

## Role of Statistics in Monitoring Rainfall

Harsha S

Research Scholar Department of PG Studies and Research in Statistics Mangalore University, Karnataka, India

---

**Abstract:** Rainfall is one of the climatological data which is widely analyzed for a long time. The present study focuses on the analysis of rainfall data in Mangalore, a coastal city of the state of Karnataka, India for the years 1970 to 2011. In addition to the classical procedures for the analysis of rainfall data, Run test and Kolmogorov-Smirnov two sample tests is innovatively used to study the random fluctuations in the yearly rainfall data and to study whether there is climate changes in the recent years. Poisson distribution is fitted to the number of rainy days for each month of the rainy season across the years and fits well to the data except for the month of July. Run test is used to study whether the number of dry and wet spells is same across the years and is a new application of the Run test.

**Keywords:** Kolmogorov-Smirnov two sample test, Rain fall, Run test, Poisson distribution.

---

### I. Introduction

Rainfall is one of the climatological data which is widely analyzed for a long time. Analysis of rainfall data is important as it facilitates policy decisions regarding the cropping pattern, sowing date, construction of roads and providing drinking water to urban and rural areas. The papers on the analysis of monthly and yearly rainfall data across the regions and the study on the pattern of dry and wet spells are vast. [1] used gamma and Poisson-gamma distribution for fitting the monthly rainfall data for 99 rain gauge stations in Catalonia. Further they used principal component analysis to group the stations into different hydrological and climatological regions. [2] studied the relationship between rainfall extremes and mean annual precipitation. Using Monte-Carlo simulation a model was developed for predicting storms in northern central Italy. [3] addressed the statistical analysis of Indiana rainfall data. The objectives of the study is to test the goodness of fit of different probability distributions. They quantify the variability in rainfall intensity estimates. Further they discussed at length the statistical procedures for the analysis of daily, monthly and yearly rainfall data. The monograph gives a vivid description of the statistical procedures for meteorological and climatological analysis and illustrated various graphical procedures for the presentation of data. [4] presented a theory of independent and identically distributed random cascades. Three sets of radar-derived rainfall data in space and time are analyzed to estimate the probability distribution of the generator.

Another aspect of rainfall data which received the attention of many researches is the analysis of dry and wet spells. [5] gives a detailed report on rainfall events and inter event spell properties in the cities Adelaide and Melbourne, Australia. For both cities mean, standard deviation, skewness coefficient, kurtosis coefficient were calculated and six theoretical distribution have been fitted to the observed spells. [6] presented a research report on features of wet and dry spells and their extremes across India. [7] did a study to provide an outlook for the occurrence of wet and dry spells in reference to the Australian monsoon.

The present study focuses on the analysis of rainfall data in Mangalore, a coastal city of the state of Karnataka, India for the years 1970 to 2011. In this study, in addition to the classical procedures for the analysis of rainfall data, Run test and Kolmogorov-Smirnov two sample test is innovatively used to study the random fluctuations in the yearly rainfall data and to study whether there is climate changes in the recent years. Poisson distribution is fitted to the number of rainy days for each month of the rainy season across the years and fits well to the data except for the month of July. Run test is used to study whether the number of dry and wet spells are same across the years and is a new application of the Run test. Following are the specific objectives of the study.

- 1) Time series analysis of weekly rainfall data across 42 years. The focus of this analysis is to study the decadal changes in the average weekly rainfall and the variability in the monthly rainfall. Nonparametric regression is used to identify the seasonal effect and cyclical changes.
- 2) To study the changes in rainfall pattern (if any). This objective becomes important in the light of global warming and its effect on the annual rainfall.
- 3) To develop models for the prediction of weekly rainfall and the number of rainy days. Multiple regression for the number of rainy days is used to develop a predictive model.

The report is organized in the following manner. Section 2 describes the various statistical procedures used for the analysis of rainfall data along with a brief description of the data set. The results are presented in section 3 and section 4 provides a discussion of the results. Concluding remarks are presented in section 5.

## II. Statistical Methodology

This section provides a description of the statistical procedures used for the analysis. Depending on the objectives the procedures can be classified into several groups and the following description is arranged accordingly.

### 2.1 Rainfall pattern

The rainfall pattern for the 42 years can be analyzed in terms of weekly and monthly rainfall. In the present analysis rainy season is divided into standard weeks (not the standard weeks as defined by the department of Meteorology). In Mangalore city the onset of monsoon is during the first week of June. Thus June 1 to June 7 is considered as the first standard week. Every standard week following this consists of 7 days. The last standard week comprises the dates from 12 to 18<sup>th</sup> October. The reason to call these weeks as standard weeks is that some of the standard weeks consists of the dates from the preceding month and the succeeding month of a year and does not necessarily refer to a particular month.

To study the changes in the average weekly rainfall across the years, the years are classified into 4 decades namely 1970-79, 1980-89, 1990-99 and 2000-2011. The last decade consists of 12 years; the years 2010 and 2011 do not fall in the decade 2000-2009 and it was decided to include these 2 years in the analysis rather than discarding them. For each decade the two way analysis of variance is carried out to test the hypothesis of equality of mean weekly rainfall; Two way ANOVA is used to eliminate variation between the weeks. ANOVA F-test is robust against the violation of normality assumption ([8]). The non parametric analogue of the two-way analysis of variance is the Friedman's test ([9]). The same tests were carried out for the number of rainy days (a day with total rainfall is more than 2.5mm). The number of rainy days for each month across the years follows Poisson distribution and square root transformation is used for the calculation of ANOVA F-test ([3]). Since the standard text books like those quoted previously provide a good description of these procedures and to save space we have omitted the details.

An important measure in the analysis of rainfall data is to study the variability in the rainfall. For the study of variability in rainfall, the appropriate duration is a month. Monthly variability in rainfall is estimated using standard deviation and decadal changes in the monthly variances of the rainfall is examined using Bartlett's test ([10]).

### 2.2 Seasonal and Cyclical Fluctuations in Rainfall

Seasonal and cyclical fluctuations of rainfall data provides useful information regarding the effect of a season and the cycles that exist in the time series of rainfall. If a particular area is drought prone cycles indicates the duration for the successive occurrence of droughts. The traditional approach to calculate seasonal indices and cyclical fluctuations is the method of moving averages. In the present study week is considered as a season and for studying the seasonal fluctuations, moving average of 20 week duration across 42 years is used. To study the cyclical fluctuations yearly data were used. For this purpose 5 and 7 years moving averages were computed.

In the recent years non-parametric regression is used in the exploratory data analysis. In this approach rather than using a particular regression model, one can use a smooth function of the predictors to link it to the response variable. For a single predictor the non-parametric regression equation is given by  $Y_i = m(X_i) + \epsilon_i$ ,  $i=1,2,\dots,n$ , where  $Y_i$  and  $X_i$  refers to the response variable and predictor variable respectively and  $\epsilon_i$ 's are independent and identically distributed  $N(0,1)$ .  $m(X_i)$  is a smooth function of  $X_i$ . In the present analysis we have used non-parametric regression to study the cyclical pattern. In the literature several kernel functions  $m(X_i)$  are proposed. The software R uses local polynomial to estimate  $m(X_i)$ . In our analysis for each week non-parametric regression is fitted for 42 years along with the total rainfall for the years. This is an innovative application of the non-parametric regression to study the cyclical fluctuations and to the best of our knowledge has not been used in the past. A good source of reference for non-parametric regression is [11].

### 2.3 Changes in Rainfall Pattern

Since the last decade global warming has received the attention of many scientists ([12]). It is believed that global warming has influence on rainfall and causes change in the rainfall pattern and there by the climate. We have tried to verify this hypothesis. If there is no climate change, the fluctuations in the yearly rainfall should be the random fluctuations. Any deviation from the random fluctuations indicates a systematic change in the yearly rainfall and one of the causes for this may be the change in the climate including global warming. To test the hypothesis of randomness, the data for 42 years were segmented into two parts, each part consisting of 21 years. Run test is used to test for the randomness. [9] suggests two types of Run test namely "Runs up" and "Runs down". In the Runs up test "positive" sign is assigned if the current observation is less than the succeeding value; otherwise "negative" sign is assigned. The total number of Runs is the test statistic for testing the hypothesis of randomness. This is a pioneering attempt to use the Runs test for studying the climate change.

Another way to study the change in the weekly rainfall is the use of Kolmogorov-Smirnov two sample test ([9]). For this purpose, the each week average rainfall for the years 1970-2006 were computed. This average weekly rainfall was next compared with a weekly rainfall for the next five years (2007-2011), comparison being made with each individual year. In the Kolmogorov-Smirnov two sample test the null hypothesis is  $H_0: F(x) = G(x)$  where  $F(\cdot)$  and  $G(\cdot)$  refer to the cumulative distribution function of the two populations under consideration. If the hypothesis is accepted, it means that the two distributions are the same. In the present case, if the hypothesis is accepted, it amounts to the equality of probability distribution of weekly rainfall for the year under consideration and the average.

Cluster analysis ([13]) is also used to see the change in the rainfall pattern(if any). Generally the conjecture of the scientists is that the global warming would induce a change in rainfall in the recent years. Cluster analysis groups the rainfall data in such a way that within each group the years have the same rainfall pattern and between the groups there is a large difference in the rainfall pattern. If the recent years are clustered in a single cluster, it would support the presumption of the scientists; otherwise it indicates that there is no change in the rainfall pattern. There are various procedures for the formation of clusters and for more details refers [13].

Another important aspect in the analysis of rainfall data is the study of wet and dry spells. A wet spell is defined as a sequence of wet days preceded and succeeded by a dry day. In a similar manner one can define a dry spell also. The change in the total number of wet and dry spells and the length of the longest wet spell (or dry spell) speaks about the changes in the rainfall pattern. A wet or dry spell is similar to a Run. In the present analysis Z-test is used to test for equality of total number of wet and dry spells during the rainy season of any two years. The test statistic is given by

$$Z = \frac{[(R_1 - R_2) - (E(R_1) - E(R_2))]}{\sqrt{[V(R_1) + V(R_2)]}}$$

where  $R_1$  and  $R_2$  refers to total number of dry and wet spells in the two years under consideration and  $E(\cdot)$  and  $V(\cdot)$  refer to mean and variance. Under the null hypothesis  $Z$  has an asymptotic standard normal distribution. Expression for mean and variance are available in [9] and not reported here.

### 2.4 Predictive Models

Multiple regression is used to develop a predictive model for the weekly rainfall. For the current year, the average weekly rainfall of the preceding years and rainfall of the preceding month are considered as predictors.

## III. Results

### 3.1 Characteristics of the Rainfall Data

As mentioned in section 1 the rainfall data corresponds to the city of Mangalore for the years 1970 to 2011. Mangalore has four rain gauge stations and the present data refers to the rain gauge at the premises of the District Commissioner's office.

The analysis starts with the box plots of the weekly rainfall for all the years. Fig3.1 (a, b, and c) presents the box plots for the starting year of the series namely 1970, the middle year (1990) and the most recent year (2011).

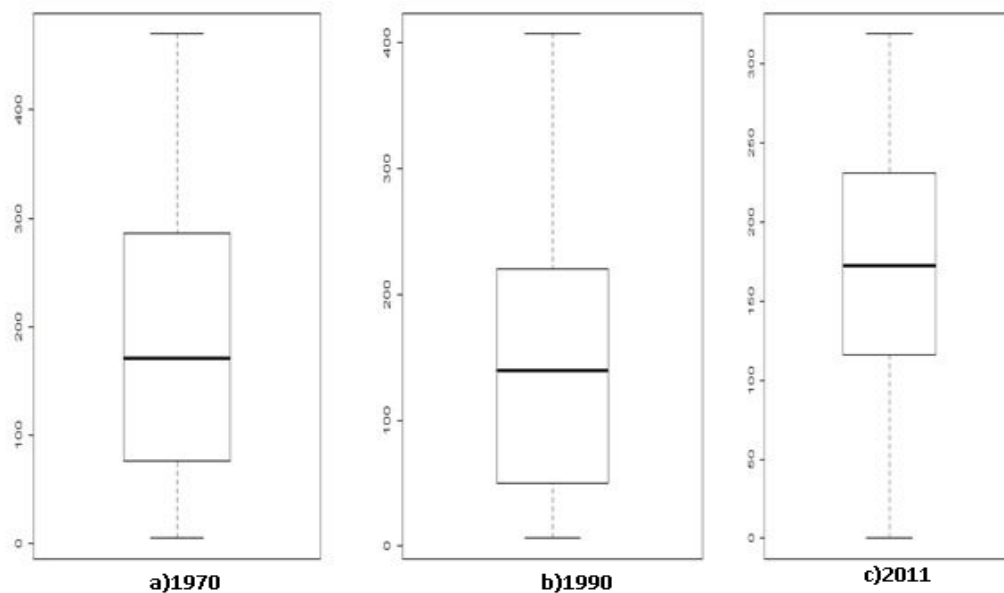


Figure 3.1: box plots for the weekly rainfall for the years 1970, 1990 and 2011

From the box plot presented in Fig3.1, it follows that for 1970 and 1990 the median rainfall is same. For 2011 the median rainfall is slightly high. To identify the distribution of weekly rainfall we have computed coefficient of skewness  $\beta_1$  (Pearson's coefficient of skewness) and coefficient of kurtosis ( $\beta_2-3$ ) (Pearson's coefficient of kurtosis ). Table 3.1 presents the values of  $\beta_1$  and  $\beta_2-3$  for each year using the daily rainfall data.

**Table3.1:** The values of coefficient of skewness  $\beta_1$  and coefficient of kurtosis  $\beta_2-3$  for the years 1970-2012

Year	Skewness( $\beta_1$ )	Kurtosis( $\beta_2-3$ )	Year	Skewness( $\beta_1$ )	Kurtosis ( $\beta_2-3$ )
1970	0.44	-1.08	1991	1.68	3.39
1971	0.90	-0.69	1992	0.51	-1.21
1972	1.66	1.93	1993	1.17	0.29
1973	0.30	-0.98	1994	0.73	-0.77
1974	1.07	0.43	1995	0.72	-0.66
1975	0.53	-0.66	1996	0.65	-1.28
1976	0.30	-1.12	1997	0.85	-0.22
1977	1.39	1.32	1998	1.15	0.79
1978	0.20	-1.58	1999	0.79	-0.96
1979	0.81	-0.94	2000	1.15	0.23
1980	0.34	-1.15	2001	0.34	-1.26
1981	0.60	-0.92	2002	0.58	-0.80
1982	0.78	-0.87	2003	1.58	1.82
1983	0.92	-0.09	2004	1.14	0.55
1984	2.13	5.26	2005	0.99	0.14
1985	1.39	2.41	2006	1.20	0.85
1986	1.28	0.66	2007	0.65	-0.75
1987	1.77	3.19	2008	1.31	0.46
1988	1.15	0.78	2009	0.95	-0.06
1989	0.53	-0.52	2010	0.39	-0.85
1990	0.47	-0.66	2011	-0.23	-0.87

From the table it is clear that the rainfall follows positively skewed distribution for all the years, except for the year 2011 where it is negatively skewed with a value of  $\beta_1 = -0.23$ . For other years the value of  $\beta_1$  ranges from 0.20 (1978) to 2.13 (1984). The value of coefficient of kurtosis ( $\beta_2-3$ ) ranges from -1.58 (1978) to 5.26 (1984). For the years 1984, 1987 and 1991 the value of ( $\beta_2-3$ ) are 5.26, 3.19, and 3.39 respectively.

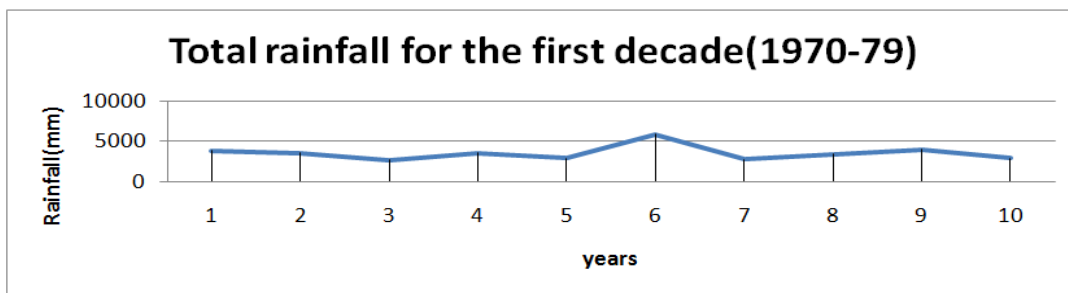
For the first week(Jun1-Jun7) value of  $\beta_1$  is equal to 0.64 and  $\beta_2-3$  is equal to -0.65. For the fifth week (Jun29-July5) the value of  $\beta_1$  and  $\beta_2-3$  are 0.52 and -0.75 respectively. Similarly the corresponding values for the 10<sup>th</sup> (Aug3-Aug9) and 20<sup>th</sup> (Oct12-Oct18) weeks are 0.71 and -0.38 & 2.72 and 9.18 respectively. From the table 3.1 and 3.2 it is clear that the distribution of weekly rainfall across the years are less skewed compare to the intra-yearly rainfall for the weeks. The value of  $\beta_1$  and  $\beta_2-3$  do not indicate a shift in the probability distribution for the weekly rainfall as there is no systematic pattern in the coefficient of skewness and kurtosis. We have also computed coefficient of skewness and kurtosis using the transformed observations  $\log Y$  and  $1/Y$  whenever the weekly rainfall is greater than zero. The transformation do not changes the value of  $\beta_1$  and  $\beta_2-3$  drastically and are not shown here. The overall conclusion is that the weekly rainfall distribution is neither normal nor lognormal.

**Table 3.2:** presents the value of coefficient of skewness  $\beta_1$  and coefficient of kurtosis ( $\beta_2-3$ ) for each week across the years.

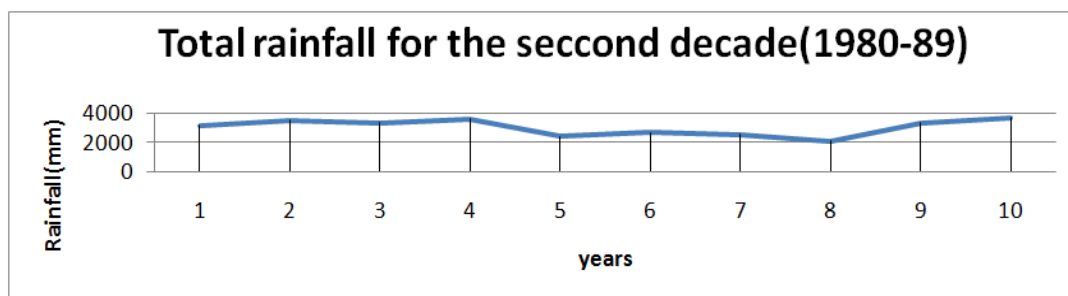
Week number	Skewness ( $\beta_1$ )	Kurtosis ( $\beta_2-3$ )	Week number	Skewness ( $\beta_1$ )	Kurtosis ( $\beta_2-3$ )
Jun1-Jun7	0.64	-0.65	Aug10-Aug16	0.85	-0.51
Jun8-Jun14	0.07	-1.17	Aug17-Aug23	0.55	-0.25
Jun15-Jun21	0.46	-0.75	Aug24-Aug30	0.42	-1.14
Jun22-Jun28	0.93	0.21	Aug31-Sep6	0.82	-0.57
Jun29-July5	0.52	-0.75	Sep7-Sep13	1.84	3.32
July6-July12	1.23	1.54	Sep14-Sep20	1.55	1.59
July13-July19	0.32	-0.61	Sep21-Sep27	1.66	3.04
July20-July26	1.002	0.77	Sep28-Oct4	1.33	2.29
July27-Aug2	0.73	-0.33	Oct5-Oct11	1.14	0.21
Aug3-Aug9	0.71	-0.38	Oct12-Oct18	2.72	9.18

### 3.2 Rainfall Pattern

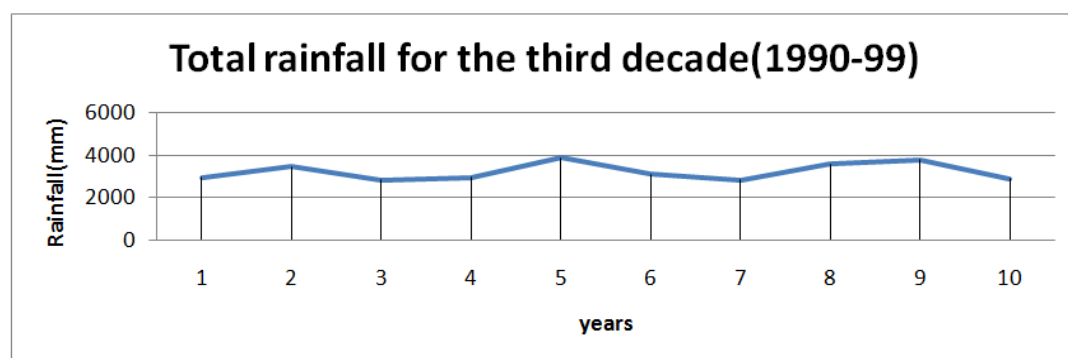
An attempt has also been made to determine whether there is any change in the total rainfall across the years. For this purpose two way analysis of variance F-statistic and Friedman's test statistic are computed for each decade. For the decade (1970-79) ANOVA F and Friedman's tests are significant. For the last decade 2000-11 ANOVA F was not significant but the Friedman's test was significant. For a time series data the significance may occur due to the fluctuations across the years and caution is needed to interpret the result.



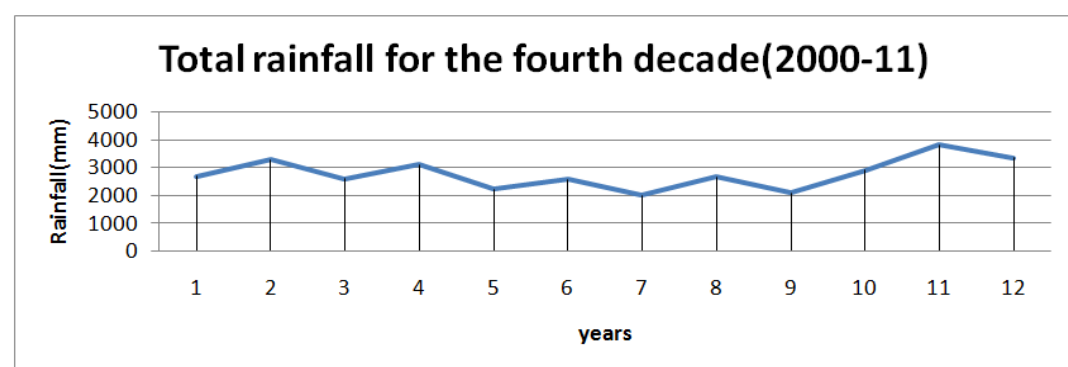
3.2(a)



3.2(b)



3.2(c)



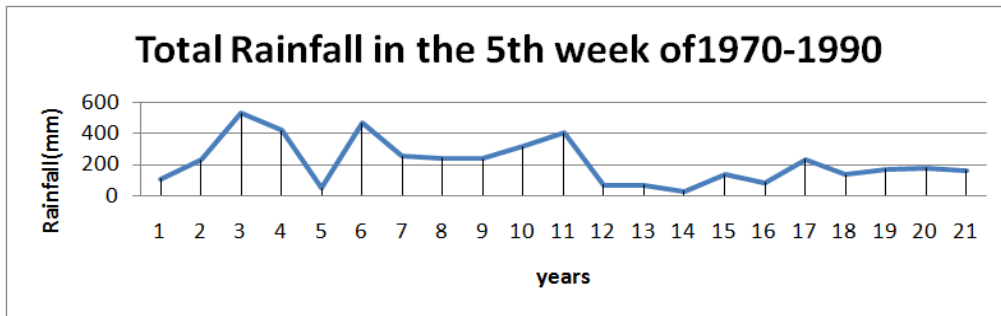
3.2(d)

**Figure 3.2:** total yearly rainfall for the year 1970-2011 classified as decades

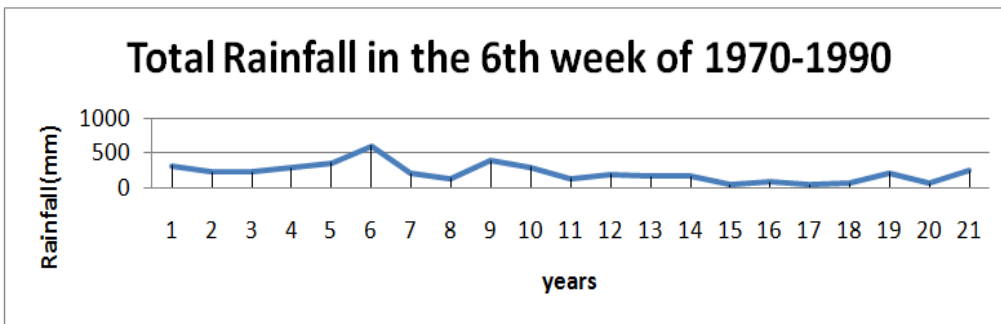
The Fig 3.2 explains partially the reasons for the significance of F and Friedman’s test in the decade 1970-79 and the significance of Friedman’s test for the decade 1980-89

For the number of rainy days in each month across the years, Poisson distribution was fitted for the months of June, July, August, September and October. Chi-square test for goodness of fit was carried out. Except for the month July the p values are not significant indicating that underlying distribution for the number of rainy days is Poisson. For the month of July mean and variance are 27.35 and 9.64 respectively indicating that the distribution of rainy days is under dispersed.

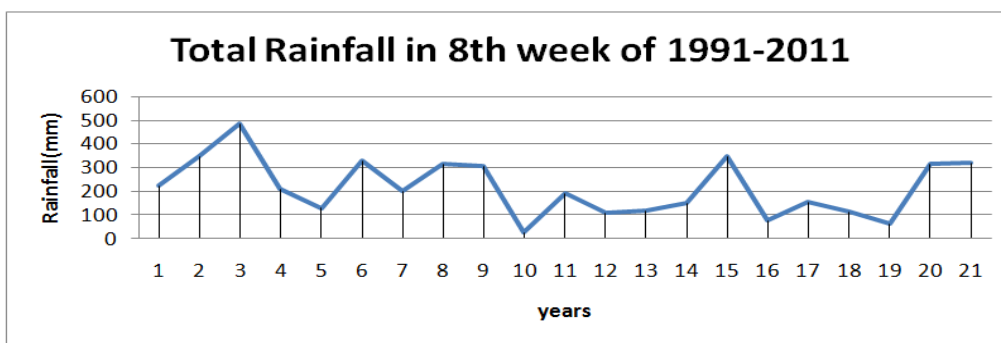
To check whether the fluctuations in rainfall are random, Run test was carried out for the two periods namely 1970-90 and 1991-11. The reason to combine the consecutive decades is that when the number of years is small, table values are not available to test the significance of the Run test and normal approximation is not satisfactory. During the period 1970-1990 for the 5<sup>th</sup> (Jun29-July5) and 6<sup>th</sup> (July6-July12) standard weeks the Run test was significant. The corresponding p-values are 0.041 and 0.043 respectively. The Fig3.3 (a and b) presents the weekly rainfall for these weeks across the years. For the period 1991-2011 the run test was significant for the 8<sup>th</sup> (July20-July26) standard week with p-value=0.044. The Fig 3.3(c) presents the weekly rainfall for this week across the years.



3.3(a)



3.3(b)

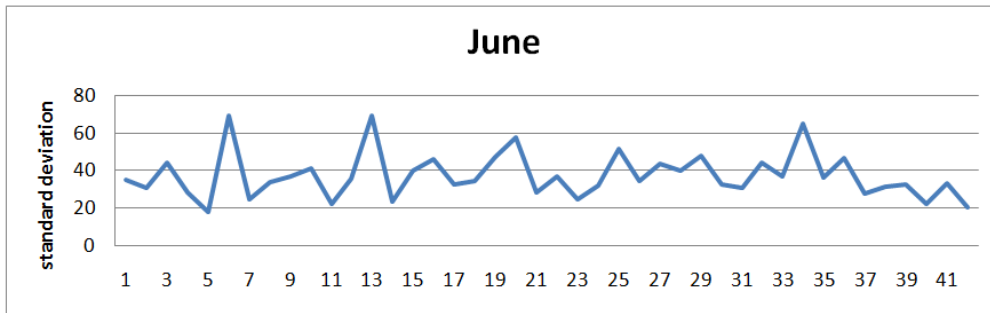


3.3(c)

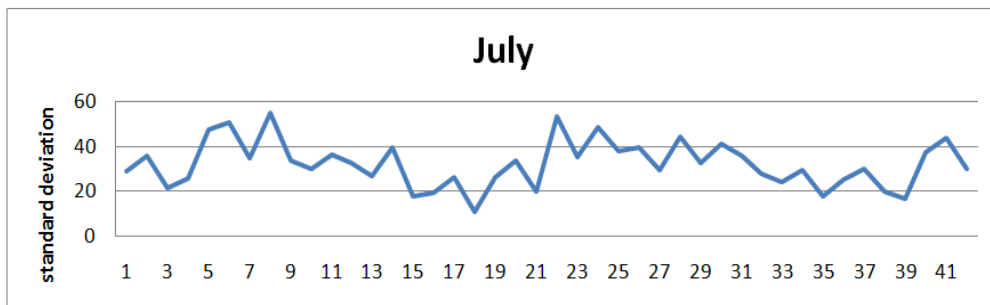
**Figure 3.3:** fifth and sixth weekly rainfall for the decade 1970-1990 & eighth weekly rainfall for the decade 1991-2011

The significance of the Run test indicates that the fluctuations in rainfall for these weeks across the years are not random, which indicates some climate changes. For the other weeks the fluctuations are random and therefore overall conclusion is that there is no influence of climate change on the rainfall.

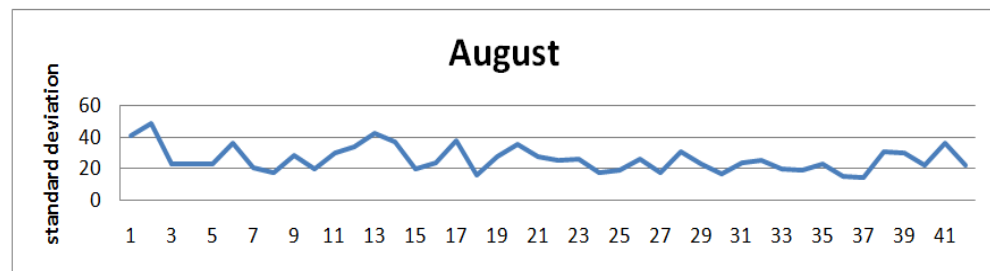
In the study of rainfall data, the amount of rainfall across weeks/months plays a crucial role for the agricultural activities. Another component which also plays a dominant role is the variability in the rainfall. If variability is large, even if the total rainfall is same, excess or deficit rainfall at the crucial periods of the farming operations may cause problems to the farmers. Thus an attempt is also made to study the variability in daily rainfall for the months June to the middle of the October. Fig3.4 (a, b, c, d and e) presents the standard deviation of the daily rainfall for these months across the years. Bartlett's test of homogeneity of variances indicates that the variability in daily rainfall remains the same across the years for each month.



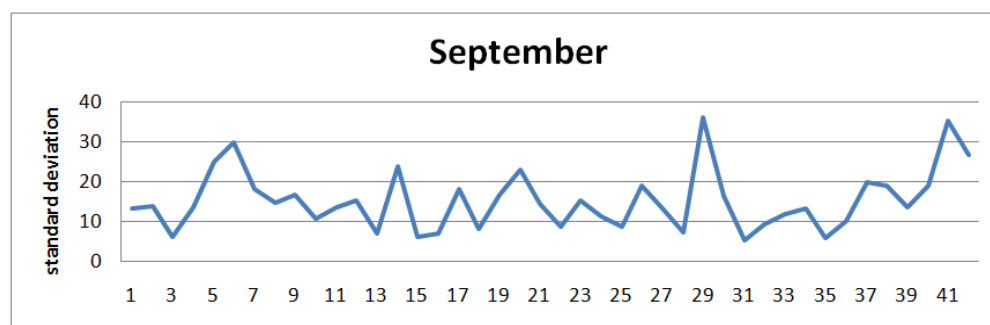
3.4(a)



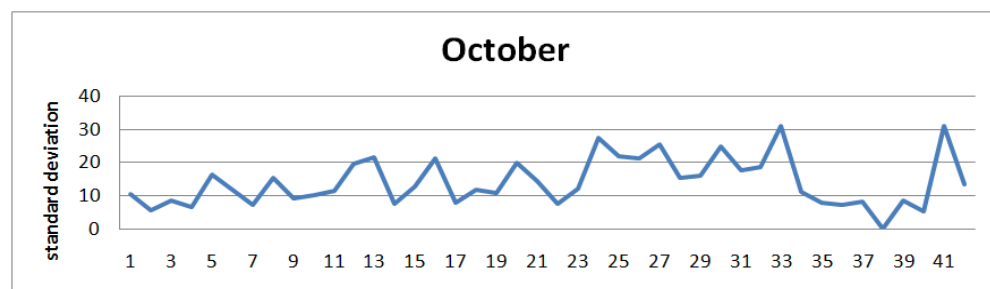
3.4(b)



3.4(c)



3.4(d)

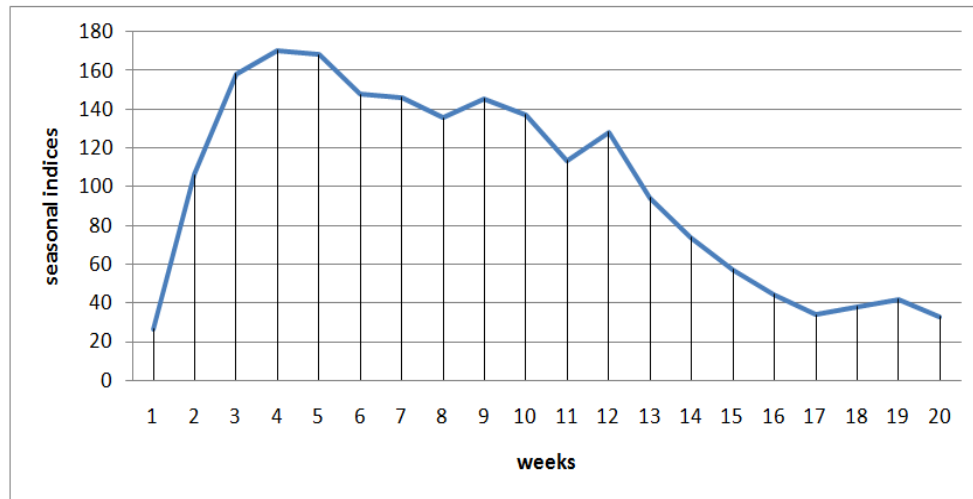


3.4(d)

Figure 3.4: standard deviation of daily rainfall for 5 months across the years

### 3.3 Seasonal and Cyclical Fluctuations in the Rainfall

In the analysis of rainfall series an important aspect of study is the seasonal and cyclical variations. In the present study, weeks constitute the seasons and using 20 point centered moving average, the seasonal indices have been computed and are presented in Fig3.5 For the standard weeks 2 (Jun 8-Jun 14) to 12(Aug 17-Aug 23) the indices are more than 100. These weeks may be considered as rainy weeks. The first standard week (Jun 1-Jun 7) and the other standard weeks 13 (Aug 24-Aug 30) to 20 (Oct 12-Oct 18) are considered as relatively sparse weeks. The period from June 8 to August 23 may be considered as the rainy period and any departure of rainfall during these days affects agricultural activities.

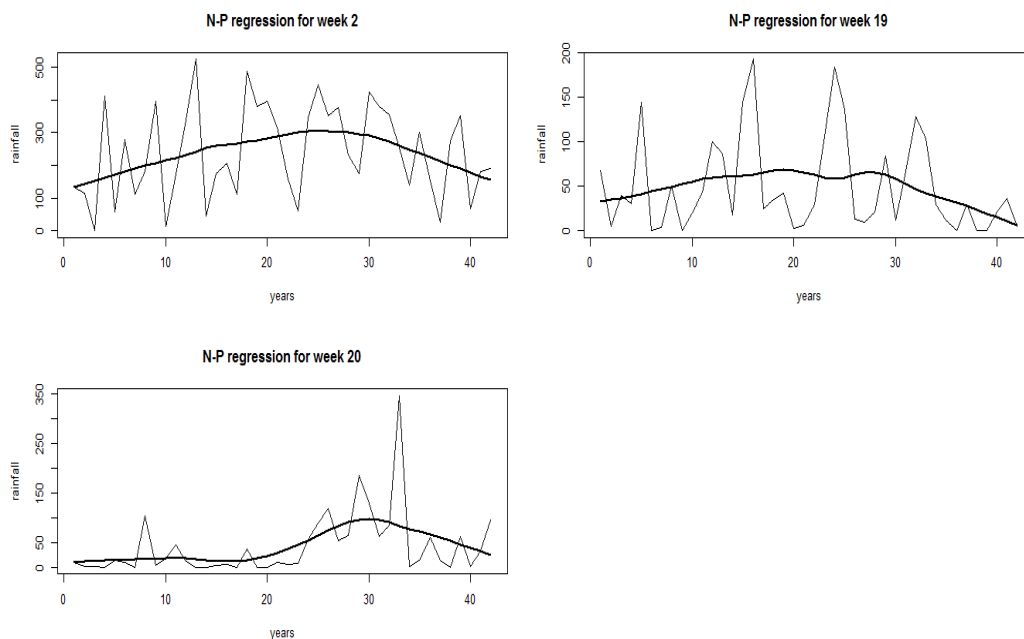


**Figure 3.5:** seasonal indices for the weekly rainfall data over period 1970-2011

For determining the cyclical pattern of total rainfall for the rainy season of 42 years, 5 point and 7 point moving averages are computed. An attempt is made to identify the cycle. From the figures indicating the ratio of the observation to the moving average it was difficult to identify any cyclical fluctuations and therefore it is not presented here.

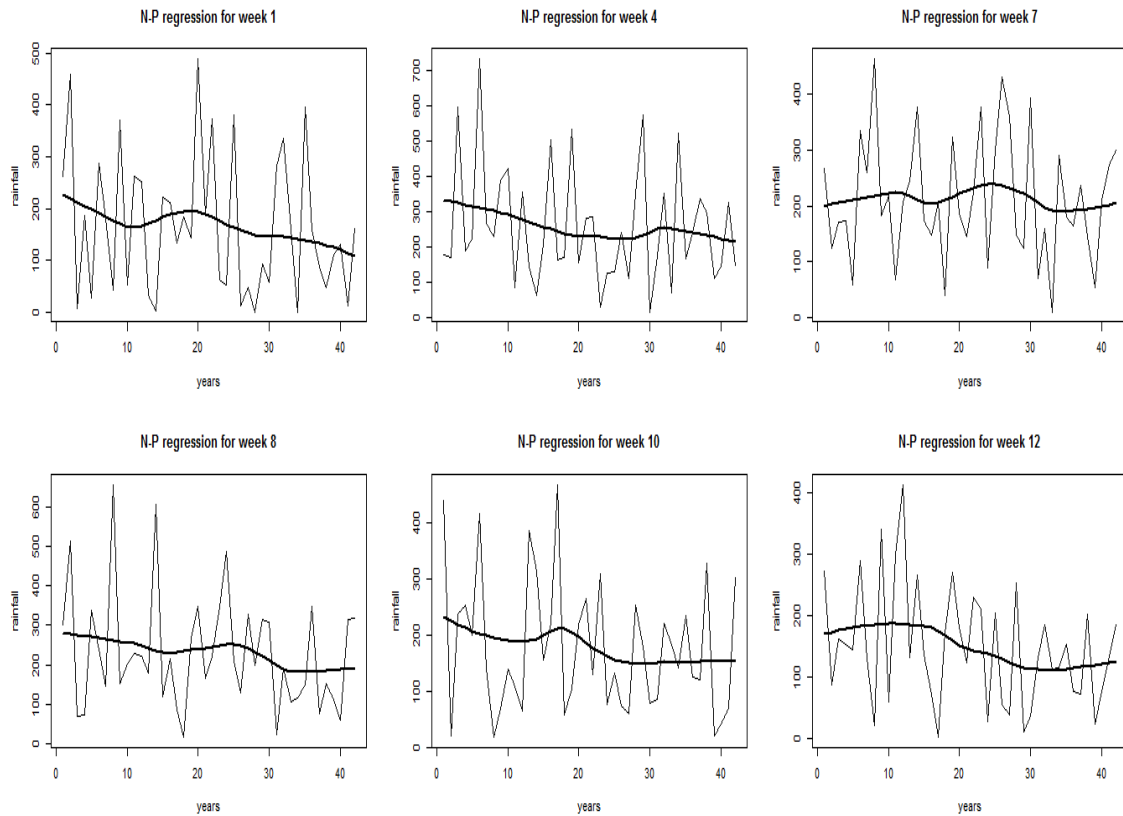
Nonparametric regression is an effective tool to estimate a sequence of observations through a smooth function. In this project, nonparametric regression is used to estimate trend and cyclical fluctuations in the weekly rainfall series for 42 years, separately for each week. The computations are done using the R-software. Local polynomial is used as kernel function.

Based on the rainfall patterns observed these are grouped into four sets and depicted in Fig 3.6(a, b, c and d)

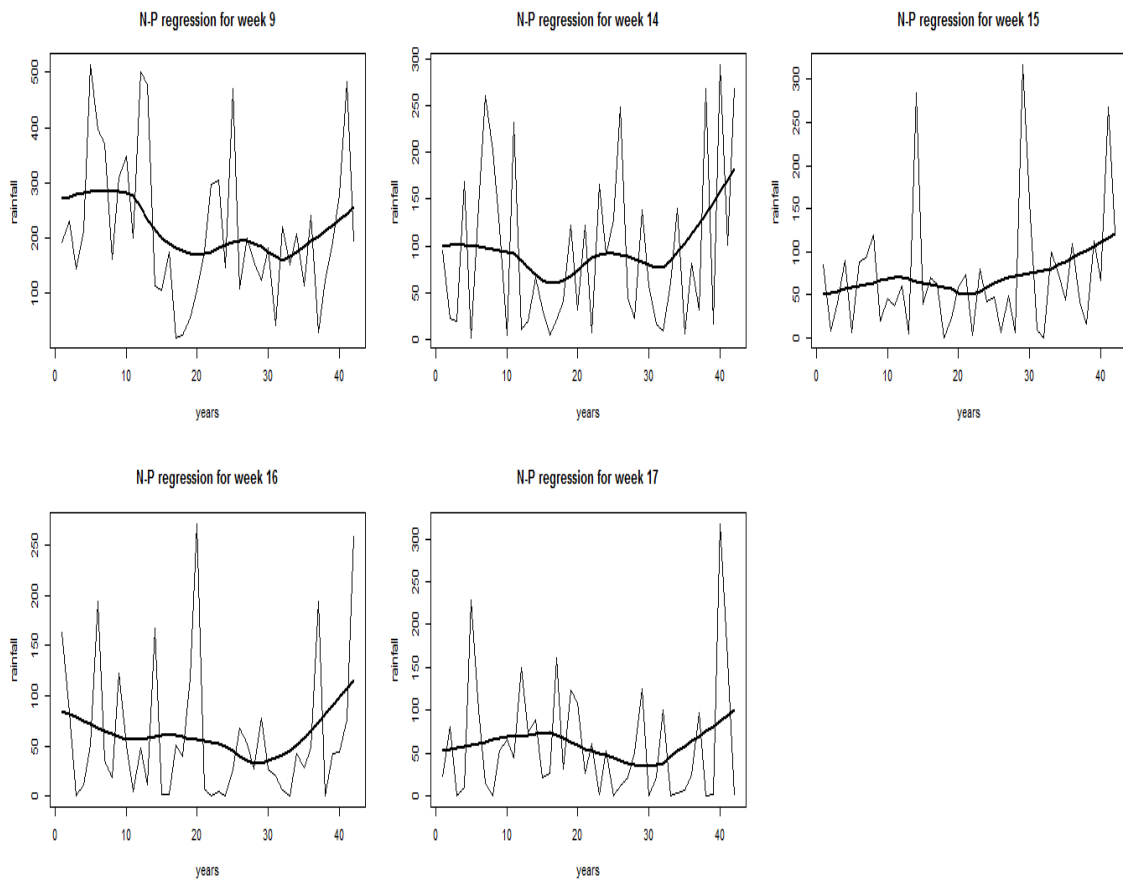


**3.6(a)**

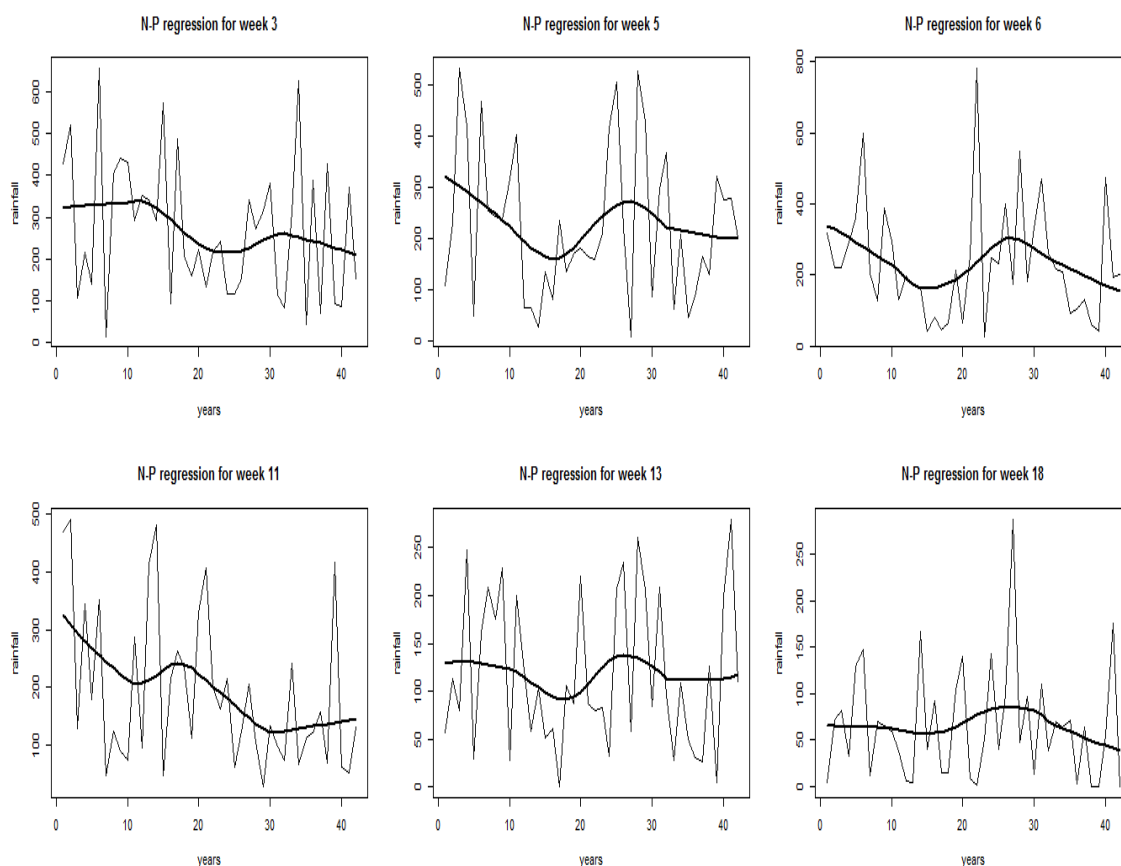




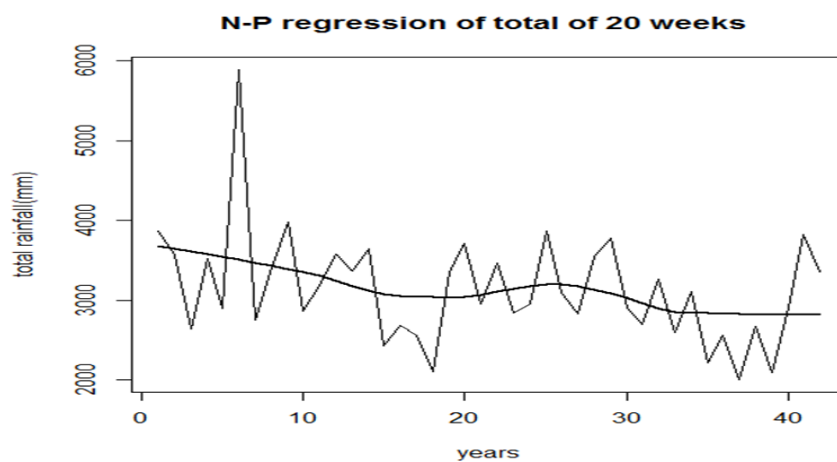
3.6(b)



3.6(c)



3.6(d)



3.6(e)

**Figure3.6:** groups of rainfall pattern of different weeks observed by non parametric regression

A brief description of these patterns is as follows.

1. For week number 2(Jun8-Jun14), 19(Oct5-Oct11) and 20(Oct12-Oct18) the rainfall pattern shows an increasing trend in the initial years and later on a decreasing trend. A decreasing trend is observed from 30<sup>th</sup> year (1999) onwards. (see Fig3.6a)
2. For week numbers 1(Jun1-Jun7), 2(Jun8-Jun14), 4(Jun22-Jun28), 7(July13-July19), 8(july20-July26), and 10(Aug3-Aug9) the rainfall pattern over the years is somewhat similar. Except for 7<sup>th</sup> week (July13-July19) we find a decreasing trend in the rainfall. For week number 12(Aug17-Aug23) a cycle of 34 years is observed which starts from the year (1973) and ends during the year (2007). For other weeks cycle of 10 to 20 years is observed except for week number 4 (Jun22-Jun28) where cyclical pattern is not very clear (see Fig 3.6b).

3. A common feature in the rainfall pattern of weeks 9(July27-Aug2),14(Aug31-Sep6),15(Sep7-Sep13),16(Sep14-Sep20), and 17(Sep21-Sep27) is that the rainfall starts increasing from around 32<sup>nd</sup> year(2001).Some kind of cyclical nature is observed during the period (1970-1999).The period of the cycle ranges from 10 to 20 years.(see Fig 3.6c)
4. For week numbers 3(Jun15-Jun21), 5(Jun29-July5), 6(July6-July12), 11(Aug10-Aug16) and13 (Aug24-Aug30) more clear cut cycles of rainfall pattern is observed. The length of the cycles varies from 10 to 20 years. For week number 18 (Sep28-Oct4) no such clear cut cycle is noticed. Once again we find a decreasing trend in the rainfall during the last decade except for week 11 (Aug10-Aug16) and 13 (Aug24-Aug30) (see Fig 3.6d).
5. The total rainfall for the rainy season shows a decreasing trend over the last 42 years (see Fig3.5e). It ranges from around 3604mm to 2972mm. A cycle of twenty years duration is observed between the years (1981-2001).

However there is no indication of cycles being repeated in the future. Hence we cannot arrive at a definite conclusion on cyclical nature in total rainfall.

### 3.4 Change in Rainfall

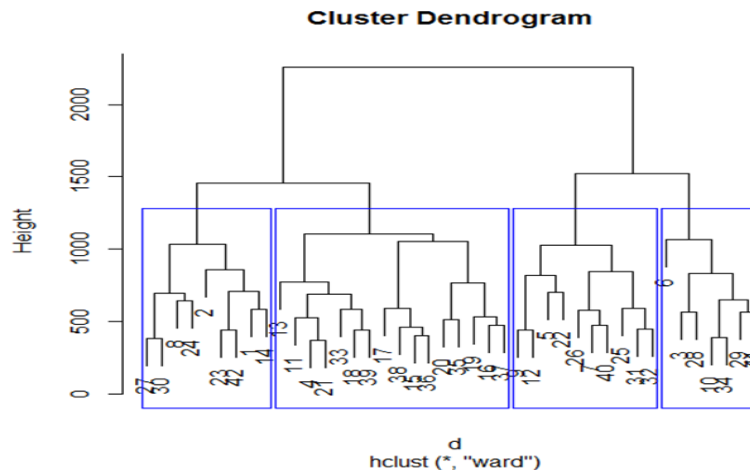
A conjecture of scientists is that, global warming affects rainfall and causes a change in the pattern of rainfall (Global Warming-Wikipedia). Some of them are opinioned that the evidence can be found in the rainfall pattern in the recent years .To examine whether the conjecture is true, Kolmogorov-Smirnov two sample test is used. The weekly rainfall for each year of the recent five years is compared with the weekly average of the rainfall for the first 37-years (long term trend, 1970-2006). If Kolmogorov-Smirnov two sample test is significant, it indicates that the weekly rainfall pattern has changed compared to the long term average. The Kolmogorov-Smirnov two sample test is significant only for the year 2008 with a p-value 0.02.

To strengthen the finding of Kolmogorov-Smirnov two sample test, cluster analysis is also done using the R-software by the Ward’s method of hierarchical clustering. Fig 3.7 presents the dendrogram corresponding to the cluster analysis. Initially 5 clusters were formed. On examination of dendrogram we decided to reduce the number of clusters to 4 only. Table 3.3 presents the classification of years into 4 clusters. Among the recent years, 2005, 2006, 2007 and 2008 fall into a single cluster (cluster number 2 with mean weekly rainfall of 137.9mm). This partially confirms the finding that rainfall pattern show a decreasing trend in the recent years.

**Table 3.3:** Classification of years into different clusters and their mean

Cluster Number	1	2	3	4
Years in each cluster	1970,1971, 1977,1983, 1992,1993, 1996,1999,2011	1973,1980, 1982,1984, 1985,1986, 1987,1988, 1989,1990, 2002,2004, 2005,2006, 2007,2008	1974,1976, 1978,1981, 1991,1994, 1995,2000, 2001,2009	1972,1975, 1979,1997, 1998,2003, 2010
Mean	163.0672	137.3900	162.4210	183.1621

If the majority of the recent years were to form a single cluster, it would have shown a change in the rainfall pattern for these years compared to other years which constitute the remaining clusters. This analysis support the conclusion obtained from Kolmogorov-Smirnov test.



**Figure 3.7:** cluster dendrogram using ward’s method

**3.5 Wet and dry spells**

In hydrological studies the major components of interests are a) daily, weekly or monthly amount of rain b) Number of rainy days and c) The wet and dry spells. Usually in the study of wet and dry spells, daily rainfall data is considered for the entire year. In the present analysis we have studied wet and dry spells for the rainy season only. The components of interest in this analysis are the number of wet and dry spells and the maximum length of the wet or dry spells. Table 3.4 Presents these details for the rainy seasons of the years 2007 to 2011. From the table it is clear that the total number of wet and dry spells ranges from 11(2007) to 24(2010). The maximum length of the wet spell is 16(2010) followed by 13(2009).

As mentioned previously total number of wet and dry spells are similar to the total number of Runs. Thus an innovative application of Run test is used to see whether the total number of wet and dry spells is the same between any two years. There are  $5C_2 = 10$  comparisons. The Z-test for the equality of total number of Runs is significant between the years 2007 and 2010 (p value is less than 0.001) and between the years 2007 and 2011 (p value is less than 0.01). Thus the total number of wet and dry spells do not follow a systematic pattern and thus does not support the hypothesis of change in rainfall pattern.

**Table 3.4:** Number of wet and dry spells and the maximum length of wet and dry spells of the rainy seasons (2007-2011)

Year	Wet spells	Dry spells	Total number of wet and dry spells
2007	8	3	11
2008	10	13	23
2009	13	10	23
2010	16	8	24
2011	10	5	15

**3.6 Predictive model**

Multiple regression equations is of the form  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$  were tried. Here Y represents the response variable and X1 & X2 are the explanatory variable. In our model fitting we used rainfall of the current week as Y and X1 representing the rainfall of the previous week and X2 represent the average rainfall of the last 10 years for the previous week. Multiple regression model fitted produced poor fit with  $R^2$  value around 0.25. Such type of models are not useful for the prediction purpose. Thus they are not reported. It may be due to the fact that the rainfall of a particular week depends on so many other climatic factors for which data was not available for us.

**IV. Discussion**

The significance of Friedman’s test for the years 1970-1979 and 2000-2011 only indicates that the median weekly rainfall for these years is not equal. It does not imply that the rainfall pattern has changed during this period. The Run test for the period 1991-2011 is significant only for the 8<sup>th</sup> week (July20-July26). Thus during this period the fluctuations in average weekly rainfall for the remaining 19 weeks is random. The Kolmogorov-Smirnov two sample test and the cluster analysis do not support the hypothesis of change in rainfall pattern in the recent years. The nonparametric regression reveals that there is a shift in the rainfall pattern for some of the weeks where it is decreasing for the last five years. The overall conclusion from all the analysis can be summarized and given below

1. There is no change in the long term total rainfall for the years 1970-2011 and the yearly fluctuations are by and large are random.
2. The variability in the daily rainfall for the months June to September does not show a change.
3. No clear cut cycles are observed in the total rainfall during the years 1970-2011
4. Although there is partial evidence that there is change in weekly rainfall in the recent years, it is too early to conclude that rainfall pattern is changing. Rainfall data for the next five years may provide evidence for the change in rainfall pattern.

**V. Conclusion**

In the present project an extensive analysis of daily and weekly rainfall is carried out for the 42 years. For the rain series of Catalonia region Lana et al. (2009) could able to arrive at a stationary series after 110 years. Since 42 years is small period to obtain a stationary time series of rainfall we have not used autoregressive or autoregressive moving average procedure for the analysis of rainfall. In this project several innovative applications of the statistical procedures are suggested for the analysis of hydrological data. For measuring the changes in hydrological data such as rainfall, more than one statistical procedure is required to prove the conjecture of the scientists. This project is a modest beginning in this direction. Due to the shortage of time the detailed investigation of wet and dry spells is not carried out. Although Markov models are developed for predicting the rainfall in this project we could not undertaken this.

### Acknowledgement

The author is great full to Prof. Aruna Rao, Professor of Statistics, Mangalore University for his valuable suggestions. Also the author would like to thank Prof. Ismail B, Chairman, Department of Statistics, Mangalore University for his support. Further the author would like to thank Ms. Preema Pereira and Mrs. Rashmi Shetty for their help during this work.

### References

- [1]. Lana, X., Burgueno, A., Martinez, M.D. and Serra,C.(2009) A review of statistical analysis on monthly and daily rainfall in Catalonia. *Journal of weather and climatic of the western Mediterranean*. 6, 15-29.
- [2]. Baldassarre, G. D., Castellarin, A. and Brath ,A.(2006) Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy. *Hydrology and earth system Sciences*.10, 589-601.
- [3]. Rao, A. R. and Kao, S. C.(2006) Statistical analysis of Indiana rainfall data. JTRP Technical Report.
- [4]. Over, T. M. and Gupta, V. K.(April,1994) Statistical analysis of mesoscale rainfall: Dependence of a random cascade generator on large scale forcing. *Journal of Applied Metireology*. 33, 12. 1526-1542.
- [5]. Chowdhury, R. K. and Beecham, S.(2009) Characterization of rainfall spells for urban water management. *Wiley Online Library*.
- [6]. Singh, N. and Ranade, A. A. (2009). Climatic and Hydroclimatic Features of wet and dry spells and their Extreans across India. Research report, *Indian Institute of Tropical Meteorology Pune*.
- [7]. Krishnamurti, T. N., Sang-ok Han. And Misra, Vasubandhu. (2007) Prediction of the dry and wet spell of the Australian monsoon. *Wiley Online Library*.
- [8]. Lehmann E. L. and Romano, J. P (2005) Testing Statistical Hypotheses,3<sup>rd</sup> edition. Springer Texts in Statistics
- [9]. Gibbons, J.D.(1971) Non parametric statistical inference. *Mcgraw-HILL Kogakusha, Ltd*.
- [10]. Kendall, M. and Stuart,A. (1976).The advanced theory of statistics, 4<sup>th</sup> edition Volume-1.*Charles Griffin & Company Ltd, London& High Wycombe*.
- [11]. Hardle, W.(1990) Applied nonparametric regression, *Cambridge university press, Cambridge*.
- [12]. Global Warming- Wikipedia, the free encyclopedia
- [13]. Johnson, R. A. and Wichern, D. W.(2001)*Applied Multivariate statistical Analysis*. Prentice-Hall, Inc., Englewood Cliffs