Statistical Analysis of Climatic Factors and Their Relative Influence on Economic Factor in Agriculture for a Selected District of Terai Zone in West Bengal

THESIS
SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (SCIENCE)
OF THE
UNIVERSITY OF NORTH BENGAL
1995

By

Swaraj Kumar Mukhopadhyay

DEPARTMENT OF AGRICULTURAL STATISTICS,
FACULTY OF AGRICULTURE
NORTH BENGAL CAMPUS
BIDHAN CHANDRA KRISHI VISWAVIDYALAYA
PUNDIBARI, COOCHBIHAR
WEST BENGAL, INDIA

DEDICATED TO

MY

BELOVED PARENTS

ACKNOWLEDGEMENT

My indebtedness to Professor Satya Pada Pal and Mr. Sanjeeb Basu, thesis supervisors is indeed inexpressible. They not only intimate me to the entire course of investigation but also provide me with all necessary help required to complete this research work. I have learnt a great deal under their benevolent guidance.

I am also grateful to all teachers of the Department of Mathematics for constant encouragement.

I would like to acknowledge, the competent authority of I.M.D., Pune and Cooch Behar Center, and C.T.R.I.,Dinhata and also State Agricultural Department, West Bengal for supplying the meteorological observations and yield data of rice for Cooch Behar. I would also like to thank the authority of CSSC,ISI,Calcutta for analyzing data by electronic computer.

Sincere thanks are also due to Sri Raju Saha for carefully typing the thesis.

I would like to acknowledge my deep sense of gratitude to Dr. Apurba kumar Mukhopadhyay and Sri Adar Kumar Mukhopadhyay, my elder brothers, Sri Tarun Kumar Mukhopadhyay, my younger brother Kalpana,my wife and other members of my family whose overwhelming encouragement and irreparable sacrifices have paved the way for my study.

Swaraj Kumar Mukhopadhyay

(Swaraj Kumar Mukhopadhyay)

Date : RajaRammohanpur,

Siliguri.

The 30th November..1994.

C O N T E N T S

CHAPTER -I

INTRODUCTION

CHAPTER - I

INTRODUCTION

1.1. Introduction

The population of our country is estimated to reach thousand million mark in 2000 A.D. It is implied that to meet the basic necessities, for the existence of mankind, the agricultural production should have increased rapidly enough to keep pace with the population growth. But our estimates show that the irrigated area is not expected to be more than forty two percent of the total cropped area in the ensuing time. Therefore, judicious utilization of the direct precipitation would have to be thought of for increasing agricultural production.

The Director General of Food & Agriculture Organization (FAO) said in a statement to the press on February 1st,1973, "It is intolerable that this world of the 1970's with all its scientific progress and its slowly growing sense of common purpose,should go on enduring a situation in which the changes of enough decent food for millions of human beings may simply depend on the whim of one year's weather". The statement would be contemplated in the sense ofournation. Consideringthe economic importance of agriculture, the important role of weather and climate is patent.

Earlier Ramdas (1967) had pointed out "Weather exercises a profound influence on farm work as well as crop growth. Success or failure of farming is very intimately linked up with the prevailing weather conditions. Farming in India is largely

gamble in weather conditions...................A sound knowledge of climatic factors and the effects of weather on crop growth and yield is , therefore, essential for every farmer........... "
Doubtless,thepractical difficulties of specifying the rainfall variables more precisely account forthis gap in the literature. And rainfall totals are important in defining the climate of a region.

Rainfall is chosen as a climatic variable in this thesis because it is the climatic element of maximum significance specially in this region where the inter-annual variations of other other climatic elements are not so large.

The report of the National Commission on Agriculture (1976) made recommendations for future research thrust in these aspects,some of them are as follows :

i) The method of working district normals should be reviewed.

ii) Work in the field of Mathematical modeling should be reviewed.

iii) Crop-weather studies should be intensified and formulae should be developed on a district wise basis.

Thus, a detailed examination of the systematic study of precipitation,particularly of a relatively small region may however, be suggested for regional peculiarities of the climatic system. The present investigation concentrates on the district of Cooch Behar, formerly an important region of the independent Kingdom of Coach , in the state of West Bengal . We admit that our geographical coverage is limited, but the studies of smaller region make it possible to investigate in greater details over the Terai agroclimatic zone of West Bengal.

Agricultural systems in this region i.e.the district of Cooch Behar, are susceptible to vagaries of rainfall received to a very large degree. The precipitation brings both benefits and problems to the ecosystem of this district.

The rainfall,with extremes bring drought,in meteorology,on the one hand and flood i.e. excess of this phenomenal rate on the other hand. A surfeit of water may be a problem, if it causes flooding but itis not always experienced, drought however, is invariably a problem in agriculture. So,the occurrence of droughts and floods, in meteorology in this district may have a systematic pattern which together with the interplay between these two factors could therefore be analyzed.

During the peak monsoon season,the rain does not occur on all days at a station. Rain occurs in spells. The prior knowledge of dry spells , wet spells and weather cycles, is very useful for crop planning. Investigation of the probability of occurrence of these spells and cycles has been undertaken in this study .

The present study is intended to seek empirical evidence to determine the suitability of probability distribution model for representing the character of the daily rainfall occurrence.

The nature of the daily precipitation has a complete lack of confidence in their occurrence which creates the well known uncertainty of the nature of the daily rainfall. In this study attention has been paid to the measure of this uncertainty.

The dominant characteristic of the climate is the marked rhythmic recurrence of precipitations, and the droughts or floods naturally found, its echo in a marked rhythm of their occurrence in this region. Such a study of rainfall variability in this region has attracted a good deal of attention in this study.

Agriculture and meteorology remain self contained disciplines since very ancient time. But the research in this subject requires a thorough knowledge of both the meteorology and its influence on crop production. For estimating the impact of weather on crop production, the use of total rainfall within a week or a month is not significant to predict the yield of a crop. The relation between the yield of winter rice and the nature of daily rainfall during the sensitive months in this district has been of great interest to study .

All these apprehensions need to be analyzed through empirical investigations of the actual behaviour of precipitation in the district of Cooch Behar.

It is surprising to find that a detailed study of the behaviour of annual and daily anomalies of the atmospheric circulation pattern in this region has not been previously undertaken.

The studies examining the behaviour of nature of daily rainfall and the possible periodicities in the rainfall record as well as the relationship between yield of winter rice and the nature of daily rainfall during the crop sensitive months are for many years quite rare in the literature and most of the works published are studies on the variability of climatological nature in this meteorological sub-division.

This thesis is essentially oriented by applied mathematics which depends on the understanding of theoretical principles of empirical statistical techniques.

1.2: OBJECTIVES

The specific objectives set forth in this present study are as follows :

a) To determine the distribution function of inter arrival time of drought years and flood years.

b) To derive the one day dependent weather model and to estimate average length of dry spells and wet spells and weather cycles and to determine the probability distribution of spells as well as weather cycles.

c) To derive the stochastic dependence model on the nature of the weather of a day and to assess the measure of uncertainty.

d) To detect the hidden periodicities in the recorded time series of annual rainfall.

e) To develop the relation between the winter rice yield and the nature of daily rainfall during the sensitive months on the life time of that crop.

1.3 : PLAN OF THE STUDY

This study is organized into nine chapters including this chapter which states the problem and the scope of the study.

Chapter-II reviews past studies on relative objectives.

Chapter-III discusses the materials and methods used in this study. It also presents the short profile of the district of Cooch Behar, West Bengal.

Chapter-IV deals with probability distribution of inter arrival time of drought years and flood years for two identical rain gauge stations in Cooch Behar district.

Chapter-V provides the preliminary analysis of spell distribution as well as the weather cycles of Cooch Behar station.

Chapter-VI also provides the measure of uncertainty of daily rainfall with the help of information theory. It also deals with the stochastic dependence of daily rainfall and it is tested by entropy.

Chapter-VII investigates to detect the hidden periodicities in the annual rainfall at two stations of Cooch Behar district.

Chapter-VIII is designed to establish the relationship between the yield of winter rice and the derived categories of daily rainfall at Cooch Behar.

The summary and conclusions drawn from the present study are recorded in Chapter-IX. Based on these conclusions certain suggestions also have been proposed to identify the scope of future research to strengthen the present study.

---------------*************-----------------

CHAPTER - II

REVIEW OF LITERATURE

CHAPTER II

REVIEW OF LITURATURE

2.1. Introduction:

This chapter reviews the methodology, conclusions and limitations of the studies undertaken in the past on the behaviour of precipitation in our country and abroad .

The past studies are classified into five categories according to methodologies adopted by the respective objectives. This sequence of methodology also broadly reflects the chronological order in which this dissertation has been undertaken.

2.2 Variations of precipitation :

Walker (1914) in a study presented a procedure for classifying the years of deficit rainfall in three categories i.e. years with rainfall deficiency between 30 to 45 percent, 45 to 60 percent and over 60 percent of the mean annual rainfall and called them years of "Large", "Serious" and "Disastrous" deficiency.

Walker (1919) used the seasonal rainfall data during the period 1841 to 1908 in his study. However, he had some doubt about the degree of accuracy of the data. Drought concept was applied, perhaps, for the first time to agriculture by Ramdas (1967) in his study in 1950, where he defined drought to have occurred if the weekly rainfall was less than double the mean deviation. He identified 1877, 1899 and 1913 as years of outstanding agricultural drought while in 1920 the drought was only partial, affecting North, West and Central parts of India.

Mallik (1958) examined the occurrence of droughts in Bihar and Uttar Prades. He used the definition of drought which occurred when the actual rainfall during a week was less than or equal to half the normal rainfall.

Dhar and Changrancy (1966) attempted to study the meteorological situations associated with floods in Assam during the monsoon months. It was observed that the floods and the associated meteorological conditions were responsible for some spells of heavy rainfall in the region.

Sarkar (1979) studied the drought in India and their predictability. He mentioned the three facets of drought. These were drought in meteorology,drought in hydrology,and drought in agriculture. He tabulated areas which received less than 75 percent of the normal rain during the monsoon season. The years when the affected area exceeded 25 percent of the arable land were termed as drought years. From the table it was observed that the worst drought years, in terms of rainfall were 1877,1891,1901,1904 and 1972 during the study period in the meteorological region of Sub-Himalayan West Bengal.It was further found that the droughts on account of rainfall,were very rare in Gangetic West Bengal but it might occurr once in the period of 6-8 years in Sub-Himalayan West Bengal.

Chowdhury and Abhyankar (1984) attempted to compile and collect drought climatology of India. They used the seasonal rainfall deficiency of more than 25 percent for identifying meteorological drought over various sub-divisions of India.

It was also observed that the occurrence of good and bad monsoons was a random phenomenon though therewas weak tendency for a good monsoon year to recur in the following year. They observed that the meteorological drought might occur once in six to eight years in the region of Sub-Himalayan West Bengal. It was also revealed that using the power spectral analysis, none of the peaks was found significant at 95 percent level. However, a peak of 2.8 years cycle was found to be significant at 90 percent level. This, possibly, could be associated with a quasi-biennial oscillation.

Mooley *tl(1984) studied the droughts which affected the peninsular India during the period 1861 to 1980. They used a criterion for identification of well-marked drought. The criterion was expressed as an index of standard variate. It was observed from the study that Cox's test for randomization was applied to test the existence of trend on the time interval between the successive occurrences of well-marked droughts. The result of the test confirmed the randomness of the occurrence of well-marked droughts over the peninsula. They also concluded that the more frequent trough position over the foot-Himalayas during drought years were seen to be associated with large scale drought over peninsula.*

Ganesan and Rao (1986) in a study attempted to analyze the climatological conditions and their local difference. They analyzed the rainfall characteristics over a small area in Bangalore. Bangalore City centre and Bangalore Air-Port (within a distance of 10 K.M.) were used for the study. It was revealed that the amount of rainfall in the city was in general larger except in post monsoon season than that in the Air-Port area. They also concluded that even within a limited area the rainfall varied considerably.

Elizabeth *tl.(1988) examined the changing rainfall patterns in Western Sudan. Particular attention had to be paid to the period since 1965 when annual rainfall total began to show a marked decline. It was observed from the study that the 20 years period (from 1965 – 85) was a period of progressive deterioration in relative rainfall.*

Gragam Farmer (1988) examined the rainfall anomaly for the short rains on the Kenyan coast during the period 1901 to 1984. It was revealed that there was a notable difference in the variability of rainfall series for the two sub-periods,1901-42 and 1943-84. A one-tailed variance ratio,F-test was performed for seasonal rainfall series and he concluded that the rainfall series of sub-period 1943 – 84 had significantly larger variance than the sub-period 1901-42But the interesting feature was that if the rainfall of 1967 was removed, then the two sub-periods might be considered homogeneous. Following on, from F-test,the means of the two sub-periods were tested by Berhrens-Fisher test. None of the series tested showed a significant difference between the means for two time periods.

Chowdhury t l(1989) attempted to examine the distribution of time intervals between the successive droughts.The droughts in meteorology defined by the Drought Index of rainfall which was expressed as a standard variate. The index was divided into two equal parts and student's t-test was applied to ensure the homogeneity of the observations. They observed that the Drought Index series was homogeneous.

They also examined the intervals between the successive drought years which were very important in drought incidence. It was revealed from the study that the frequency distribution of time intervals between the occurrence of successive droughts followed the Poisson distribution. In the study,they observed that the largest time interval of droughts among the independent time intervals was not statistically significant.

They also observed that Drought Indices did not follow the normal distribution. The nature of distribution of the series was found to be positively skewed leptokurtic.

Prasad and Lal (1989) studied the long term variations of the annual and seasonal rainfalls at Jalpaiguri in North Bengal. The study of the annual and seasonal (June to September) rainfall variations for the period 1901-80 indicated that the highest annual rainfall (4546 mm) and seasonal rainfall (3997 mm) occurred in 1938 while the lowest annual rainfall of 1884 mm and seasonal rainfall of 1676 mm occurred in 1978.

It was also found that long term means of annual and seasonal rainfalls were 3276 mm and 2640 mm respectively for the eighty years rainfall series. The annual and seasonal coefficients of variations were 16.1 percent and 18.2 percent respectively. It was also observed that the 30 years mean annual and seasonal rainfalls were highest for the period 1916-45. There was a steep rise in the mean annual and seasonal rainfall during this period in comparison to the mean rainfall of study period which decreased there after continuously till 1980.

Lakshmanaswamy and Jindal (1990) attempted to study the variability of area weighted annual and monsoon rainfall for the 35 meteorological sub-divisions of India for the period from 1901 to 1988 with a view to finding any increase or decrease of rainfall over India. It was revealed from the study that the peak annual rainfall occurred in 1917 followed by another peak in 1956. The least rainfall occurred in 1972. The entire period was divided into three sub-periods ,*iz. 1901-30,1931-60 and 1961-88. It was found that the area weighted average rainfall of the sub-period 1901-30 was greater than those of the other two sub-periodswhile the standard deviation of the sub-period 1961 8 was greater than that of the other two sub-periods in the meteorological region of Sub-Himalayan West Bengal and Sikkim. It was also observed from the analysis of the area weighted monsoon and annual rainfall during study period over India that there was a steady increase of rainfall in the period 1901-50 and gradual fall of rainfall from 1950.*

Mukhopadhyay (1992a) studied the long term variation of annual rainfall in Coochbehar district. The time series were made from the average of the annual rainfalls of two stations. He identified the flood and drought years by an index. It was observed from the study that the years 1980 and 1921 were the lowest and the highest rainfall years in the study period. F-test and t-test were applied to establish the homogeneity of the time series. The tests were applied considering the two sub-periods. It is observed that the frequency distribution of the time intervals between the successive drought years was found to follow the Poisson distribution whereas the same for flood years,the Mixture of Poisson's distribution.

Mukhopadhyay (1992b) attempted to determine the probability distribution function of inter-arrival times of successive dry and wet years of two stations in Cooch Behar district separately . It is observed from the study that no significant trend for the occurrence of successive wet and dry years could be detected in the series of the two stations. The frequency distribution of the inter-arrival times for the dry and wet years could be considered to follow the negative binomial distribution. It was also observed that the wet sequence as well as the dry sequence for two places having negative binomial distribution could be considered as drawn from the same negative binomial population.

2.3. Pattern of rainfall.

Sundararaj and Ramchandra (1973) in a study determined the Markov-based geometric model for wet spells and dry spells,and the lengths of weather cycle. It was revealed that the model was best fitted to the wet spells but the dry spells could not be considered to fit that model to the experimental observations. In the study the dry-wet cycles and wet-dry cycles were also computed and the behaviour of these cycles did not appear to confirm the Markov dependant geometric model.

Agnihotri et al. (1984)attempted to predict the behaviour of dry spells and wet spells as well as weather cycles in Chandigarh. The study period was 23 years of daily rainfall. The Markov chain model was fitted to the daily rainfall data. The parameters of two-state Markov chain model were used to fit the distribution of spells as well as weather cycles. It was also observed that the observed frequencies of weather cycles of various durations and dry/wet spells were found to follow the geometric distribution.

Chowdhury and Abhyankar (1984) briefly analyzed the seasonal rainfall by Markov chain method. They used the two-state Markov chain model. They studied the drought spells and non-drought spells for the various meteorological sub-divisions in India and also computed the climatic cycle of the regions. They came to the conclusion that non-drought cycles did not conform to the geometric model and the length of climatic cycle could not have much significance.

Aneja and Sribastava (1986) derived a model for occurrence of the daily rainfall under the assumption that successive day's rainfall did not occur independently but were inter-dependent. The model was represented by two-state Markov chain. They also computed the average length of dry spells and wet pells and expected length of weather cycle.

2.4. Measure of uncertainty

Basu (1988) in a study evaluated the nature of rainfall with the help of information theory. It was revealed that the weighted entropy values over the transition probabilities during the beginning and closing months of monsoon at Maithon were less in comparison with the other monsoon months. The weighted entropy of the active monsoon months were higher. It was also revealed that the difference in uncertainty between the Markovian model and the random model were small,almost negligible. He observed that during the monsoon months the degree of uncertainty of heavy rainfall and very heavy rainfall were more than the light,moderate and non-rainy rainfalls. He used the redundancy test to determine the favourableness and unfavourableness of Markovian system.

Mukhopadhyay (1992c) attempted to examine the Markov dependence by entropy considering the three states of outcome. There normal, bad and good years of rainfall are accordingly defined. Redundancy test was applied to test the stochastic dependence on the one-step 3 x 3 Markov chain model. But it was observed from the study that the redundancy test could not favour the system of Markovian dependence. So,he came to the conclusion that the occurrence of individual state might not depend on the previous state of occurrence. Then he concluded that the occurrence of annual rainfall was a random phenomenon.

2.5. Periodicity in rainfall.

Rao et al.(1973)ina study attempted to examine the periodicity of droughts in India for the period of sixty years. The power spectrum method was used to identify the periodicity. It was revealed from the study that there was a peak of the period 4.6 years which was significant at 95 percent level. And there was also a peak of the period 3.5 years which was significant at 90 percent level. The study was conducted for the various meteorological regions. The above result was observed for the region of Gangetic West Bengal. The same result was observed in meteorological region of Bihar plateau. They also concluded that the overall picture did not encourage one to look with confidence for the periodicity of droughts in India.

Pathack (1982) in a study investigated the periodicity of rainfall observations over the Island of Mauritius. Spectral analysis was used to determine the periodicity of the climatic time series for the various stations of that island.He observed that the climatic series contained the fluctuation corresponding to period of the order to record length. It was also observed in the study that the presence of periodicity of 10-13 years might be associated with the solar cycles,and some other less pronounced peaks were noted in the period ranges of slightly over two years, 3-4 years and 16-20 years.

Ogallo (1984) investigated to study the existence of the periodic or quasi-periodic fluctuations in the seasonal rainfall over East Africa. He used the Spearman rank correlation test to examine the trend of the time series. The spectral analysis technique was used to detect the periodicity of the seasonal rainfall on the yearly basis. It was observed from the yearly time series that no significant trend was detected at 5 percent level in most of the time series. The result of the spectral analysis indicated the dominance of short period fluctuations. The most prominent spectral peaks were centered around 2-3 years and 5-6 years. This analysis also indicated that a weak peak of a period 10-11 years appeared in many time series.

Bhalme *tl(1984) attempted to studytheperiodicity of Drought Area Index in India. Power spectrum analysis was applied to detect the significant periodicities of droughts in monsoon season in India. It was observed from the analysis that the power spectrum of the Drought Area Index series showed quasi periodicity of about 2.5 to 3.5 years. The study also suggested that the principal cause of recurrence of large scale droughts over India, with most common period in the range of 3-6 years periodicity might possibly be the influence of the southern oscillation of monsoon.*

Ogallo (1986) examined the hidden periodicity of the regional annual rainfall in East Africa. Spectral analysis was used to detect the dominant spectral peaks. It was observed from the study that the dominant spectral peaks were centered around 2.2-2.8, 3-3.7, 5-6 years and 10-13 years. It was also revealed that similar temporal characteristics were observed in the annual rainfall records of the individual stations within these regions.

Currie and O'Brien (1988) examined the 136 yearly total precipitation record in the North Eastern United States of America. The record length distribution of 136 series prior to maximum entropy spectrum analysis were used mostly between 80 to 89 length. In terms of spectra the evidence was found for strong band limited 19 years luni-solar signal and a weaker term of period 10-11 years on the yearly total precipitation on record. They established that the variations in climate data did not show the 'pseudo periodic' behaviour.

Olapido (1989) examined the spectra in the derived growing season Bhalme-Mooley drought index series for the Interior Plains of North America. He used the non-harmonic spectral technique for estimating spectra for frequencies in order to detect periodic and quasi-periodic components in the rainfall time series. It was further examined for evidence of periodic variation which had similar frequency of occurrence in adjacent stations. It was also found that the quasi-biennial,quasi-triennial and quasi-five year oscillations were concentrated in almost all the stations of the study region. It was further observed that the periodicities greater than the quasi-five year frequency bands was not significant in terms of the number of stations.

Fong Chao (1990) presented an interesting study on the use of maximum entropy spectrum in harmonic analysis of time series. This study critically examined the amplitude of the maximum spectral peak. It was observed that the maximum entropy power spectrum contained no information about the amplitude of the spectral peak . He suggested that the maximum entropy power spectrum could be used to estimate the complex frequency, but not the complex amplitude.

Currie and O 'Brien (1990) in a study presented the maximum entropy spectrum analysis of 120 yearly precipitation records in the corn belt region of United States of America. It was observed from the analysis that the 18.6 years luni-solar terms in 109 records and the 10-11 year solar cycle signal in terms of 100 records were found in the region.

Mitra t l *(1991) analyzed the rainfall data in North-West India, the Plains in Uttar Pradesh and North Central India. They applied the method of maximum entropy spectral analysis for the period of 50 years rainfall data. The result of the analysis indicated that 91 out of 115 rain-gauge stations appeared to respond to 18.6 years luni-solar signal and the statistical average value of the signal periodicity in rainfall proved to be 18.3 ± 1.6 years. The result also observed the existence of a 10-11 year solar cycle in 77 out of 115 rain gauge stations in the region.*

It was revealed that the presence of an 8-9 year component periodicity indicated in almost every stations in the experimental region.

2.6. Relation between rainfall and rice yield.

Jain tl (1980) in a study attempted to understand the effects of climatic variables on rice yield and to forecast the rice yield using the climatic variables. They observed that the maximum temperature,relative humidity and sunshine had small beneficial effects in general through out the crop season. But effect of rainfall was in general beneficial through out the crop season.

Agarwal tl (1983) in a study examined the joint effect of climatic variables on rice yield. They found that the rise of temperature and humidity had small beneficial effects,in general,throughout the crop season. But the increase in rainfall was beneficial for the yield of rice except during the ripening stage of the crop.

Agarwal tl (1986) attempted to study the individual effects of weather factors on rice yield. They concluded that above average rainfall had beneficial effect throughout the growth phase and detrimental effect during the ripening phase of the rice crop in general. The effects of rainfall and increase in the number of rainy days were fluctuating up to vegetable stage but beneficial during reproductive and detrimental during ripening stage of the crop.

Mongia and Gajja (1986) examined the influence of rainfall on productivity performance of paddy crop in Andaman & Nikobor Islands. They introduced mainly two independent variables of which one (X_1) was normal rainfall when rainfall was above normal otherwise actual rainfall and other (X_2) was the difference between actual and normal rainfall in case of rainfall above normal otherwise zero. The estimated regression co-efficient revealed that it was positive and significant at various levels i.e.1 percent level5 *percent level in all the four regions under study. It was further revealed that impact of seasonal deviation of the rainfall on the yield of paddy was estimated positive and highly significant in the region.*

Mohendra dev (1987) examined that growth performance took explicit account of the impact of rainfall on crop production. It was revealed from the study that the variation of rainfall alone could help explain in a substantial measure the variation of crop production. A comparison of unadjusted and adjusted rainfall growth rate revealed that weather adjusted growth rate was higher in the states like West Bengal etc.

Prasad and Dudhane (1989) in a study investigated the relationship between rainfall and rice yield in the region of Gangetic West Bengal. The impact of agricultural technology on rice yield was analyzed also by modified model using technological trend as an independent variable. A systematic methodology was adopted to develop the model to forecast the rice yield in that region from the direct and derived parameters of rainfall data. It was revealed that all independent variables emerged significant functional relationship with the rice yield of that region. It was also observed that the satisfactory performance of the suggested model could forecast, successfully,the rice yield from the derived parameters of rainfall and agricultural technology.

CHAPTER - III

MATERIALS AND METHODS

CHAPTER -III

MATERIALS AND METHODS

3.1. SHORT PROFILE OF COOCH BEHAR DISTRICT

Cooch Behar is a district which is situated at the north eastern fringe of Jalpaiguri division in the state of West Bengal. In shape it forms an irregular triangle having an area of 3404 sq.K.M. and is bounded by Jalpaiguri district in the north and west, in the east partially by Goalpara district of Assam and partly by Rangpore district of Bangladesh and the south is entirely bounded by Bangladesh. One of the striking features of Indo-Bangladesh border is the presence of innumerable "CHHITMAHAL"s spread throughout the border line.

The district lies between the parallels 26° 50' 40" and 25° 32' 20" North latitudes and is bounded on the east and west by 89° 47' 40" and 89° 54' 35" East longitudes.

The district of Cooch Behar geographically forms a part of Himalayan Terai of West Bengal. The district belongs to the Terai agro-climatic zone of West Bengal. This district also belongs to the Meteorological sub-division of Sub-Himalyan West Bengal. It is also situated in the high atmospheric pressure zone. The atmospheric pressure is 1000 millibar.

The climate of this district is basically per-humid in nature with distinctive characteristics of high rainfall,high relative humidity and low temperature. About three-fourth of the total annual rain fall is received during the south western monsoon (June to September). The first intermonsoon period (March to May) and the second intermonsoon period (October to November) have a relatively low intensity of rainfall.

The number of rainy days are recorded highest in the months of June & July. The number of rainy days often crossed twenty days in the months of June and July followed by 15 to 20 days in August and September. But in the months of April,May and October it varied in maintaining a central tendency towards 10 days in a month. During the season, minimum and maximum temperatures recorded are in the range $5^{\circ}C$ to $33^{\circ}C$. The temperature reached the maximum in August and the lowest temperature is recorded in the month of January. During the rainy season, the relative humidity appears around 90% in this district. In winter season it ranges from 60 to 75 percent. The overall climate of Cooch Behar is damp and not so hot. The normal annual rainfall of this district is 3292 millimeters.

The district is a net-work of rivers and small streams, beds of which have formed the soil of Cooch Behar. The principal rivers are the Teesta,Torsha,Jaldhaka,Kaljani,Raidak, and Mansai etc. . The majority of these rivers take rise in the Himalayas. After passing through Duars these rivers enter this district from the northern part and flow into Bangladesh on their way to join the Brahmaputra.

The net cropped area is about 75 percent of the total area of this district. The main crop of this district during

Kharif season is Winter Rice(i.e. Aman paddy). A little over 47 percent of gross cropped area is shared by Winter Rice (Aman) in this district. The aman paddy is basically rain fed crop here because only 10% of cultivated land are under irrigation scheme. But the most of facility of irrigation concerned is in the Rabi crop and to some extent in the pre-Kharif corp.

The majority of the people of Cooch Behar depend on agriculture and it is their principal source of livelihood.

3.2. The nature of Data used :

It is pointed out earlier that the Cooch Behar district has been selected for the study. The crucial variables used in this study are rainfall and yield of winter rice. On the whole the categories of data required are as follow :

3.2.1. Rainfall (Monthly) :

There are more than seven rain gauge stations in Cooch Behar district where rainfall is recorded. However there are immense irregularities in the maintenance of proper records.

The total precipitation is not uniform throughout this district. It is the highest in the northern part and as one proceeds south wards it decreases.

The northern part of this district has an annual precipitation rate more than the normal rainfall (3292 mm), while the southern part of this district has less than normal rainfall. Therefore,on the assumption that the rainfall at the center of the northern part is highly correlated with that at other centers of this part and as the same assumption holds for the southern part of the district,we have chosen two rain gauge stations for which proper information is available. The chosen two rain gauge stations are Cooch Behar and Dinhata. The monthly rainfall data are collected from published observations from sources and some of them are collected from the rain gauge stations.

The monthly rainfall data of the two stations have been collected for the period from 1901 to 1988. The data consist of monthly sum. During compilation of monthly data some difficulties have been faced due to missing records from the stations. However,the difficulties in filling up the missing data for some months,have been overcome by considering monthly normals of the stations or by using the respective long term monthly mean values for the station. Meteorologically more significant annual totals are derived from the monthly sums of rainfall of the respective stations.

3.2.2. DAILY RAINFALL

Daily rainfall data for the Cooch Behar station have been collected from the Meteorological, Center Cooch Behar. The daily rainfall data of Cooch Behar are available for the period from 1971 to 1988.

In this context,it is noted that the daily rainfall data for Dinhata cannot be available for long period. This discrepancy imposed some limitations to the study.

3.2.3. Estimated Rice yield :

Aman paddy singularly accounts for about 70% of total rice area of this district. It is grown extensively as a rain fed crop. The winter paddy,too,is predominated by local varieties. A little over 80 percent of total area under winter paddy is occupied by local varieties.

The climatic factor, particularly rainfall, is used to estimate the yield of winter rice. Winter rice yield data in Kg/hectare have been collected for the period,1972 to 1988 from the record published by the Govt. of West Bengal.

The rice yield data consist for the district of Cooch Behar. The yield of winter rice has been included as dependent variable in the study for crop-weather relationship.

3.3. SOURCES OF DATA

The monthly sums of rainfall data for the raingauge stations at Cooch Behar and Dinhata are collected from India Meteorological Department by some published records and from its local Meteorological center. Some relevant rainfall data have been collected from the raingauge station at Central Tobacco Research Institute, Dinhata, Cooch Behar.

The yield estimates of winter rice for the Cooch Behar district have been collected from "Statistical Abstract" published by the Department of Applied Economics and Statistics, Govt. of West Bengal and estimates of area and production of principal crops in West Bengal published by Socio-Economic and Evaluation

Branch, Govt.of West Bengal.

Some relevant observations have been collected from the "Key Statistics" published by the District Statistical officer,Bureau of Applied Economics and statistics, Government of West Bengal, Cooch Behar.

Some statistics regarding the physical features of this district have been collected from "Agriculture in Northern Districts of West Bengal-Profile and Prospects" published by North Bengal campus,Bidhan Chandra krishi Viswavidyalaya, Cooch Behar, West Bengal.

3.4. METHODOLOGY USED

The statistical method is based on the analysis of real data. A particular technique is more clearly understood when it is presented in numerical as well as mathematical term.

The basic statistical theory and methods are freely followed from time to time throughout the study. These are mean,standard deviation and variance,mean deviation etc.. The graphical analysis is also employed to familiarize with the mathematical concepts.

The two basic discrete distributions i.e.,negative binomial and Markov-dependent Geometric model,the most common kind of discrete variables,are considered for probability model. The geometric probability model arising from the Markovian probabilities are considered to fit the distribution for time varying variable. The variables are frequency variables and the use of this model have some a "prior" justification in terms of modeling of rare events occurring over periods of time.

The stochastic process has been employed by utilizing the Markovian property. The transition probabilities constitute the Markovian model. Here we confine to the homogeneous chain system. We have also considered the unit-step transition with two-state and five-state Markov chains. The entropy defined by Shannon has been used to measure the uncertainty of the transition probabilities and to test the Markovian dependence of daily rainfall.

To measure the oscillatory character in the historic time series of rainfall,that is,to search for hidden periodicities within the instrumental time frame,spectral analysis is used to recognize this rhythm. The autocovariance, auotocorrelation are also utilized to set up the spectral analysis with its initial parameters with the help of Box-Jenkin's method.

The multiple regression analysis is also applied in this study whenever this is found to be meaningful,particularly,to establish the empirical crop-weather relationship in the district of Cooch Behar.

The maximum likelihood method,least squares method and the method of moments are adopted to aim at the evaluation of the unknown basic parameters of the populations from the information of sample or realization under study.

To establish the relationship between the distribution of population and the distribution of sample realization statistics ,the sampling theory are employed to measure the reliability of statistics. The basic sampling theory is concerned here mainly with the four sampling distributions.

These distributions are :

i) Standard Normal distribution.

ii) Student's t-distribution.

iii) Snedecor's F-distribution,and

iv) Chi-Square distribution

with it's limiting forms also.

These are the distributions of corresponding statistics,e.i. the z,t,F and X^2 statistics. These statistics are applied whenever these are required and meaningful.

Besides these tests, we have employed some other tests like Geary test and Durbin-Watson test etc.

The statistical tables of the sample distribution and the statistical tables for Geary test have been used comprehensively to get the theoretical values of the relevant test criteria.

For the analysis of Chapter VII, we have used the statistical software program of BMDP and the statistical procedure have been explained duly,CSSC,Indian Statistical Institute provides the Statistical Computer program for appropriate analysis.

In Chapter VIII,we have used the statistical program of the Micro-Data processor,Personal Computer (PC-FA-11) of N.B.Campus,Bidhan Chandra Krishi Viswavidyalaya,Cooch Behar.

CHAPTER -IV

VARIATION OF ANNUAL PRECIPITATION

CHAPTER IV

VARIATION OF ANNUAL PRECIPITATION

4.1. INTRODUCTION

In the Terai Agro-Climate zone,above 75% of its annual rainfall occurs during June to September. The major share of the water need of this zone during the entire year has to be met with the rainfall that is received during these four monsoon months. Large variations in rainfall distribution have been observed from year to year in this zone. Deficient and excessive rainfall are the result of extremes of the rainfall distribution.Cooch Behar is one of the districts which receive this type of variations of rainfall.

Very few studies have been undertaken to statistically examine the interval between successive drought years and flood years. Here we assume that the drought and flood, in meteorology may be considered as the deficit and excess of average rainfall over the historic time series of rainfall. Studies of drought and flood from this angle would bring out more diagnostic feature which could be useful in deficit and excess rainfall in this district.

The main objectives of this chapter are given below :

i) To test the normality of time series of rainfall.

ii) To determine the probability distribution of time intervals between successive deficit rainfall years.

iii) To determine the probability distribution of time intervals between the successive excess rainfall years.

iv) To study the distribution of largest intervals of drought and flood years.

v) To test the homogeneity of the distribution of same event at different places.

4.2. INDEX :

More than 75% of the annual rainfall in Cooch Behar occurs in the deep monsoon season months i.e. from June to September. For the purpose of identifing the excess and deficit years of rainfall,we introduce an index, which is the standardized annual rainfall to the time series data. The index,in the present study,serves the requirements to identify and to quantify the behaviour of the rainfall character. This criterion of rainfall series would express a standardized normal variate and is defined as

$$\text{Index} = \frac{(X_i - \overline{X})}{\sigma}$$

where X_i is the rainfall in the ith year in terms of annual.

$\overline{X}$ is the mean of annual rainfall over the time frame of the study period.

σ is the standard deviation of annual rainfall over the study period.

A desirable condition for the proposed index is that it should be a dimensionless number with negative and positive sign. The index should take into account the year to year variability of rainfall in this district.

The index would be positive when the ith year rainfall is greater than the mean rainfall and these would be regarded as the excess year of rainfall. And when the index is negative, the year would be treated as deficient year or drought year.

The index and its intensity would also be considered to identify the intensity of drought and flood year.

The classification of drought and flood intensities are shown in Table 4.I.

4.3. YEAR OF DROUGHT AND FLOOD

The drought years along with their intensities and ranking order to index values are given in Tables 4.2(a) and 4.2(b) for the stations,Cooch Behar and Dinhata respectively. All the drought years for each of the above mentioned stations are arranged in ascending order of magnitude of the index to their intensities and are given the rank number. The value of the index for individual year is considered for the ranking. The years which experienced drought of moderate intensity, are 1901, 1930, 1933, 1939, 1942, 1947, 1957, 1972, 1979 and 1980 at the Cooch Behar station where it is ascertained in the northern part of Cooch Behar district. In this area, 1930 appears to be the worst affected drought year. At the station Dinhata the moderate drought years are experienced in the years 1901, 1907, 1914, 1919, 1930, 1933, 1939, 1950, 1957, 1978 and 1980. 1978 experiences to be the worst drought affected year.

The occurrence of drought in two consecutive years is observed on eight occasions at Cooch Behar and four occasions at

Dinhata. These are 1903-4, 1914-15, 1922-23, 1925-26, 1936-37, 1939-40, 1946-47 and 1972-73 at Cooch Behar and 1914-15, 1929-30, 1950-51, and 1970-71 at Dinhata.

The consecutive occurrence of drought in three years is observed two times at Cooch Behar viz. 1928-30 and 1952-54 while it appears at Dinhata in four times viz. 1917-19, 1932-34, 1945-47 and 1980-82.

Occurrence of drought in five and eight consecutive years is observed once each time respectively and these years are 1959-63 and 1975-82 at the Cooch Behar station. But at Dinhata, the occurrence of drought in four, five and six consecutive years appeared once each time respectively and the years are 1959-62, 1936-40 and 1906-11.

The flood years along with their intensities and the ranking order of index values are given in Tables **4.3(a)** and **4.3(b)** for Cooch Behar and Dinhata stations respectivly.

The years which experienced flood of moderate intensity are 1902, 1905, 1906, 1907, 1910, 1916 and 1974 at Cooch Behar and that at Dinhata are 1916, 1927, 1954, 1956, 1973, 1977 and 1988. The severe flood years experienced at Cooch Behar are 1920 and 1988 while at Dinhata are 1969, 1974, 1984 and 1987. At the Cooch Behar station 1921 appeared to be the extreme flood affected year.

The occurrence of worst flood in two consecutive years is observed on three occasions at Dinhata and two occasions at Cooch Behar. These are 1927-28, 1973-74 and 1987-88 at Dinhata and 1910-11 and 1920-21 at Cooch Behar.

The consecutive three years of moderate flood years are observed once at each station, 1920-22 at Dinhata and 1905-07 at Cooch Behar.

The annual rainfall series of Cooch Behar are illustrated in Figure 4.1 and that of Dinhata are given in Figure 4.2.

It is interesting to note that the ariel distance of two rain gauge stations is very small, about 15 K.M. but the intensity of rainfall differs from each other.

4.4. STATISTICAL PROPERTIES OF INDEX

We wish to learn something of the naturally occurring variability of time-averaged mean by calculating the standard deviation of the time average means determined from different realizations. For these standard deviations to be truly representative of naturally occurring variability, time averaged means determined from one realization should be independent. Therefore, it is important that the assumption of the independence of yearly realization is to be reasonable one. Therefore, the index from the realization of annual rainfall is also considered as independent.

Before applying any statistical test to the index, it is necessary to ensure the homogeneity of the data. For this purpose this index series is divided into two parts, viz. 1901 to 1944 and 1945 to 1988. And Standard Normal test and Snedecor's F-test have been applied to establish the homogeneity of the whole series .The tests are applied to the index of annual rainfall data for the two raingauge stations in Cooch Behar district. Test of significance for difference of means and test of significance for difference of standard deviations are given herewith.

4.4.1. TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS.

Let $\overline{X}_1$ be the mean of a sample of size n_1 with standard deviation S_1 and let $\overline{X}_2$ be the mean of a sample of size n_2 with standard deviation S_2, Thus, null hypothesis is that the means of the two sub-periods are equal. Under the null hypothesis the test statistic becomes

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

The calculated values of mean and standard deviation are given in Table **4.4.** for both the rainfall series. The calculated values of Z for Cooch Behar and Dinhata are 1.37and 1.75 respectively.Calculated values of Z for both cases are less than the critical value i.e. 1.96 at 5% level of significance. The values of Z, then, are not significant. So, it is, as such, considered that the two sub-periods for each time series do not significantly differ from each other.

4.4.2. TEST OF SIGNIFICANCE FOR EQUALITY OF VARIANCES .

Suppose we want to test whether the two independent estimates of the variance are homogeneous or not.

Under the null hypothesis that the standard deviations of two subperiods are equal the statistic is given by

$$F = \frac{S_1^2}{S_2^2}$$

$$\text{if} \quad S_1^2 > S_2^2$$

The greater of the two variances is to be taken in the numerator. The calculated values of F-statistic for the two stations are also given in the same table. It is observed from Table 4.4. that the F-Values are 1.38 and 1.20. The tabular value of F is 1.58 with the degrees of freedom 43 and 43 at the 5% level of significance. The calculated value of F is less than the tabular value of variance ratio at 5% level of significance with 43 and 43 degrees freedom. So,it is not significant,we cannot reject the null hypothesis that the standard deviations are equal for these two sub-periods. The results are the same for the two rainfall indices. Considering these results of the two tests i.e., the mean test and the standard deviation test,the homogeneity of the data can be ensured. So, the two index series of 88 years are taken to be homogeneous to the first and second half period.

4.4.3. NORMALITY TEST

Before studying the long term variations in the precipitation data each rainfall series is examined for its normality aspect. Though the assumption of the normal distribution of climatic time series may not be strictly true, for a large number of variables such an assumption is reasonable. However, we may apply a test statistic which is known as Geary's test. The test criterion depends on the ratio of the mean deviation to the standard deviation under the null hypothesis that a normal distribution is a valid model to the long term rainfall data.

The test statistic is

$$G = \frac{\Sigma(|X_i - \overline{X}|)/N}{\sqrt{\Sigma(X_i - \overline{X})^2/N}}$$

Where X_i = *observation of the series*

i = 1,2,...............N

$\overline{X}$ = mean of the series

N = Total number of observations

From the table of the G-statistic the critical values of the test for different number of observations can be shown and *the null hypothesis may be tested. The normality test is carried out to the two time series data i.e., the annual rainfall at the two raingauge stations at Cooch Behar and Dinhata.*

We have to begin with a hypothesis to be tested, namely, that a particular data set is correctly modeled by a specified distribution. If the G value is less than 0.80 as in the case of annual rainfall at Cooch Behar, the nature of the distribution is skewed and leptokurtic while if the value of G is greater than 0.80, the nature of distribution as in the case of annual rainfall at Dinhata is symmetric and platykurtic. But it may be assumed that approximately the nature of the distribution in the latter case is symmetric and mesokurtic.

So, it may be concluded that the rainfall time series of the two raingauge stations approximately follow the normal distribution.

Table **4.5.** gives the statistics of the two annual rainfall time series where the mean, the standard deviation, the mean deviation and the value of Geary test with their tabular values are shown. The normal rainfall at Cooch Behar and Dinhata are 3533 mm and 2895 mm respectively. The mean and the standard deviation at Cooch Behar are greater than those at Dinhata but in the case of the mean deviation it is reversed. The normality test is carried out to the two time series data of annual rainfall. We have for the annual rainfall data of Cooch Behar whose mean deviation and standard deviation are 525.2 and 710 respectively and the value of G = 0.74. And the values of the mean deviation, the standard deviation and G for Dinhata centre are 450.58, 560 and 0.81 respectively.

The total number of observations N = 88 here, the table value gives a lower and upper 2.5 percent critical values of G-test. These are approximately 0.74 and 0.85 respectively. So, our calculated values of G are within this acceptance region. Then we may come to the conclusion that the normal distribution is shown to be an acceptable model. Table **4.4.** gives the values of G and other relative statistics.

4.4.4. RANK CORRELATION TEST

Considering the magnitude of each index ranking have been placed for the sets of the flood year as well as the drought year sequence. Spearman rank correlation has been applied to all the sequences of droughts and floods. The time of occurrence of drought and flood spells may be considered as a variable. And another variable is the rank of the observed data which can be evaluated on the intensity of the index of drought and flood sequence. The Spearman rank correlation (r_s) is defined by

$$r_s = 1 - \frac{6\Sigma(R_i - \quad {}^2}{n(n^2 - 1)}$$

Where R_i *is the rank of the h event and n is the number of pairs of variables.*

The rank correlation coefficient is nothing more than the product moment correlation coefficient with the rank values treated as numerical variable values in the correlation coefficient formula. The standard error of the rank correlation is obtained as $\frac{1}{\sqrt{n - 1}}$.

The test statistic is defined as

$$Z = r_s \sqrt{n-1}$$

which is distributed as standard normal distribution.

We have the drought sequence for the two stations and the flood sequence for the two stations also. The Spearman rank correlation has been applied to both the sequences. The calculated rank correlation co-efficients are given in Table 4.6. with their standard errors. The calculated Z values and its hypothetical values are also given in that table.

The rank correlation coefficients of the drought sequence at Cooch Behar and of the flood sequence at Dinhata are very low and their standard errors are quite high in relation to their respective correlation coefficient. This shows that the rank correlation coefficient in above mentioned cases are statistically non-significant.We may come to the conclusion that the flood sequence of Dinhata and the drought sequence of Cooch Behar do not indicate any trend to the occurrence of flood at Dinhata and drought at Cooch Behar. Thus the drought at Cooch Behar and the flood at Dinhata have occurred randomly.

The rank correlation coefficient for the drought sequence of Dinhata is equal to its standard error subsequently the Z-value is one,so we may come to the conclusion as described in the above case. Thus the drought sequence does not indicate trend as it is also non-significant.

The rank correlation coefficient of the flood sequence at Cooch Behar is slightly greater than its standard error but the Z-value cannot give any significant evidence that we may come to any conclusion reversely.

The analyses of trend and homogeneity are carried out in this sub-section in using the Spearman rank correlation,Standard Normal-test and Snedecor's F-test. These analyses indicate that there is no significant change in rainfall amounts over Cooch Behar district during last 88 year.

The analyses are in agreement with Rao and Jagannathan (1963) who observed no significant change in rainfall over India during last 100 years.

4.5. INTERARRIVAL TIME OF DROUGHT AND FLOOD

The index has the dual purpose for assessing drought and flood years also with their varying intensity. Here,we introduce the flood as success event and the drought as failure event in the statistical point of view. On the basis of the value of the index,we have obtained the basic data of the drought sequence separated by drought free intervals of varying lengths. And applying the same method we would have the flood sequence separated by flood free intervals of varying length. Present sub-section of this study introduces some statistical techniques in the analysis of such time intervals. The equality of the mean and the variance is an important characteristic of the Poisson distribution. Whereas the negative binomial distribution provides an excellent model because the distribution has a variance larger than the mean. An interesting point is that the mean and the variance are measured from the origin zero,the mean alone is affected. But the variance remains unchanged. The distribution of interarrival times of drought or flood sequences, the zero class missing, is assumed to be meaningless when we consider to fit the probability distribution on the time interval between the successive occurrence of drought or flood sequences. Many research workers have used different statistical methods to study frequency distributions of drought occurrences in a fixed interval of time or in an interval of varying length with the omission of zero class. And the data have been fitted to the Poisson distribution. Here the mean value of the Poisson process has a significant role in determining the density function for successive occurrence of drought and flood sequences. In this process we consider the mean of the Poisson process as a constant or as a stochastic process.

Now, the parameter of a Poisson distribution is a random variable having gamma distribution then the process is a mixed Poisson process. Barn Droff-Neilson (1.1969) *confirmed the result. But the problem arises when the behaviour of the parameter of Poisson process is unknown.*

In the sampling procedure, Kendall and Stuart (1977) suggested that the arbitrary assumptions about the distribution of the parameter of the Poisson process are of no use to fit the data on negative binomial distribution. Now we proceed on this angle of direction.

Let f(x : r,p) denote the probability that there are x failure years preceding the r-th success year in r + x trial. Now the last trial must be a success whose probability is p. In the remaining (x+r-1) trials (years) we must have (r-1) success whose probability is given by

$$\binom{x+r-1}{r-1} p^{r-1} q^{x}$$

where $q = 1-p$

Therefore, by compound probability theorem

f(x:r,p) is the product of these two probabilities, i.e.

$$\binom{x+r-1}{r-1} p^{r-1} q^{x} p = \binom{x+r-1}{r-1} p^{r} q^{x}$$

Which is the required density function of negative binomial distribution.

4.5.1 ESTIMATION OF PARAMETERS OF NEGATIVE BINOMIAL DISTRIBUTION

The method of moments and the maximum likelihood method have been applied to estimate the parameters of this distribution. The estimators are unbiased also. The estimation of r is not so straight forward. However, the method of moments can be applied in an indirect fashion.

We have to estimate the parameters for the negative binomial distribution which are given by

$$\text{Mean} \quad (\mu) = \frac{rq}{p}$$

$$\text{Variance} = \frac{rq}{p^2}$$

Thus, p, q and r have been calculated from the observed frequency distribution by using the estimates of the mean and the ariance of the distribution.

The recurrence relation for fitting the negative binomial distribution is given by

$$f(x+1) = \frac{r+x}{x+1} \quad f(x)$$

$$\text{where} \quad x = 0,1,2,\ldots\ldots\ldots\ldots$$

$$\text{and,} \ f(0) = p^r$$

Subsequently, we get the probabilities of each class. The calculated probability of each class is multiplied by the total number of frequency. We get the expected frequencies of each class which are presented in Tables **4.7.**(a & b) for both drought and flood sequences of the stations.

4.5.2. GOODNESS OF FIT TEST

A very powerful test for testing the significance of the discrepancy between the theory and experiment is Chi-square test of goodness of fit.

If, O_i (i = 1,.......n) is a set of experimental frequencies and E_i (i = 1,.......n) is the corresponding set of theoretical frequencies, then Karl pearson's Chi-square given by

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad \text{where} \quad \sum_i O_i = \sum_i E_i$$

follows Chi-square distribution with (n-1) degrees of freedom.

Tables **4.2.(a & b)** show the years of drought or flood. A perusal of the occurrence of drought or flood years show that the sequence for both categories are rather random which have been tested earlier.

Following the procedure mentioned earlier we get the frequency distribution of interarrival times of drought years as well as interarrival times of flood years for both the raingauge stations.

The observed frequency distribution of drought intervals for Cooch Behar and Dinhata are shown in Tables **4.7.(a)**,and the observed frequency distribution of flood intervals for both the stations are given in Table **4.7(b)**. The observed frequency distribution of time intervals between the occurrence of successive droughts and floods are illustrated in Figures **4.3.(a)** and **4.4.(a)** for Coochbehar station and in Figures **4.3.(b)**, and **4.4.(b)** for Dinhata station.

The frequency distributions of time intervals between the occurrence of successive droughts and floods are of the nature of exponential. The observed data are fitted to the negative binomial distribution. The parameters of the distribution are estimated from the observed frequency distribution. The parameters for the interarrival times of drought years for both the stations are given in Table **4.7(a)**. And that for the interarrival times of flood years for both the stations are given in Table **4.7(b)**. Introducing the recurrence relation we get the expected value of the corresponding frequencies which are also given in these tables. The observed and expected frequencies are denoted by 'O' and 'E' in the tables. All the frequency distributions for the interarrival times of flood or drought years are fitted to the negative binomial distribution.

Chi-square goodness of fit test has been applied to test the validity of the negative binomial model. The observed and expected frequencies are given in Table **4.7(a)** for the interarrival times of droughts and in Table **4.7(b)** for the interarrival times of floods at both the stations. After applying the chi-square goodness of fit test to the expected and observed frequencies,we get the calculated value of chi-square. These values are also given in the tables. Tables **4.7(a & b)** also show the critical values with suitable degrees of freedom.

The computed values of Chi-square are not found significant at 5 percent level for all cases. We,therefore,accept the null hypothesis that there are no significant differences between the observed values and the theoretical values of the time interval for both the events. So,the distribution of interarrival times of droughts at Cooch Behar and Dinhata can be considered to follow the negative binomial distribution. And the distribution of interarrival times of floods, the negative binomial distribution may be judged a valid model. Therefore the distribution of interarrival times of drought and flood years for Cooch Behar and Dinhata are overall best fitted to negative binomial model.

Hence,we may come to the conclusion that the negative binomial distribution is a good fit to the experimental data.

4.5.3. DISTRIBUTION OF THE LARGEST INTERVAL.

Fisher's 'g' statistic is introduced here to test cyclic order of the largest interval of drought or flood sequences. Fisher's 'g' statistic is the ratio of largest interval between occurrence of an event to the sum of such intervals. If t_L be the largest among n independent time intervals $\bar{t}$ be the mean of the intervals then the test statistic 'g' is defined as

$$g = \frac{t_L}{n\bar{t}}$$

For testing significance of the largest interval,it is assumed that each of the n intervals contributes a certain fraction to the total sum of squares and 'g' is taken to be the largest of these fractions. The probability of 'g' exceeding any given value has been worked out by Fisher as described by Priestly (1981).

If the value at 5 percent level of significance is $g_{.05}$ then the largest interval t_L is significant if

$$t_L \geq g_{.05}\, n\bar{t}$$

The significance test described above is based on the assumption that the probability that the largest of the n intervals should exceed 'g' is given by

$$\alpha = n(1-z)^{n-1}$$

where α is the level of significance.

z is the critical value of the test.

n is the total number of frequencies.

The procedure outlined above may also be used to test the largest interval for its occurrence on drought and flood aspect.

The computed value of 'g' and the theoretical value of 'g' are given at the end of Table **4.7(b)**. The maximum interval between droughts at Cooch Behar and Dinhata are 6 years and 5 years respectively. The computed values of 'g' are 0.15 and 0.14 and the hypothetical values of 'g' are 0.13 and 0.14 respectively for the two stations as given in Table **4.7(a)**. The hypothetical values of the maximum interval for droughts are 5.07 and 5.05 years for Cooch Behar and Dinhata respectively. The hypothetical value of the drought interval for Dinhata is greater than the observed interval of 5 years. So this largest interval between droughts is not significant,so the maximum drought interval at

Dinhata may not maintain any cyclic order. Thus this may occur purely randomly. But the largest interval between droughts for Cooch Behar is something different. Here the hypothetical value of t_L is less than the observed interval of 6 years. Hence,the largest interval is significant. Therefore,the largest drought interval at Cooch Behar may be expected to occur periodically by 6 years.

The analysis of the largest interval between droughts at Cooch Behar may recognize in agreement with Sarkar (1979) and chowdhury and Abhyankar (1984) that drought may experience once in 6 to 8 years in Sub-Himalayan West Bengal region. The hypothetical values of the largest interval for flood are 7.54 and 6.74 for Cooch Behar and Dinhata respectively. The hypothetical value of thelargest interval between floods is less than the observed interval of 8 years at Cooch Behar. So it seems to be significant. Therefore,the largest interval between floods may occur cyclically by 8 years. So,it may be concluded that the flood can be expected to occur once in 8 years. The hypothetical value of the largest interval between floods is greater than the observed value of the flood interval of 6 years. Hence,it shows statistically non-significant.

Studies with this technique of flood incidence in any place of India are also not available. The hypothetical values of drought and flood intervals for Cooch Behar and Dinhata are given in Tables **4.7(a & b)**. So,we come to the conclusion that drought and flood may be expected to occur once in 6 years and 8 years respectively. And drought can occur more frequently than floods at Cooch Behar.

4.6. EQUALITY OF DISTRIBUTIONS OF SAME EVENT :

A problem of great importance is that whether several random realizations of the same meteorological event can be considered as drawn from the same population of rainfall. We have two distributions of the same meteorological event like drought and flood. We may consider that the distributions of drought intervals and flood intervals come from the same population of the respective events. We also have that the interarrival times of drought years as well as flood years are considered to follow the negative binomial distribution. We like to test here that the two distributions of interarrival times of drought or flood years are homogeneous. To test this,we set the null hypothesis that the two negative binomial distributionsfor each event are same.

We group the frequencies of each class and the last group includes all frequencies greater than 4. The groups and their corresponding frequencies are given in Table **4.8.** By using the method of moments and the maximum likelihood method,we have to estimate the parameters of the common negative binomial distribution of each event like drought and flood separately.

The expected frequencies of each group are estimated by recurrence relation as usual. To estimate the expected frequency of each group,we have multiplied the fitted probabilities by the average number of observations of the respective event. The observed frequencies and the expected frequencies are given in

Table **4.8**. The hypothesis can be tested in a variety of ways one of which is the limiting chi-square distribution as given below :

$$G_{2k} = \sum_{i=1}^{2} \sum_{j=0}^{K} \frac{(N_{ij} - n_i p_j)^2}{n_i p_j}.$$

has a limiting chi-square distribution with (2K-2) degrees of freedom. Here 2 degrees of freedom are lost due to two parameters being estimated from the observed values.

Where K is the number of groups, N_{ij} is the frequency of the j-th.group in the *distribution.*

$n_i p_j$ is replaced here by the estimated frequencies of the common distribution.

*The calculated value of the test statistic and tabulated value of the Chi-square with suitable degrees of freedom at 5 percent level of significance are given in Table **4.8.** It is found that the calculated value of Chi-square for the time interval of drought years is less than the tabulated value of Chi-square with 6(8-2) degrees of freedom at the 5 percent level of significance. So,we may accept the null hypothesis. We come to the conclusion that there is no evidence from the data to support that the two distributions of interarrival time of drought years for two different places are different negative binomial populations. The calculated value of test statistic for the time interval of flood years is less than the tabular value of Chi-square with 6(8-2) degree of freedom at 5 percent level of significance. So,it is statistically non-significant,hence,we cannot reject the null hypothesis that the two distributions of interarrival time of flood years are drawn from the same population.*

4.7 CONCLUSION

The annual rainfalls recorded in this region are independently and normally distributed. The drought and flood years,in this district appear to have no trend i.e., they occur randomly.The frequency distribution of time interval between the occurrence of successive drought and flood years are considered to follow the negative binomial distribution. The behaviour of occurrence of successive drought as well as flood,in meteorology are identical for both the places. Drought and flood may be expected to appear once in 6 years and 8 years respectively in the district of Cooch Behar.

TABLE - 4.1

CLASSIFICATION OF DROUGHT AND FLOOD YEARS CONSIDERING THEIR INTENSITIES.

CLASS	INTENSITY
DROUGHT	
- .01 to - 0.99	Slight
- 1.0 to - 1.99	Moderate
- 2.0 to - 3.00	Severe
FLOOD	
+ .01 to +0.99	Slight
+ 1.0 to +1.99	Moderate
+ 2.0 to +3.00	Severe
+ 3.00 and above	Extreme

TABLE : 4.2(a)

YEAR OF DROUGHT AT COOCH BEHAR

YEAR	INDEX VALUE	CATEGORY	RANKING
1901	1.059	Moderate	13
1903	0.171	Slight	42
1904	0.191	Do	41
1908	0.459	Do	31
1912	0.126	Do	46
1914	0.868	Do	16
1915	0.875	Do	15
1917	0.746	Do	19
1919	0.337	Do	36
1922	.284	Do	38
1923	0.623	Do	24
1925	0.467	Do	30
1926	0.282	Do	39
1928	0.352	Do	35
1929	0.695	Do	22
1930	1.780	Moderate	01
1933	1.056	Do	11
1934	0.136	Slight	44
1936	0.49	Do	29
1937	0.725	Do	21
1939	1.116	Moderate	07
1940	0.905	Slight	13
1942	1.07	Moderate	18

1944	0.655	Slight	23
1946	0.96	Do	12
1947	1.216	Moderate	05
1950	0.409	Slight	32
1952	0.902	Do	14
1953	0.103	Do	47
1954	0.086	Do	48
1957	1.287	Moderate	04
1959	1.065	Do	09
1960	0.733	Slight	20
1961	0.853	Do	17
1962	0.613	Do	25
1963	0.518	Do	27
1970	0.171	Do	43
1972	1.166	Moderate	06
1973	0.246	Slight	40
1975	0.30	Do	37
1976	0.779	Do	80
1977	0.372	Do	34
1978	0.398	Do	33
1979	1.419	Moderate	03
1980	1.441	Do	02
1981	0.499	Slight	28
1982	0.061	Do	49
1984	0.128	Do	45
1986	0.535	Do	26

TABLE 4.2(b).

YEARS OF DROUGHT AT DINHATA

YEARS	INDEX VALUE	CATEGORY	RANKING
1901	1.133	Moderate	13
1904	0.949	Slight	90
1906	0.757	Do	22
1907	1.081	Moderate	14
1908	1.717	Do	03
1909	0.366	Slight	33
1910	0.560	Do	30
1911	0.106	Do	43
1914	1.049	Moderate	15
1915	0.742	Slight	23
1917	0.830	Do	20
1918	0.209	Do	38
1919	1,124	Moderate	09
1923	0.348	Slight	34
1929	0.687	Do	25
1930	1.235	Moderate	07
1932	0.332	Slight	35
1933	1.03	Moderate	16
1934	1.128	Do	08
1935	1.394	Do	05
1937	1.185	Do	10
1938	0.592	Slight	29

1939	1.16	Moderate	11
1940	1.722	Do	02
1942	0.983	Slight	17
1945	0.171	Slight	41
1946	0.107	Do	44
1947	0.621	Do	27
1950	1.115	Moderate	12
1951	0.175	Slight	40
1953	0.170	Do	42
1957	1.263	Moderate	06
1959	0.617	Slight	28
1960	0.318	Do	36
1961	0.955	Do	18
1962	0.512	Do	31
1966	0.792	Do	21
1968	0.009	Do	45
1970	0.191	Do	39
1971	0.225	Do	37
1975	0.726	Do	22
1978	1.847	Moderate	01
1980	1.676	Do	04
1981	0.476	Slight	32
1982	0.626	Do	26

TABLE 4.3(a)

YEAR OF FLOOD AT COOCHBEHAR

YEARS	INDEX VALUE	CATEGORY	RANKING
1902	1.395	Moderate	10
1905	1.748	Do	07
1906	1.758	Do	06
1907	1.889	Do	04
1909	0.088	Slight	35
1910	1.409	Moderate	09
1911	1.302	Do	11
1913	.249	Slight	30
1916	1.794	Moderate	05
1918	0.364	Slight	25
1920	2.74	Severe	02
1921	4.25	Extreme	01
1924	0.825	Slight	15
1927	0.542	Do	90
1931	0.60	Do	16
1932	0.575	Do	17
1935	0.349	Do	26
1938	0.367	Do	23
1941	0.088	Do	36
1943	0.365	Do	24
1945	0.214	Do	31
1948	0.311	Do	27

1949	0.87	Do	37
1951	0.105	Slight	34
1955	0.034	Do	39
1956	0.171	Do	33
1958	0.84	Do	14
1964	00.392	Do	22
1965	0.049	Do	38
1966	0.294	Do	28
1967	0.419	Do	21
1968	.445	Slight	20
1969	.26	Slight	29
1971	.549	Do	18
1974	1.416	Moderate	08
1983	0.176	Slight	32
1985	0.963	Do	12
1987	0.884	Do	13
1988	2.596	Severe	03

TABLE - 4.3(b).

YEAR OF FLOOD AT DINHATA

YEAR	INDEX VALUE	CATEGORY	RANKING
1902	0.96	Slight	19
1903	0.353	Do	34
1905	0.387	Do	31
1912	0.342	Do	35
1913	0.99	Do	18

1916	1.32	Moderate	10
1920	1.558	Moderate	06
1921	1.599	Do	05
1922	1.422	Do	09
1924	1.535	Do	07
1925	0.412	Slight	29
1926	0.359	Do	33
1927	1.192	Moderate	12
1928	1.008	Do	17
1931	0.939	Slight	21
1935	0.02	Do	42
1941	0.896	Slight	22
1943	0.571	Slight	26
1944	0.161	Do	37
1948	0.764	Do	23
1949	0.714	Do	24
1952	0.002	Do	43
1954	1.03	Moderate	15
1955	0.6	Slight	25
1956	1.027	Moderate	16
1958	1.119	Do	13
1963	0.09	Slight	40
1964	0.093	Do	39
1965	0.405	Do	30
1967	0.155	Do	38
1969	2.55	Severe	04

1972	0.951	Slight	20
1973	1.105	Moderate	14
1974	2.659	Severe	02
1976	0.387	Slight	32
1977	1.224	Moderate	11
1979	0.039	Slight	41
1983	0.294	Do	36
1984	2.761	Severe	01
1985	0.564	Slight	27
1986	0.430	Do	28
1987	2.580	Severe	03
1988	1.517	Moderate	08

T A B L E - 4.4

Z VALUE AND F VALUE OF THE SUB-PERIOD OF INDEX SERIES

STATISTICS	COOCH BEHAR		DINHATA	
	1st Half	2nd Half	1st Half	2nd Half
n	44 (n_1)	44 (n_2)	44 (n_1)	44 (n_2)
Mean	0.163	- 0.151	- 0.145	0.238
Standard	1.157	0.9836	0.974	1.0667
Calculated value of Z		1.37		1.75
Hypothetical value at 5%		1.96		1.96
Calculated value of F	1.38			1.20
Table value F at 5%	1.58			1.58
D.F.	(43,43)			(43,43)

TABLE = 4.5.

BASIC STATISTICS OF TWO SERIES

	COOCH BEHAR	DINHATA
N	88	88
Mean	3533	2895
Standard deviation	710.00	560.00
Mean deviation	525.2	450.58
Calculated value of G	0.74	0.80
Theoretical value of G		
(Lower & Upper)	0.74 - 0.85	

T A B L E = 4.6.

RANK CORRELATION OF DROUGHT AND FLOOD SEQUENCE

COOCH BEHAR	DROUGHT	FLOOD
No. of pairs.	49	58
Correlation coefficient.	-.027	0.17
S.E.of Correlation coefficient.	0.144	0.164
Calculated Z-value.	0.19	1.034

DINHATA	DROUGHT	FLOOD
No. of pairs	45	43
Correlation of Coefficient	0.15	- 0.04
S.E.of correlation Coefficient.	0.15	0.15
Calculated Z value	1.00	.26
Hypothetical value at 5%	1.96	1.96

T A B L E – 4.7(a)

THE OBSERVED AND EXPECTED FREQUENCY OF DROUGHT SEQUENCE.

INTERVAL	COOCH BEHAR		DINHATA	
	O	E	O	E
0	23	24.66	25	23.32
1	18	13.03	9	10.44
2	4	5.04	5	5.15
3	2	3.67	4	3.5
4			0	1.04
5			1	0.82
6	1	1.60		
Total (n)	48	48	44	44
$\bar{x}$	0.8125		.82	
σ^2	1.1942		1.37	
p	0.68		0.59	
q	0.32		0.41	
r	1.7265		1.2	
x^2(cal)	3.20		0.35	
DF	1		1	
X^2tab(5%)	3.84		3.84	
g^*(cal)	0.15		0.14	
g(theoretical)5%	0.13		0.14	
Hypothetical t_L	5.07		5.05	

T A B L E = 4.7(b)

THE OBSERVED AND EXPECTED FREQUENCY OF FLOOD SEQUENCE.

INTERVAL	COOCH BEHAR		DINHATA	
	O	E	O	E
0	14	16.03	21	20.07
1	12	10.03	10	10.66
2	8	5.62	4	5.51
3	2	3.02	4	2.82
4	0	1.59	1	1.51
5	1	1.00	1	1.00
6	0		1	0.43
7	0			
8	1	0.71		
Total(n)	38	38	42	42
$\overline{X}$	1.24		1.07	
σ^2	2.44		2.16	
p	0.506		0.496	
q	0.494		0.506	
r	1.267		1.053	
x^2(cal)		2.50		0.76
DF		1		1
x^2-tab(5%)		3.84		3.84
g^*(cal)		0.17		013
g (theoretical)5%		0.16		.15
Hypothetical t_L		7.54		6.74

T A B L E = 4.8.

DISTRIBUTION OF INTERVAL OF TIME FOR TWO PLACES

DROUGHT

	COOCH BEHAR OBSERVED	DINHATA OBSERVED	COMMON	EXPECTED
0	23	23	48	22.78
1	18	9	27	14.67
2	4	5	9	5.84
3	2	4	8	2.89
4	0	0	–	–
5	0	1	–	–
6	1	0	–	–
Total	48	44	92	46
X^2 (cal)				4.71
DF				6
X^2 at 5% (tab)				12.6

FLOOD

	COOCH BEHAR OBSERVED	DINHATA OBSERVED	COMMON	EXPECTED
0	14	21	35	16.28
1	12	10	22	12.91
2	8	4	12	6.58
3	2	4	11	4.23
4	0	1		
5	1	1		
6	0	1		
7	0	0		
8	1	0		
Total	38	42	80	40
X^2 (cal)			5.55	
DF			6	
X^2 at 5%(tab)			12.6	

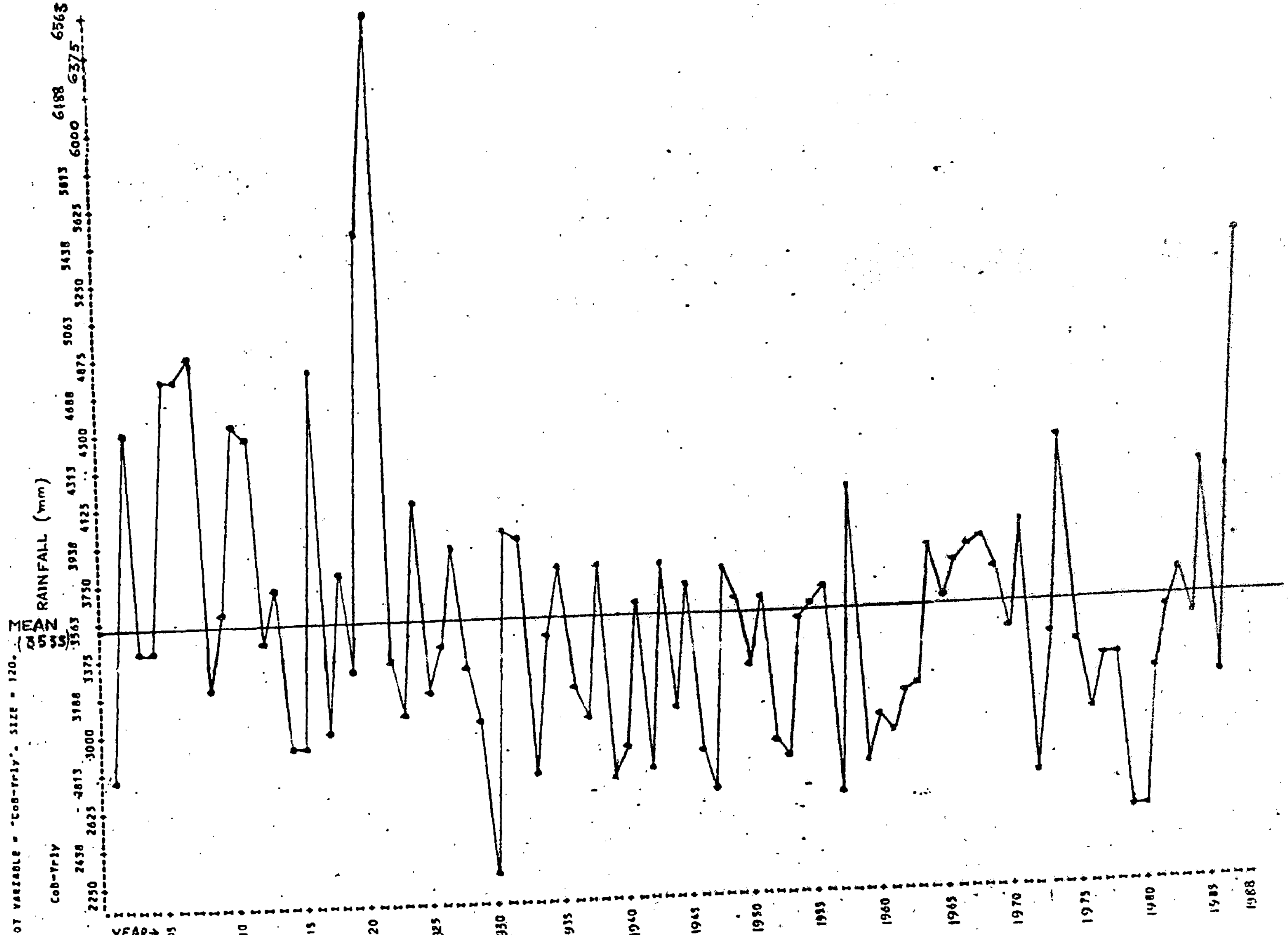


Fig.4.1. Annual rainfall of Cooch Behar from 1901 to 1968 .

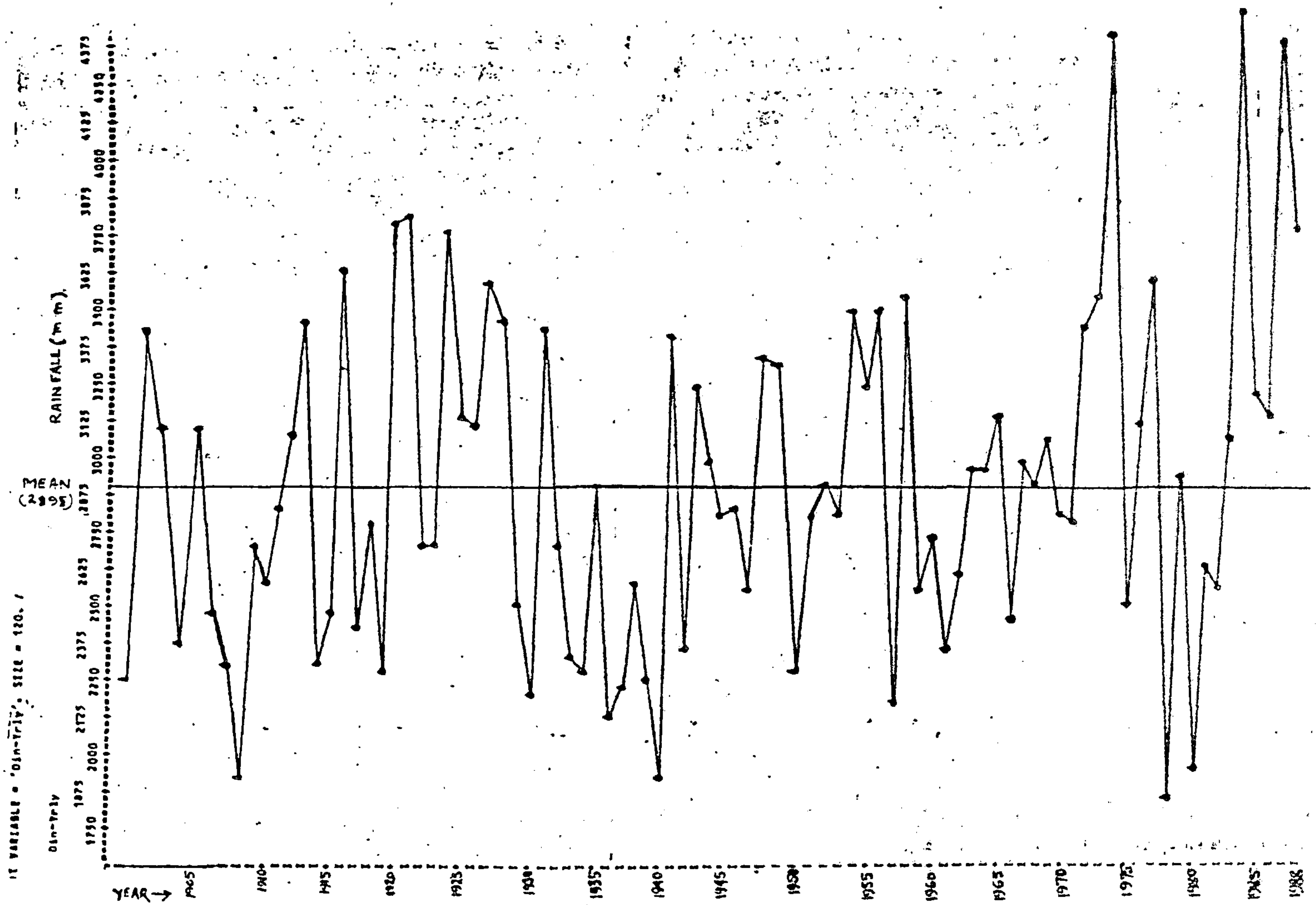


Fig.4.2 Annual rainfall of Dirhat from 1901 to 1988.

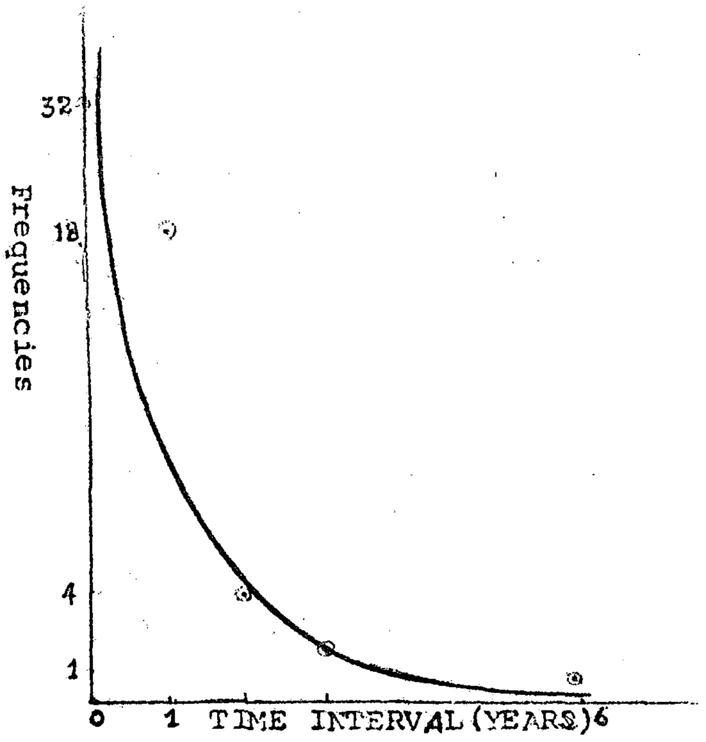


Fig.4.3 (a) DROUGHT TIME INTERVAL (COOCHBEHAR)

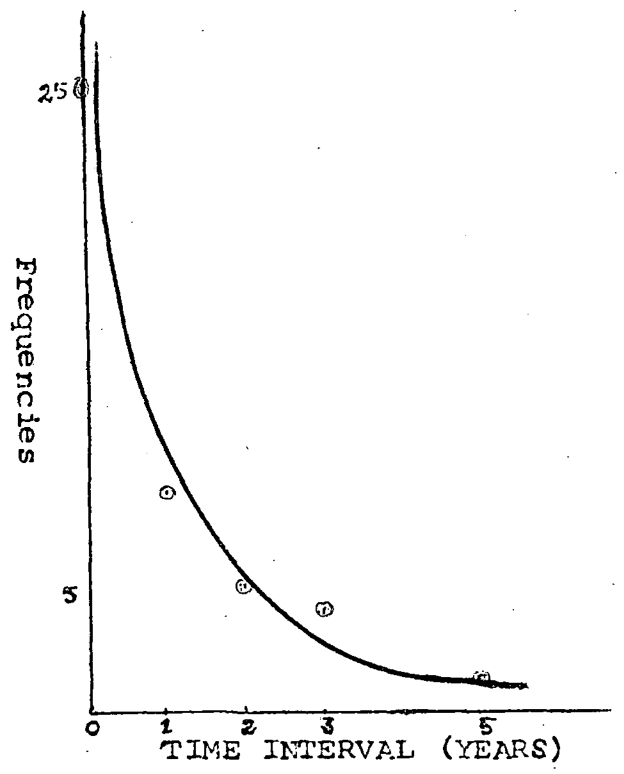
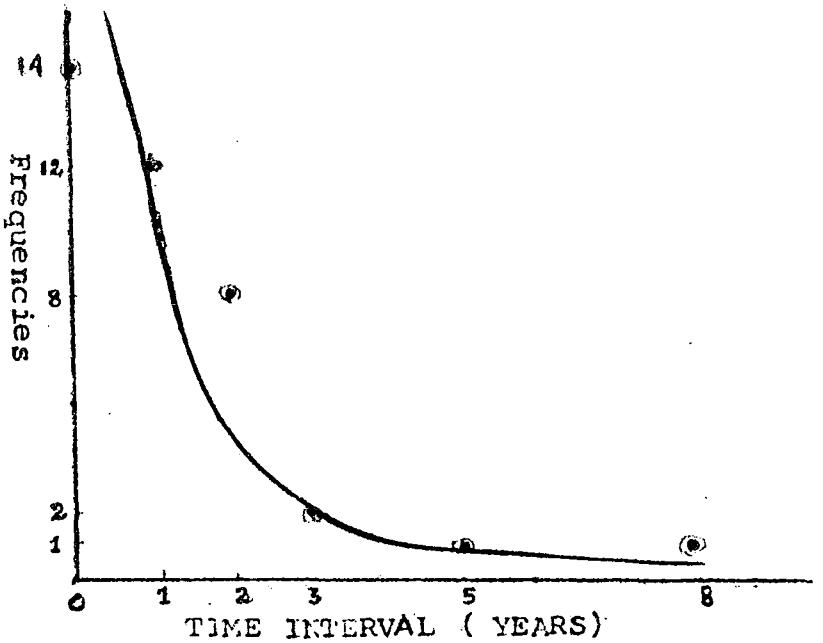


Fig.4.3 (b) DROUGHT TIME INTERVAL (DINHATA)

Fig.4.4.(a) .FLOOD TIME INTERVAL (COOCHBEHAR)



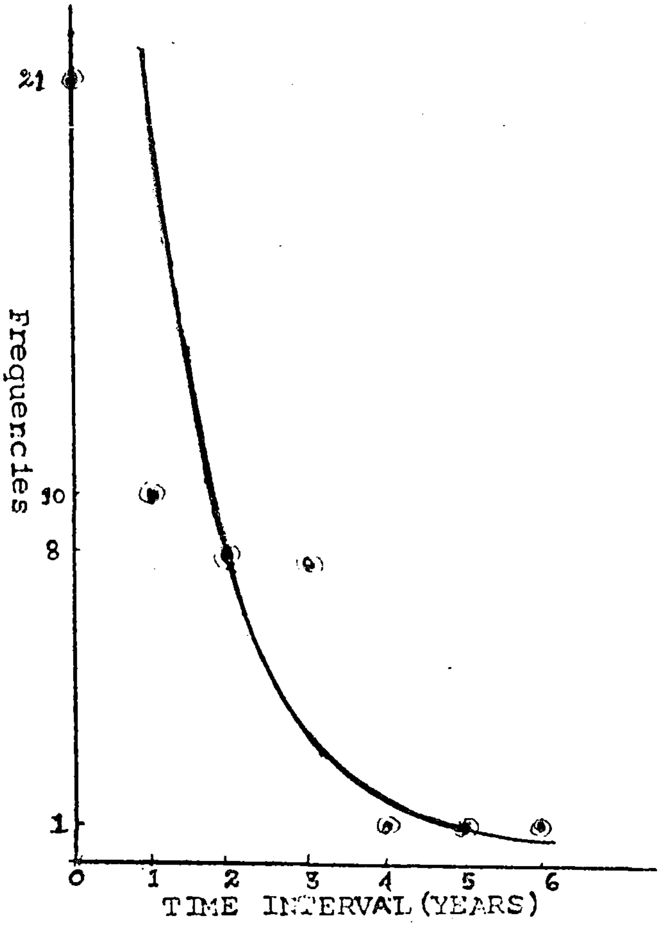


Fig.4.4.(b) FLOOD TIME INTERVAL (DINHATA)

CHAPTER - V

PATTERN OF DAILY RAINFALL

CHAPTER -V

PATTERN OF DAILY RAINFALL

5.1 INTRODUCTION

The knowledge of weather,in general,of rainfall in particular for performing the pre-cultivation operations and sound crop planning needs much emphasis. Mostly the work in this direction is confined to the total rainfall,maximum and minimum amount of rainfall and its ranges etc. However,not only the inadequacy of rainfall but also excess or deficiency is harmful to a particular crop . It is also natural to imagine that for the total agricultural production the total amount of rainfall is not of much importance but the pattern of its occurrence such as spell of rainy and dry days, expected number of dry days between two rainy days and their repetitions are of much use for the rainfed crops in this region. The repetitive behaviour of weather during seasonal months has always fascinated meteorologists and Statisticians. Meteorologists seek physical explanations for such phenomena and statisticians explore possibilities of model building to explain random phenomena. Such models serve the important function of providing an orderly basis and permit further use of the deductive power of mathematics to reach conclusions that may even provide clues to physical understanding of the complex phenomena.

Among the earlier studies may be mentioned the work of Cochran (1938) who proposed a probability model, based on the "Theory of Runs" to study the "persistancy" behaviour of rainy days while others like Gabriel and Neumann(1957) suggested geometric distribution as a suitable model for wet and dry weather spells. Gabriel and Neumann (1962) considered empirical evidence to find out the suitability of the pattern of occurrence of rainfall with the help of Markov chain model. With respect to the precipitation phenomena,a major thrust of this stochastic approach has been to predict the behaviour of spell distribution and subsequently the weather cycle under the assumption that the probability of occurrence of precipitation of any day depends upon the previous observation and only on it. So,Markov chain model has been used to serve this purpose.

To be precise,the main objectives of this section are :

i) To fit the Markov chain model to daily rainfall data.

ii) To determine the distribution of dry spells and wet spells and dry-wet and wet-dry cycles.

iii) To estimate the average lengths of dry spells and wet spells and the expected length of weather cycle.

5.2. MARKOV CHAIN MODEL OF DAY'S WEATHER.

For the purpose of the present study this daily rainfall data are based on the daily rainfall during the period 30th April to 30th September in each year.Each day is classified as wet day if the day receives the amount of rainfall greater than or equal to 2.5 mm or a dry day if the amount of rainfall is less than 2.5 mm.

This classification gives a sequence of wet and dry days which can be regarded as a two-state Markov chain with wet and dry days. Each day of this sequence is classified as one of the following possibilities.

i) A dry day preceded by a dry day.

ii) A wet day preceded by a dry day.

iii) A dry day preceded by a wet day.

iv) A wet day preceded by a wet day.

Thus , for each year,the 30th April which is regarded as the initial day may be considered for classifying the 1st May for each year. Thus, for each year, the nature of a day is classified as one of the four possibilities depending on the previous day is dry or wet.

Repeating the process each cell frequencies for the above four possibilities are obtained.

Let these frequencies be f(D/D),f(W/D), f(D/W),f(W/W) respectively with

$$f(D/D) + f(W/D) = n_1$$

and

$$f(D/W) + f(W/W) = n_2$$

The cell frequencies are arranged in the matrix form for the month of May , June,July,August and September based on daily rainfall data at Cooch Behar. These cell frequencies for each month separately are given in Table 5.1. And the corresponding cell frequencies of all the months are pooled and we get a new form of matrix computed by pooling the corresponding frequencies for May, June, July, August and September.

These pooled or total frequencies are also reported in the same Table.

It is obvious that the two conditional probabilities,

p_{12} and p_{22} which have to be estimated,are required for describing the two state Markov chain model.

The maximum likelihood methods have been applied to estimate these parameters of the model.

The conditional probabilities can be estimated as below :

$$P_{12} = \frac{f(W/D)}{n_1} \quad \text{and} \quad P_{22} = \frac{f(W/W)}{n_2}$$

So, under the two-state Markov chain model the other two conditional probabilities are easily obtained as

$$P_{11} = 1-P_{12} \quad \text{and,} \quad P_{21} = 1-P_{22}$$

Given that the previous day is dry,let the transition probabilities of a day being dry and wet be respectively

P_{11} and P_{12} with $P_{11} + P_{12} = 1$.

Similarly given that the previous day is wet, let the transition probabilities of a day being dry and wet are respectively P_{21} and P_{22} with $P_{21} + P_{22} = 1$

The transition probability matrix,in such a case can be arranged as

$$P = \begin{vmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{vmatrix}$$

Since the sequence of dry and wet days can be regarded here as a finite sequence on the time axis,we can take the starting day i.e. the 30th April as an initial day, as dry or wet. The process is then expected to settle down to a two-state Markov chain model for each month as well as their pooled estimates. The stochastic matrix of each month i.e. May, June, July, August and September and the pooled one are also reported in Table **5.2.** It is indicated in Table 5.2 that the transition probability of the occurrence of a wet day preceded by a wet day for each month at Cooch Behar is always the highest among all the conditional probabilities during the monsoon months. But the conditional probability for the occurrence of a dry day preceded by a dry day is the highest among the other conditional probabilities in the month of May at Cooch Behar.

5.2.1. TEST OF INDEPENDENCE :

For determining whether,the occurrence of a wet day and a dry day depends only on the immediately preceding day's weather, we test the hypothesis of independence against Markov dependence.

Now under the null hypothesis the likelihood ratio statistic has been used for the testing. The test statistic is given by

$$\lambda^{*} = -2 \operatorname{Log} \lambda = 2 \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \log \frac{(n) n_{ij}}{n_{i} n_{j}}$$

Usual notations have been used here,i.e. n_{ij} = cell frequency of ith raw and jth column, n_{i} = total of ith row, n_{j} = total of jth column and $n = \sum^{2} n_{i} = \sum_{j}^{2} n_{j}$

The test statistic has an asymptotic Chi-square distribution with one degree of freedom for the two-state Markov chain model. The calculated values of test statistic (λ^{*})are given in Table 5.3. for each of the months and for the pool of these months. The test of independence strongly suggests that the probability of the occurrence of rainfall of any day is dependent on the weather of the previous day as the calculated values of the test statistic are much greater than the tabulated value of chi-square at one percent level of significance. This indicates that the Markov chain model is very much effective on the weather of a day at Cooch Behar during the premonsoon month like May and effective monsoon months June to September as well as the pooled observations for all these months.

Therefore,we come to the conclusion that the hypothesis of independence of weather of consecutive days is to be rejected and therefore,the occurrence of a wet or dry day is influenced by the previous day's weather only.

5.3. SPELL AND WEATHER CYCLE DISTRIBUTION

A rainfall event is described by two sets of meteorological data representing wet and dry phases. In a sequence of wet and dry days, there are possibilities of the occurrence of wet days and dry days in succession. A number of successive wet days preceded and followed by a dry day is regarded as the length of a wet spell. Subsequently a dry spell is defined in the similar way. A weather cycle is defined as the combination of a wet spell with the immediate successive dry spell (wet-dry cycle) or a dry spell with the immediate successive wet spell (dry-wet cycle). The following criteria are fixed in the analysis for classification of a wet day,wet/dry spell and weather cycle.

i) A day is defined to be a wet day if it receives rainfall more than or equal to 2.5.mm .

ii) A wet spell is included in a month if any day of this particular spell falls within that month,no matter if the wet spell does not end in the month.

iii) A dry spell is included in a month only if the immediately following wet spell is included in that particular month. In case,a month ends with a dry day the dry spell including that dry day should be accounted in the following month.

iv) In case of weather cycles,the whole length of the cycle which overlaps two adjacent months is assigned to that month which shares more than half its length.In case of equality,it is arbitrarily assigned to the previous month.

In adopting above rules for the assignment of spells and cycles to all the months,we are likely to introduce little or no bias in the long run as a result of the random characteristics associated with the mid-points of the spells and cycle lengths.

Following the above procedure the observed frequencies of dry and wet spells of various lengths during the rainy months are shown in Tables 5.4 and 5.5 respectively. It is seen from the Table 5.4 that the observed frequency for the duration of a dry spell is seen up to eleven days in the months of May and August but seven,eight and twelve days in the months of July,June and September respectively. The dry spells of twelve days during the month of September are observed twice while the dry spells of eleven days during the months of May and August occur thrice and twice respectively. However,it is mostly observed that the frequency of the duration of a dry spell decreases as the size of the dry spell increases for all the rainy months at Cooch Behar.

From Table 5.5 the observed frequency of wet spells is seen up to sixteen days in the month of July followed by thirteen days in the month of September. The same for twelve days in the months of May and June and that of nine days in the month of August are also observed during the study period. Though a wet spell of sixteen days during the month of July is observed once but in this month,the long wet spells are observed frequently.

Table 5.6 shows the observed frequency of dry-wet cycles for all the rainy months. It is seen from Table 5.6 that dry-wet cycles of longer durations occur in the month of July followed by the months of May,August and then by september and June respectively. This is probably because of longer wet spell during these rainy months.

From Table 5.7 the wet-dry cycle of longer duration occur once in 25 days during the month of September followed by July,June , August and May. This is due to the longer wet and dry spells during these months.

The probabilities of obtaining these three events, a wet spell of length w-day,a dry spell of length d-days and a weather cycle of length n-days can be constructed from the Markov dependent geometric models. The probability generating expression for wet spell length (x) , dry spell length (y) and the weather cycle length (z) are given below :

$$P\ (x=w) = (p_{22})^{w-1}(1 - p_{22}).$$

$$P(y=d) = (p_{11})^{d-1}\ (1-p_{11})$$

$$P(z=n) = (p_{12}\ X\ p_{21})\ X \left[\frac{p_{22}^{n-1} - p_{11}^{n-1}}{p_{22} - p_{11}} \right]$$

Here

$P(x=w)$ = Probability of a wet-spell of length w days.

$P(y=d)$ = Probability of a dry-spell of length d days

$P(z=n)$ = Probability of a weather cycle of length n days

p_{11}, p_{12}, p_{21} and p_{22} are defined in sub-section 5.2.

The conditional probabilities p_{11}, p_{12}, p_{21} and p_{22} which serve as the basic elements to construct the Markov dependent geometric model, are already estimated separately for each month May to September as well as the pooled estimates of these months. These values are given in Table **5.2.** These conditional probabilities are based on 18 years daily rainfall observations at Cooch Behar during these rainy months.

We can employ these transition probabilities into the foregoing formulae as expressed in this sub-section. The relative probabilities as generated by the Markov dependent model are estimated for each of the rainy months as well as the total of these months. These relative probabilities are multiplied by the appropriate sample sizes,we get the expected frequencies by lengths of wet and dry spells and wet-dry and dry-wet cycle for each of the rainy months as well as the total of these months.

The observed frequencies along with their expected frequencies are given in Table 5.4 for dry spell distribution and Table-5.5 for wet spell distribution.

Now we discuss here separately wet spell,dry spell and dry-wet cycle and wet-dry cycle for the respective months.

5.3.1. DRY SPELL :

Table 5.4 shows the distribution of dry spell for the months May, June, July, August and September and the pooled distribution of these rainy months. The Chi-square goodness of fit test has been applied to test the validity of the model. The values of the test statistic are calculated separately for each distribution of dry spell. These values of the Chi-square are given in Table 5.4 with suitable degrees of freedom along with the table values at five percent level of significance. It is seen that the observed and expected frequencies for dry spells of varying lengths for each distribution fit well since the calculated values of Chi-square are nonsignificant at five percent level. The observed and expected frequencies of dry spells during the months May to September i.e.total of rainy months are shown in Figure.5.1.

Therefore,we come to the conclusion that the dry spells with varying lengths can be considered tofollow Markov dependent geometric model. So,this model is a valid one to the dry spell distribution at Cooch Behar.

5.3.2. WET SPELL

The Chi-square goodness of fit test hasbeen considered to test the reliability in agreement between the observed and the expected frequencies of wet spells to all the distributions of wet spells with varying lengths. The calculated values forthe test statistic are given in Table 5.5 with suitable degrees of freedom at the 5 percent level of significance. The calculated values of Chi-square are less than the table values at 5 percent level with suitable degrees of freedom for all the distributions of wet spell with varying lengths. So,the values of test statistic are not significant at 5 percent level. Hence,there is good agreement between the observed and the expected frequencies for wet spells of varied lengths for each distribution of wet spells under consideration. The observed and the expected frequencies of wet spells during the premonsoon month (May) and the active monsoon months i.e. the total of these months are shown in Figure 5.2.

Therefore,the distribution of wet spells for each month and the rainy months can be considered to follow the Markov dependent geometric model at Cooch Behar.

5.3.3. WEATHER CYCLE :

The distribution of the length of weather cycles consists of two categories as mentioned earlier i.e.dry-wet and wet-dry cycles. Table 5.6 and Table 5.7 report the observed and the expected frequencies for dry-wet and wet-dry cycles. However,it may be noted that due to symmetry of the appropriate formulae by using the p_{11} and p_{22} the probabilities for different lengths of the cycles remain the same for dry-wet and wet-dry cycles, although the actual expected frequencies can be different due to the marginal differences of the sample size. Table 5.6 and Table 5.7 show the observed and the expected frequencies of thesetwo types of weather cycles separately for the consolidated periods.

The calculated values of Chi-square are less than the tabulated values of Chi-square at five percent level of significance. These values are given in Table 5.6 for dry-wet cycles and in Table 5.7 for wet-dry cycles for all the distributions. So,the values of the test statistic are nonsignificant. Figures 5.3 and 5.4. show the observed and the expected frequencies of dry-wet and wet-dry cycles of the period May to September respectively.

As judged solely from the test values,the dry-wet and wet-dry cycles appeared to confirm to follow the Markov dependent geometric model. The findings are similar in the case of dry spells and wet spells.

As observed elsewhere in the case of dry spells, wet spells and weather cycles,the expected frequencies at the long length are adjusted corresponding to the observed ones.

However,for large mass of data of the present type,accumulated over a long period of eighteen years, the influence of many extraneous factors affecting the reliability and uniformity of data cannot be ruled out. Putting all these type of factors together and bearing in mind,some of the inadequacies in the data base itself, the fundamental Markovian assumption and Markov dependent geometric distribution model for weather spells and cycles are not too unreasonable.

Therefore,it may be concluded that the average length of expected dry spell is 2.5 days while that of a wet spell is 3 days. And they constitute a weather cycle of 5.5 days which is as such,the average value of the observed length of a weather cycle. Hence,the spells as well as weather cycles have conformed to the Markov dependent geometric model.

5.4. LENGTH OF SPELLS AND WEATHER CYCLES :

In the foregoing analysis,we have tested the fundamental Markovian assumption involving the daily weather at Cooch Behar for different months. In this section we use the parameters of two State Markov chain model on the daily rainfall for the rainy months and pooled period at Cooch Behar to estimate the length of wet and dry spells and also the length of weather cycles. It is assumed that the geometric model also enables us to examine the number of dry or wet spells and weather cycles. But we have already established that Markov-dependent geometric models are best fitted to the distribution of wet and dry spells and also weather cycles.

Moreover,considering the distribution the expected length of a dry spell of length 'd' days is given by

$$E(d) = \frac{1}{P_{12}}$$

The expected length of a wet spell of 'w' successive wet days followed by a dry day, then is given by

$$E(w) = \frac{1}{(1-p_{22})}$$

The expected length of a weather cycle i.e. a dry spell followed by a wet spell or vice-versa is then given by

$$E(c) = E(d) + E(w)$$

$$= \frac{1}{p_{12}} + \frac{1}{(1-p_{22})}$$

p_{12} and p_{22} are the usual notations as given in earlier sub-section.

Employing these transition probabilities, the expected length of dry and wet spells and also the expected length of weathercycles can be evaluated as shown above. The observed values of lengths of dry and wet spells and the lengths of weather cycles are computed from the observed frequency distribution of the spells and weather cycles for each month from May to September at Cooch Behar,during the study period of eighteen years.

The observed and the expected lengths of spells of dry and wet spells and the lengths of weather cycles are given in Table **5.8.** The observed values of weather cycles are estimated by taking the average of the means of two cycles i.e. wet-dry and dry-wet cycles. These values are estimated for all the rainy months i.e. May to September at Cooch Behar.

It is seen from that table that the variation in the lengths of wet spells is large for all months. It is also observed that the expected length of dry spells varied from 1.96 to 2.91 while in the case of wet spells it varies from 2.54 to 4.02. A note worthy feature is that the expected lengths of dry spells are greater than the observed lengths in all the months expect in

August where it is reversed. And the expected lengths of wet spells are greater than the observed lengths in all the months except in September where it is also reversed.

The observed lengths of weather cycles fluctuated from 5.2 days in the months of June and August to 6.2 days in the month July ,whereas the expected lengths varied from 5.34 in the month of August to 6.35 days in the month of July at Cooch Behar.

5.5 CONCLUSION :

Markov chain models have been fitted to the daily rain fall during the months,May to September separately in this district.The distribution of dry and wet spells as well as dry-wet and wet-dry cycles of the period May to September and rainy season as a whole have been fitted to Markov dependent geometric model over this area. The average expected lengths of dry spells and wet spells are 2.5 and 3 days respectively and they constitute the 5.5 days of the weather cycle which is nearly the same as that of the average observed weather cycle.

T A B L E– 5.1.

TRANSITION FREQUENCIES

(For five months and total)

Month	f(D/D)	f(W/D)	Total(n_1)	f(D/W)	f(W/W)	Total(n_2)
MAY	203	106	309	106	143	249
JUNE	126	96	222	94	224	318
JULY	80	83	163	90	305	395
AUGUST	172	105	277	104	177	281
SEPTEMBER	153	97	250	94	196	290
TOTAL	734	487	1221	488	1045	1933

GRAND TOTAL = (1221 + 1933) 2754

TABLE - 5.2.

TRANSITION PROBABILITIES

For six stochastic matrices.

	P_{11}	P_{12}	P_{21}	P_{22}
MAY	.656	.344	.426	.574
JUNE	.576	.424	.296	.704
JULY	.491	.509	.228	.772
AUGUST	.621	.379	.37	.63
SEPTEMBER	.612	.388	.324	.676
POOLED	0.6012	.3988	.3183	.6817

TABLE = 5.3.

VALUES OF TEST STATISTICS

MONTH	CALCULATED TEST STATISTIC	D.F	TABLE VALUE OF
MAY	13.02*	1	X^2 with 1 df.at
JUNE	17.45*	1	1% level =6.64
JULY	15.72*	1	
AUGUST	15.41*	1	
SEPTEMBER	19.71*	1	
POOLED	96.60*	1	

* Significant at 1% level.

T A B L E = 5.4.

Observed (O) and expected (E) frequencies of dry spells :

*	MAY O	MAY E	JUNE O	JUNE E	JULY O	JULY E	AUGUST O	AUGUST E	SEPTEMBER O	SEPTEMBER E	TOTAL O	TOTAL E
1	48	37.80	48	43.40	46	43.90	41	40.30	45	37.40	228	209.6
2	20	24.80	21	24.80	24	21.50	25	25.00	22	22.80	112	109.7
3	10	16.30	11	14.10	09	10.50	12	15.50	09	13.90	051	71.8
4	12	10.70	13	8.00	02	05.00	10	09.60	03	8.50	040	43.1
5	08	7.00	02	4.60	01	2.80	09	6.70	07	5.20	027	25.9
6	03	4.60	02	2.60	02	1.30	02	3.70	03	3.20	012	15.50
7	01	3.00	03	2.00	02	1.00	01	2.30	04	2.50	011	9.30
8	01	2.50	01	1.50			01	1.40	01	1.50	004	5.60
9	04	2.00					02	.90			006	3.40
10	00	1.00					01	.70			001	2.50
11	03	.50					02	.60			005	1.60
12									02	1.00	002	1.00
Total :	110	110	101	101	86	86	106	106	96	96	499	499
X^2 (Cal)	6.60		5.65		2.41		2.36		7.87		12.18	
D.F.	5		4		4		5		5		8	
X^2 (Tab) at 5%	11.07		9.49		9.49		11.07		11.07		15.507	

* = Dry Spell (Day)

TABLE = 5.5.

Observed and expected frequencies of WET SPELL

	MAY		JUNE		JULY		AUGUST		SEPTEMBER		TOTAL	
*	O	E	O	E	O	E	O	E	O	E	O	E
1	48	45.95	36	31.90	28	22.30	28	22.60	32	30.4	172	150.10
2	26	26.40	20	22.40	16	17.20	13	14.20	20	20.7	95	102.00
3	14	15.15	18	15.80	13	13.25	08	18.95	21	14.05	74	69.40
4	10	8.70	12	11.20	07	10.20	03	5.65	07	9.60	39	47.20
5	5	5.00	5	7.80	5	7.85	1	3.60	07	6.5	23	32.10
6	1	2.90	5	5.50	6	6.05	2	2.25	02	4.4	16	21.80
7	2	1.65	2	3.80	7	4.85	1	1.50	01	3.5	13	14.85
8			3	2.70	5	3.80	4	1.25	03	2.85	15	10.10
9			3	2.50	1	2.80	1	1.00	1	2.00	6	6.90
10	1	1.25	2	2.00	3	2.50					6	4.70
11			1	1.40	2	2.10					3	3.80
12	1	1.00	1	1.00	3	2.00					5	3.00
13									1	1.00	1	2.05
14					1	2.00					1	1.00
Total	108	108	108	108	97	97	61	61	95	95	469	469
X^2 (cal)	0.79		2.18		3.99		2.80		6.70		10.97	
D.F.	5		6		6		4		5		9	
X^2 (at 5% leel)	11.07		12.59		12.59		9.49		11.07		16.92	

* = Wet Spell (Day)

T A B L E = 5.6.

Observed and expected frequencies of Dry-Wet cycle :

*	MAY O	MAY E	JUNE O	JUNE E	JULY O	JULY E	AUGUST O	AUGUST E	SEPTEMBER O	SEPTEMBER E	TOTAL O	TOTAL E
2	22	13.87	17	11.65	13	10.21	14	14.30	13	14.30	79	59.55
3	17	17.08	17	14.89	17	12.86	20	17.98	14	14.50	85	76.25
4	7	15.77	8	14.33	13	12.35	16	16.85	16	14.10	65	73.45
5	9	12.92	11	12.30	08	10.71	19	14.08	15	12.17	62	62.70
6	14	10.45	09	9.92	05	8.87	08	10.95	08	9.75	44	50.20
7	08	7.6	05	7.71	05	7.10	06	8.25	05	7.55	29	37.02
8	04	4.75	07	5.84	04	5.60	02	6.00	08	5.75	25	29.20
9	03	3.8	03	4.34	06	4.40	02	4.30	03	4.25	17	21.30
10	02	2.85	07	3.20	03	3.40	04	3.01	01	3.15	07	15.80
11	02	1.90	03	2.32	02	2.70	03	2.15	02	2.25	02	11.20
12	02	1.24	00	1.67	02	2.10	01	4.45	01	1.60	06	7.90
13	00	.75	01	1.20	03	1.80	05	1.20		1.80	09	5.60
14	01	.51	01	.87	03	1.60	00	.80	02	1.50	08	4.20
15	02	.50	02	.76	1	1.50			01	1.00	06	3.50
16	00	.50							01	1.00	01	2.80
17	02	.51					01	.80		0.60	03	2.50
18											00	
19					02	1.8					02	1.85
Total	95	95	91	91	87	87	102	102	90	90	465	465
X^2(cal)	12.08		7.44		5.96		8.31		4.32		20.40	
Df	6		7		7		7		7		12	
X^2 at 5% level	12.59		14.07		14.07		14.07		14.07		21.03	

* = Dry-Wet Cycles(Days)

T A B L E = 5.7.

Observed and expected frequencies of Wet-Dry cycle.

	MAY		JUNE		JULY		AUGUST		SEPTEMBER		TOTAL	
*	O	E	O	E	O	E	O	E	O	E	O	E
2	15	13.87	12	11.52	12	10.00	21	14.10	17	11.15	77	58.70
3	18	17.08	20	14.72	17	12.60	21	17.60	11	14.35	82	75.20
4	17	15.77	16	14.20	14	12.10	13	16.50	14	13.00	74	72.50
5	14	12.92	09	12.10	8	10.00	12	13.80	13	11.95	56	62.70
6	11	10.45	5	9.80	9	8.60	7	10.75	7	9.65	39	49.50
7	2	7.6	7	7.30	6	6.90	9	8.10	5	7.50	29	36.70
8	4	4.75	7	5.80	6	5.50	4	5.80	8	5.70	29	28.90
9	3	3.80	3	4.80	4	4.30	1	4.20	5	5.30	16	21.10
10	2	2.85	6	4.00	1	3.40	2	2.95	0	0	11	15.60
11	0		3	3.00	3	2.80	3	2.10	5	4.20	14	11.10
12	4	2.16			2	2.50	1	1.40	0		7	7.80
13	2	1.50	1	1.86	2	2.00	3	1.00	2	3.20	10	5.50
14	1	1.00			4	1.80	1	0.70	0		6	4.10
15									1	2.00	1	2.60
16	1	0.75					1	0.50			2	2.00
17			1	1.00	1	1.50					2	1.80
18					1	1.00					1	1.50
23	1	0.50					1	0.50			2	1.20
25									1	1.00	1	.50
Total	95	95	90	90	85	85	100	100	89	89	459	459
X^2 (Cal)	4.50		5.40		2.90		6.98		6.72		18.14	
D.F.	6		7		7		7		8		12	
X^2 (at 5%)	12.59		14.07		14.07		14.07		15.507		21.03	

* = Wet-Dry Cycles(Days)

T A B L E = 5.8

OBSERVED AND EXPECTED LENGTHS OF DRY SPELL, WET SPELL AND WEATHER CYCLES

(Observed and expected length (in days)

	DRY SPELL		WET SPELL		WEATHER CYCLE	
	Oi	Ei	Oi	Ei	Oi	Ei
MAY	2.83	2.91	2.31	3.35	6.06	6.26
JUNE	2.23	2.36	3.16	3.38	5.2	5.74
JULY	1.86	1.96	4.02	4.39	6.2	6.35
AUGUST	2.72	2.64	2.54	2.70	5.20	5.34
SEPTEMBER	2.51	2.58	2.77	2.09	5.35	5.67
AVERAGE	2.43	2.49	2.96	3.18	5.6	5.87

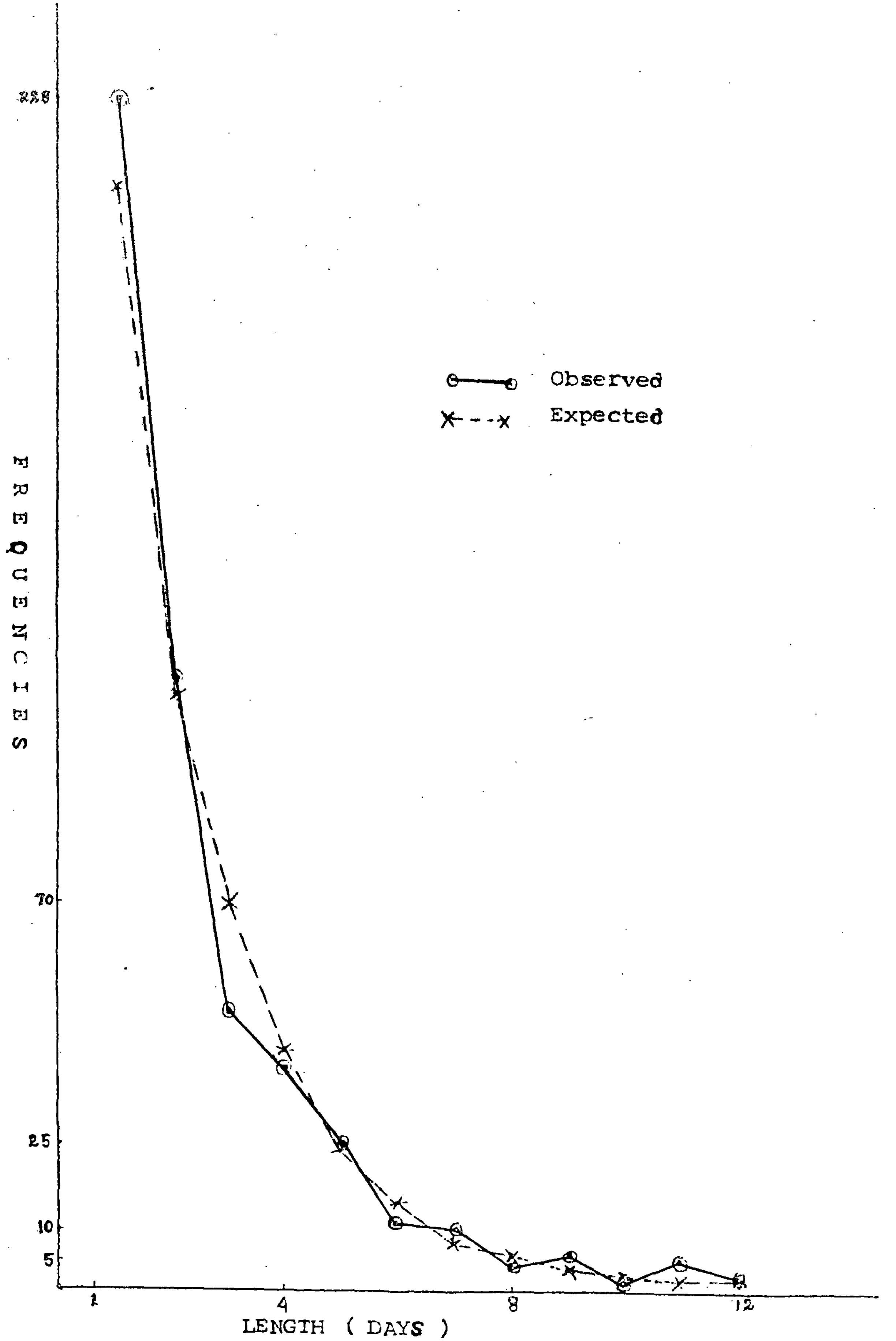


Fig.5.1. Observed and expected frequencies of dry spell (May to September)

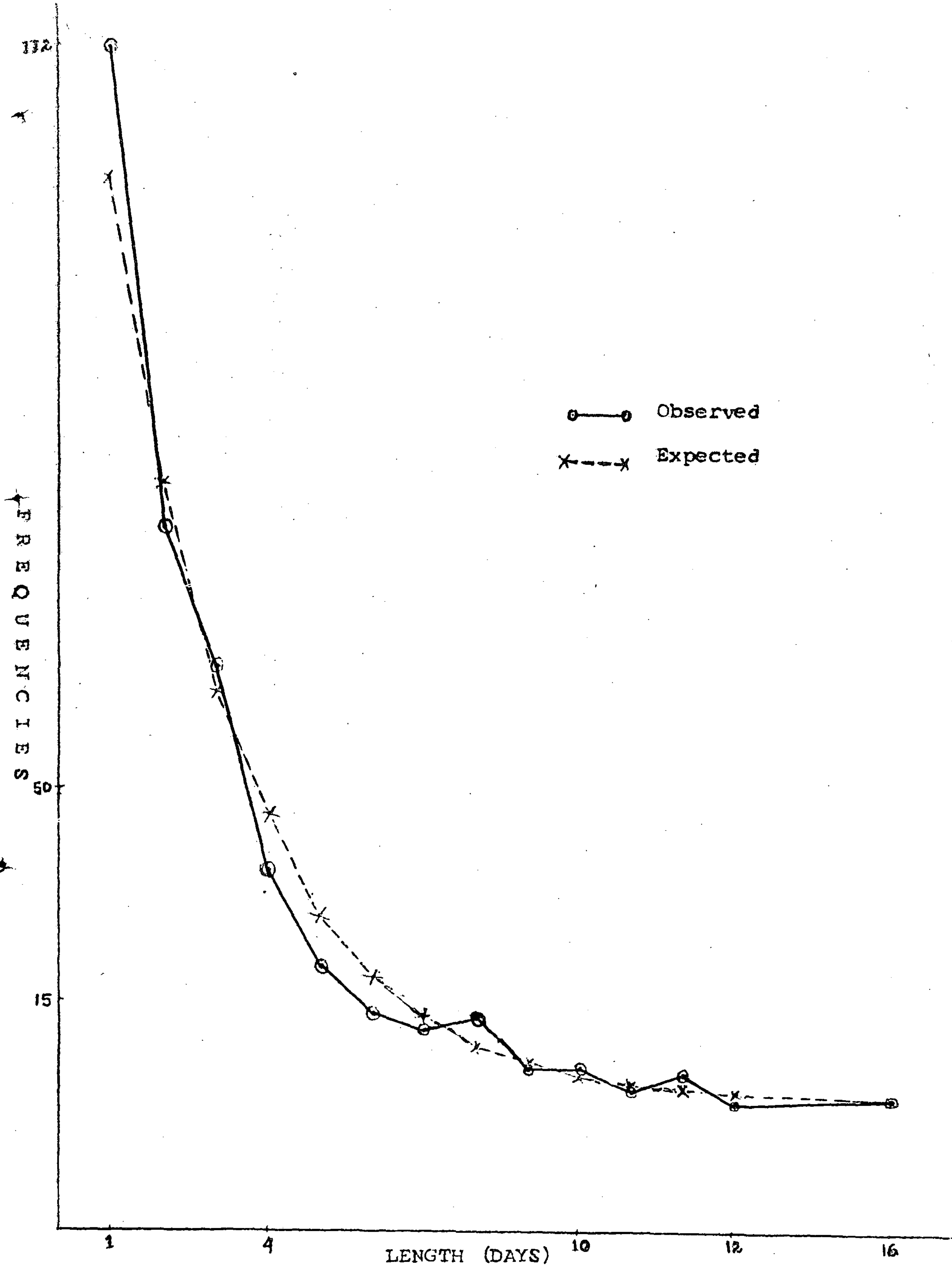


Fig.5.2. Observed and expected frequencies wet-spell (May to September)

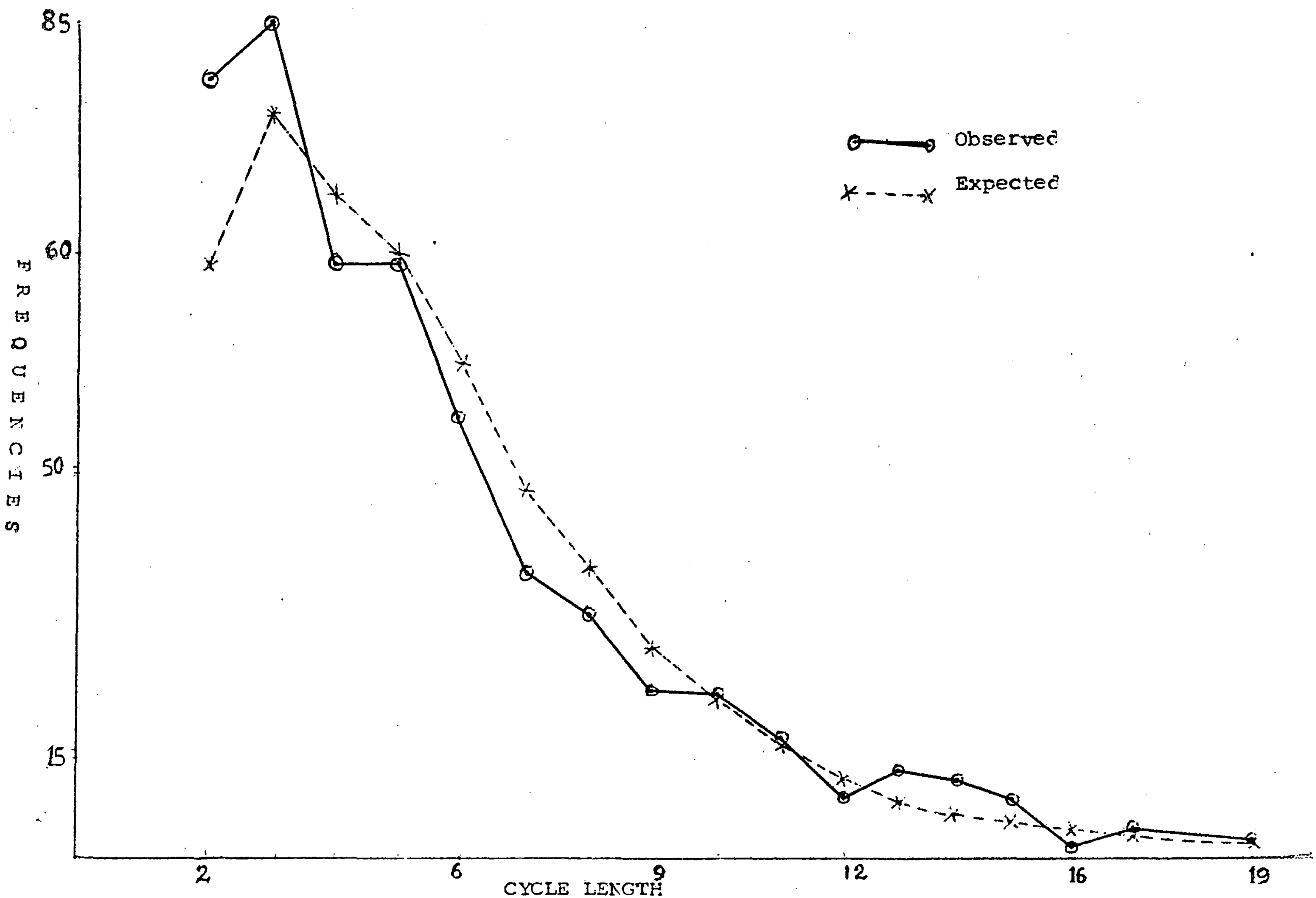


Fig.5.3. Observed and expected frequencies dry-wet cycles during May to September.

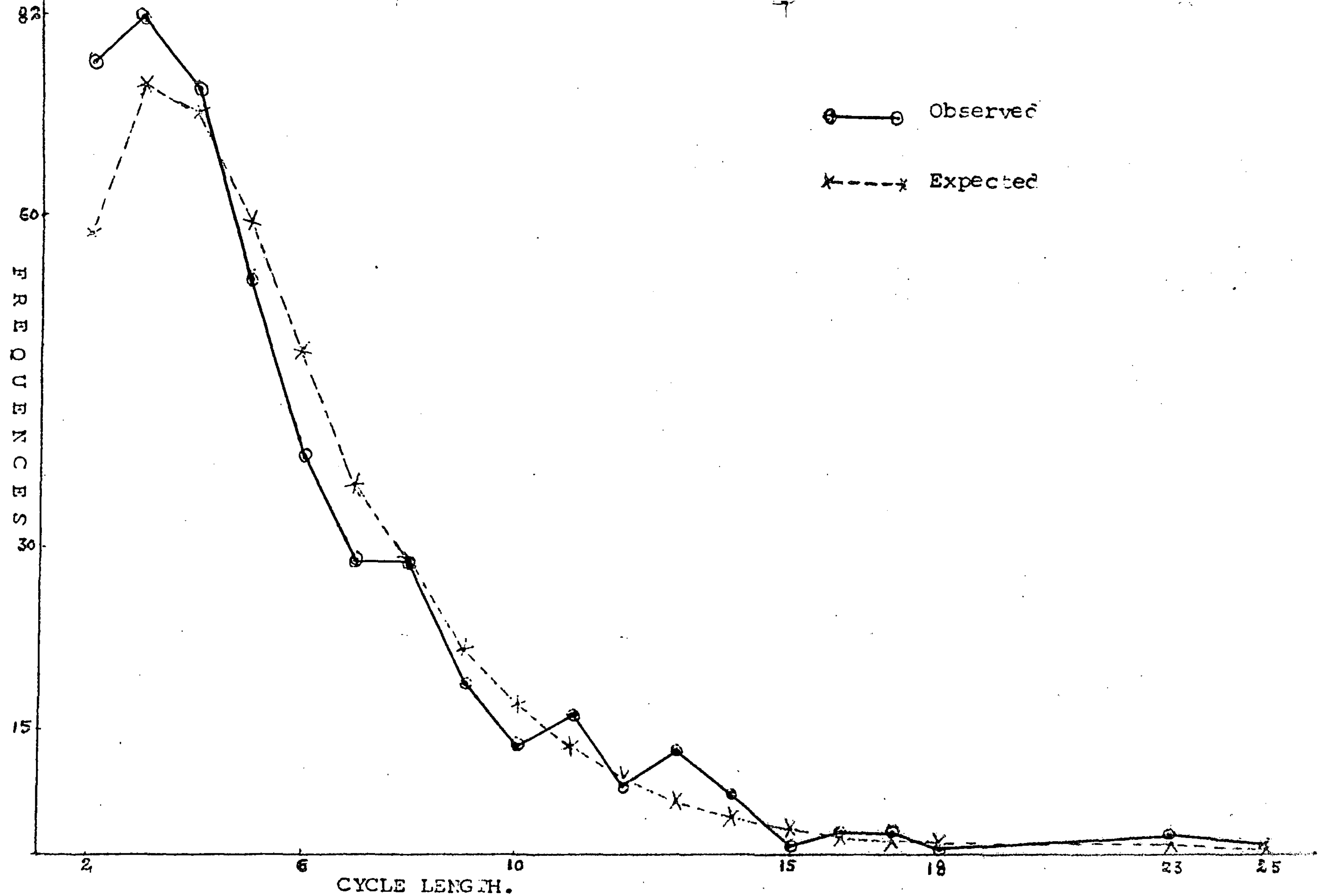


Fig.5.4. Observed and expected frequencies of wet-dry cycles during May to September.

CHAPTER - VI

MEASURE OF UNCERTAINTY ON DAILY RAINFALL

CHAPTER - VI

MEASURE OF UNCERTAINTY ON DAILY RAINFALL

6.1. INTRODUCTION :

Many scientists and meteorologists have studied the probabilities of the occurrence of dry and wet days only by fitting a Markov chain model and geometric and Markovian-geometric spell distribution. But the studies of uncertainty about the nature of a day's rainfall are rarely available. Every probability distribution has some uncertainty associated with it. The concept of entropy is introduced here to provide a quantitative measure of this uncertainty and to test the stochastic dependence.

The main objectives of this chapter are :

i) To examine the quantitative values of degree of uncertainty.

ii) To study the nature of rainfall event by Markov chain method.

6.2. STOCHASTIC MATRICES FOR EACH MONTH

A one step 5 × 5 transition probability matrix from one transitional state to another has been formed for each of the rainy months,May to September from the frequency of the occurrence of daily rainfall amount at Cooch Behar.

The nature of the daily rainfall has been classified according to the intensity of daily rainfall. The categories are given below :

Category No.	Rainy days	Intensity of daily rainfall (in m.m.)
0	Non-rainy day	0 - 2.4
1	Light rainy day	2.5 - 10.0
2	Moderate rainy day	10.1 - 50.00
3	Heavy rainy day	50.1 - 124.9
4	Very heavy rainy day	125 and above

The mode of classification yields a sequence of non-rainy,light rainy,moderate rainy,heavy rainy and very heavy rainy days which can be regarded as a five-state Markov chain. Thus,for each year,transition to 1st May is classified as one of the twenty five possibilities depending on the weather of 30th April which is a non-rainy day or other type of the nature of rainy days. Repeating the process each year, the cell frequencies for the above twenty five possibilities are obtained. Here we use the daily rainfall data from 1971 to 1988. In this study the daily rainfalls are utilized from 30th April to 30th September for each year. So,each year consists of 153 observations and the total number of observations is 2754.

Let the cell frequencies of these twentyfive possibilities be denoted by n_{ij} where

$i \ \& \ j \in S \ ; \ S = (0,1,2,3,4)$ and $i \rightarrow j$

where 'S' is the state-space.

These frequencies are arranged in the form of a matrix for each monsoon month i.e., June, July, August, September and one pre-monsoon month i.e. May. The cell frequencies are arranged in matrix form for the five months which are given in Table 6.1. The conditional probabilities which can be estimated by maximum likelihood method are defined as

$$P_{ij} = P\left\{ X_k = j \mid X_{k-1} = i \right\}$$

$$i \;\&\; j \in S$$

$$S = (0,1,2,3,4)$$

The conditional probabilities p_{ij} are calculated by dividing the cell frequencies of one state by the total frequencies of that state. This yields the estimation of each cell probability and the cell probabilities are given by

$$P_{ij} = \frac{n_{ij}}{n_i}$$

where $$n_i = \sum_{j=0}^{4} n_{ij}$$

subject to the restrictions

$$\sum_j P_{ij} = 1 , \quad \text{for each } i$$

$$\text{and} \qquad P_{ij} \geq 0 \quad \text{for all } i \;\&\; j.$$

Each row constitutes the probability vector and it is convenient to give the set of all such vectors as a matrix. This matrix is known as the transition probability matrix or the matrix of transition probabilities of the Markov chain or simply the Markov matrix. This is a square matrix (5x5) with non-negative elements and unit row sum.

The one-step transition probability matrix for each monsoon month and one pre-monsoon month are given in Table 6.2(a).

The estimate of the probability of individual state of occurrence (π_i) is given by

$$\pi_i = \frac{n_i}{n} \qquad \text{where} \quad n = \Sigma\, n_i,\ \Sigma\pi_i = 1,\ i \in S.$$

The values of n_i and n are given in Table **6.2(b)** and the values of π_i are given in Table **6.2(b)**.

6.3. REDUNDANCY TEST

The concept of entropy given by Shannon, in his mathematical model serves as a measure of uncertainty of the transition probabilities (P_{ij}) and this is defined by

$$H_i = - \sum_{j=0}^{4} p_{ij} \log_{10} P_{ij}$$

for each i.

Where H_i denotes the entropy of the ith state of occurrence.

The values of H_i for different states and for different months are given in Table **6.3.**

The weighted entropy value for each of the months is calculated as a sum of the entropy values of all the different categories with the probabilities of the corresponding states.

The weighted entropy H can be calculated as

$$H = - \Sigma\, \pi_i \quad H_i$$

Where π_i is the probability of individual state of occurrence i.e. in the ith state.

And H_i is the entropy of the i th state. The values of the weighted entropy are given in Table **6.3.**

The weighted values over the transition probabilities during the monsoon months and the pre-monsoon month have showed a typical nature. The weighted entropy value for the month of July is the highest and it is followed by June. It is observed that

June and July are considered to be the most active monsoon months in Cooch Behar district. The most interesting matter is that the weighted entropy for the month of May has a significant role for May to be considered as a monsoon month. The weighted entropy of this month is likely to be equivalent to that of the month of August. Now we may consider May as the beginning month of monsoon at Cooch Behar.

The entropy of the stationary distribution i.e. the entropy of the individual state of occurrence is given by

$$H_{ii} = -\Sigma \; \pi_i \log \pi_i$$

Where π_i is the probability of the ith state of occurrence.

The values of the entropy of the individual state of occurrence have been given in Table **6.3**. The entropy has an important use in measuring the uncertainty as well as in testing the hypothesis of Markov dependence.

A measure of uncertainty,M,of the stationary model is obtained from the individual states of occurrence over the Markovian model in the system and is given by

$$M = H_{ii} - H = \left(-\Sigma \; \pi_i \log \pi_i \right) - \Sigma \; \pi_i \; H_i$$

The values of M for all the months are given in Table **6.3.**

The measure of uncertainty of stationary model and that of the Markovian model have some difference which is very small,almost negligible. The difference in uncertainty between the Markovian model and the stationary model at Cooch Behar lies between + 8% to −12 percent.

The redundancy of the state of occurrence, R, is obtained as the difference from one of the ratio of the weighted entropy value H to the maximum possible entropy (H_{max})

Here, H_{max} = Log5

as we have here only 5 states of occurrence as Thail (1973).

So,

$$R = 1 - \frac{H}{H_{max}}$$

This redundancy value is used to determine the favourableness or unfavourableness of the Markovian system. As the redundancy value, R, tends to 1, the Markovian system tends to maximum favourable condition i.e. almost certain. Now on the light of this argument, we may examine the Markovian dependency on the monsoon months or rainy months. From Table 6.3 it is observed that the redundancy value of the month of May is the maximum but it is very low in comparison to the value one. The redundancy values of the monsoon months lie between 0.14 and 0.29. Considering these values, we may conclude that the one-day dependence cannot be considered as explaining the rainfall pattern at Cooch Behar.

In the next sub-section, we would verify this conclusion by adopting another method to test the Markovian dependence through informatrix of the rainy months.

6.4. LIKELIHOOD-RATIO TEST BY ENTROPY

In this section we use the informational measure to test the hypothesis of Markovian dependence.

The mathematical model of Shannon is used to obtain the measure of entropy in individual state of occurrence which we have calculated in the previous section, denoted by H_{11} and also the transitional probabilities of informatrices. We have H_{11} as given in Table **6.3.**

The average conditional uncertainty can be measured by

$$H_{21} = - \Sigma \pi_i p_{ij} \log p_{ij}$$

$$= \frac{1}{n} \left[\Sigma n_i \log n_i - \Sigma n_{ij} \log n_{ij} \right]$$

This is the same as the weighted entropy which is stated in the previous section. These values are also given in Table **6.4** for each of the months.

The hypothesis testing, involving Markov chains has been considered by several ways. Mainly the Chi-square and the likelihood ratio-criterion have been used for testing the hypothesis of independence of the random variable. Here we introduce a test criterion which involves the entropy but this is equivalent test of the likelihood-ratio criterion. The test statistic is $T_1 = 2n\ (H_1 - H_{21}) = 2\ \Sigma\ n_{ij} \log \frac{n_{ij}}{n_i n_j / n}$

$$= 2\ \Sigma\ n_{ij} \log \frac{n \cdot n_{ij}}{n_i n_j}$$

That is T_1 is the same as the likelihood-ratio criterion. The test statistic has a limiting Chi-square distribution with $(m - 1)^2$ degrees of freedom (here m = 5) and the large values of the statistic correspond to rejection of the hypothesis.

Before applying this test, we now set up the null hypothesis: the occurrence of one day rainfall is independent against Markov dependence. The test statistic has been applied to each of the months separately. The test procedure followed here has been described in Basawara & Rao (1980). Here we have to indicate that the total numbers of observations (n) are different for different months. The total number of observations for each ofthe months of May, July and August is (31×18) = 558 , ignoring the initial observation,the last day of April. And the total number of observations for each of the months of June and September is (30 × 18) =540 each. These total numbers of observations of each month are given in Table **6.4.**

The tabulated value of Chi-square with $(m-1)^2$ = 16 degrees of freedom,at 5% level of significance is 26.30 and that at 1% level is 32.00. It is also shown in Table **6.4.** The calculated value of T_1 for each month is also shown in the same table. The computed value of T_1 for the month of May is 19.64 which is less than the theoretical value of Chi-square with 16 degrees of freedom at 5% level of significance. This is non-significant. Therefore,we cannot reject the null hypothesis . Hence,null hypothesis is accepted. Then we may come to the conclusion that the occurrences of daily rainfall in the month of May are really independent. In the month of May, the day's precipitation does not depend on the precipitation of the previous day. Hence,we may say that the daily weather occurs randomly in the month of May at Cooch Behar.

Now we consider the case of active monsoon months which indicates an interesting result. The computed values of thetest statistic for these four months are greater than the tabulated value of Chi-square with 16 degrees of freedom at 5 percent level of significance. So,these are all significant. Thus,we can reject the null hypothesis for each of the four months. The active monsoon months except September are also significant at one percent level of significance.

Above test suggests that the weather of a day is influenced by the immediately preceding day's weather only during the monsoon months at Cooch Behar.

The analysis of the behaviour of the daily weather has established that our result is in good agreement with that of Medhi (1976). But he considered only the two state Markov chain model of the daily rainfall in Guwahati, Assam.

6.5 CONCLUSION.

The daily weather pattern of Cooch Behar during the monsoon months has been established to follow the Markov chain model.The weighted entropy for the month of July is the highest during the monsoon months. Among the probabilities of individual states of occurrence the probabilities of non-rainy days are highest, followed by moderate rainy days during the monsoon season except the month of July where this feature is reversed.

The validity of the redundancy test is verified by using the test equivalent to the likelihood-ratio test. So,it is observed that the redundancy test is not so powerful test of Markov dependence. But the test equivalent to the likelihood-ratio test has shown better result against the Markov dependence on the same observations. But considering the empirical result of Medhi (1976), it would have been suggested that the latter test is more powerful test as well as the most appropriate test against the Markov dependence of daily rainfall.

TABLE - 6.1.

Transitional cell frequencies in matrix form for rainy months at Coochbehar

May

0	1	2	3	4
203	46	51	8	1
36	11	31	3	0
56	23	55	10	0
14	3	7	0	0
0	0	0	0	0

June

0	1	2	3	4
128	39	43	9	3
26	9	38	4	1
55	19	61	22	11
10	8	20	12	4
3	3	9	3	0

July

0	1	2	3	4
80	33	36	9	5
37	45	30	10	3
33	30	68	33	6
14	10	25	15	9
6	2	8	9	2

August

0	1	2	3	4
172	44	46	13	2
46	21	15	6	5
40	20	23	19	2
15	7	20	14	7
3	1	6	8	3

September

0	1	2	3	4
153	39	44	13	1
37	28	22	11	3
46	28	42	12	2
10	7	12	9	6
1	2	7	2	2

'0' - Non-rainy day, '1'-Light rainy day,
'2' - Moderate rainy day,
'3'- Heavy rainy day,
'4' - Very heavy rainy day.

TABLE : 6.2(a).

One-step(5×5) stochastic matrix for rainy months at Cooch Behar.

May

P_{ij} =	0	1	2	3	4
0	.66	.15	.16	.03	00
1	.44	.14	.38	.04	00
2	.39	.16	.38	.07	00
3	.58	.13	.29	00	00
4	00	00	00	00	00

June

P_{ij} =	0	1	2	3	4
0	.58	.18	.19	.04	.01
1	.33	.12	.49	.05	.01
2	.33	.11	.36	.13	.07
3	.19	.15	.37	.22	.07
4	.17	.17	.50	.16	.00

July

P_{ij} =	0	1	2	3	4
0	.49	.20	.20	.06	.03
1	.30	.36	.24	.08	.02
2	.19	.18	.40	.19	.04
3	.19	.14	.34	.21	.12
4	.23	.07	.30	.33	.07

August

P_{ij} =	0	1	2	3	4
0	.62	.16	.17	.04	.01
1	.50	.23	.16	.06	.05
2	.39	.19	.22	.18	.02
3	.24	.11	.32	.22	.11
4	.14	.05	.29	.38	.14

September

P_{ij} =	0	1	2	3	4
0	.61	.16	.17	.05	.01
1	.36	.28	.22	.11	.03
2	.35	.22	.32	.09	.02
3	.23	.16	.27	.20	.14
4	.07	.20	.47	.13	.13

(ij = 0,1,2,3,4) where,

'0' – Non-rainy day, '1'–Light rainy day,

'2' – Moderate rainy day, '3'– Heavy rainy day,

'4' – Very heavy rainy day.

T A B L E - 6.2(b)

Stationary vectors and probabilities of rainy months at Cooch Behar.

STATIONARY VECTOR

	0	1	2	3	4	Total
May	309	81	144	24	0	558
June	222	78	168	54	18	540
July	163	125	170	73	27	558
August	277	93	104	63	21	558
September	250	101	130	44	15	540
				GRAND TOTAL	=	2754

STATIONARY PROBABILITY

	0	1	2	3	4
May	.55	.15	.26	.04	0
June	.41	.15	.31	.10	.03
July	.30	.22	.30	.13	.05
August	.50	.17	.19	.11	.03
September	.46	.19	.24	.08	.03

- '0 ' – Non-rainy day, '1' – Light rainy day,
  '2 ' – Moderate rainy day,
  '3 ' – Heavy rainy day,
  '4 ' –Very heavy rainy day.

TABLE : 6.3.

MEASURE OF ENTROPY AND REDUNDANCY TEST

(For the experimental months)

	May	June	July	August	September
Non-Rainy Day(H_o)	.415	.484	.555	.341	.474
Light Rain (H_1)	.492	.506	.587	.563	.610
Moderate Rain(H_2)	.527	.620	.623	.609	.591
Heavy Rain (H_3)	.422	.646	.669	.663	.697
Very Heavy Rain(H_4)	.0	.540	.624	.620	.605
Entropy (H_i)	.47	.58	.48	.57	.57
Weighted Entropy	.45	.54	.60	.49	.55
Redundancy	.348	.217	.140	.288	.215
M	.02	.04	- .12	.08	.02

TABLE 6.4.

Entropy of individual state, weighted entropy and the value of test statistic for the experimental months.

	May	June	July	August	September
H_{1i}	.4744	.5847	.6398	.5767	.574
H_{2i}	.4568	.5477	.601	.5344	.549
n	558	540	558	588	540
X^2 (cal)	19.64	41.04**	43.30**	47.21**	27.0*
X^2 5%	26.30				
DF	16				

* Significant at 5% level.

** Significant at 1% level

CHAPTER - VII

PERIODICITY IN YEARLY RAINFALL

CHAPTER - VII

PERIODICITY IN YEARLY RAINFALL

7.1. INTRODUCTION

The rhythmic behaviour of precipitation in India and elsewhere has more importance now a days. While this information is of interest, it gives an insight into the temporal variation of the rainfall process. A study of the time scale variations would start with time-domain and frequency domain analyses of data in time series. These are also the initial steps in bulding stochastic models of time series. The nature of relationship between these rhythmic behaviours in a time series of rainfall data becomes more clear through this type of analysis. The first step of the prediction of rainfall with time series analysis technique involves a search for periodicities in the historical records. Periodicities ,if present,represent potential predictive power for the nature of rainfall during a future time period. Major researches for periodicities of rainfall time series and their association with periodicities,luni-solar and solar-cycle phenomena are now very common in climatological investigation. Other than these periodicities several short term cycles which are known as quasi-biennial,quasi-triennial and quasi-five year etc. periodicity show significant results in mostof the raingauge stations or meteorological sub-divisions in India. Suitable statistical analysis of time series may prove to be useful in the sense that the fluctuations are unexplainable on the physical basis, past history may give some indication about the future.

In this study an attempt has been made to examine the actual nature of the historical precipitation time series and to detect the periodicity in yearly rainfall data at the two stations in Cooch Behar district.

7.2. Autocorrelation :

The most conventional method of computing the power spectrum through the autocovariance function is known as the lagged product method, referred to equally spaced time series of finite length. In this section, the time domain analysis has been applied to examine the randomness in the annual rainfall of the two stations in Cooch Behar district **viz.** Dinhata and Cooch Behar. In this situation, available records of this natural phenomenon are limited in length by 88 years. These are spaced at equal time interval as in year.

We have already discussed in chapter IV of this study that the annual rainfall at the two stations i.e. at Cooch Behar and at Dinhata can be considered to follow the normal distribution. Maximum likelihood estimators have been applied to estimate the autocorrelations for lags 0 to K (K<N) and are given by

$$r_m = \frac{C_m}{C_o}$$

Where $$C_m = \frac{1}{N} \sum_{t=1}^{N-m} (X_t - \overline{X})(X_{t+m} - \overline{X})$$

and $$\overline{X} = \frac{1}{N} \sum_{t=1}^{N} X_t$$

r_m = autocorrelation coefficient for lag m.

N = total number of yearly rainfalls

X_t = variate representing tth year's rainfall

$\overline{X}$ = mean of variate X

m = period of lag takes the values from 0,1,2......K.

Tables - 7.1 and 7.2 show the autocorrelation coefficients for the two time series of rainfall data separately. We have computed here the autocorrelation coefficients up to lag 25 which are used to test the randomness of the annual rainfall of the two stations in Cooch Behar district.

The standard errors of sample autocorrelation function are estimated for each of the coefficients and are defined by

$$\text{Estimate of S.E}(r_m) = \frac{\left[1 + 2\sum_{i=0}^{m-1} r_i^2\right]^{1/2}}{\sqrt{N}}$$

The computed standard errors of each of the autocorrelation coefficients are also given in Tables- 7.1 & 7.2 for the annual rainfall at Cooch Behar station and Dinhata station respectively.

7.2.1. TEST OF SIGNIFICANCE OF AUTOCORRELATION.

Table 7.1 shows the standard errors of autocorrelation coefficients which are based on the order of N. On the basis of the asymptotic normal distribution, the approximate two-tailed critical values for the autocorrelation are $\pm\ 2\sqrt{1/N}$ which are known as Bartlett band at the 5 percent level of significance. Here we have N = 88,so the approximate Bartlett band is 2.13 at 5 percent level. The calculated values of the autocorrelation coefficients do not fall out side this band. The autocorrelation function has the same shape throughout. This is the shape as characterized by a random process. But in this process,the critical values are the same throughout. And if we consider the 2 SD band the critical values may be changed according to the order of lag of autocorrelation. However in this case also,we have the standard error of each order of lag of autocorrelation coefficient as stated in the table. So,we can easily get the 2 SD of each order of lag of autocorrelation coefficient and it is seen that the same results are obtained. It is to be noted that the two series appear to satisfy the stationarity condition.

7.3. PERIODOGRAM :

To investigate the existence of any periodic or any quasi-periodic fluctuations, the series is subject to periodogram analysis. The usual procedure adopted in periodogram analysis is to locate the frequencies $\{w_k\}$ by using search technique based on a function called the periodogram.

For N observational data $X_1, X_2, \ldots\ldots, X_N$ the periodogram I(w) is defined for all w in thw range $-\pi \leq w \leq \pi$ by

$$I(w) = \frac{2}{N}\left| \sum_{t=1}^{N} X_t e^{-i2\pi wt} \right|^2$$

or alternatively $I(w) = \frac{2}{N}| \{A(w)\}^2 + \{B(w)\}^2 |$

where $A(w) = \sum_{t=1}^{N} X_t \cos 2\pi wt,\qquad B(w) = \sum_{t=1}^{N} X_t \sin 2\pi wt$

Although $I(w)$ is defined for all w in $(-\pi, \pi)$ we cannot evaluate it numerically as a continuous function of w. We compute it only at a discrete set of frequencies .e.,e evaluate

$$I_k = I(w_k), \quad w_k = \frac{k}{N}, \quad k = 0,1,2,\ldots\ldots,\frac{N}{2}$$

and then plot I_k against $\frac{k}{N}$

[We should note here that some authors call it spectogram. Theycall the graph of I_k againet the periods $\frac{2\pi N}{k}$ the periodogram]

Now we consider the standardized form of the periodogram

$$I^{*}(W_k) = I(W_k)/S^2$$

Where S^2 is the variance of the time series and the standardized periodogram is also plotted against the angular frequencies, $\frac{k}{N}$. The true frequencies may be located by noting the positions of the peaks in the graph of the standardized periodogram ordinates.

Figure **7.1**(a) shows the standardized periodogram of the annual rainfall time series of Cooch Behar,plotting the standardized periodogram versus the angular frequencies. And Figure **7.1**(b) shows the same for Dinhata station. From this graphical representation ,there are observed five peaks on the spectrum of the annual raifall at Cooch Behar as well as at Dinhata station. Besides these peaks there are also some small peaks.

7.3.1. TEST OF SIGNIFICANCE OF SPECTRAL PEAKS.

The significance test of spectral power estimates may be determined by several methods.

For testing the significance of spectral peak, we consider the test criterion, suggested by Fuller (1976). Therefore for testing the hypothesis for detecting the periodicity at α percent level of significance and holding simultaneously for standardized periodogram values is given by

$$b_m = -\pi^{-1} \log_e \beta$$

Where $\beta = 1 - (1 - \alpha)^{1/m} \approx \frac{\alpha}{m}$

If we put $\alpha = 0.05$ $m = 44$, we get,

$$b_{44} = 2.158$$

If any standardized periodogram estimate is greater than the test value, i.e. 2.158, it indicates the presence of significant periodicity in the time series. And the period, P, corresponding to any spectral estimate is given by

$$P = \frac{1}{W_k} = \frac{N}{K}$$

Where W_k is the angular frequency corresponding to the significant spectral peak in question.

A list of significant spectral peaks with their corresponding periods, identified by spectral analysis for both the stations are given in Table **7.3** and Figures 7.1 (a & b) illustrate the standardized periodogram of annual rainfall at Cooch Behar and Dinhata respectively. The standardized periodogram values of the remarkable spectral peaks are 3.201, 2.626, 2.293, and

2.299 corresponding to the periods 2.667,3.52,4.889 and 22.0 years respectively. All the values of standardized periodogram are greater than the hypothetical value, 2.158. Hence these show the presence of peaks, significant at 5 percent level,at the station of Cooch Behar corresponding to the wave periods in the range 2.667 to 22 years. These are indications of the existence of prominent periodicities in the time series of rainfall at Cooch Behar station.

Now we come to consider the wave period in the time series of rainfall at Dinhata where five remarkable peaks are under considerationin this analysis.The spectral peaks,in consideration,with their corresponding periods for Dinhata station are given in the same table (Table **7.3**) . Here the values of standardized periodogram of remarkable spectral peaks are 2.301, 2.275, 5.631, 4.117 and 3.438 corresponding to the periods, 2.75, 3.25, 3.52, 14.67 and 29.33 years respectively. All these values of the standardized periodogram are greater than the hypothetical value (2.158). These are highly significant at the 5 percent level. The significant spectral peaks indicate the prominent periodicities in the yearly rainfall at Dinhata.

The extent and significance of 2.0 - 2.9 years frequency band have concentrated in this region. The significant peaks at the approximately quasi-biennial oscillation (2.2 years) are particularly characteristics of Cooch Behar district.

The significant quasi-triennial (3.0 - 3.9 years) periodicity is concentrated also. Stations with significant quasi-quadrennial (4.0 to 4.9 years) oscillation is also identified in the time series. The quasi-quadrennial oscillation is approximately quasi-five year oscillation as the 4.889 years has been considered as 5 years oscillation.

The spectral analysis has revealed some interesting evidence for the 14.67 and 29 years periodicities in the yearly rainfall at Dinhata while 22.0 years periodicity in the annual rainfall at Cooch Behar. But the existence of the 14.67 and 29 years periodicities at Cooch Behar and that of 22.0 years periodicity at Dinhata is not statistically significant.

However the periodicity of 14.67 years in the annual rainfall series at Dinhata and that of 22.0 years at Cooch Behar have been found statistically significant in this homogeneous meteorological region.

So, slightly different pattern for the 10-11 year solar cycle occurs at the location at Dinhata but the same wave band is not significant at Cooch Behar station. The yearly rainfall records at Cooch Behar have yielded an evidence for slightly different pattern of luni-solar cycle.

The presence of 29 years wave is significant at the 5 percent level in the rainfall of Dinhata station while the presence of such a period at the other station is not significant. Moreover, some neighbouring spectral estimates may also be considered significant but to overcome this problem, we try to employ the procedure of smoothing the periodogram in the next sub-section.

7.4. SMOOTH SPECTRUM

Smoothing the periodogram is to smooth the neighbouring autocovariances with a time domain moving average. Every frequency domain window (moving average) has a time domain representation and reversely,every time domain window has a frequency domain representation.

However,the new spectral estimates are obtained by smoothing the raw estimates of periodogram by a set of weights.

Humming estimates the periodogram by Tuckey-Hamming window and the estimated smooth periodogram is given by

$$P(W) = \frac{1}{\pi} \left(C_o + 2 \sum_{K=1}^{N} \lambda_k C_k \operatorname{Cos} 2\pi w\right)$$

Where λ_k is the lag window with truncation point M defined at discrete points k. In this computation the Tukey-Hanning window has been used as

$$\lambda_k = .54 + .46 \operatorname{Cos} \left(\frac{\pi k}{M}\right), \; K \leq M .$$

Various band widths have been used in this computation. As various wave lengths are detected in the previous section, progressively larger band widths are employed to estimate the smoothed spectral density.

The chosen default is Q^2 where Q is the cube root of the number of observations. Here Q^2 is approximately 20.

Following the computation procedure mentioned above,we have obtained the smooth spectral estimates at the angular frequency W_k.

The smooth spectrum has been standardized. So, we have,

$$P^*(W_k) = \frac{P(W_k)}{S^2}$$

Where S^2 is the variance of the series of X_t . The standardized spectrums are used to detect the peak of the power spectram for truncation points (M) . There are three distinct wave periods in the smoothed spectral density of the rainfall at Cooch Behar station. But there are four remarkable spectral peaks in the rainfall at Dinhata station. The frequency scale is linear since the bandwidth is independent of frequency.

The standardized smooth spectral density functions corresponding to their angular frequencies are illustrated in Figures **7.2(a & b)** for the stations Cooch Behar and Dinhata respectively.

Table **7.4** also presents the remarkable high spectral peaks corresponding to their respective periods for both the stations.

In the next sub-section ,we adopt a test statistic to test the hypothesis for the presence of periodicity in the individual rainfall series.

7.4.1. TEST OF SIGNIFICANCE OF SPECTRAL PEAKS.

For testing the presence of periodicity in the rainfall series, an α significance level holding simultaneously for M values of $P^*(w_k)$ is given by

$$b_M = \frac{\chi^2_{d,1-\beta}}{\pi d}$$

where $$\beta = \frac{\alpha}{M}$$

and 'd' denotes the degrees of freedom of the spectral window. The degrees of freedom of the lag window is 2.5164(N/M). The test criterion is used to compute the critical limit under the null hypothesis. The test has been used by Helmut (1986).

Now in our assignment, we have to put $\alpha = 0.05$, $M = 20$ and $N = 88$, then we obtain

$$b_{20} = 0.$$

The standardized smooth spectral densities are employed here, to detect the presence of periodicities in the individual rainfall series.

If any of the spectral estimates lies outside this critical point, it means the presence of significant periodicity in the rainfall series.

The standard smoothed spectrum density values of the remarkable spectral peaks are 1.771, 1.265 and 1.257 corresponding to the periods 2.667 , 3.52 and 5.176 respectively in the annual rainfall series at Cooch Behar station. All these values are greater than the hypothetical value of Chi-square at the 5 percent level of significance . The results give evidence for low band-limited signals near quasi-biennial, quasi-triennial and quasi-five year in the rainfall of this station. It is obvious from the Figure **7.2(a)** that periodicities greater than quasi-five year bands are not statistically significant in the climatological time series at Cooch Behar station.

But it is interesting to note that the evidence to be presented later is more convincing for the presence of high band-limited signal near 18.6 year luni-solar cycle. The standardized smoothed spectral density values of the remarkable spectral peaks are 0.958, 1.672, 0.805 and 1.999 corresponding to the periods 2.378, 3.52, 8.80 and 17.60 respectively in the annual rainfall series at Dinhata. All these values of the spectral peaks are also greater than the hypothetical value of Chi-square at the 5 percent level of significance. So,this smoothed spectrum consists of one signal induced by the highly resonant 17.6 year quasi-standing wave and additionally there are three smaller cycle terms of near 9 years, quasi-triennial and quasi-biennial in the annual rainfall at Dinhata station.

In this analysis,there is a band-limited power near 17.6 years periodicity at Dinhata station which yields the evidence of slightly different pattern of 18.6 luni-solar cycle but it is not identified in the rainfall at Cooch Behar. And the analysis of the smooth power spectrum has reported evidence for a signal with period 8.8 years in the rainfall at Dinhata ,but this period is not found in the annual rainfall at Cooch Behar station while the existence of the quasi-five year oscillation in the rainfall at Cooch Behar has been reported by this analysis.Most of the procedures of analysis are discussed by Jenkins & Watts (1968), Priestley (1981) and Chatfield (1982).

So,there are some significant differences observed between the simple periodogram and the smooth spectral estimates. But it is also indicated that the dominance of short duration fluctuations has been confirmed by the two methods of analysis in the individual rainfall series of the two stations. While a long duration periodicity cannot be ruled out in the annual rainfall in

Cooch Behar district.

7.5. CONCLUSION.

The spectral analysis technique has a very long history during which various approaches have been developed for the computation of the spectral estimates of the observed data. The simple periodogram analysis and the autocorrelation transformation technique have been used in this study.

The quasi-biennial oscillation in the annual rainfall in this region has been indicated prominently. This is in conformity with the well known fact that quasi-biennial oscillations exist in the several meteorological phenomena over India. The quasi-triennial and quasi-five year periodicities are also identified in the annual rainfall series at these homogeneous meteorological stations.

The spectral analysis of yearly total precipitation records at Cooch Behar yields evidence for 17.6 year cycle signal. The periodicity of 17.6 years may approximately be considered as presence of 18.6 years luni-solar cycle. The 10 -11 years solar cycle in the rainfall series have not been detected as such in the analysis though approximately similar result,the periodicity of 14.67 years has been found as evidence for a significant signal.

TABLE 7.1. Autocorrelation function & its plot of yearly rainfall at Coochbehar

AUTOCORRELATIONS

1- 12	.19	0.0	.17	.10	.11	-.02	.01	-.06	-.11	.11	.20	-.16
ST.E	.11	.11	.11	.11	.11	.12	.12	.12	.12	.12	.12	.12
13-24	.04	.24	.09	.08	.05	.02	.02	-.01	-.02	.04	-.08	-.02
ST.E	.12	.12	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13
25-25	-.C2											
ST.E.	.13											

PLOT OF AUTOCORRELATIONS

```
                          -1.0 -0.8 -0.8 -0.4 -0.2  0.0 0.2 0.4 0.6 0.8 1.0
LAG          CCRR.          +....+....+....+....+....+...+...+...+...+...+

1            0.190                              +    Ixxxx
2           -0.004                              +    I    +
3            0.171                              +    Ixxxx
4            0.105                              +    Ixxx +
5            0.115                              +    Ixxx +
6           -0.022                              +   xI    +
7            0.007                              +    I    +
8           -0.059                              +   xI    +
9           -0.114                              + xxxI    +
10           0.107                              +    Ixxx +
11           0.204                              +    Ixxxxx+
12          -0.161                              +xxxxI    +
13           0.045                                   Ix   +
14           0.201                              +    Ixxxxx+
15           0.093                              +    Ixx   +
16           0.081                              +    Ixx   +
17           0.052                              +    Ix    +
18           0.015                              +    I     +
19           0.024                              +    Ix    +
20          -0.009                              +    I     +
21          -0.019                              +    I     +
22           0.042                              +    Ix    +
23         - 0.055                              +   xI     +
24         - 0.017                              +    I     +
25         - 0.017                              +    I     +
```

TABLE 7.2. Autocorrelation function and its plot of yearly rainfall at Dinhata.

AUTOCORRELATIONS

1- 12	.12	.09	.20	.09	-.05	-.21	.04	-.08	-.16	.09	.12	-.13
ST.E.	.11	.11	.11	.11	.11	.11	.12	.12	.12	.12	.12	.12
13-24	-.C7	.02	.01	-.09	-.07	-.02	.01	-.06	.07	-.02	.02	-.08
ST.E.	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13
25- 25	.C6											
ST.E	.13											

PLOT OF AUTOCORRELATIONS

```
                  -1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2  0.4  0.6  0.8  1.0
LAG      CCRR.    +....+....+....+....+....+...+...+....+....+....+
                                             I
 1        0.121                         +    Ixxx+
 2        0.093                         +    Ixx +
 3        0.203                         +    Ixxxxx
 4        0.092                         +    Ixx    +
 5       -0.052                         +   xI      +
 6       -0.205                         +xxxxI      +
 7        0.040                         +     Ix    +
 8       -0.082                         +  xxI      +
 9       -0.163                         +xxxxI      +
10        0.085                         +     Ixx   +
11        0.123                         +     Ixxx  +
12       -0.135                         + xxxI      +
13       -0.066                         +  xxI      +
14        0.023                         +     Ix    +
15        0.015                         +     I     +
16       -0.094                         +  xxI      +
17       -0.068                         +  xxI      +
18       -0.016                         +     I     +
19        0.014                         +     I     +
20       -0.059                         +    xI     +
21        0.067                         +     Ixx   +
22       -0.019                         +     I     +
23        0.015                         +     I     +
24       -0.076                         +   xxI     +
25        0.064                         +     Ixx   +
```

TABLE : 7.3.

Spectral peaks of standardized periodogram

	Cooch Behar		Dinhata	
	Periodogram values	Period	Periodogram values	Period
1st peak	2.299*	22.00	3.438*	29.33
2nd peak	2.293*	4.889	4.117*	14.67
3rd peak	2.626*	3.52	5.631*	3.52
4th peak	3.201*	2.667	2.275*	3.25
5th peak			2.301*	2.75

* Significant at 5% level.

TABLE : 7.4

Spectral peaks of standardized smooth spectrum.

	Cooch Behar		Dinhata	
	Spectrum values	Period	Spectrum values	Period
1st peak	1.257*	5.176	1.999*	17.60
2nd peak	1.265*	3.52	0.805*	8.80
3rd peak	1.171*	2.667	1.672*	3.52
4th peak			0.958*	2.378

*Significant at 5% level.

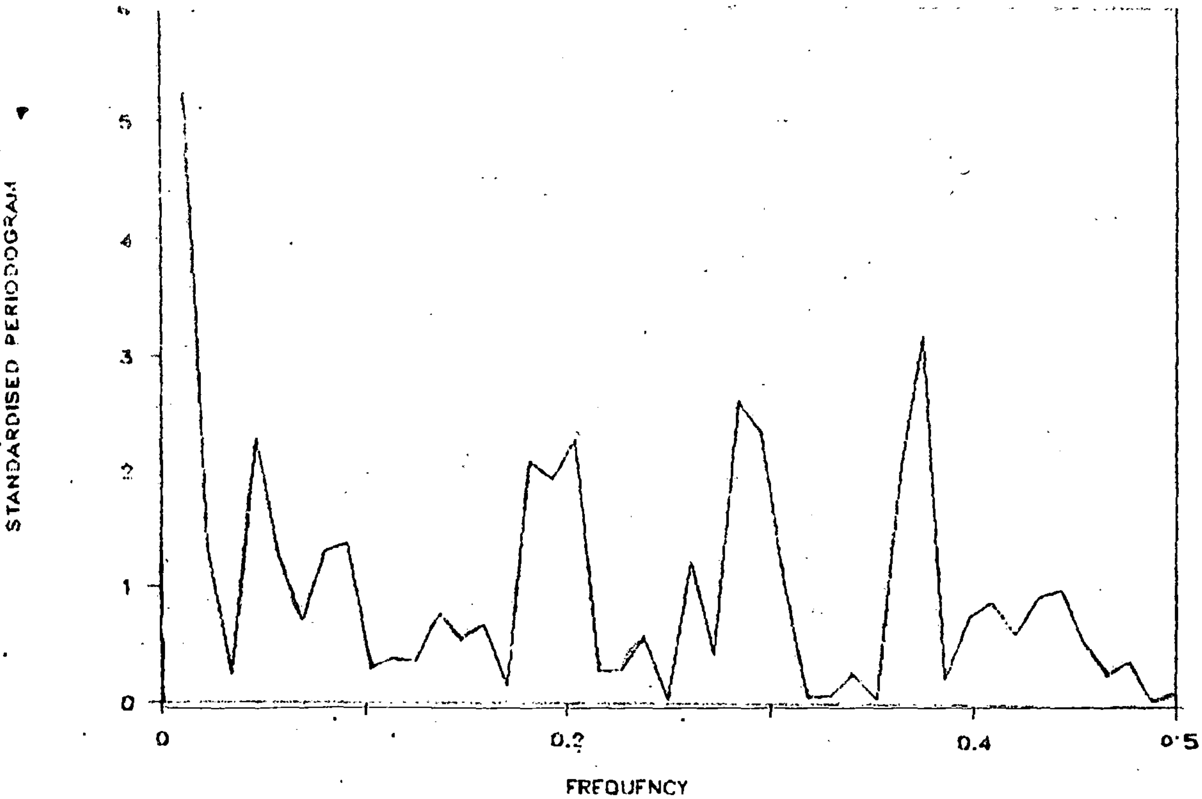


Figure : 7:1(a). Standardised Periodogram of annual rainfall at Cooch Behar.

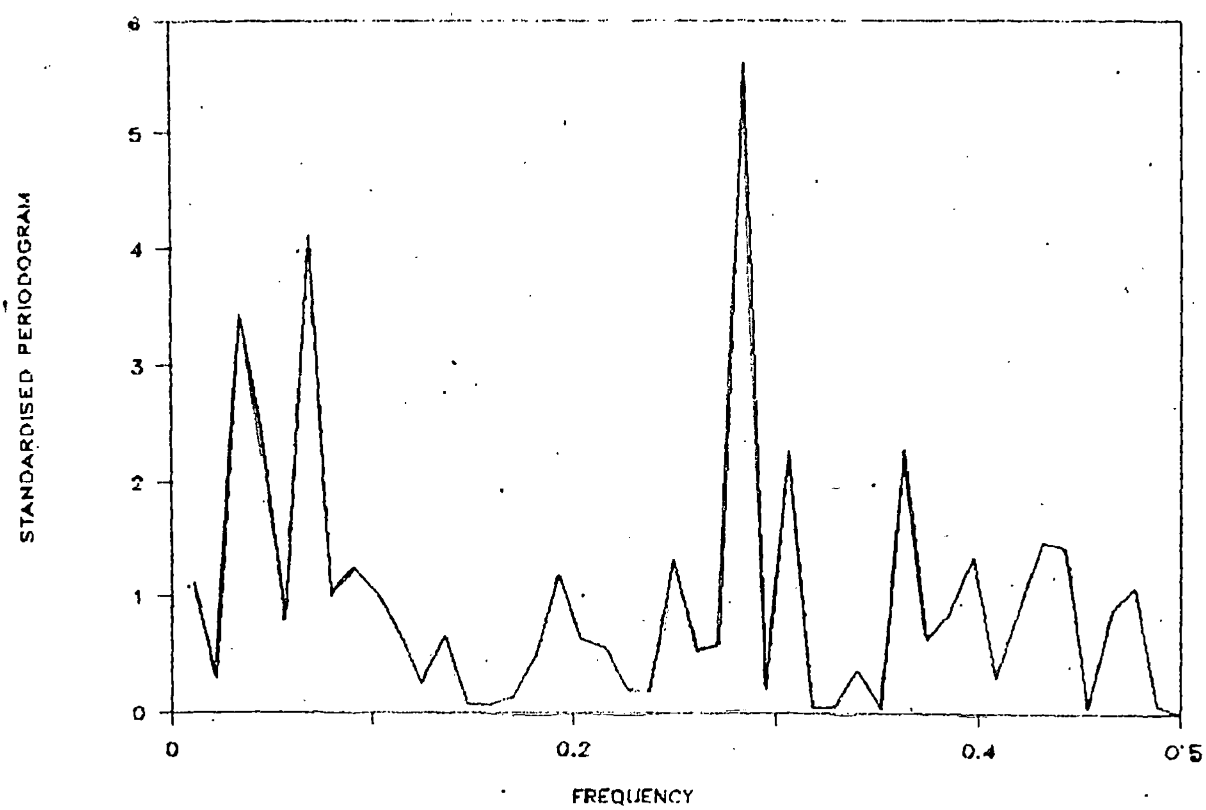


Figure. 7.1(b). Standardised Periodogram of annual rainfall at Dinhata.

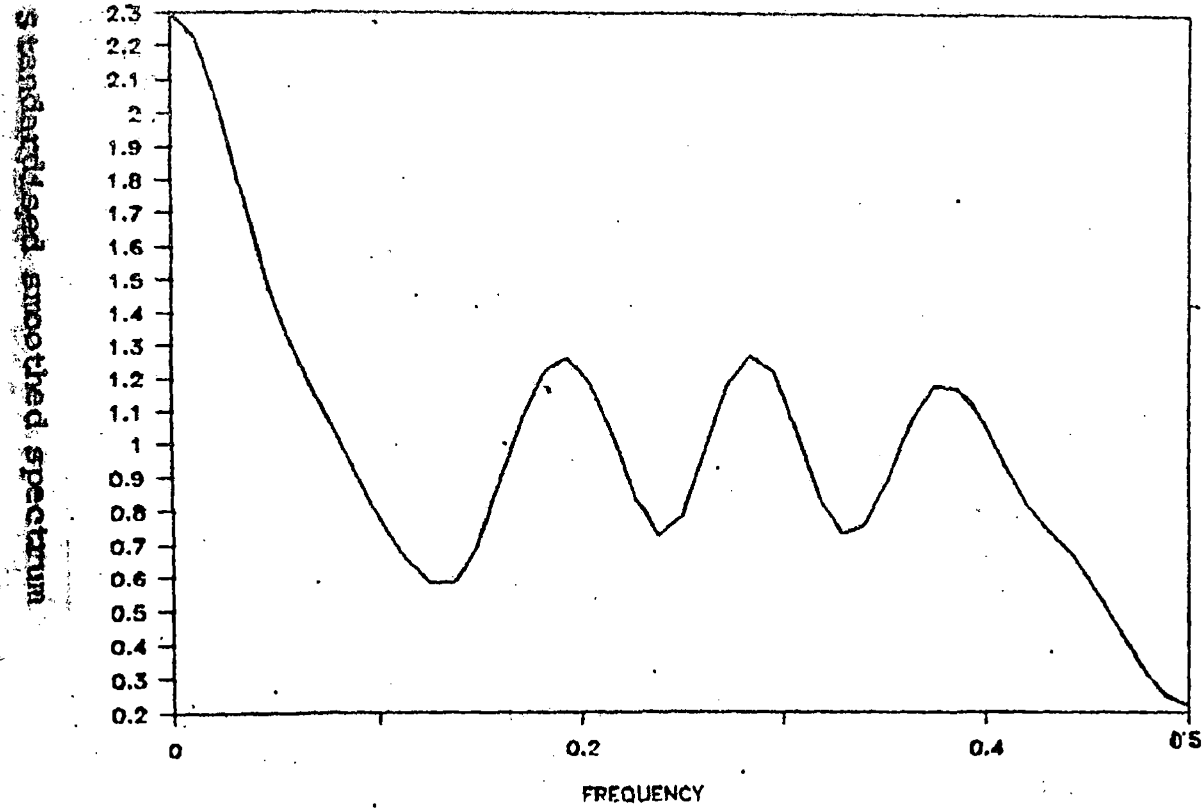


Figure. 7.2(a). Standardised smoothed spectrum of annual rainfall at Cooch Behar.

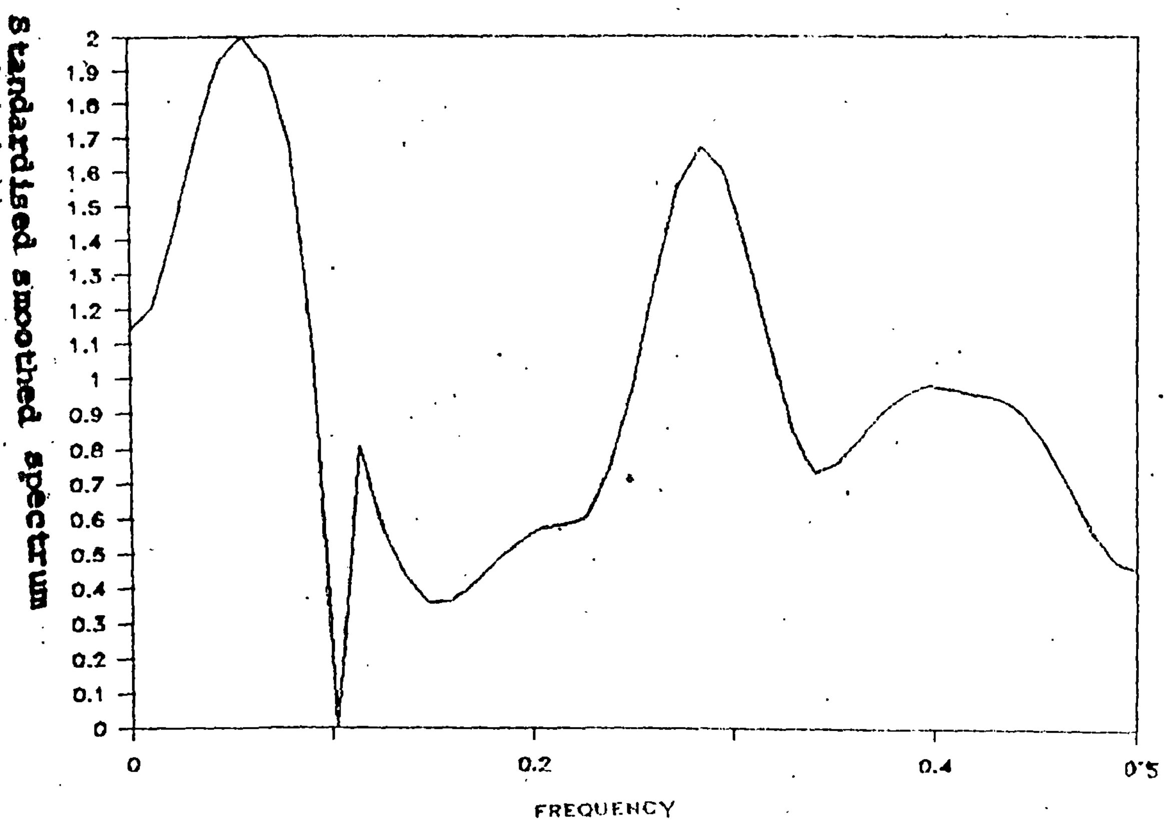


Figure. 7.2(b). Standardised smoothed spectrum of annual rainfall at Dinhata.

CHAPTER - VIII

RELATIONSHIP BETWEEN RAINFALL AND RICE YIELD

CHAPTER - VIII

RELATIONSHIP BETWEEN RAINFALL AND RICE YIELD.

8.1. INTRODUCTION

Agriculture is considered as the primary sector for economic development of our country .In tern agricultural productivity can be defined in different senses of which agriculture yield per hectare is one .

The influences of weather element,particularly rainfall on the agricultural operations and crop yield,can affect favourable and unfabourble conditions during the different stages of development. So,rainfall is the factor determining the success or failure of agricultural enterprises. This is because cultivators have no control over this natural random phenomenon.

In this area farming is done under unirrigated conditions as such it depends mostly on weather and climate. The impact of weather and climate on agricultural productivity is of vital importance. Rainfall is the most important weather parameter that affects paddy crop in this region where irrigation plays a minor role. Paddy is the principal crop of the kharif season. It is mostly rain fed crop. The total absolute rainfall of a particular season and the total production of crop are considered for most of the economic analysis on food grain production. But the nature of intensity of daily rainfall has much influence on variations of yield of this particular crop. These effects are manifested through plant characteristics like height, number of tillers,leaf area and number of ear heads which ultimately affect the yield of the crop.For the purpose of our thesis ,we concentrate on the effect of this climatic factor to the yield of rice.And it has been considered in this study as economic factor in agriculture .

This chapter attempts to establish the relationship between rice yield and the nature of daily rainfall in this district.

8.2. VARIABLES

A time series of crop yield may be divided into three major sources of variation in the yield of rice over any region. These are identified as (i) technological change,(ii) meteorological variability and (iii) random "noise". The technological change in yields creeps in due to the recent advancement in agricultural technology. A suitable linear time-scale variable is introduced in crop-yield model to account for the technological trend. Among meteorological parameters rainfall is one of the most important elements which affect crop yield significantly. The third component has the impact and contribution of random noise which is very little.

Of all the factors affecting agricultural production, availability of water is undoubtedly the most important one. For rainfed crops,rainfall constitutes water gain while the nature of rainy days are of much importance during different stages of development of crop. The nature of daily rainfall data,derived from it,are used in this study. The classification is made by comparing the amount of rainfall received during the past 24 hours. The daily derived rainfall parameters are classified into the following categories :

i) Rainy Days (RD) Days when the rainfall is equal to or more then 2.5 mm.

ii) Restricted Rainy Days (RRD). Number of rainy days except when preceded by a day having 50 mm or more rainfall.

iii) Crop Rainy Days (CRD) Number of days when the rainfall is 5 mm or more. But a day having rainfall 3.8 mm is considered as a crop rainy day with a gap of two non rainy days.

The daily derived parameters as described above are based on rainfall,during the period 1972 to 1988 at Cooch Behar. The monthly derived values from June to the middle of October of each year have been computed. The nature of the day's rainfall for the months of sensitive period of crop life time are considered as the explanatory variables for this study. The average yield of winter rice in this district is used as the dependent variable in this study. A data series of average yield of winter rice for the same period i.e.1972 to 1988 has been used in the present study to develop the multiple regression model.

The yield rate of winter rice has been considered as the dependent variable throughout this chapter. The other variables are potential explanatory variables.

Now we first consider the model with fourteen explanatory variables namely :

I) <u>Rainy Days (RD)</u> :

X_1 = Rainy days during July.

X_4 = Rainy days during August.

X_7 = Rainy days during September.

X_{10} = Rainy days during October.

X_{11} = Rainy days during June

II) Restricted rainy days (RRD)-

X_2 = Restricted rainy days during July

X_5 = Restricted rainy days during August

X_8 = Restricted rainy days during September

X_{12} = Restricted rainy days during June

III) Crop rainy days (CRD) -

X_3 = Crop rainy days during July

X_6 = Crop rainy days during August

X_9 = Crop rainy days during September

X_{13} = Crop rainy days during June

X_{14} = Technological trend in time-scale variable takes from 1 to 17.

8.3. DATA MATRIX OF EXPLANATORY VARIABLES

The individuals of 17 years and 14 variables are mainly related to the derived variables from the daily rainfall data. These variables are separated by rainy months included in the life cycle of the crop. We denote the number of individuals by n, the number of variables by k. So,there are nk variable values.The individual variables are denoted by the subscripts : $X_1,......X_k$. Mathematically,a set of values are arranged in rows and columns in the matrix form (data matrix). The single symbol X stands for the numerical information in this data matrix as a whole . The data,in the form of a matrix are given in the Appendix.

A preliminary summarisation of the data is provided by calculating the mean and standard deviation of each variable. These basic statistics of the variables are given in Table **8.1.** This table also shows the standard error of mean of each variable.

It is also seen from the table that except the dependent variable, the means of the independent variables are in the range from 3.765 to 21.412 days while their standard deviations arefrom 2.322 to 3.704. But those of the dependent variables are 1194.882 and 140.265 respectively. It is also observed that the standard errors of the means of the variables, X_{11} and X_{14} are greater than one while those of the other variables are less than unity. The standard error of the means of the dependent variable is 34.019. So, there are high variability on the observations of the yield of winter rice in this district.

Besides these basic statistics,the other measurements of association between each pair of explanatory variables are correlation coefficients. The Table **8.2** reports the correlation matrix between the explanatory variables as well as the dependent variable with each explanatory variable. The correlation matrix is a symmetric matrix.

The most strongly and positively correlated pair of variables are X_{11} and X_{13} with a correlation coefficient 0.90 and the weakest are X_5 and X_9 with a correlation coefficient .005, virtually zero.

Table **8.2** also shows the correlation between the dependent variable and each of the explanatory variables in which the dependent variable is positively correlated with X_3, X_4, X_5, X_9, X_{10}, X_{13} and X_{14} whereas it has negative correlation with the variables X_1, X_2, X_6, X_7, X_8, X_{11} and X_{12}.

8.4. MULTIPLE REGRESSION MODEL.

An agroclimatic study of the relationship between the crop yields and weather parameters is carried out with the help of empirical statistical multiple regression model. This model is also generally employed for making quantitative crop yield forecast on operational basis.

The technique has been utilized in the present study to develop linear multiple regression model to establish the crop weather relationship. The relationship between yield per hectare (in Kg) of paddy crop and the nature of daily rainfall is to be established employing the following functional form :

$$y = b_o + \sum_{i=1}^{k} b_i x_i + u_i$$

Where y is the dependent variable i.e. yield of winter rice in Cooch Behar.

X_i's are the derived parameters from the nature of daily rainfall for the sensitive period by month. These are also known as nonstochastic components of the model.

u_i is the stochastic component. This is also known as disturbance term.

8.4.1. ASSUMPTIONS

The linear multiple regression model is based on some assumptions . Some of them are related to the distribution of stochastic variables, some to the relationship between nonstochastic and stochastic variables and finally some refer to the relationship between these nonstochastic variables.

a) Assumptions about the stochastic variables.

Assumption -I.Normality-the stochastic variables are normally distributed with zero mean and constant variance. Regarding constant variance,for all values of X, u' s would show the same dispersion round their mean.

i.e.

$$E\ (u_i) = 0$$

$$\text{and } E\ (u_i^2) = \sigma_u^2 \quad \text{(Constant)}$$

Assumption 2. Nonautocorrelation or serial independence of the u's. This assumption states that any two values of disturbance term (stochastic) u_i and u_j are not linearly correlated i.e.

$$E\ (u_i u_j) = 0 \quad i \neq j.$$

(b) Assumption 3.Assumption for stochastic and nonstochastic variables.

Independence of u_i and x_i. This assumption states that every stochastic term is independent of the explanatory variables i.e.

$$E\ (u_i\ x_{ij}) = 0$$

Where X_{ij} is the element of the data matrix,i.e.,in the ith variable and the jth individual year.

This assumption is not very critical. It is automatically fulfilled if the explanatory variable is nonstochastic.

Assumption 4. Assumption about the relationship between the explanatory variables themselves.

This assumption states the closeness of explanatory variables to each other. This is the case of multicollinearity.

8.4.2. ESTIMATION OF COEFFICIENT OF REGRESSION MODEL

We try to use the rules of thumb by which we can derive

(a) The regression co-efficients ,

(b) The variance of the co-efficients

and

(c) The coefficient of multiple determination,

The population regression space involving dependent variable Y and k explanatory variables :

$X_1, X_2, \ldots\ldots, X_k$ may be described as -

$$y_i = B_o + B_1 X_{1i} + B_2 X_{2i} \ldots\ldots B_k X_{ki} + u_i$$

Where B_o is the intercept , B_1 to B_k are the regression (slope) coefficients, u_i is the disturbance term the suffix i indicates the ith observation of the population which comprises n observations. We should therefore, have n number of linear equations. These equations can be put in a compact form using matrix notation :

$$\underset{\sim}{y} = \underset{\sim}{x} \underset{\sim}{B} + \underset{\sim}{u} .$$

where

$\underset{\sim}{y}$= (nx1) column vector of observations on the dependent variables y_i

$\underset{\sim}{X}$= nx(k+1) matrix giving n observations on k explanatory variables. The first column of the matrix is unity representing the intercept term.

$\underset{\sim}{B}$= $\{(K+1)x1\}$ column vector of the regression coefficient.

$\underset{\sim}{u}$= (nX1) column vector of n observations on disturbance term i.e.,

$$\underset{\sim}{Y} = \begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{vmatrix}$$

$$\underset{\sim}{X} = \begin{vmatrix} 1 & X_{11} & X_{21} & X_{k1} \\ 1 & X_{12} & X_{22} & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & X_{kn} \end{vmatrix}$$

$$\underset{\sim}{B} = \begin{vmatrix} B_o \\ B_1 \\ \vdots \\ B_k \end{vmatrix} \qquad \underset{\sim}{u} = \begin{vmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{vmatrix}$$

The ordinary least square method has been applied to estimate the regression coefficients. This is given by

$$\underset{\sim}{b} = (X'X)^{-1} \underset{\sim}{X}'\underset{\sim}{Y}$$

Here the estimators follow the Gauss-Markov property.

In this model the ordinary least squares estimators are those which minimize Σe_i^2 with respect to b_i where e_i is the n element column vector of n residuals in the sample regression space and it can be measured by

$$e_i^2 = (Y_i - b_o - b_1 x_{1i} - \dots\dots\dots\dots b_k X_{ki})^2$$

The total variation in the dependent variable (Y) can be partitioned into variation due to chosen regressors and the residual variance. The partitioning is obtained as follows :

$$\sum_i^n (Y_i - \bar{Y})^2 = \Sigma (\hat{Y}_i - \bar{Y})^2 + \Sigma e^2$$

$$\text{i.e.,} \Sigma y_i^2 = \Sigma \hat{y}_i^2 + \Sigma e_i^2$$

Where,

Σy_i^2 = Total sum of square (TSS)

$\Sigma \hat{y}_i^2$ = Regression sum of square (RSS)

and

Σe_i^2 = Error sum of square

The coefficient of multiple determination or the squared multiple correlation coefficient is the ratio of regression sum of square to total sum of square and is denoted by R^2 .

The estimate of R^2 can be obtained as

$$R^2 = \frac{\text{Regression SS}}{\text{TSS}}$$

$$= \frac{\Sigma\ \hat{y}_i^2}{\Sigma\ y_i^2}$$

When the explanatory variables in the function are large with comparison to the number of observations, the multiple determination can never reduce but it will rise usually, because the numerator will increase in the expression of R^2 while the denominator remains the same. To correct this defect we adjust R^2 by considering the degree of freedom. So, the adjusted coefficient of multiple determination is

$$\bar{R}^2 = 1 - (1 - R^2)\ \frac{(n - 1)}{(n - k)}$$

$$= 1 - \frac{\Sigma\ e^2/(n - k)}{\Sigma\ y^2/(n - 1)}$$

Where R^2 is the unadjusted multiple correlation coefficient, n is the number of sample observations and k is the number of parameters estimated from the samples.

The adjusted value of multiple correlation coefficient $\bar{R}^2$ when multiplied by 100 gives the percentage of total variation in Y explained by the regression.

The method outlined in this sub-section is concerned to develop the suitable multiple regression model. Multiple regression analysis, in practice, require the use of appropriate computer programme. The methods employed in our analysis are supported by the software package of Data matric. The computer out puts of multiple regression analysis are presented there step by step.

The multiple regression equation can be written as

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \;........\; + b_{14} \; x_{14} + u$$

The extract form of the computer output for the fourteen explanatory variables model is shown in Table **8.3.** For this model,the explanatory variables, X_1, X_2, X_6, X_7, X_8, X_{11} and X_{12} have marginal negative effects and X_3, X_4, X_5, X_9, X_{10}, X_{13} and X_{14} have the marginal positive effect on the dependent variable. The table also shows the value of t-statistics i.e.,b_i/SE(b_i) of each of the explanatory variables as well as the constant i.e. the intercept.

In this equation the degree of freedom of t-statistic is 2,i.e, [17 - (14 + 1)]. It is observed that the regression coefficients b_4,b_{11} and b_{14} have the t-value less than 2 i.e. the hypothetical value of standard-error test which takes the constant value 2 to compare the calculated value. But we cannot consider this procedure to detect the unwanted variables in the equation.

However,to get the accuracy of the predictive model we have to omit the variable which may contribute little to predict the dependent one.

Now,the step-down method has been applied to solve this problem. One of the rules of this method is to omit the unwanted variable (x_u) if

$$T = b_u^2 \;/\; s^2 < 1$$

Where b_u is the regression coefficient of the unwanted explanatory variables and s^2 is square of standard error of that coefficient.

Now, we consider the multiple regression equation of fourteen explanatory variables. Here the variable x_{14} is detected as unwanted variable where the value of T for the variable x_{14} is .02224 (<1). The variable x_{14} i.e. time scale variable is omitted from our multiple regression equation.

As the technological trend i.e. time-scale variable is omitted from the multiple regression equation, it is indicated that the advances in agricultural technology have very little contribution on the yield of winter rice in the district of Cooch Behar.

Moreover, the technological trend caused by the use of high yielding varieties, large scale use of fertilizers, better irrigation facilities, use of insecticides and improved management practices has no significant contribution on the production of winter rice at the district of Cooch Behar.

The calculated value of R^2, adjusted $\bar{R}^2$ and F-value of the multiple regression equation and the regression coefficients along with respective values of standard error and t-values are given in Table 8.3.

Now, the regression of y on the remaining (14-1) i.e., thirteen explanatory variables are to be computed. The computed new multiple regression equation is also given in Table 8.4, and the same rule is followed to identify the unwanted variables.

Here the value of T for the value X_{11} is observed less than one i.e.,

$$T\ (x_{11}) = 0.0627 \quad (< 1)$$

The values of R^2, adjusted $\overline{R}^2$ and F- value and model of multiple regression equation with the values of regression coefficient corresponding to their standard errors and the t-values are given in Table **8.4.**

So, this variable X_{11} is omitted from the equation.

This is indicated that the rainy days during the month of June have no significant effect on the production of winter rice in Cooch Behar district.

Here we also note, that the value of R^2 is decreased by a small quantity whereas the value of adjusted $\overline{R}^2$ and the F value have been increased due to the omission of the variable X_{14} that is the variable of technological change.

By following the same method we now compute the multiple regression equation by eliminating the variable X_{11} i.e. rainy days during the month of June.

We next compute the regression :

$$y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{12}, X_{13})$$

Now, omitting the variables X_{11} and X_{14}, the regression equation of y on the remaining twelve explanatory variables is computed. The latest model of multiple regression equation is given in Table **8.5** which finally consists of one dependent variable and twelve explanatory variables. The variable X_{11}, rainy days during the month of June and the variable X_{14}, the technological trend are excluded from the multiple regression equation by adopting the step-down method. But in the new equation,there is no variable which we can omit by applying the step-down method. So,this equation now has been employed to develop the relation between the winter rice yield and the derived variables of rainfall of different sensitive months.

So,finally,we fit multiple linear regression model on yield of winter rice.

However the calculated values of R^2,adjusted $\overline{R}^2$ and F-statistic of the equation and the required regression coefficients along with their standard errors and the corresponding t-values are given in Table 8.5.

The estimates of the multiple correlation coefficient and regression coefficient which are defined through least squares estimators,are essential to examine the reliability precision of the estimates.

8.4.3. TEST FOR THE SIGNIFICANCE OF THE MULTIPLE CORRELATION COEFFICIENT

The test performed in regression analysis is a test concerning the overall explanatory power of the regression as measured by R^2.

The null hypothesis of this test is that the multiple correlation coefficient in population is zero and the alternative hypothesis is that it is greater than zero.

The F ratio is a test of significance of R^2 ,the statistic :

$$F^* = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k}$$

Confirms to F-distribution with (k,n-k-1) degrees of freedom.

So,we can apply the test of this hypothesis under consideration,knowing only the value of observed multiple correlation.

The values of R^2,adjusted $\bar{R}^2$ and F^* ratio are given in Table **8.5**.

The calculated R^2,the squared coefficient of multiple correlation is 0.9832 and the adjusted $\bar{R}^2$ is 0.9327. The calculated value of F^* ratio is 19.4674. The tabulated value of F-ratio with the degrees of freedom 12 and 4 is 14.37 at 1 percent level of significance. Hence,it is statistically highly significant. Therefore ,we consider that multiple regression equation is well fitted to the derived parameters of rainfall in Cooch Behar district.

Now consider the value of adjusted $\bar{R}^2$ which is 0.9327. We multiply the value of $\bar{R}^2$ by hundred,we get it 93 percent.

The model,therefore, can explain 93 percent of total variation of rice yield with multiple correlation coefficient as 0.98. Hence,all the explanatory variables may be emerged significantly at 1 percent level.

8.4.4. <u>THE MEAN AND VARIANCE OF THE PARAMETER ESTIMATES</u>

Since the least square estimators are linear combinations of independent variables. The intercept and slopes are normally distributed with

$$\text{Mean} : E(b_0) = B_0$$

$$\text{and} : E(b_i) = B_i$$

And the variance of any estimator may be derived from the principal diagonal of dispersion matrix $(X'X)^{-1}$ and then multiplied by σ^2 .

An unbiased estimator of σ^2 is denoted by $\hat{\sigma}^2$ and is given by

$$\hat{\sigma}^2 = \frac{\Sigma e_i^2}{(n-k-1)}$$

Where k is the total number of regressors,or independent explanatory variables of the model.

so,standard error of $b_i = \sqrt{\hat{\sigma}^2 \; C_{ii}}$

where C_{ii} is the diagonal element of the dispersion matrix $(X'X)^{-1}$.The procedure described above has been suggested by Rao (1984).

The value of b_i,s and their standard error are given in Table 8.5 which also shows our best fitted multiple regression model.

8.4.5. TEST OF SIGNIFICANCE OF PARAMETER ESTIMATES

When the sample is small $(n < 30)$ and the parent population is normal,we may apply a test statistic which is based on Student's t-distribution. To perform a two tailed test of this test criterion we must adopt the null and alternative hypothesis with desired level of significance and suitable degree of freedom.

We assume that the least squares estimates are normally distributed.

The null hypothesis is that there is no linear relationship between the explanatory variables and the dependent variable y i.e.

$$H_o : B_i = 0$$

and is tested against the alternative hypothesis,

$$H_1 : B_i \neq 0$$

In this case the t-statistic reduces to

$$t^* = \frac{b_i}{\hat{\sigma}(b_i)}$$

The sample value of t^* is estimated by dividing the estimate b_i by its standard error.

Where

b_i = least squares estimate of B_i

$\sigma^2(b_i)$ = estimated variance of B_i

n = Sample size

k = Total no. of estimated parameters.

The value of t^* is compared to the theoretical (tabular) values of t which defines the critical region in two-tailed test with {n − (k + 1)} degrees of freedom.

If the calculated values of t^* are greater than the tabular value of t i.e. $t^* > t$,we may reject the null hypothesis. This means that we may accept the estimated b_i which is statistically significant at chosen level of significance.

Therefore,it is essential to examine the reliability or the precision of the least square estimators.

Table 8.5 reports the value of the estimates of the parameters and their respective standard errors and the computed t^*-value of each of the variables.

The observed value of the t^*-ratio would be compared to the theoretical value of t,obtainable from the t-table with degree of freedom {n - (k + 1)} = 17 - (12 + 1) = 4 which is the suitable degree of freedom for the fitted model.

The statistical reliability of the estimated parameters of the equation as shown in Table 8.5 is discussed here one by one parameter.

The calculated t-value of the regression coefficient of b_1 is 5.995 which is greater than the tabulated value of t with 4 degree of freedom at one percent level of significance (4.604). Since we reject the null hypothesis with 99 percent confidence the variable X_1,the rainy days during July,has a significant effect on the yield of winter rice in the district of Cooch Behar. But the regression coefficient bears negative sign and the negative coefficient is also statistically significant. Hence the yield of winter rice is found to be inversely related to the rainy days in the month of July.

The same result has also been observed on the coefficient of variable X_2, the restricted rainy days in the month of July but the coefficient is significant at 5 percent level (t = 2.776). This suggests that this variable has a significant and negative influence of the yield of winter rice at this place.

The coefficient of the variable X_3 i.e.,the crop rainy days in the months of July has positive sign and it is statistically highly significant at 5 percent level. The result shows that the variable has a positive influence on the rice yield.

The coefficient of the variable X_4, rainy days in the month of August, has a positive sign but it is not statistically significant, showing that this variable is not a determinant of the yield of winter rice at Cooch Behar.

The estimated coefficient of the variable X_5, the restricted rainy days in the month of August, has a positive sign and it is found to be highly significant, in influencing the yield of rice. It is significant at one percent level of significance.

The coefficient of the variable X_6, the crop rainy days in the month of August, has a negative sign and it is also statistically significant at 5 percent level. Hence, it is inversely related to the yield of winter rice.

The estimated coefficients of the variables x_7 and x_8, the rainy days and the restricted rainy days in the month of September respectively, have negative sign and these are found to be highly significant at one percent level of significance. So, these variables are also inversely related to the yield of winter rice.

The estimated coefficients of the variables X_9 and X_{10}, the crop rainy days in the month of September and the rainy days in the month of October, have positive sign and are statistically highly significant. These are significant at one percent level of significance. The results show that the variables have positive influence to the yield of winter rice at this place.

The estimated coefficient of the variable X_{12}, restricted rainy days in the month of June, has a statistically highly significant at 0.1. percent level (t=8.61) it has a negative influence on the rice yield.

The coefficient of the variable X_{13}, the crop rainy days in the month of June, has a positive sign and it is found to be highly significant at 0.1 percent level. This suggests that the variable has a significant influence on the yield of winter rice in Cooch Behar.

8.5. SECOND ORDER TEST OF FITTED EQUATION

If the assumptions of an econometric method are violated in any application, the estimates obtained from the method do not possess some or all of the optimal properties. Now we will develop the economic criteria or second order test for judging the goodness of the estimates. These criteria provide some evidence about the validity of the assumptions of the classical linear regression model. The major problems likely to be encountered at this stage of analysis are autocorrelation and multicollinearity. While multicollnearity is the problem concerning closeness of explanatory variables to each other, the other pertains to the behaviour of estimated residuals.

8.5.1. AUTOCORRELATION.

One of our assumptions of ordinary least squares specified earlier is that the successive values of disturbance term (u) are independent which is stated as

$$E\,(u_i u_j) = 0 \qquad i \neq j, i \;\&\; j = 1,2,\ldots\ldots,n$$

When this assumption is not satisfied,the problem of autocorrelation arises. For testing the presence of autocorrelated errors,we use the residual terms by estimating the ordinary least squares.

Darbin-Watson have suggested a test for autocorrelated errors which is applicable to small samples also.

The test may be outlined as follows :

The null hypothesis is

$Ho : \rho = 0$

or , Ho: the U's are not autocorrelated. This hypothesis is tested against the alternative hypothesis

$$H_1 : \rho \neq 0$$

i.e. H_1 : the U's are autocorrelated

To test the null hypothesis,we use the Durbin-Watson statistic :

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Where e_t are sample residuals.

It can be shown that the value of d lies between 0 and 4 and that when d = 2, then $\rho = 0$,

Thus, testing Ho : $\rho = 0$ is equivalent to testing Ho : d = 2. Expanding the 'd' statistic we, therefore, obtain

$$d \approx 2(1 - \rho)$$

From this expression, it is clear that the values of d lies between 0 and 4.

If there is no autocorrelation i.e,

$$\rho = 0 \text{ i.e. } d = 2$$

Thus if from the sample data we get d = 2 and we can not reject the null hypothesis that there is no autocorrelation in the function.

So, it should be clear that in the Durbin-Watson test, the null hypothesis of zero autocorrelation is carried out indirectly when the explanatory variables are large, by testing the equivalent null hypothesis, Ho : d = 2

The computed D -W value (d) is 2.042 which is likely equivalent to 2. Thus, from the sample data it is found that

$$d = 2$$

Hence, we accept the null hypothesis that there is no autocorrelation in the multiplied regression equation. The calculated values of D-W statistic is given in Table **-8.5.**

Therefore, the autocorrelation is not found to be significant and hence, it may be concluded that statistically, the yields in winter rice in Cooch Behar district do not persist from year to year during the time frame of this study.

8.5.2. MULTICOLLINEARITY

The multicollinearity in econometrics is defined as a situation of high correlation between the explanatory variables. Thus,the problem of multicollinearity is always present except for an extreme case where the explanatory variables are called orthogonal. By very definition,two variables are called orthogonal if the correlation between them is zero. But such a case is never encountered where the real world data are handled.If multicollinearity is always present,we have to ascertain whether it is serious i.e.,whether it is of such magnitude as to bother the analyst.

To ascertain the presence and extent of multicollinearity in the data simple zero order correlation matrix of all the explanatory variables are worked out. If none of the correlation coefficient is high,the problem of multicollinearity is not considered serious. In general,if the correlation coefficient is 0.8 or more,it is to be considered high but if it is less than 0.70,it can be considered as low.

In this connection Klein has suggested that multicollinearity should not be considered serious if the simple correlation between a pair of variables is less than the multiple correlation coefficient. That is Klein argues that collinearity is harmful if

$$r^2_{x_i x_j} \geq R^2_{y.x_1 \ldots\ldots\ldots\ldots x_k},$$

where $r^2_{x_i x_j}$ is the simple correlation coefficient between any two explanatory variables (X_i, X_j) and R^2 is the multiple correlation coefficient of the function.

In using Klein's method of detecting multicollinearity,we should consider the value of the estimated multiple correlation coefficient. The simple correlation matrix of explanatory variables are given in Table **8.2.** The multiple correlation coefficient (R^2) has been estimated as 0.98. Then,according to Klein,the simple correlation coefficient between explanatory variables cannot be considered high because these are all less than 0.98.

Hence,multicollinearity is not a problem in the fitted model of multiple regression.

8.6 CONCLUSION :

The linear form of the multiple regression equation is found to be empirically appropriate as it possesses high explanatory power. The results show that the derived parameters of rainfall in the months,June to October (middle) have been found to be highly significant in influencing the yield of winter rice in Cooch Behar. The model explains more than 93 percent of total variation in yield with multiple correlation coefficient exceeding 0.98.

All the independent variables except the variable X_4,the rainy days in the month of August,have emerged significant ,at least,at 5% level and reveal stability of regression coefficients. This model may forecast the rice yields moderately within tolerable limits of statistical significance. From the result of Durbin-Watson test for autocorrelations,it may be concluded that the successive values of the disturbance term i.e. stochastic term are independant. Hence,the assumption of ordinary least squares estimate that the successive values of the random variable i.e. disturbance term,are independent,has been established. From Klien's approach,multicollinearity is not a problem in the fitted multiple regression equation. Hence a crucial condition for the application of least squares that the explanatory variables are not perfectly linearly correlated is established.

The satisfactory performance of well fitted model of multiple regression equation are developed to establish the crop-weather relationship in the district of Cooch Behar.

T A B L E – 8.1

MEAN, STANDARD DEVIATION AND STANDARD ERROR OF MEAN OF INDEPENDENT AND DEPENDENT VARIABLES

VARIABLES	MEAN	STD.DEV	STD.ER.
X_1	21.412	3.374	0.818
X_2	16.706	2.823	0.685
X_3	18.706	3.496	0.848
X_4	15.421	3.432	0.832
X_5	11.529	2.322	0.563
X_6	13.706	3.704	0.898
X_7	15.882	3.238	0.785
X_8	13.235	2.705	0.656
X_9	14.059	2.680	0.650
X_{10}	3.705	2.705	0.656
X_{11}	17.294	4.469	1.084
X_{12}	14.235	3.419	0.829
X_{13}	15.235	3.632	0.881
X_{14}	9.000	5.050	1.225
Y	1194.882	140.265	34.019

X_i's–Independent variables (Nature of rainfall) Y–dependent variable.

T A B L E – 8.2.

CORRELATION MATRIX

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	Y
X_1	.16	.62	.38	–.05	.33	.38	.40	.18	–.07	–.12	–.12	–.07	.41	–.396
X_2	+	.37	.33	.07	.23	–.03	–.02	–.05	.16	.02	.18	.09	–.13	–.005
X_3	+	+	.30	–.01	.15	.32	.38	.03	–.31	.28	.35	.31	.12	.076
X_4	+	+	+	–.02	.70	.05	.09	.15	–.24	–.05	–.14	–.02	.30	.082
X_5	+	+	+	+	.03	.03	.03	.005	–.31	.13	.46	.21	–.37	.059
X_6	+	+	+	+	+	.09	.08	.20	–.14	–.27	–.32	–.21	.31	–.075
X_7	+	+	+	+	+	+	.64	.69	–.38	–.18	–.18	.04	.32	–.038
X_8	+	+	+	+	+	+	+	.64	–.45	–.06	–.07	.15	.27	–.213
X_9	+	+	+	+	+	+	+	+	–.30	–.21	–.17	–.03	.29	.140
X_{10}	+	+	+	+	+	+	+	+	+	–.07	–.11	–.22	–.13	.106
X_{11}	+	+	+	+	+	+	+	+	+	+	.85	.90	–.21	–.162
X_{12}	+	+	+	+	+	+	+	+	+	+	+	.69	–.35	–.030
X_{13}	+	+	+	+	+	+	+	+	+	+	+	+	–.26	.240
X_{14}	+	+	+	+	+	+	+	+	+	+	+	+	+	.400

Y–dependent variable and X_i–explanatory variables.

TABLE - 8.3.

MULTIPLE REGRESSION EQUATION WITH FOURTEEN EXPLANATORY VARIABLES

Parameter	Estimate Co-efficients	Std.error of coefficients	t-Value(calculated)
b_0	849.1006	183.949	4.616
b_1	-34.3120	9.230	3.717
b_2	-17.1002	8.231	2.078
b_3	70.4048	11.836	5.948
b_4	30.0341	21.970	1.367
b_5	66.1988	15.083	4.389
b_6	-48.3085	19.483	2.479
b_7	-66.0784	19.271	3.429
b_8	-42.8111	10.267	4.170
b_9	104.2897	18.419	5.662
b_{10}	44.3591	12.112	3.662
b_{11}	-2.9575	12.498	0.237
b_{12}	-108.9854	16.937	6.435
b_{13}	69.7930	13.804	5.056
b_{14}	0.6180	4.144	0.149

$R^2 = 0.9837$ $\bar{R}^2 = 0.8694$

$F(14,2) = 8.6099$ $b_u = 14$ $T = 0.02224$

TABLE - 8.4.

MULTIPLE REGRESSION EQUATION WITH THIRTEEN EXPLANATORY VARIABLES.

Parameter	Estimated Co-efficient.	Std.error of coefficient	t-value(calculated)
b_0	857.4954	143.781	5.964
b_1	-33.5842	6.433	5.221
b_2	-16.7674	6.505	2.578
b_3	69.4451	8.156	8.514
b_4	29.0431	17.193	1.689
b_5	64.8592	9.948	6.520
b_6	-47.1222	14.603	3.227
b_7	-65.0304	14.732	4.414
b_8	-43.0244	8.347	5.154
b_9	103.6104	14.653	7.071
b_{10}	43.3822	8.365	5.186
b_{11}	-2.0467	9.803	0.246
b_{12}	-108.0116	12.830	8.418
b_{13}	68.7763	9.855	6.979

$R^2 = 0.9835$ $\bar{R}^2 = 0.912$

$F\ (13,3) = 13.7529$

$b_u = 11$ $T = 0.06027$

TABLE - 8.5.

MULTIPLE REGRESSION EQUATION (FITTED MODEL)

Parameter	Estimate co-efficient.	Std.error of coefficient	t-value(calculated)
b_0	846.9494	120.019	7.057
b_1	-33.3476	5.563	5.995 **
b_2	-15.9109	4.802	3.313 *
b_3	69.2490	7.100	9.754 ***
b_4	27.2104	13.547	2.009 (NS)
b_5	65.2518	8.588	7.598 **
b_6	-45.7636	11.820	3.872 *
b_7	-64.7982	12.859	5.039 **
b_8	-43.1232	7.293	5.913 **
b_9	103.7087	12.812	8.095 **
b_{10}	42.6943	6.894	6.193 **
b_{12}	-109.2711	10.286	10.623 ***
b_{13}	67.1473	6.374	10.535 ***

R^2 = 0.9832 $\bar{R}^2$ = 0.9327

F(12,4) = 19.4674** D-W Stat = 2.042

* - Significant at 5 percent.

** - Significant at 1 percent.

*** - Significant at 0.1 percent

NS - Non-significant.

CHAPTER - IX

SUMMARY AND CONCLUSIONS

CHAPTER : IX

SUMMARY AND CONCLUSIONS

9.1. SUMMARY OF FINDINGS :

In the district of Cooch Behar, about 10 percent of the total crop area has enjoyed the facility of irrigation. In the remaining area, crops are grown under unirrigated conditions and as such depend on rainfall. For Scientific crop planning and operation adjustment, it is necessary to have a complete knowledge of the various aspects of precipitation. In order to get a complete information on this aspect, suitable statistical methodology is required. At a particular time and space, the occurrence of a given amount of rainfall is beset with uncertainty. So, the statement about it, can only be made in terms of probabilities. Some efforts are already under way in our country. This study is a step in that direction. The present study has provided some quantitative estimates of parameters and tested the relevant hypotheses.

To develop statistical methodology for studying the behaviour of rainfall in this region, the annual rainfall for Cooch Behar and Dinhata stations spreading over 88 years and daily rainfall data for 18 years for Cooch Behar station only have been utilized in this study. The yield data of winter rice for Cooch Behar district are also used in the present study.

The important results from this investigation have been presented here one by one.

The years of large scale deficient and excess rainfall have been identified in this study. The year 1920 was the worst flood year and 1930 appeared to bethe most drought year,in meteorology, in the district of Cooch Behar. The occurrence of droughts and floods, in meteorology,has appeared to be a random phenomenon.

However, the probability distribution of interarrival times between the successive years of drought and the probability distribution of interarrival times between the successive years of flood are considered to follow negative binomial distribution. It is also indicated that the probability distribution of interarrival times for drought and flood years at the two places can be considered as drawn from the same negative binomial population. Droughts and floods may be expected to occur once in 6 years and 8 years respectively and droughts may appear more frequently than floods in this region.

An attempt has been made to study the repetitive behaviour of the weather of rainy months by using the Markov-dependent geometric model. Two-State Markov chain model is fitted to the daily rainfall data for Cooch Behar. The present study has investigated the empirical validity of Markov-dependent geometrical models for wet spells, dry spells and weather cycles of Cooch Behar. The average expected lengths of dry spell and wet spell are 2.5 days and 3 days respectively and they constitute the 5.5 days of weather cycle which is, as such, the average length of observed weather cycle.

The present study has also provided the daily weather pattern of Cooch Behar during the monsoon months. One-step (5×5) Markov chain model is fitted to the daily rainfall data at Cooch Behar. The degree of uncertainty has been measured with the help of information theory. It is also revealed that the test based on entropy is most appropriate rather than the redundancy test which is also based on entropy,against the hypothesis of Marknovian dependence.

The present study has also investigated the spectrum of annual precipitation of two stations in Cooch Behar district and it is found that there are evidences of various band-limited signals. The first signal is quasi-biennial (2.0 to 2.6 years) signal. The second signal has a period near quasi-triennial (3.1 to 3.5 years). The third signal has a period near quasi-five year oscillations. The fourth signal has a period of 9 years. The fifth signal has a period of 17.5 years which is nearly in phase with lunisolar cycle. So,it is not surprising that the time series of meteorological data displays a great deal of variability from year to year. In order to detect the presence of hidden periodicities in the annual precipitations we have found by power spectral analysis,the evidence of highly statistically significant values of the five signals.

Since a regional approach has been considered, a linear multiple regression equation has been developed to predict the yield of winter rice in Cooch Behar district. The regression equation is appropriate if different physical mechanisms are involved in the nature of daily precipitation. The yield of winter rice in Cooch Behar district has been considered as the dependent variable and the derived categories of daily rainfall have been considered as explanatory variables. The multiple regression equation has been fitted to establish the crop-weather relationship. The yield regressed equation is verified in all intermediate states from fourteen explanatory variables to the final twelve-term equation. The combination of twelve explanatory variables has been selected for which these variables are appeared physically and statistically realistic. The coefficient of determination indicates that more than 93 percent of the total variation in yield is explained by the explanatory variables. It is,also noted that the values of the yield of winter rice are independent of the earlier values.

9.2. FUTURE SCOPE OF RESEARCH WORK.

There are several bottlenecks in the extension of this study, the first and foremost is advanced mathematical and statistical techniques.

Though the scope of the study is restricted to Cooch Behar district,the results are valid for many areas having similar agroclimatic condition as well as meteorological sub-division. Moreover,meteorological behaviour has limits of accuracy defined by physical nature of the region and also the rainfall amount can exhibit substantial variability over very small distance. A more sophisticated approach such as the probability of precipitation due to ariel variability needs to be examined towards facilitating the principal types of local adjustment. This apparent incoherent spatial and temporal behaviour of precipitation in the region of study may have complicated the theoretical development of its causal mechanisms. However,it is a matter of investigation. The study needs to be replicated under different climatic conditions. And the appropriate statistical techniques would be an important field of research work.

Higher order entropy may be introduced to test the stochastic dependence by the theoretical development and this is one of the fields of absolutely new informational measure on the probabilistic system.

The knowledge of amount of rainfall and its behaviour is not necessarily sufficient for crop planning and operational adjustment. For some region, measures of other climatic variables like soil moisture,temperature,humidity,sunshine,evaporation and average wind speed etc. are also essential to consider for better establishment of crop-weather relationship. Moreover,considering their interactions, a multiplicative model is to be needed for the better understanding of crop-weather relationship and it would be one important field of future research work for various agroclimatic regions.

B I B L I O G R A P Y

BIBLIOGRAPHY

Agnihotri.Y, Bansal.R.C. and Singh.P.(1984) "Spell distribution and weather cycle at Chandigarh" Mausam 35 (1),99-102

Agarwal.R,Jain R.C. and Jha.M.P. (1983). "Joint effect of weather variable on rice yield". Mausam 34 (2). 189-194.

Agarwal.R,Jain R.C. and Jha M.P. (1986) "Models for standing rice crop-weather relationship". Mausam 37 (1). 67-70.

Aneja.D.R. and Srivastava.O.P.(1986) "A study technique for rainfall pattern" . Aligarh.J.Statist. 6. 26-31.

Barndroff.N and Yeo.G.F. (1969). "Negative binomial processes". J.Appl.Prob. 6.633-647.

Basawa.I.V. and Prakas Rao.B.L.S (1980) "Statistical inference for stochastic processes",Academic Press. London.

Basu.G.C. (1988) " A study of monsoon daily rainfall at Maithon by Markovian model and information theory". Mausam. 39 (1). 83-86.

Bhalme.H.N.,Mooley.D.A and Jadav.S.K. (1984) "Large-scale pressure index connected with the southern oscillation and its potential for prediction large-scale droughts over India". Mausam. 35 (3). 335-360.

Bhargava.P.N.,Narain.P.,Singh.D. and Saksena. A.(1977) "Monograph on statistical studies on the behaviour of rainfall in a region in relation to a crop". I.A.R.S.(ICAR) New Delhi.

Biswas.B and Bhadram.C.V.V. (1984) "A study of major rainstorms of the Teesta basin". Mausam. 35(2) 187-190.

Box G.E.P and Jenkins G.M.(1970) " Time series Analysis forecasting and control" Holden day. San Franshisco.

Chatfield C. (1982) "The Analysis of Time Series: An Introduction".Chapman and Hall,London.

Chowdhury.A. and Abhyankar.V.P. (1984) " On some climatological aspects of drought in India ". Mausam 35 (3). 375-378.

Chowdhury.A. and Abhyankar.V.P. (1984) " A Markov chain model for the probability of drought incidence in India". Mausam. 35(3).403-405.

Chowdhury.A.,Dandekar.M.M. and Raut.P.S. (1989). "Variability in drought incidence over India - a statistical approach". Mausam. 40(2). 207-214.

Cochran W.G. (1938) "An extension of Gold's method of examining the apparent persistence of one type of weather". Quart.J.R.Met.Soc. 64(227). 631-634.

Cooper.R.A. and Weekes.A.J. (1988). "Data,models and statistical analysis". Heritage publishing Co. New Delhi.

Currie.R.G. and O'Brien.D.P. (1988) "Periodic 18.6 year and cyclic 10 to 11 year signals in north-eastern United states precipitation data". J.Climatol. 8 255-281.

Currie R.G. and O'Brien.D.P.(1990). "Deterministic signals in precipitation records from the American corn belt." Int.J.climatol. 10. 179-189.

Dhar.O.N. and Changrancey.T.G.(1966). "A study of meteorological situations associated with major floods in Assam during the monsoon months." Indian.J.Met.Geo.Phys. 17. 111-118.

Domenico.P.A.(1972). "Concept and model in ground water hydrology". Mc Grow -Hill. New york.

Elizabeth.E.,Khalil.S.E.S.,Nicholds.N.,Abdalla.A.A. and Ryd Jeski.D (1988). "Changing rainfall patterns in Western Sudan". J.Clamatol.8. 45-53.

Fong Chao.B. (1990) . "On the use of maximum entropy/Autoregressive spectrum in harmonic analysis of time series ". Pageoph. 134 (2). 303-311.

Fuller.W.A. (1976) "Introduction to Statistical time series". Wiley, New York.

Gabriel.K.R. and Neumann.J. (1957). "On distribution of weather cycles by length". Q.J.R.Met.Soc. 83. 375-380.

Gabriel.K.R. and Neumann.J. (1962) "A Markov chain model of daily rainfall occurrence at Tel-Aviv". Q.J.R.Met.Soc. 88.90-95.

Ganesan.G.S. and Rao.H.R.A. (1986) "A comparative study of rainfall spells in Bangalore". Mausam. 37(2). 223-230.

Government of India. (1976) "Climate and Agriculture " Report of the National Commission of Agriculture. Part-IV. Ministry of Agri.Irri. New Delhi.

Graham.F. (1988) "Seasonal forecasting of the Kenya Coast short rains". Int.J.Climato 1.8. 489-497.

Gregory.S. (1979) "The definition of wet and dry periods for discrete regional units". Weather 34. 363-369.

Helmut.P (1986) "A note on time series analysis of yearly temperature data". J.Statist.Soc. 149. 174-185.

Jain.R.C., Agarwal.R. and Jha.M.P.(1980) "Effect of climatic variables on rice yield and its forecast". Mausam . 31(4). 591-596.

Jenkins.G.M. and Watts.D.G. (1968) "Spectral analysis and its applications". Holden day,SanFransisco.

Kendall.M. and Stuart.A. (1979) "The advanced theory of statistics". Vol.I & II Charles Griffin & Co.Ltd. London.

Kendall.M.Stuart.A.and Ord.J.K.(1983) "The advanced theory of statistics".Vol.III Charles Griffin & Co.Ltd. London.

Koutsoyiannis.K.(1979) "Theory of Econometrics." The Macmillan Press. New Delhi.

Lakshmanaswamy B.and Jindal G.P. (1990)". "Statistical analysis of the variability of area weighted rainfall over India". Mausam 41(4). 569-574.

Ljung.G.M and Box G.E.P.(1978). "The Likelihood function of stationary autoregressive moving average models." Biometrika 66. 265-270.

Mohendra Dev.S.(1987) "Growth and instability in food grain production-An inter-state analysis." Economic & Pol.Weekly 39 (XXII) A-82-92.

Mallik.A.K.(1958) "Proc.symp.on met. and Hydrol. aspects of floods and droughts in India ". I.M.D.New Delhi. 65-70.

Mavi.H.S.(1986) "An introduction to agrometeorology." Oxford & IBH Pub. Co. Calcutta.

Mc Quigg. J.D. (1975) "Economic impacts of weather variability". University of Missouri-Columbia,Department of Atmosphere Science.

Medhi J.(1976) "A Markov chain model for the occurrences of dry and wet days." In.J.Met.Hydrol.Geophys.27. 431-435.

Mitra.K.,Mukherji,S.and Dutta.S.N.(1991) "Some indication of 18.6 years luni-solar and 10-11 year solar cycles rainfall in North West India . The Plains of Uttarpradesh and North Central India". Int.J.Climatol. II.645-652.

Mood A.M. Graybill F.A. and Boes D.C.(1974) " Introduction to the theory of Statistics ". Mc Graw hill International Co., Singapore.

Mooley.D.A. and Parthasarathi.D.(1984) "Fluctuation in All India summer monsoon rainfall during 1971-1978". Climatic Change. 6 .287-301.

Mongia. A .D. and Gajja B.L. (1986) "Influence of rainfall on productivity performance of in Andaman and Nikobar Islands." Mausam.37(4) 542-543.

Mukhopadhyay. S. K. (1992a) "Deficiencies and excesses of annual rainfall over Cooch Behar, West Bengal." Envior. and Ecol. 10(1). 185-187.

Mukhopadhyay.S.K.(1992b) "Distribution of inter arrival time for dry and wet years". Envior. and Ecol. 10(2) 424-426.

Mukhopadhyay.S.K. (1992c) " Entropic measure of Markovian model in rainfall ". Envior. and Ecol. 10(2) 448-450.

Ogallo. L.. (1984) " Temporal fluctuations of seasonal rainfall patterns in East Africa". Mausam. 35(2). 175-180.

Ogallo. L. A. J. (1986) " Stochastic modeling of regional annual rainfall anomalies in East Africa". J.Appl.Statist. 13(1). 49-56.

Olapido E.O. (1989) "Non-integer (non-harmonic)Spectral analysis of a drought index for the America Great Plains."

Mausam.40(1). 19-28.

Pathack. B. M. (1982) " An analysis of Mauritian winter rainfall". Mausam. 33(3). 361-372.

Priestley. M. B. (1981) "Spectral analysis and time series." Vol.I & II Academic Press. London.

Prasad. R and Dudhane. S. N.(1989) "Forecasting rice yields in Gangetic West Bengal using rainfall and agricultural Technology". Mausam.40(4),441 -446.

Prsad.S and Kashyap.R .L.(1990) "Spectral analysis in one and two dimensions." Oxford & IBH Pub.Co.Calcutta.

Prasad. S.K.and Ram.L.C.(1990) "Long term variations of the rainfall at Jalpaiguri in North Bengal".Mausam.43(3) 341-342.

Ramdas. L. A.(1967)." Weather & crops." Handbook of Agriculture. ICAR. New Delhi.

Rao.C. R. (1984) " Linear Statistical Inference and Its Applications." Wiley Eastern Ltd.,New Delhi.

Rao.K.N. and Jagannath. P.(1963) Proc. UNESCO-WMO. Rome. Symposium on change of climate. UNESCO Paris,XX. 53-66.

Rao.K. N. George, C. J. Morey.P.B. and Mehta N.K.(1973) "Spectral analysis of drought index (Palmer) for India". Ind.J.Met.Hydol.Geophys. 24.257-270.

Sarkar. R. P.(1979) " Droughts in India and their predictability Proc. International Symposium of Hydrological aspects of droughts" I.I.T. New Delhi. 33-40.

Sundararaj. N. and Ramchandra.S.(1973) "Markov dependent geometric models for weather spells and weather cycles-A study". Ind.J.Mat. Hydrol.Geophys. 26(2). 221-226.

Thail H.(1971). " Principles of Econometrics " Jhon. Wiley & Sons.London.

Trendberth. K. E.(1976). " Spatial and temporal variations of the southern oscillation" **Quart.J.R.Met.Soc. 101** 55-74.

Walker. G.T. (1914)." Memorites" Part-3. India. Met, Department, Calcutta.

Walker. G. T.(1924) " Memoites " Part. 9, India Met. Department, Calcutta.

World Meteorological Organization (1966) " Climatic change" Tech.Note. No. 79.Geneva.

World Meteorological Organization (1977) " Crop-weather models and their use in yield assessment" Tech.Note.No.151. Geneva.

****

APPENDIX

NATURE OF DAILY RAINFALL VARIABLES (EXPLANATORY VARIABLES)

NATURE OF DAILY RAINFALL VARIABLES AND TIME-SCALE VARIABLE (EXPLANATORY) COOCH BEHAR

YEAR	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1972	15	14	13	13	13	10	15	12	12	2	20	19	19	1
1973	16	15	12	15	13	15	13	11	13	6	20	16	20	2
1974	23	18	23	19	13	17	20	15	19	3	20	18	17	3
1975	26	18	23	14	12	12	20	17	15	5	17	14	17	4
1976	18	13	16	15	10	11	9	9	9	5	24	17	18	5
1977	23	20	19	16	11	14	12	10	10	6	16	13	13	6
1978	18	13	16	10	10	8	15	12	15	2	11	11	10	7
1979	22	17	18	14	10	14	13	10	11	10	11	11	9	8
1980	23	19	20	17	14	16	12	12	11	2	15	14	13	9
1981	21	14	18	16	13	16	19	19	17	0	14	11	13	10
1982	24	16	23	17	17	14	15	13	13	0	23	21	20	11
1983	20	15	15	13	11	10	19	16	16	3	16	14	15	12
1984	25	16	24	15	13	14	18	13	16	5	13	10	12	13
1985	23	12	22	12	9	11	19	14	14	2	20	14	18	14
1986	20	19	19	12	10	10	17	14	16	8	22	18	17	15
1987	27	20	24	22	7	18	17	16	15	2	19	13	18	16
1988	20	17	17	23	10	23	17	12	16	3	11	8	10	17

	RD	RRD	CRD
July	X_1	X_2	X_3
August	X_4	X_5	X_6
September	X_7	X_8	X_9
October	X_{10}	–	–
June	X_{11}	X_{12}	X_{13}

X_{14} = Time-scale variable from 1 to 17