73. Robertson GW: Dry and wet spells. Project Field Report No A-6, UNDP/FAO and FELDA Malaysia, 1976.

74. Roldan, J. and Woolhiser, D.A., 1982. Stochastic daily precipitation models, 1. A comparison of occurrence processes. *Water Resour. Res.*, **18**, 1451–1459.

75. Runge-Metzger, A. and Diehl, L. 1993. Farm household systems in Northern Ghana. A case study in farming systems oriented research for the development of improved crop production systems. NAES-Research Report, GTZ.

76. Salimi MA, Sadiq M, Nabi G: Weather Risk Assessment for Punjab. Dept of Agri Meteorology, University of Agriculture Faisalabad.

77. Salimi MA, Nabi G, Muhammad F: Rainfall probabilities. Abst 1st Isl Count Conf, p 58, 23-31 August, Lahore, Pakistan, 1988.

78. Shenton, L. R. and Bowman, K. 0. (1977) *Maximum Likelihood Estimation in Small Samples.* London: Griffin.

79. Sen, Z. (1980) Critical drought analysis of periodic-stochastic processes. J.*Hydrol.,* 46., 251-263.

80. Siriwardena, L., Srikanthan, R. and McMahon, T.A. (2002), Evaluation of Two Daily Rainfall Data Generation Models, Cooperative Research Centre for Catchment Hydrology, Technical Report 02/14, 75 pp.

81. Small, M.J. and Morgan, D.J., 1986. The relationship between a continuous-time renewal model and a discrete Markov chain model of precipitation occurrence. *Water Resour. Res.*, **22**, 1422–1430.

82. Smith, J.A., 1987. Statistical model of daily rainfall occurrences. *Water Resour. Res.*, **23**, 885–893.

83. Stearn and Coe [Part 1, Lloyd, E. H. (1979) Stochastic storage problems. In *The Mathematics of Hydrology and-water Resources*

84. Singh, S. V., Kripaluni, R. H., Shaka, P., Ismail, P. M. M. and Dahale, S. D. (1981) Persistence in daily and 5-day summer monsoon rainfall over India. *Arch. Met. Geoph. and Biokl.* Ser. A, 30,261-277.

85. Srikanthan, R. and McMahon, T.A., 1985. *Stochastic generation of rainfall and evaporation data.* AWRC Technical Paper No.84, 301pp.

86. Stern, R. D. (1982) Computing a probability distribution for the start of the rains from a Markov chain model for precipitation. J.*Appl. Meteor.*, 21,420-423.

87. Stern, R. D. and Coe, R. (1982) Use of rainfall models in agricultural planning. *Agric. Meteor.*, 26, 35-50.

88. Stern, R. D., Dennett, M. D. and Dale, I. C. (1982) Analysis of daily rainfall measurements to give agronomically useful results. I. Direct methods. *Expl. Agric.*, 18, 223-236.

89. Stern, R. D., Dennett, M. D. and Garbutt, D. N. (1981) The start of the rains in West Africa. J. *Climat.*, 1, 59-68.

90. Todorovic, P. and Woolhiser, D. A. (1975) A stochastic model of n-day precipitation. J. *Appl. Meteor.*, 14, 17-24.

91. UNDP, (1991), Human Development Report 1991, United Nations Development programme.

92. Wang, Q.J. and Nathan, R.J., 2000. A daily and monthly mixed model algorithm for stochastic generation of rainfall time series. Unpublished manuscript.

93. Waymire, E. and Gupta, V. K. (1981) The mathematical structure of rainfall representation. I. A review of the stochastic rainfall models. *Water Resources Res.*, 17, 1261 -1 272.

94. Weiss, L.L., 1964. Sequences of wet and dry days described by a Markov chain model. *Mon. Weath. Rev.*, 92,169–176.

95. Williams, C.B., 1947. The log series and its applications to biological problems. *J. Ecol.*, **34**, 253–272.

96. Williams, C. R. (1952) Sequences of wet and dry days considered in relation to the logarithmic series. *Quart.* J. *Roy. Meteor. Soc.*, 78, 91-96.

97. Winstanley, D. (1974) Seasonal rainfall forecasting in West Africa. *Nature*, 248,464.

98. Woodhead, T., Waweru, E. S. and Lawes, E. F. (1970) Expected rainfall and Kenya agriculture-confidence limits for large areas at minimum cost. *Expl.Aqic.* 6, 87-97.

99. Woolhiser, D. A. and Pegram, G. G. S. (1979) Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. J.*Appl. Meteor.*, 18, 34-42.

100. Woolhiser, D.A. and Roldan, J., 1982. Stochastic daily precipitation models, 2. A comparison of distribution of amounts. *Water Resour. Res.*, **18**, 1461–1468.

101. Yevjevich, V. and Dyer, T.G.J., 1983. Basic structure of daily precipitation series. *J. Hydrol.*, **64**, 49–67. Young, K.C., 1994. A multivariate chain model for simulating climatic parameters from daily data. *J. Appl. Meteorol.*, **33**, 661–671.