Statistic and Probabilistic Variations and Rainfall Predictions of TRNC

Mohammadbagher Amjadi

Submitted to the
Institution of Graduate Studies and Research
in partial fulfilment of requirements for Degree of

Master of Science
in
Civil Engineering

Eastern Mediterranean University
August 2015, Gazimağusa
North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Serhan Çiftcioğlu
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Civil Engineering

Prof. Dr. Özgür Eren
Chair, Department of Civil Engineering

Certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Civil Engineering.

Assoc. Prof. Dr. Mustafa Ergil
Supervisor

Examining Committee

1. Assoc. Prof. Dr. Mustafa Ergil ______________________
2. Assoc. Prof. Dr. Umut Türker ______________________
3. Asst. Prof. Dr. Bertuğ Akıntuğ ______________________

ABSTRACT

This thesis deals with the monthly rainfall of six meteorological regions and TRNC (North Cyprus) as a whole for the hydrologic years from September 1975 to August 2014 period. In order to study these gathered monthly data statistically, other than the minimum required sample sizes for each region, the quality check tests (homogeneity, consistency, normality, independency, stationarity and trend) were as well carried out based on different parametric and/or non-parametric tests. To determine the most representative probability distribution models among the two widely used Normal and Log-Normal distributions for each region were use, since the gathered raınfall was based on monthly averages. In order to predict 5 years ahead of the yearly rainfall of each meteorological region and TRNC, three different time series models (Markov, Auto-regressive (AR) and Holt-Winter Multiplicative) were used. For this reason, the rainfall of hydrologic years from 1975-76 to 2003-04 were used for training and from 2004-05 to 2013-14 were used for forecasting (testing) the trained data. The best representative time-series model for each region was selected based on the standardized averages of four statistical error checking measures (MAPE, MAD, MSE and RMSE). The selected model for each region was then used to predict (estimate) the rainfall for five successive hydrologic years ahead from 2014-15 to 2018-19. To investigate the wetness or dryness characteristics of each regions and TRNC (North Cyprus), the hydrologic yearly averaged and the common monthly (from September to May) rainfall data sets were studied separately. Interestingly for all the months of all the regions, the dryness was controlling.

Key words: rainfall, forecasted data, time series models, TRNC, wet or dry spells.

ÖZ

Bu tez, KKTC toplamı ile altı meteorolojik bölgenin Hidrolojık yıl Eylül 1975 ile Ağustos 2014 dönemini kapsayan aylık yağış donelerini kapsamaktadır. Elde edilmiş bu verilerle statistiksel çalışılabilinmesi için, her bölge için ihtiyaç duyulan en az örnek sayı miktarının belirlenmesi yanında done kalite testleri (homojenite, konsistensi, normalite, staşinarite, independensi ve trend) parametrik ve/veya parametrik olmayan farklı testler kullanılarak uygulanmıştır. Her bölgeyi ifade edebilen en iyi olasılık fonksiyon dağılımı mevcut örneklemeler aylık ortalamalardan oluştuğunda, literatürde en çok kullanılan Normal ve Log-Normal dağılımları arasından belirlenmiştir. İleriye dönük veri değerleri belirlenebilmesi için, her meteorolojik bölge ve KKTC için üç değişik zaman seri modeli (Markov, Auto-regressive (AR) ve Holt-Winter Multiplikatif) kullanılmıştır. Bu amaç için hidrolojik yıla göre düzenlenmiş yağış değerlerinin 1975-76 ile 2003-04 yılları aralığındakiler alıştırma ve 2004-05 ile 2013-14 yılları aralığındakiler deneme için kullanılmıştır. Stardartize edilmiş dört statistiksel hata testi (MAPE, MAD, MSE ve RMSE) kullanılarak her bölge için en uygun zaman serisi modeli seçilmiştir. Her meteorolojik bölge için seçilen bu en uygun model kullanılarak gelecek peşpeşe beş yıldaki (2014-15 ile 2018-19 arası) olası ortalama yağış değerleri türetilmiştir. Her bölge ve KKTC için nemlilik veya kuruluk dönemleri ortalama yıllık ve benzer aylar (Eylül'den Mayıs'a kadar) ayrıca ayrı ayrı çalışılmıştır. Her ay ve tüm bölgelerin kuru aralığın etkisinde olduğu ilginç bir bulgu olarak saptanmıştır.

Anahtar kelimeler: yağış, ileriye dönük veri, zaman serisi modelleri, KKTC, nemli veya kuru aralık.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my family for letting me fulfil my dream to be a Master student here in EMU. I like to dedicate this study to them.

I also would like to thank Assoc. Prof. Dr. Mustafa Ergil for his continuous support and guidance in the preparation of this study. Without his invaluable supervision, all my efforts could have been short-sighted.

I would like to thank Meteorology Office of TRNC for allowing me to use the data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\bar{x}_{Ar}$	Arithmetic Mean
$\bar{x}_{Geo}$	Geometric Mean
$\bar{x}_{Har}$	Harmonic Mean
$\bar{x}_{W}$	Weighted Average
$\bar{x}_{Med}$	Median
$\bar{\mathrm{X}}_{\mathrm{mod}}$	Mode
d_x	Mean Deviation
C_{dx}	Coefficient of Mean Absolute Deviation
s_x^2	Variance
Sd	Standard Deviation
Cv	Coefficient of Variance
C_s	Skewness
γ_k	Kurtosis
α	Confidence Interval
C_k	Autocovariance function
$\mathrm{d}f$	Degree of freedom
r_k	Autocorrelation factor
τ	Mann-Kendall test
χ^2	Chi-square

Chapter 1

INTRODUCTION

The Intergovernmental Panel on Climate Change Fourth Assessment Report (IPCC-AR4) indicates significant summer warming in south-eastern Europe and the Mediterranean, while downward trends are associated with the mean annual rainfall (Christensen, 2007). The combined effect of high temperatures and low rainfall poses challenges to many economic sectors as well as significant threat on desertification (Giorgi, 2006; Gao and Giorgi, 2008). For instance, the IPCC-AR4 highlights that, water stress will increase in southern Europe, and hence agriculture will have to cope with increasing water demand for irrigation (Alcamo et al., 2007). In addition, the observed climate changes are likely to enhance the frequency and intensity of extreme events' occurrence, such as heatwaves and droughts (Meehl et al., 2007) which may critically affect the society and economy of small island countries, like Cyprus. There is therefore a need for more accurate climate model predictions that will provide meteorological information on national level and enable relevant climate change impact studies to assist adaptation strategies.

The characteristic summer aridity of the region has significant implications in several socio-economic sectors. Cyprus is facing its worst ever water shortage in the last few decades. Climate models are widely used to project present and future changes of climate variables. Although the ability of models has improved, systematic biases can be found in model simulations. Therefore it is recommended the accuracy of model simulations of past or contemporary climate to be evaluated by comparing the results with observations.

Weather forecasting plays an important role in our daily life. Especially in engineering, it shows itself more significantly. Among meteorological data, mainly the rainfall variations are the subject that the researchers are interested a lot. Although rainfall has a high positive effect on ecological sustainability of the living organisms, but can cause disasters like flooding or drying up of the existing reservoirs due to global warming. Hence, estimating the daily, monthly, seasonally and even the yearly amount of rainfall values for different locations may guide the researchers to some extent, for their future strategies.

1.1 Literature review

From the available literature, it was observed that, not too many studies have been carried out on the rainfall distribution patterns of North Cyprus. Ismail and Goymen (1985) discussed the general outlook of rainfall in North Cyprus by considering the yearly averages of rainfall for TRNC (North Cyprus) from 1976 to 1985 whereas Kypris 1995, studied the hydrologic yearly rainfall averages of Cyprus from 1901-02 to 1992-93 attempted to find diachronic changes using thirty years moving average where he determined the rainfall shifts during the last century. Biyikoglu 1995, gathered the annual average rainfall of TRNC (North Cyprus) from 1975 to 1994 and used it for determination of water budget of North Cyprus. Altunc 1995 and 1997, published two conference papers for the water problems of TRNC based on the basic meteorological parameters and only uses TRNC yearly average rainfall values from 1976 to 1993. Tayanc 1997, studied the opportunity of making cloud seeding over Cyprus and details the high risks components. Altan and Sen 2000, gathered the annual rainfall of TRNC from 1980-93 for agricultural studies where Pashiardis 2003, studied the records of monthly rainfall of South Cyprus from 1967 to 2001 for agricultural planning needs where only the total yearly rainfalls were used. Kimyaci 2004, examined the rainfalls of Lefkoşa Station from 1975 to 2003 and gathered the extreme (maximum) intensities for each year where he used to establish the intensity–duration and frequency curves for Lefkoşa and North Cyprus. Sharifi 2006, studied in detail, the basic hydro-climatological variations and trends of N. Cyprus where he used the hydrologic yearly average monthly rainfall for each region and TRNC from 1975-76 to 2004-05. In that study, he also studied the regional variation of temperature and wind velocities. Recently, Seyhun and Akıntuğ 2013, studied the trend analysis of monthly rainfall in North Cyprus based on 20-stations through non-parametric tests and attempted to determine if a trend exists.

1.2 Study Area

Cyprus is an island, being located in the north-eastern part of the Mediterranean Sea, and is the third largest island with a surface area of 9251 km2. It is bounded by latitudes of $35^0$45’ and $34^0$15’ N, and by longitudes of $32^0$15’ and $34^0$30’ E. The island lies about 64 km south of Turkey, 97 km west of Syria and 402 km north of Egypt’s Nile Delta and 380 km south east of Greece. Islands total coastline is 782 km in length (Kypris, 1995).

After the peace operation in 1974, TRNC was established in 1983, as a separate unilateral state on the northern one third of the island, where the remaining part is under the control of so called Cyprus Government. Its capital is Nicosia (Lefkoşa) being the unique divided capital in the world. The population of the whole island based on recent census in 2,000 was 748,000 of which 68 percent is Greek, 27 percent is Turkish and the remaining 5 percent belongs to various minorities (Kimyaci 2004).

The central Troodos massif, raising to 1951 meters and to a less extend, the long narrow Kyrenia mountain range, with peaks of about 1000 meters, play an important role in defining the weather condition of Cyprus. The predominantly clearer skies and high sunshine amounts give large seasonal and daily differences between temperatures of the sea and the interior of the island. At latitude 35° North and longitude 33° East, Cyprus has a change in day length from 9.8 hours in December to 14.5 hours in June. Since Cyprus lies at the eastern end of the Mediterranean Sea, it belongs to the Mediterranean climate zone, therefore it experiences mild winters and hot dry summers. Island of Cyprus intense Mediterranean climate with a typical seasonal rhythm strongly marked with respect to temperature, rainfall and weather in general. Winters, rather changeable are mild, with some rain and snow on Troodos Mountain, are from November to mid-March and separated by short autumn and spring seasons of rapid change in weather conditions. In summer, the extension of the summer Asian Thermal Low is evident throughout the eastern Mediterranean in all seasonal circulation patterns (Kostopoulou and Jones, 2007a, b), associated with high temperatures and abundant sunshine with hot dry summers from mid-May to mid-September. Hence, in summer, the island is mainly under the influence of a shallow through of low pressure extending from the great continental depression cantered over southwest Asia. It is a season of high temperatures with almost cloudless sky. Rainfall is almost negligible but isolated thunderstorms sometimes occur which give rainfall amounting to less than 5% of the total in the average year. In winter, Cyprus is near the track of fairly small depression that cross the Mediterranean Sea from west to east between the continental anticyclone of Eurasia and the generally low-pressure belt of North Africa. These depressions give periods of disturbed weather usually lasting from one to three days and produce most of the annual rainfall (Pashiardis, 2003).

The wet season extends from November to March, with most of the rain falling between December and February (approximately 60% of the annual total). Rainfall is generally associated with the movement of moist maritime flows to the North, occurring particularly over areas of high elevation (Kostopoulou and Jones, 2007a). Winter rainfall is closely related to cyclo-genesis in the region (Pinto et al., 2006). Nevertheless, it is not uncommon for isolated summer thunderstorms to occur, which however contribute to less than 5% to the total annual rainfall amount (Pashiardis, 2003).

Undoubtedly, estimating a data to a very close value is impossible, but there are statistically accepted probability distribution functions and time series models that provides reasonable solutions for the prediction of the near future data within the acceptable confidence intervals.

In this study, amending 10 more recent years monthly based regional rainfall of North Cyprus to the previously studied rainfall by Sharifi 2006, also the time series models were studied as a new chapter. Since the exact hydrologic model for any data is never known, among the popular models existing in literature, Markov, Auto-Regressive (AR) and Holt-Winter Multiplicative models were selected for this study where five successive hydrologic years averages from 2014-15 to 2018-19 were predicted.

Hence, the first objective, is to analyse the monthly available rainfall of North Cyprus from 1975 to 2014 for each meteorological region and for TRNC as a whole, and secondly to identify the most representative model(s) giving the most likelihood statistical indices based on the existing data so as to predict relevant data for the near future for each region.

Island of Cyprus is meteorologically grouped into 14 main geographical regions as shown in Fig1.1, but due to political reasons, no official communication based on exchanging, sharing or using the gathered relevant data of any documents is possible hence, for this small island, the southern part excludes the northern part in any study including hydro-meteorological studies so as the northern part.

Along the north, TRNC State Meteorology Department, with simple regional modifications along the regional boundaries and renumbering of the existing meteorologically divided map, establishes its own meteorological regions. Hence, along the geographical occupation of TRNC, there are 6 meteorologically grouped geographical regions as shown in Fig. 1.2:

a) I North Coast and Beşparmak Mountains (1),
b) II West Mesaria (~4),
c) III Central Mesaria (~5),
d) IV East Coast (part of 7),
e) V East Mesaria (~6) and
f) VI Karpaz (~2).

The values within the parenthesis imply the regional numbering suggested for the whole Cyprus and still in use as detailed in Figure 1.1, whereas the Roman numbering font is used by TRNC meteorology department so as to reduce the confusion due to the regional boundary modifications.

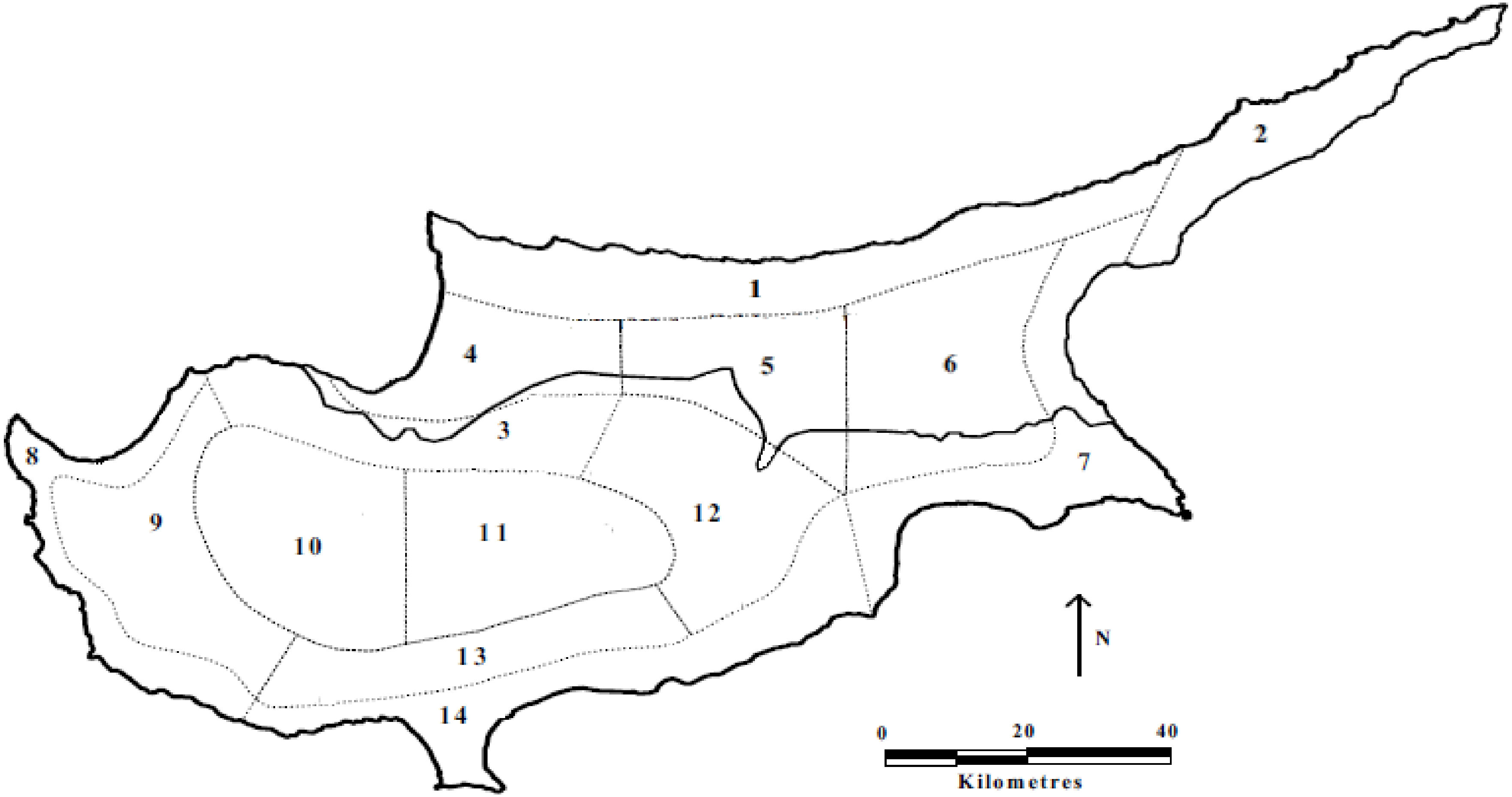


Figure 1.1: Geographical regions map of Cyprus based on meteorological aspect (obtained from Meteorology Office, TRNC).

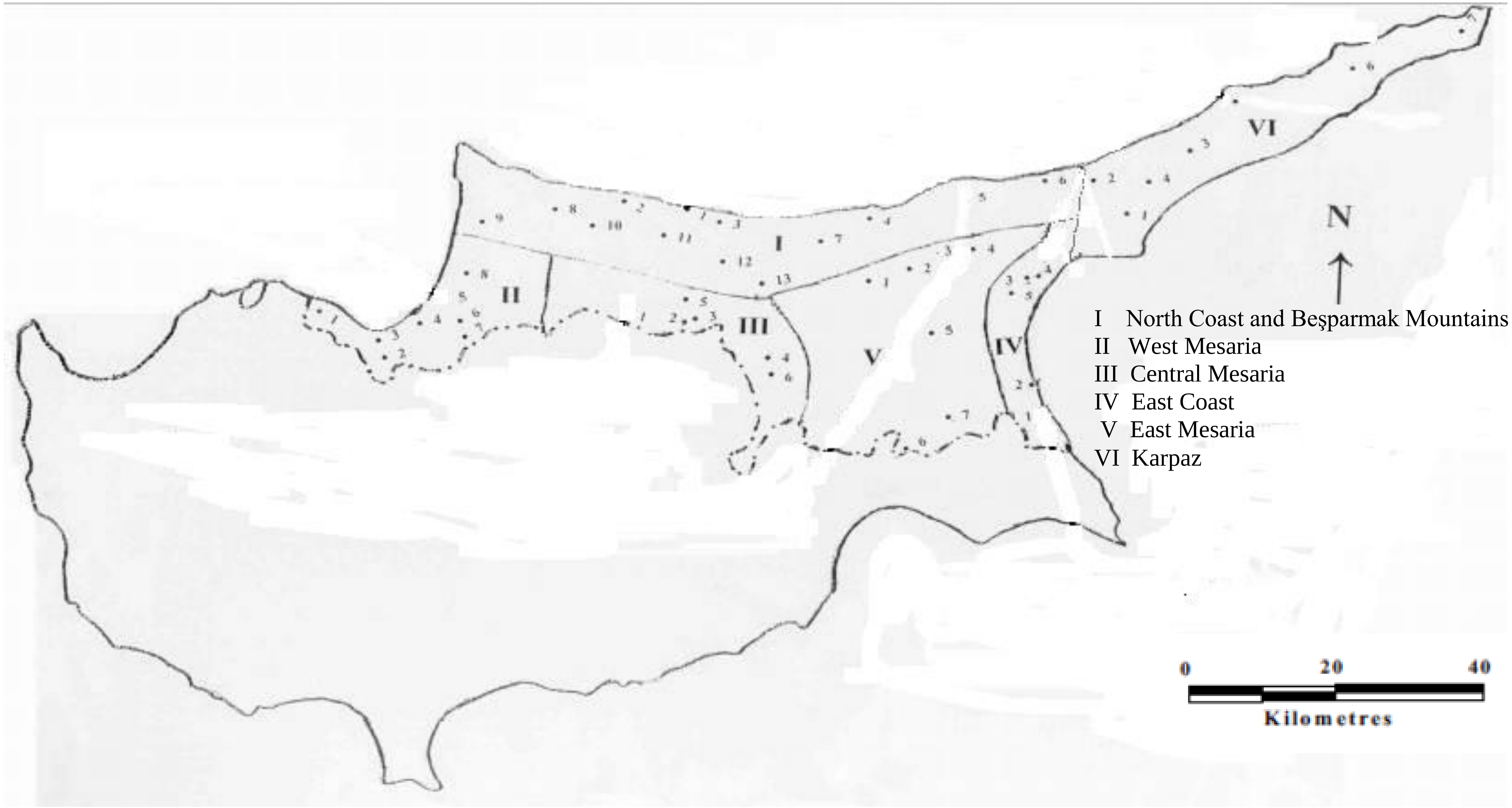


Figure 1.2: Geographical regions map of TRNC, based on meteorological aspect and the locations of their representative stations (obtained from Meteorology Office, TRNC).

1.3 Rainfall

Rainfall is any product of the atmospheric water that falls under the action of gravity on our planet. Among the hydro-climatologic parameters, the liquid phase of this rainfall, i.e. the rainfall; was examined in this study and hence, the monthly variations of rainfall of the six meteorologically divided geographical areas of North Cyprus, as well as for TRNC as a whole unit were compiled. Due to ongoing construction of various stations within each region since 1974, some of the observation data records had late starts. In order to overcome this weakness, the regional averages were used in this study. Table 1.1 details each meteorological region of TRNC that is characterized by different number of meteorological stations. All these gathered data were statistically examined through appropriate statistical measures and indices.

1.4 Objective of Study

The objective of this study is to examine the variations of monthly rainfall gathered from six meteorological regions and TRNC. For this reason first the gathered data quality (Homogeneity, Normality, Consistency, Trend, and Stationarity) statistically checked. Later based on time series analysis validated equations were generated and ten years ahead rainfall for each regions and TRNC were generated. Also the wet/dry spells for each month for each region and TRNC were studied.

1.5 Outline of Thesis

This thesis consists five chapters. The details are given bellow:

In the second chapter, the basic statistical terminologies, the widely used statistical deterministic and stochastic functions, the parametric and/or non-parametric measures and indices used for checking the appropriateness of these functions were all detailed. A sample calculation of each statistical measure and parameter used in this study, was also detailed based on one of the representative region among the studied six regions.

Central Mesaria region was selected as a sample where the following analyses were applied:

- determination of minimum input data number for each regions that is required to analyse that data,
- testing normality,
- testing homogeneity (for each region and among the regions),
- testing consistency,
- examining the occurrence of trend,
- finding the best fitted distribution (among Normal and log-Normal) probability for each region and TRNC.

The third chapter consists of information about time series and its parameters definitions as well. The time series models that are used in this study were explained and relevant examples were presented. In this chapter, all different time series models were applied to each region and the estimated values from hydrologic year 2004-05 till 2013-14 were compared with the measured data of these years. After comparing the error of prediction based on statistical error measures and the measured data, the most likelihood model for each station were suggested.

In this study, among the widely used time series models, the below three models were only used:

1. Markov,
2. Auto Regressive, (AR)
3. Holt-Winters Multiplicative Model.

The testing, forecasting and prediction of time series were done mainly by using Minitab, and Excel softwares.

In the fourth chapter all the graphs and the tables of the statistical parameters and the time series models of regions were illustrated.

The last chapter gives conclusion and recommendations for relevant future studies.

Table 1.1: Meteorological regions of TRNC and their measured parameters (Meteorology Dept. TRNC)

Station of T.R.N.C.	**Measured Parameters**			
	Temperature	**Wind Speed**	**Evaporation**	**Rainfall**
I. N. Coast and Beşparmak Mount.				
1. Girne	X	X	X	X
2. Lapta	X			X
3. Beylerbeyi				X
4. Esentepe	X			X
5.Tatlısu				X
6. Kantara				X
7. Alevkaya	X			X
8. Çamlıbel	X			X
9. Akdeniz				X
10. Kozanköy				X
11. Boğazköy	X	X	X	X
12. Taşkent				X
13. Değirmenlik				X
II. West Mesaria				
1. Yeşilırmak				X
2. Lefke				X
3. Yeşilyurt				X
4. Gaziveren				X
5. Güzelyurt				X
6. Yukarı Bostancı				X
7. Zümrütköy	X	X	X	X
8. Kalkanlı				X
III. Central Mesaria				
1. Alayköy				X
2. Lefkoşa (1)	X	X	X	X
3. Lefkoşa (2)	X	X	X	X
4. Ercan	X	X	X	X
5. Yakın Doğu Üni.	X			X
6. Margo				X
IV. East Coast				
1. Gazimağusa	X	X	X	X
2. Salamis				X
3. İskele (1)				X
4. İskele (2)	X			X
5. Yeniboğaziçi				X
V. East Mesaria				
1. Serdarlı				X
2. Göndere				X
3. Geçitkale (1)	X	X	X	X
4. Geçitkale (2)	X			X
5. Dörtyol				X
6. Beyarmudu	X			X
7. Çayönü				X
VI. Karpaz				
1. Çayırova				X
2. Büyükkonuk				X
3. Ziyamet				X
4. Mehmetçik				X
5. Yenierenköy	X			X
6. Dipkarpaz				X
7. Zafer Burnu				X

Chapter 2

STATISTICAL TERMINOLOGIES, PROCEDURES AND SAMPLE CALCULATIONS

2.1 Introduction

It is a known fact that, many quantities encountered in all phases of life are treated as random variables in statistical sense. Theoretically, there should be a scientific explanation as to the occurence of every sensible being, so the physical and the engineering quantities should be mathematically formulated. Because of the three dimensional complexity and time necessarily being the fourth dimension of some phenomena, however, even the most developed organizations or individuals of exceptional dexterity are unable to mathematically depict some events such as hurricanes, many meteorological incidents, and severe earthquakes. For example, aside from snow melt, everybody knows that, an intense rainfall exceeding the infiltration capacity of a particular area causes direct overland flow ultimately results in a flood. The physical mechanism of direct runoff beginning from a thin sheet flow, passing through the rest of the drainage paths, and finally continuing its travel in a mis qualitatively explainable. There are qualitative models that accounts these complicated mechanisms with respectable accuracy through appropriate computer programs and packages but the unpredictability of the meteorological events however, brings about a serious difficulty for realistic calculations of the magnitude and spatial and temporal variation of the hydro-meteorological (i.e. rainfall, snow, evaporation, etc.) input, in the first place. Most of the case study problems in engineering dealt with these uncertainties. Even the conditions of similar cases look common and similar, their effects may be different. This is mainly due to the randomness characteristic that involves during the occurrence of the natural (real case) problems and the inappropriateness of the suggested model as well as the gathered data that is used to express this occurred phenomenon mathematically. Naturally, mankind will keep up the endeavour of making accurate meteorological forecasts for longer periods in the coming future.

Statistics is a tool that uses the data for better decision making. It is concerned with scientific methods for collecting, organizing, summarizing, presenting and analysing data as well as with drawing valid conclusions and making reasonable decisions on the basis of such data. On the other hand, probability theory and statistics deal with these randomness characteristics and their risks. The probability theory generates mathematical models so as to analyse the random variable whereas the statistics attempt to suggest most appropriate guesses by applying those mathematical models. Hence, for any problem having random variable component, through probabilistic approach, it is necessary to analyse the existing observations (data) simply adopting statistical parametric and non-parametric approaches so as to obtain meaningful magnitudes like mean, median, standard deviation etc...

Data is a set of information (observation or experimental result or numerical figure or evidence) that is gathered for examining (or using during the decision making process) from which conclusions can be obtained. The topic of statistics involves the study of how to gather, sum up, and interpret any existing data since such conclusions are essential for the decision making processes Bowerman and Oconnell, 1997. Any properly classified collection of objects about which a statistical investigation is being created is a population. So the number of individuals in any population is the size of that population which can be finite or infinite. A finite set of items taken from the population with a specific plan is called a sample. The total number of individuals in a sample is called the sample size. Generally if the data are less than or equal to 30, in statistics is referred as sample.

The engineering problems in general and the hydrologic cycle especially contain quantity of events such as rainfall, runoff, infiltration, evaporation, etc. that can be explained through above mentioned approach where the time component as well interferes. Usually, the number of available data in engineering are small in size, so the sample statistics are used during analyses. So the hydrologic variables that are collected based on time and/or space can be grouped as:

i. Historical or chronological,
ii. Field collected,
iii. Experimental (laboratory level),
iv. Simultaneous measurements of two or more variables

2.2 Statistical Measures of Central Tendency and Dispersion

Statistical parameters (magnitudes) of any random data, helps us to define the centre of that data and also how the remaining data spread around this centre value, i.e. the variation, the skewness and the kurtosis.

If the population of the data does not known, (which is the case in most of the engineering problems), the statistics even helps to estimate the above mentioned case through the sample statistics approach.

To determine it, two basic approaches that are most widely in use are:

i- the statistic moments (parametric/analytic statistics) and

ii- the ranking (non-parametric/non-analytic) statistics.

In most of the studies, it is believed that, the population data obeys the normal distribution character. This is valid if the magnitudes of any data are not deviating too much from one sample to another, (having minor risk of sampling error) hence, the statistic moments approach can safely be used. But, if the sample size is rather small and/or the distribution of the data is skewed (not obeys the normal distribution) and/or even within the data there are outliers (at least there is a value which is very big or very small compared with the remaining data) then, those above mentioned statistical magnitudes show high variations. Therefore, for this type of data, instead of using the parametric approach, the non-parametric (quintile or so called the ranking) statistical approach should be adopted. Nonparametric tests are also called distribution free tests (Maidment 1993).

2.2.1 Central Parameters

A particular value that can be considered as characteristic or representative of a set of data and about which the observation can be considered as the centre or middle, is called the average. It is the best common characteristic of a data that illustrates the central tendency. It can be determined for parametric and for non-parametric cases.

2.2.1.1 Analytical Means

There are different approaches that use simple mathematics to define the mean i.e. the average:

- **Arithmetic Mean ($\bar{x}_{ar}$):** It is the widely used and simpler way of finding the average. It is obtained by summing of all the data and dividing it over the total number of data that forms that group data ‘n’.

$$\overline{x}_{ar} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad 2.1$$

- **Geometric Mean ($\overline{x}_{geo}$):** Geometric mean is the logarithmic average of the data. Not defined if even any of the data is negative and zero. (Since its value is reasonably close to the median, can be sometimes used instead).

$$\overline{x}_{geo} = \sqrt[n]{x_1 x_2 \ldots x_n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = \log^{-1}\left(\sum_{i=1}^{n} \frac{\log(x_i)}{n}\right) \qquad 2.2$$

- **Harmonic Mean ($\overline{x}_{har}$):** Harmonic mean the reciprocals average of the data, and not defined if even anyone of the data is negative and zero:

$$\overline{x}_{har} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad 2 \qquad 2.3$$

- **Weighted Average ($\overline{X}_w$):** This method is used in order to get a more representative average (mean) value of any specific data that is taken from different measuring periods of different stations or regions. Therefore, for any station or region the average value determined from that specific measuring period is added with the average of the other stations' based on their measuring period and will be repeated for the whole stations or regions that are supposed to be involved in that averaging process. The result is obtained by dividing the summation of reciprocal squares of the involved stations or regions over the weighted averages based on different measuring periods (Usul 2005).

$$\overline{x}_w = \frac{\frac{\overline{x}_i}{n^2{}_i} + \frac{\overline{x}_j}{n^2{}_j} + \ldots + \frac{\overline{x}_n}{n^2{}_n}}{\frac{1}{n^2{}_i} + \frac{1}{n^2{}_j} + \ldots + \frac{1}{n^2{}_n}} \qquad 2.4$$

- **Root Mean Squares (*RMS*):** It is the square root of the individually squared and then added of all data.

$$\text{RMS}=\sqrt{\frac{1}{n}(x_1^2+x_2^2+...+x_n^2)}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2} \qquad 2.5$$

- **Root Mean Cubes (*RMC*):** It is the cubic root of the individually cubed and then added of all data.

$$\text{RMC}=\sqrt[3]{\frac{1}{n}(x_1^3+x_2^3+...+x_n^3)}=\sqrt[3]{\frac{1}{n}\sum_{i=1}^{n}x_i^3} \qquad 2.6$$

2.2.1.2 Non-Analytical Means

- **Median ($\overline{X}_{med}$):** It is the central item of the ranked (sorted in ascending or descending order) data. In other words, the median is that value of the variable which divides the group into two equal parts, where one part representing all values greater and the other all values lesser than the median. It is not affected by outliers. Depending on the total number of data ‘n’ that forms the data (odd or even), median is determined:

$$x_{med}=\text{Select the } i^{th}\text{term from the rank} \quad x_i \begin{cases} i=\dfrac{n}{2} & \text{if} \quad 'i' \quad \text{is} \quad \text{odd} \\ i=\dfrac{n}{2}+1 & \text{if} \quad 'i' \quad \text{is} \quad \text{even} \end{cases} \qquad 2.7$$

- **Mode** ($\overline{X}_{mod}$): The most repeated data within the data is called the mode. If two or more data within the same data having the same number of maximum repeated value, than the mode is not defined. It is not a good representative data in engineering studies.

2.2.2 Dispersion (Spread) Parameters

The measures of central tendencies (i.e. means) indicate the general magnitude of the data and locate only the center of distribution measures. But they do not establish the degree of variability or the spread out or scatter of the individual items and their deviations (or the difference with) the means. It is obvious that, even two statistical data having common mean, median and mode values may differ widely in the scatter or in their values about the measures of central tendencies. Noting also that an average alone does not tell the full story unless the manner in which the individual items are scattered around the central tendency are well defined. The parameters that observe how data within the data group spreads around the analytical (parametric) and non-analytical (non-parametric) central tendencies (mean) are:

- **Range:** It is the difference between the largest (l) and the smallest (s) values within the studied data. Range = $x_l - x_s$
- **Relative Range of a Dispersion (R_r):** It is the ratio of range and the mean.

  $R_r = (x_l - x_s) / (x_l + x_s)$
- **Mean Deviation** (d_x)**:** It is the averaged positive value that represents how the remaining data within the data is scattered (deviated) from the arithmetic mean (mainly) for parametric and from the median for non-parametric case. Noting that, the absolute value used for the determination of the mean deviation is to some extent desperate from mathematical view. The mean (or median) absolute deviation can be calculated as:

$$d_x = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}_{ar}\right| \text{ (for analytical mean)} \qquad 2.8a$$

$$d_x = \text{Median}\left|x_i - \overline{x}_{med}\right| \text{ (for non-analytical mean)} \qquad 2.8b$$

- **Coefficient of Mean Absolute Deviation** (**C_{dx}**)**:** It is the ratio of mean (median) absolute deviation with the mean (median).

$$Cd_x = \left[\frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}_{ar}\right|\right] / \overline{x}_{ar} \text{ (for analytical mean)} \qquad 2.9a$$

$$Cd_x = \left[\text{Med}\left|x_i - \overline{x}_{med}\right|\right] / \overline{x}_{med} \text{ (for non-analytical mean)} \qquad 2.9b$$

- **Variance (σ_x^2):** The absolute value inserted within the mean deviation is slightly inconvenient from mathematical point of view. Hence, to remove this inconvenience the squired deviation from the mean (or median) is taken as a starting point for a measure of spread. The result obtained is referred as variance. The replacement of “n” to “n – 1” is done due mathematical reasons so as to correct the formula for the sample instead of population where the symbol s^2 is usually used to indicate it. The variance of a sample is defined as

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_{ar})^2 \quad \text{(for analytical mean)} \qquad 2.10a$$

The term variance was used to describe the square of the standard deviation. To eliminate the disadvantages of different dimensions of variance and the original observations, the square root of the variance is taken and is referred as the standard deviation. The standard deviation of a sample is:

$$s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_{ar})^2} \quad \text{(for analytical mean)} \qquad 2.10b$$

(The above term is the standard deviation which is basically equals to the root mean square from the mean). For the non-analytical case the standard deviation is named as the interquartile range where the median (50%) value within the given data interferes indirectly instead of the mean.

$$\text{Percentile Range (PR)} = x_{\%90} - x_{\%10} \quad \text{(for non-analytical mean)} \qquad 2.10c$$

- **Coefficient of Variance (**Cv**):** It is the ratio of standard deviation (or interquartile range) with the sample mean (or median).

$$Cv = \frac{S_d}{\overline{x}_{ar}} \qquad \text{(for analytical mean)} \qquad 2.11a$$

$$PCv = \frac{x_{90\%} - x_{10\%}}{\overline{x}_{med}} \qquad \text{(for non-analytical mean)} \qquad 2.11b$$

When the mean is close to zero value, the coefficient of variance is out of use, so it should only be applied when all the observations are always positive or always negative. Note that the coefficient of variation is always positive

(Birpınar 2003). Determination of coefficient of variance helps the researchers to investigate the existence of say inter-annual variability of annual totals over the study area. When CV is less than or equal to 1 implies stable trend, otherwise unstable.

2.2.3 Asymmetry or Skewness (C_s)

Skewness is the degree of asymmetry of a distribution which is a dimensionless value. It gives how the studied data is skewed from the normal distribution. If a distribution is symmetrical, the value of skewness is zero. Hence it can be used to detect if the data deviates from the normality. If it is positively distributed, it has a long tail at right side and similarly if it is negatively distributed it has a long tail at its left side. For the analytical mean case, it is referred as the coefficient of skewness and is expressed for a sample as:

$$C_S = \frac{n\sum_{i=1}^{n}(x_i - \overline{x}_{ar})^3}{(n-1)(n-2)s_d^{\,3}} \quad \text{(for analytical mean)} \qquad 2.12a$$

Note that $C_s = 0.00$ implies normal distribution otherwise skewed.

For the non-analytical mean, the coefficient of skewness is referred as the percentile skewness coefficient and is given as:

$$PCx = \frac{x_{90\%} - x_{50\%} - (x_{50\%} - x_{10\%})}{x_{90\%} - x_{10\%}} \quad \text{(for non-analytical mean)} \qquad 2.12b$$

2.2.4 Peakedness or Kurtosis (γ_k)

Basically means ‘Bulginess’ in Greek language, where kurtosis implies the degree of ‘flatness’ or ‘peakedness’ of the data. Hence it can be used to detect if the data deviates from the normal (bell-shaped) curve. For a sample, it is given as:

$$\gamma_k = \frac{n(n+1)\sum_{i=1}^{n}(x_i - \overline{x}_{ar})^4}{(n-1)(n-2)(n-3)s_d^4} \quad \text{(for analytical mean)} \qquad 2.13a$$

Note that γ_k = 3.000 implies normal distribution

$$P\gamma_k = \frac{(x_{75\%} - x_{25\%})/2}{x_{90\%} - x_{10\%}} \quad \text{(for non-analytical mean)} \qquad 2.13b$$

Note that $P\gamma_k$ = 0.263 implies normal distribution.

2.3 The Probability Distribution Functions (PDF)

Any quantity which is defined as a random variable can be mathematically expressed ascribing a suitable probability distribution function to it. The simplest type of probability distribution is the uniform distribution, whose probability density function is a rectangle. Its magnitude-probability distribution is very simple but unfortunately almost none of the hydro-meteorological variables are obeying to this distribution. The most widely known continuous probability distribution is the normal distribution (normal curve, or Gaussian distribution) and all togather hundreds of different probability distributions are said to be available. Yet there may not exist a clear-cut deduction mechanism for some distributions as they may evolve as mathematical expediences. There are some special distributions which are used for statistical tests rather than depicting the probabilistic behaviour of some particular physical random quantity like Chi^2 and students' t.

Unfortunately, inspite of their analytical innocent appearance, analytical integral of the most probability density functions are impossible including the normal distributions as well. Tables were prepared to expess the numerical approximate solutions by the professional numerical analysts for many distribution functions to help the practitioners in this field.

Probability density functions curves arising in practice take on certain shapes, like symmetrical (bell-shaped), skewed (positively or negatively), J- or reverse J-shaped, U-shaped, bimodal or multimodal etc… The probability that of a random variable which is less than or equal to a specific value of x based on its cumulative data is called the cumulative density function and is mathematically obtained through the integral.

$$F(x) = \int_{-\infty}^{\infty} f(t)dt \qquad 2.14$$

Probability distribution is a function that allocates a probability to every interval of real numbers where the basic concepts in statistics are in calculating within the required confidence intervals, to determinate a reasonable distribution model by checking the hypothesis through best fitting methods (Kimyacı 2004).

2.3.1 Normal / Log-Normal Distribution Family

Under very genaral conditions, as the number of variables (i.e. observations, data) in the sum becomes larger, the distribution of the sum of random variables will approach to the normal distribution forming a bell-like shape. In short, the normal or Gaussian distribution defines those random variables which are formed by the additive effects of so many other variables and it is brifly defined as the distribution of sums. Since the random variation in many phenomena arise from a number of additive variations, owing to its analytical tractability and to the familiarity of many engineers with the distribution, the normal model is very often used in practice when there is no reason to believe that an additive physical mechanism exists. It may be sensed that, the normal distribution which is a general purpose and a popular distribution in statistics is a panacea (cure-for-every-disease) type of a distribution. Unfortunately it is not correct, especially as far as the hydro-meteorological random variables are concerened. Therefore, the normal distribution in its conventionally known form is rearly used in water resources engineering. However, it is still one of the most significant distributions, simply because first there are 2-parameter normal distribution (also known as log-Normal) and 3-parameter log-Normal version of it, and secondly, there have been quite a few attempts to convert the observed sample distribution to the normal by some sort of a mathematical transformations.

The standard equations of this family are:

i. Normal distribution $x = \bar{x} + zs_d$ 2.15

ii. Log-Normal distribution $\log x = \overline{\log x} + zs_{\log x}$ 2.16

where logx implies the logarithm of the x value, $\overline{\log x}$ is the average of the logarithmic x values and $s_{\log x}$ is the standard deviation of the logarithmic x values.

2.4 Plotting Positions

Probability of an event can be obtained with the help of plotting positions. After finding the values through the selected equations, these data should be drawn on appropriate probability graphs with the help of plotting positions. Famous plotting positions are tabulated as below (Mutreja 1990).

Table 2.1: Mostly used plotting positions

California	**Modified California**	**Hazen**	**Chegodayev's**	**Weibull**	**Blom**	**Gringorten**	**Tukey**
$\frac{m}{N}$	$\frac{m-1}{N}$	$\frac{2m-1}{N}$	$\frac{m-0.3}{N+0.4}$	$\frac{m}{N+1}$	$\frac{m-(3/8)}{N+(1/4)}$	$\frac{m-0.44}{N+0.12}$	$\frac{3m-1}{3N-1}$

Weibull plotting positioning for rainfall of this study being automatically selected by Minitab 16® software.

2.5 Elementary Sampling Theory

Sampling theory is a study of relationships existing between a population and samples drawn from that population. From the practical viewpoint, however, it is often more important to be able to infer information about a population from samples drawn from it. Hence, determining the sample statistics and generalizing it for the population parameters is widely used in most of the engineering approaches. Although the population composed of infinite number of observation size, the sample being assumed to be the representative of that population has a finite size. Usually, if the number of observations is < 30, then this set of data is refered as the sample (Spiegel 1999 and Seyhan 1994). But as it is clear, there is no any lower limit that bounds the sample and even in some cases, observation size of 30 (being the upper limit) may not be an enough observation size so as to represent the population that it is drawn from. Although the above mentioned problems having outmost importance in statistical measures, unfortunately either less attention was paid or even ignored in some cases and most probably the gathered, obtained or extracted data gives irrelevent and/or unappropriate results and hence guiding the researchers wrongly. Hence, from the practical point of view, it is necessary to check,

a) the existing sample size appropriateness, as well as

b) the sample-population relationship appropriateness.

To overcome the required sample size appropriateness through determining the minimum sample size ‘n’ requirement which is in fact varies from population to population, there exists no rule or guideline, so to overcome this weakness, (determining the minimum required sample size ‘n_{min}’), at least one of the possible rule of thumb solutions listed below could be adopted (Sen 2003):

i- The required minimum sample size ‘n_{req}’ is reached, if no significant variation (to an acceptable level) occurs based on the mean values as the sample size number ‘n’ increases;

- n_{req} based on the means = $\frac{\bar{x}_{n+1}}{\bar{x}_n}$ < to some acceptable level say 0.1 (i.e.10 %)

ii- The required minimum sample size ‘n_{req}’ is reached, if no significant variation (to an acceptable level) occurs based on the standard deviations as the sample size number ‘n’ increases;

- n_{req} based on the standard deviations = $\frac{s_{d\ n+1}}{s_{d\ n}}$ < to some acceptable level. In this study 0.9 (i.e.90 %) is selected.

iii- The required minimum sample size ‘n_{req}’ is reached, if no significant variation (to an acceptable level) occurs based on the standardized values as the sample size number ‘n’ increases;

- n_{req} based on the standardized variables = $\frac{x-\bar{x}}{s_d}$ < to some acceptable level.

2.6 Confidence Interval (α %)

It is very often in engineering to make decisions about populations on the basis of sample information. Usually the mean and the standard deviation are the two parameters in use for comparison provided that the population and the sample are obeying the normal distribution. So is the expected that any data lies within the interval of mean and plus minus any multiples of standard deviation. A frequency curved progressed from sample data is the best approximation of the population curve. If someone was to use this distribution to hypothesize about true value, he/she could select a level of confidence about the statement and determine limits between which one could expect the through value to lie with that arbitrarily selected percent confidence. Usually this level is selected as 95 percent and hence the interval between these limits is termed as a confidence interval. Due to limited sample size, instead of a single value depending on the problem type either one-sided or two-sided confidence intervals can be developed. A two-sided confidence interval provides both upper and lower limits. For one sided confidence interval, provides either upper or lower limit value, but not the both. Hence, for the above mentioned level (i.e. 95 %), the confidence interval is 90 percent by considering both upper and limits and the range of data will be given based on this expected percent confidence level. Therefore, to express any confidence interval, the expected degree of confidence level should be first fixed for any data and then, depending on the type of the problem either one-sided or two-sided confidence intervals will be selected. So the confidence interval gives an estimated range of values which is probably to include an unknown population parameter. The estimated range is being calculated from the given set of the sample data. Confidence interval can be used for mean, standard deviation, etc. It is mostly indicated by the Greek letter ‘α’. This interval is referred as the confident region where one can expect to find any data that may exists within that range with such probability level. Usually 95 % (z=1.96) and 99 % (z=2.58) confidence levels are in practical use (Spiegel 1999).

- Confidence Interval for Mean

$$\bar{x} \pm z \frac{s_d}{\sqrt{n}} \qquad 2.17$$

- Confidence Interval for Standard Deviation

$$s_x \pm z \frac{s_d}{\sqrt{2n}} \qquad 2.18$$

2.7 Degree of Freedom (df)

In order to compute any statistical calculations usually it is necessary to use the observations obtained from a sample as well as certain population parameters. If these are not known; which is the case in most of the hydro-meteorolocial studies, they must be estimated. The number of degrees of freedom of a statistics implies the existing number of independent variables within the sample minus the number of population parameters used so as to estimate the sample. In other words, it is described as the figure of autonomous observations (n) minus the number of population parameters which are estimated from the sample observations (usually the mean and standard deviation). For example in *t*-test since there were 2 parameter for the test to be defined (mean and standard deviation), degree of freedom would be expressed as n-2, but in F-test since the standard deviation was the only used parameter, the degree of freedom should be used by n-1.

2.8 Statistical Hypotheses

There are generally the statements about the probability distribution of the populations. In many instances, a statistical hypothesis is formulated for the sole purpose of rejecting or nullifying. The whole hypothesis cannot be used to prove it is correct but instead works on rejections. The null hypothesis, denoted by H_0 is the nominal or the simple case and the alternate hypothesis denoted by H_1 is based on the departure from H_0 that most of the hydro-meteorologists expect to have. The procedures that enables to determine whether the observed samples differ significantly from the results expected and thus helps to decide whether to accept or reject hypothesis are called test of hypotheses or rules of decisions.

If one rejects the hypothesis when it should be accepted that indicates Type I error. If one accepts a hypothesis that should be rejected that is referred as Type II error. Note that in both case wrong decision in judgement has occurred. In order the decision hypotheses to be good, they must be designed so as to minimize errors in decision.

This is not a simple matter, the only way to reduce both types of errors is to increase the sample size which may or may not be possible (Spiegel 1999).

In testing a given hypothesis, the maximum probability with which one would be willing to risk a Type I error is called the level of significance of the test. This probability is often denoted by α, and is generally specified before any samples are drawn so that the results obtained will not influence the choice. In practice significance level 0.05 or 0.1 is customary, although other values like 0.01, 0.005 and 0.001 are as well used for some specific cases. If, for example, the 0.05 (5 %) significance level is chosen in designing a decision rule, then there are about 5 % chances that one would reject the hypothesis when it should be accepted; that is about 95 % confident one made the right decision. In such case it is said that the hypothesis has been rejected at the 0.05 significant level which means that the hypothesis has a 0.05 probability of being wrong (Spiegel 1999).

Depending on characteristics of the population, the hypotheses can be carried out for two-sided (two-tailed) or one-sided (one-tailed) tests. Often the hydro-meteorologists interested only in extreme values of one side of the mean (like testing one process is better than the other i.e. one-sided; which is different from testing whether one process is better or worse than the other i.e. two-sided).

In this study one-tailed test is used. It means null hypothesis is accepted if the P-value is bigger than 5%.

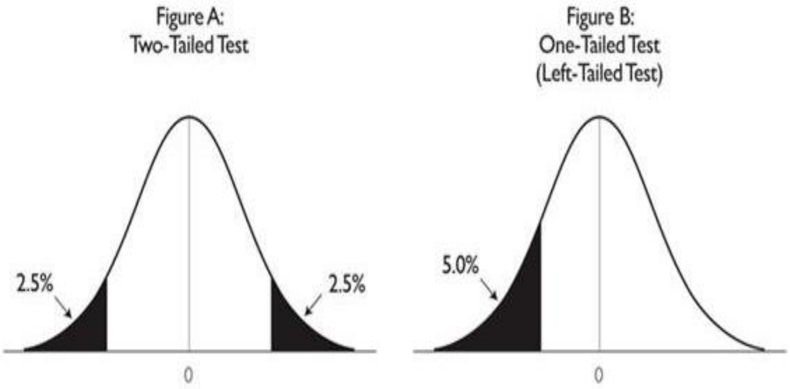


Figure 2.1: One-tailed versus two-tailed hypothesis tests

Hydrological processes are conventionally regarded as stationary process. However, there is a growing evidence of trends and long-term variability which may be related to anthropogenic influences and the natural features of the climate system. These processes are based on long-term trends. Hence appropriate parametric or parametric free (non-parametric) tests should be adopted to evaluate the significance of the trend existence. Two types of trends including monotonic trend and step (shift) change are usually considered in climatological and hydrological variables. In trend tests, the null hypothesis H_0 is that there is no trend in the population from which the data variable is drawn and H_1 implies there is a trend in the records. Parametric and non-parametric methods are usually used for trend detection. The non-parametric tests are more robust compared to their parametric counterparts. Among non-parametric tests Mann-Kendal test is the best choice for detecting monotonic trend while Mann-Whitney test is a good alternative for step change detection which is widely used for the homogeneity control of the data.

2.9 Tests for Quality Control of the Data

In the beginning of the 20th century, as the countries started to get developed their environments, realize the importance of the hydro-meteorological stations and then onwards develop different instrumentation techniques to measure different parameters.

It is known that, the management of water resources has always been subject to a variety of sources of uncertainty, not least of which has been the natural variability of the climate. Such considerations are now compounded by the possible influence of anthropogenic ally-related climate change, the investigation of which places a premium on long, homogeneous instrumental records of both hydrological and hydro-meteorological variables. Unfortunately, in many countries hydrometric networks have been subject to disruption owing to a variety of causes, ranging from rationalization in the interests of cost-cutting to civil unrest. Indeed, it was observed that inadequate and unreliable data constitute a serious constraint to efficient water management. In these circumstances, the analysis of hydrological and hydro-meteorological time series requires increasing vigilance and the application of at least a minimum amount of data screening. The procedure recommended by Dahmen 1990, consists of five steps:

1. Plotting the data for visual examination and checking the straightness of the established inclined line,
2. Testing the time series for the presence of linear trend,
3. Testing for the stability of the means and variances in split-record samples drawn from the time series,
4. Testing for the presence of significant serial correlation, and
5. Testing relative consistency and homogeneity with other data.

Visual inspection of a plot of the data in order to locate potential change points, i.e. jumps in the mean, can be assisted by the application of a non-parametric statistical tests that are simple straightforward in application, but does not address the related problem of how to proceed once the time series fail any of the tests. In such circumstances, at least initially, recourse should be made to the station metadata (the data about the data) in order to check for possible changes in instruments and their siting or changes in observational practices (inconsistency), or the station environment (non-homogeneity). Such information is not always readily available, but even if a station history is available, the adjustment of suspect records generally requires the deployment of more sophisticated algorithms. In particular, the detection of an overall long-term movement in a time series by a Spearman rank correlation test raises the further question as to its actual duration and timing.

The answer can be obtained by the application of further, more sophisticated (parametric) tests for the detection of jumps and trends even when the timing and duration of an apparent trend have been quantified, a decision is required on its authenticity: could the movement be a reflection of long-term climate variability, or is it an artefact of the instrumentation or its environment. The following case study illustrates the need for objective consideration of the results from data screening, particularly in the wider context of regional weather systems and their variability.

Long records of hydrological and hydro-meteorological variables are of inestimable value for the planning, design and management of water resource systems. Unfortunately, such long series invariably exhibit inconsistencies and non-homogeneities arising from a wide variety of causes – for example, changes in instruments and observation practices, and alterations to the general environment of the instruments themselves. Widely used hydro-meteorological data are having only limited sample sizes that are subsets of a very large population. Hence, to be able to analyse them statistically through any test, a prior importance should be taken since the establishment of each test has its own mathematical limitations. Among those limitations, to characterize the data, the below mentioned test could be used; to detect any sample-population relationship appropriateness, in literature, various statistically defined parametric and non-parametric tests are available. A test based on parametric assumption like mean, standard deviation, skewness, etc. is called parametric test such as ANOVA (*t*-test, F test) and moving average etc. A non-parametric (parameter free) test, consequently is a test that does not need parametric assumption, for instance Mann-Kendall test, Sen's estimator of Median slope etc.

Both non-parametric and parametric statistical tests are available to detect the presence of long-term movements in recorded time series. The interpretation of results from such testing has often to be carried out in the absence of sufficient station metadata, for inconsistency and non-homogeneity and should be interpreted in the context of prevailing weather systems. Parametric methods cannot be used for the analysis of rainfall in general since usually they do not obey to normal distribution, hence non-parametric methods should be adopted. The trends in rainfall totals identified could therefore be interpreted as arising from natural variability or even greenhouse gas forcing rather than from any inconsistency and non-homogeneity. It should not be forgotten that, the gathered data are having only limited sample sizes being the subsets of a very large data population hence, to be able to analyse these data statistically through any appropriate test a prior importance should be taken since the establishment of each statistical test has its own mathematical limitations. Among those limitations, the below mentioned characteristics of the data plays outmost importance:

2.9.1 Normality Test

In statistics, normality tests are used to determine if a data is well-modelled by a normal distribution and to compute how likely it is for a random variable underlying the data to be normally distributed. Hence, assessment of the normality of data is a prerequisite for many statistical tests because normally distributed data is an underlying assumption in all the parametric testing. In other words, application of most of the statistical methods requires the data to behave in a Gaussian fashion. There are two main methods of assessing normality:

i. graphical

An informal approach to testing normality is to compare a histogram of the sample data to a normal probability curve. The empirical distribution of the data (the histogram) should be bell-shaped and resemble the normal distribution. This might be difficult to see if the sample is small. In this case one might proceed by regressing the data against the quantiles of a normal distribution with the same mean and variance as the sample. Lack of fit to the regression line suggests a departure from normality. Even drawing the data either on the probability or normal distribution graph paper with the help of the plotting position and the points plotted should fall approximately on a straight line, indicating high positive correlation of normality.

ii. numerical

Numerically Normality of any data can be tested through parametric and/or non-parametric tests as given in relevant literature.

Parametric Tests

- D'Agostino's K-squared test
- Jarque–Bera test
- Coefficient of Variance (CV) [where CV < 25% implies normality]
- Comparing the mean and the median (or the logarithmic mean with its logarithmic median) values.
- Anderson–Darling test.

Non-parametric Tests

- Kolmogorov–Smirnov (K-S or KS) test
- Shapiro–Wilk test
- Pearson's Chi-square test
- Shapiro–Francia test.

In this study, for the Normality test Anderson-Darling (parametric) test is done through Minitab 16® software.

- **Anderson-Darling Test**

This test compares the ECDF (empirical cumulative distribution function) of the sample data with the distribution expected if the data were normal. If the observed difference is adequately large, the null hypothesis of the population normality should be rejected.

Because this test for each region is done by Minitab 16® 16 software, the theory is not explained here in details. If the p-value that is given by software will be equal or greater than 5%, then it is concluded that, the data set is normally distributed.

2.9.2 Homogeneity Test

Homogeneity (its opposite, heterogeneity), relates the statistical properties of any one part of an overall data are the same as any other part. Homogeneity can be studied to several degrees of complexity among them homoscedasticity which examines how much the variability of data-values changes throughout a data. It is used to determine whether frequency counts are distributed identically across different data subset groups. A test of homogeneity compares the proportions of responses from two or more populations. In other words homogeneity tests determine if within the time series data there is a specific time period at which a change within the data occurs.

Homogeneity test of any data can be tested through parametric and/or non-parametric tests.

Parametric Tests

- Alexandersson’s Standard Normal Homogeneity Test (SNHT)
- Buishand Rnge Test (BR)
- ANOVA test
- Von Neumann Test (VNR)

Non-parametric Tests

- Mann-Whitney-Pettitt test
- Pearson's Chi-square test

In this study, the rainfall of each region is checked for Homogeneity among the above mentioned 4 tests (SNHT, BR, VNR, and Pettitt). ANOVA test (t-test, F-test) is also used to check the Homogeneity of each regions data between the nearby regions.

2.9.2.1 Standard Normal Homogeneity Test

A statistic T(y) is used to compare the mean of the first y years with the last of (n-y) years and can be written as bellow:

$$T_y = y\overline{Z_1} + (n - y)\overline{Z_2},\ y = 1,2, \dots , n \qquad 2.19$$

$$\overline{Z_1} = \frac{1}{y}\sum_{i=1}^{n}\frac{(Y_i - \bar{Y})}{Sd} \quad \text{and } \overline{Z_2} = \frac{1}{n-y}\sum_{i=1}^{n}\frac{(Y_i - \bar{Y})}{Sd} \qquad 2.20$$

The year y consisted of break if value of T is maximum. To reject null hypothesis the statistic,

$$T_0 = \max T_y \qquad 2.21$$

Is greater than the critical value, which depends on the sample size (Kang 2012).

2.9.2.2 Buishand Range Test

The adjusted partial sum is defined as

$$S_0{}^* = 0 \text{ and } S_y{}^* = \sum_{i=1}^{y}(Y_i - \overline{Y}) \ , \ \ y = 1,2,\dots,n \qquad 2.22$$

When the series is homogeneous, then the value of $S_y{}^*$ will rise and fall around zero. The year y has break when $S_y{}^*$ has reached a maximum (negative shift) or minimum (positive shift) . Rescaled adjusted range, R is obtained by

$$R = \frac{\max S_y{}^* - \min S_y{}^*}{Sd} \qquad 2.23$$

The $R/\sqrt{n}$ is then compared with the critical values given by Buishand 1982.

2.9.2.3 Pettitt Test

This test is based on the rank, r_i of the Y_i and ignores the normality of series.

$$X_y = 2\sum_{i=1}^{y} r_i - y(n+1) \ , \ \ y = 1,2,\dots n \qquad 2.24$$

The break occurs in year k when

$$X_k = \max|X_y| \qquad 2.25$$

The value then is compared with the critical value by Pettitt(1979).

2.9.2.4 Von Neumann Ratio Test

It is a test that used the ratio of mean square successive (year to year) difference to the variance. The test statistic is shown as follows:

$$N = \frac{\sum_{i=1}^{n-1}(Y_i - Y_{i+1})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \qquad 2.26$$

In this study XLSTAT software is used to do these four Homogeneity tests. The software gives p-value for each region, if the p-value is bigger than 5% (significant level alpha) the time will be homogeneous (Kang 2012).

2.9.2.5 ANOVA (t-test, F-test)

In order to check the homogeneity, correlation and comparison of any two sets of data, a common method called Analysis of Variances 'ANOVA' is used. Student's *t*-test and Fisher's F-test are mostly used distributions for this purpose. The formulations of these tests are given as below; beside the formulation, the determined answers from the equations should be checked by appropriate tables of t-test and F-test given in the appendix, based on the degrees of freedom and the interested confidence intervals. If the obtained value is less than the calculated critical value, the test proves the homogeneity and the test is hence assumed to be acceptable. In fact *t*-test is comparing the means and F-test is comparing the standard deviations of the data (Salvatore 1982).

2.9.2.5.1 t-test

$$t = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{s_{d\,x}^2}{n} + \frac{s_{d\,y}^2}{m}}} \qquad 2.27$$

where $\overline{x}$ and $\overline{y}$ are the means of the data sets.

$s_{d\,x}$ and $s_{d\,y}$ are the standard deviations of data sets and

n and m are the number of data available for each data set (x and y).

2.9.2.5.2 F-test

$$F = \frac{s_{d\,x}^2}{s_{d\,y}^2} \qquad 2.28$$

The important note in F-test is that, the smaller value of the standard deviation should also be at the numerator and the larger value should be at denominator.

2.9.3 Consistency Test

Consistency is another desired property for any data. It checks whether or not any data within the data is reasonable. In other words, it checks if there is a surprise data (outlier) compared with the similar family of data. For example, records for rainfall within an area might be increased in three ways: records for additional time periods; records for additional sites with a fixed area; records for extra sites obtained by extending the size of the area. In such cases, the property of consistency may be limited to one or more of the possible ways a sample size can grow.

Parametric Test

- t - test

Non-parametric Test

- Double mass curve

To check the consistency of time series in this study, the double mass curve method is used.

2.9.3.1 Double Mass Curve

Double mass curve is a fundamental tool in data analysis. It is a plot of cumulative values of one variable against the cumulation of another quantities during the same time period. The theory of double mass curve is that, when cumulation of two quantities is drawn, they represent straight line. If there is a break in this continuous line, it means that there is a systematic error and it requires to be corrected. Conversely if, there is no break or change of slope within the line, it could be concluded that, the two sets of compared data are consistent. Correction of the data can be done by multiplying a constant ratio based on slopes.

$$P_{\text{adjusted}} = \frac{M_a}{M_0} P_{\text{observed}} \qquad 2.29$$

where M_a is the slope of the line before the abrupt change and M_0 is the slope of the systematic errors line. Following figure illustrates the error and the way the correction should be done (Usul 2005).

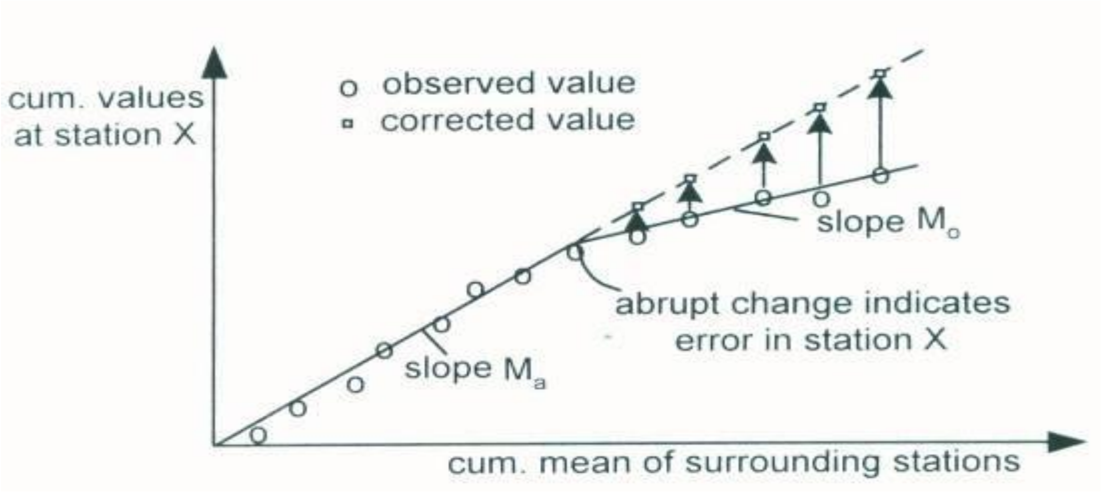


Figure 2.2: Double mass curve due systematic error and its correction method (Usul 2005).

2.9.4 Independency Test

This test is used to check if two categorical variables are from a single (common) population. The test evaluates if the existence of association between two categorical variables (uncorrelated, unrelated) are independent of each or not. Recall that two events are independent when neither event influences the other. It can be done by comparing means and/or standard deviations of the two categorical data. Independence violates the assumption of serial correlation of data and implying that there is no short-term correlation between the observations (data). A common method called Analysis of Variances 'ANOVA' is adopted where the Student's *t*-test and Fisher's F-test are the mostly used distributions that consider the degrees of freedom and the interested confidence intervals. In fact *t*-test is comparing the means and F-test is comparing the standard deviations of the data (Salvatore 1982).

Parametric Tests

- t-test (mean)
- F- test (variance)
- Portmanteau test

Non-Parametric Tests

- Pearson's Chi-square ($\chi 2$) test
- Seasonal Kendall test
- Wilcoxon signed-rank test (Mann–Whitney U test or Mann–Whitney–Wilcoxon (MWW) or Wilcoxon rank-sum test (WRS) or Wilcoxon–Mann–Whitney test)

2.9.5 Trend Test

Trend is a change in the level of data series, usually over the time but sometimes in space. It is a general increase or decrease in the observed values of random variable over a time. In most cases, it is not generally possible to detect trends that are not apparent by inspection, especially for data records of short to moderate length - say 20 years or less. Testing the existence of linear (monotonic) trend (serial correlation) within the whole time series is important in hydro-meteorological data. Testing for the existence of linear (monotonic) trend within the whole time series can be done through parametric and/or non-parametric tests (Spiegel 1999).

Parametric Test

- Linear Regression analysis (or Pearson correlation ‘r’)

Non-parametric Tests

- Mann-Kendall Rank (‘tau’ or ‘τ’) test [< ± 2 indicates absence of a significant trend]

- Theil–Sen's trend estimator (or Sen's median slope estimator, or Kendall robust line-fit method or Kendall–Theil robust line) [suggests a magnitude to the long term data, i.e. trend]
- Spearman's rho ('**ρ**' or 'rank') test.

In this study Mann-Kendall and Sen's Median slope (non-parametric) tests are used to check the trend of data .

2.9.5.1 Mann-Kendall (τ) and Sen's Median Slope Tests

Mann-Kendall test is a non-parametric test that is used to find trend in time series. It was suggested by Mann (1945) and Kendall (1975). Mann-Kendall test also referred as Kendall's Tau 'τ' test. Mann-Kendall test is used to measure the connection of two sets of data. When one set of data is time then this test is used to point out the trend (Birpınar 2003). The test statistic is founded by

$$\tau = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \text{sign}(X_i - X_j) \qquad 2.30$$

where 'τ' is approximately converging to normal distribution stated as $N_{(0,s^2)}$, if 'τ' is positive, it illustrates that the trend is increasing and if it is negative, it means the trend is decreasing. Standard deviation S_x is also described as

$$s_x = \sqrt{n(n-1)(2n+5)/18} \qquad 2.31$$

After obtaining the 'τ', Median slope should be obtained through Sen's method. Sen's method for the approximation of slope needs a time series of equally spaced data. Sen's method proceeds by calculating the slope as a change in measurement per change in time. The equation is given as below (Sharifi 2006):

$$Q = \frac{x_j - x_i}{j - i} \qquad 2.32$$

Table 2.2: Sen's method procedure (Sharifi 2006)

Time	**1**	**2**	**3**		**T**
Data	X_1	X_2	X_3	**...**	X_T
—	-----	$\frac{X_2 - X_1}{2-1}$	$\frac{X_3 - X_1}{3-1}$ $\frac{X_3 - X_2}{3-2}$ ...		$\frac{X_T - X_1}{T-1}$ $\frac{X_T - X_2}{T-2}$ $\frac{X_T - X_3}{T-3}$ . . $\frac{X_T - X_{T-2}}{T-(T-2)}$ $\frac{X_T - X_{T-1}}{T-(T-1)}$

After finding all the values, they will all be shifted to one column consecutively namely, X_1, X_2, X_3… X_T, the procedure is repeated all after each other until the data squeezes in one column. Therefore the median of this recent column is found. This number is called Sen's Median slope. In this study, XLSTAT software is used to apply Mann-Kendall and Sens Median Slope tests.

2.9.6 Stationary Test

The purpose of this test is to determine if the mean values and variances of the series vary with time. Stationary time series implies that none of the data varies with time series. In literature, the widely used stationary tests are:

- Priestley-Subba Rao (PSR) Test
- Phillips-Perron Unit Root Test
- Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests
- Augmented Dickey – Fuller (ADF) test

In this study, the ADF test is used to check Stationarity of time series.

2.9.6.1 Augmented Dickey–Fuller Unit Root Test

Augmented Dickey–Fuller unit root test (ADF) as its name refers is a test for a unit root in any time series. Being augmented implies larger and more complicated set of time series models.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta_{y_{t-1}} + \cdots + \delta_{p-1} \Delta_{y_{t-p-1}} + \varepsilon_t \qquad 2.33$$

where α is a constant, β the coefficient on a time trend and p is the lag order of the autoregressive process. By including lags of the order p the ADF formulation allows for higher-order autoregressive processes. This means that the lag length p has to be determined when applying the test. The unit root test is then carried out under the null hypothesis $\gamma \geq 0$ against the alternative hypothesis of $\gamma < 0$

In this study Autoregressive order 1 (AR1) is used, therefore:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t \qquad 2.34$$

$H_0: \gamma \geq 0$ implies that the data is stationary,

$H_1: \gamma < 0$ implies that the data is non-stationary.

2.10 Procedure and Sample Calculations

2.10.1 Determination of minimum number of data required

The minimum number of data required for any statistical study depends on the range of the available data, its average, and the expected degree of acceptable level of deviation. For this statistical study, to determine the minimum number of required data n_{min}, an empirical method suggested by Sen 2003, that bases on two simple tests were performed and the minimum value n_{min}, which satisfies both of these tests, is assumed to be the answer of the above mentioned problem. The tests are a kind of altering average tests. The test starts from number of two data (the value of n=1 and n+1=2) and compares the mean deviation of the averages and standard deviations of these n and n+1 values with the subsequently verified mean and standard deviation values that were computed based on one less than number of data within acceptable degree. In this study, the variation of the average being less than 2% and variation of standard deviation being less than 5% , were selected tentatively (Sen 2003). By comparing the

n_{min} values, the biggest will be selected so as to satisfy the limitation of the both empirical equations.

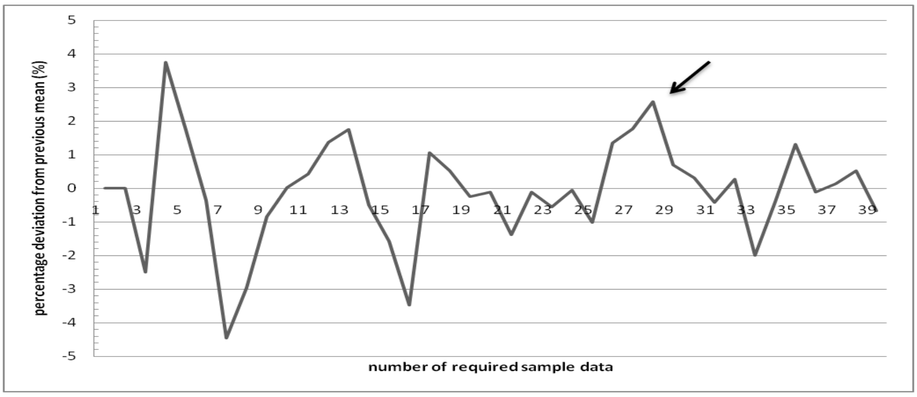


Figure 2.3: Curve showing the required number of sample size for Central Mesaria regions' rainfall, based on the percentage deviations of the mean values.

As it can be seen through the graph in Figure 2.3, minimum number of data for Central Mesaria region is 28 years, because after that, the fluctuation was less than ±2 percent.

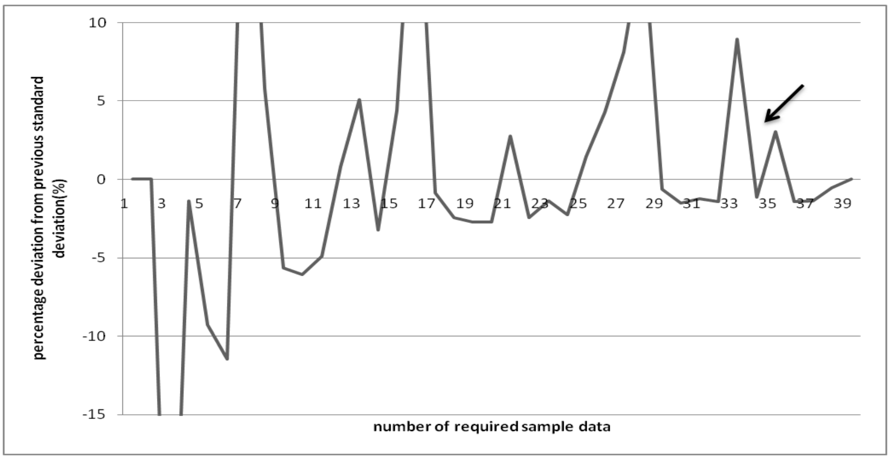


Figure 2.4: Curve showing the required number of sample size for Central Mesaria regions' rainfall, based on the percentage variation of the standard deviations

As it can be noticed in Figure 2.4, the minimum number of data based on standard deviation for Central Mesaria is chosen to be 34 years, considering the fluctuations less than ±5 percent.

Consequently, based on the comparisons of means and standard deviations, the required minimum number of data for any statistic and probabilistic study for Central Mesaria rainfall is 34 years.

Table 2.3: Appropriate rainfall sample size of Central Mesaria for statistic and probabilistic studies

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
28 < 39 OK	34 < 39 OK

2.10.2 Test of Normality

The Normality test is done by Minitab 16® software;

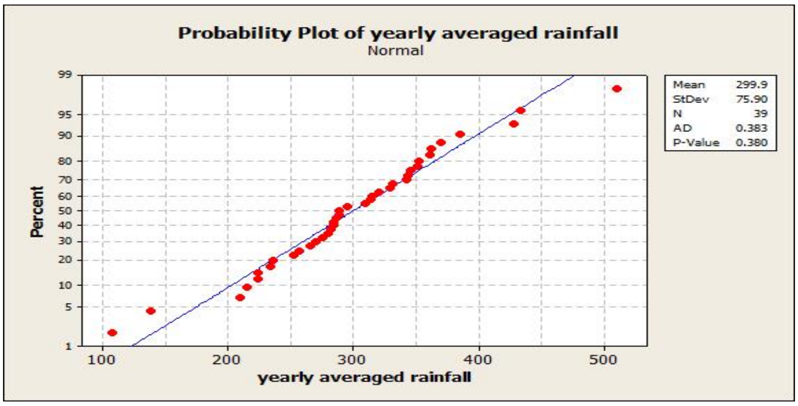


Figure 2.5: Normality test of Central Mesaria rainfall

Result: p-value = 0.38 > 0.05. Therefore, Central Mesaria rainfall is normally distributed.

2.10.3 Test of Homogeneity

In order to check the homogeneity and correlation between two sets of data, student's t-test and Fisher's F-test are used. For t-test the following steps are followed:

2.10.3.1 Procedure for t-test

1. The means of the two sets $\overline{x}$ and $\overline{y}$ were found.
2. The standard deviations of the two sets $s_{d\,x}$ and $s_{d\,y}$ were determined.

3. Apply the Equation 2.27.
4. Fixing the degree of freedom and confidence interval and referring to table of t-distribution, allowable (critical) value of t is found.
5. Comparing the t value with critical t, if the calculated value is smaller, test is acceptable and correlation exists otherwise the two sets of data are not correlated (Spiegel 1999).

2.10.3.2 Sample calculation for checking the rainfall correlations using t-test between Central Mesaria with two other regions (East Coast and North Coast)

Table 2.4: t-test results between rainfall of Central Mesaria and East Coast regions.

n,m	39,39
Central Mesaria Mean	299.9
East Coast Mean	334.7
Central Mesaria Sd	75.9
East Coast Sd	92.8
Degree of freedom	37
Level of confidence	95%
Allowable t	1.69 (Appendix A)

Obtained t= -1.5608 < 1.69 Acceptable

As it can be seen, the calculated *t* is less than allowable *t* therefore the correlation exists between Central Mesaria and East Coast. Whereas, for the data below it is found that no correlation exists between Central Mesaria and North Coast regions.

Table 2.5: t-test between rainfall of Central Mesaria and North Coast regions

n, m	39,39
Central Mesaria Mean	299.9
North Coast Mean	461.9
Central Mesaria Sd	75.9
North Coast Sd	100.62
degree of freedom	37
level of confidence	95%
Allowable t	1.69 (Appendix A)

Obtained t= -6.2458 > 1.69 Unacceptable

2.10.3.3 Procedure for F-test

This test is based on comparing the standard deviations. The procedure is:

1. Determine the standard deviations of two set s_x and s_y.
2. Obtain the deviation of smaller value over the larger value. As given in Equation 2.28.
3. Like t-test, considering the confidence level and degree of freedom, the allowable F value would be read from the appropriate table.
4. The obtained F value and the allowable F value were compared and if the calculated F is smaller than allowable F, a correlation between two sets exists. Otherwise there is no correlation between two sets (Sharifi 2006).

2.10.3.4 Sample calculation for checking the rainfall correlations using F-test between Central Mesaria and East Coast regions.

Table 2.6: F-test of rainfall between Central Mesaria and East Coast regional data and its result

n	39
Central Mesaria Sd	75.9
East Coast Sd	92.8
degree of freedom	38
level of confidence	95%
Allowable F	1.72 (Appendix B)

$F = 0.6684 < 1.72$ Acceptable

2.10.4 Test of Consistency

Checking if the data is consistent and lies within the data collected from neighboring regions or not is carried out through the double mass curve method. Steps of applying double mass curve are as follows:

- The accumulation of the desired parameter in the studied region (station) is found,

- The accumulation of the average of the desired parameter over the nearby regions (station) is calculated.
- A graph is drawn of which its x-axis is cumulative average of the parameter over nearby regions and its y-axis represents the cumulative of desired parameter over the studied region (Usul 2005).

The consistency of rainfall between Central Mesaria and rainfall averages of other nearby 5 meteorological regions of TRNC based on double mass curve method is obtained and shown in Figure 2.6.

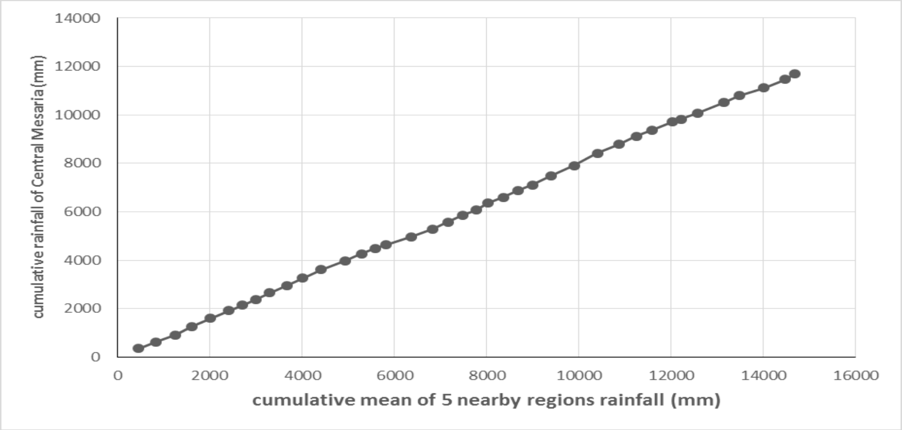


Figure 2.6: Central Mesaria region rainfall consistency check through double mass curve method with respect to nearby 5 other meteorological regions average rainfall

It is found that, Central Mesaria region rainfall is consistent with the nearby 5 meteorological regions average rainfall.

2.10.5 Test of Trend

In order to check if there is a trend within the data, non-parametric (Mann-Kendal and Sen Median Slope tests) were used in this study.

2.10.5.1 Mann-Kendall and Sen’s Median Slope Tests

This test is performed through the XLSTAT software where the p-value of Mann-Kendall was computed with a confidence interval of two sided α (alpha) = 5%.

Hypothesis of this test:

H_0: There is no trend within the series,

H1: There is a trend within the series.

Hence, p-value of Mann-Kendall greater than 5% (two sided) implies no trend within series.

The sample output of Mann-Kendall and Sens Median Slope tests for Central Mesaria rainfall obtained from XLSTAT is given below.

XLSTAT 2015.4.01.20978 - Mann-Kendall trend tests - on 8/26/2015 at 6:39:46 PM
Time series: Workbook = Book1 / Sheet = central mesaria / Range = 'central mesaria'!$A$1:$A$39 / 38 rows and 1 column
Confidence interval (%): 5
Confidence interval (%)(Sen's slope): 5

Summary statistics:

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
351.8	38	0	38	107.500	510.300	298.574	76.437

Mann-Kendall trend test / Two-tailed test (351.8):

Kendall's tau	0.046
S	32.000
Var(S)	6326.000
p-value (Two-tailed)	**0.697**
alpha	0.05

The exact p-value could not be computed. An approximation has been used to compute the p-value.

Test interpretation:
H0: There is no trend in the series
Ha: There is a trend in the series
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 69.67%.

The continuity correction has been applied.

Ties have been detected in the data and the appropriate corrections have been applied.

Sen's slope: **0.521**

Confidence interval:] -0.059 , 0.769 [

Result: p-value of Mann-Kendall trend test is found to be 0.697 which is greater than 0.05 (5%), implying that there is no trend within the rainfall of Central Mesaria although Sens slope is determined to be 0.521. Note that, Sens slope has significance once there is a trend (i.e. Sens value is not significant in this sample).

2.10.6 Sample ADF test of Central Mesaria Rainfall

Table 2.7: ADF test of Central Mesaria hydrologic years averaged rainfall obtained from Excel

Summary Output

Regression Statistics	
Multiple R	0.43
R Square	0.19
Adj. R Square	0.16
Standard Error	67.53
Observations	28

ANOVA

	df	*SS*	*MS*	*F*	*Signif. F*
Regression	1	28049.48	28049.48	6.14	0.02
Residual	26	118596.02	4561.38		
Total	27	146645.50			

	Coeff.	*Std. Error*	*t Stat*	*p-value*	*Lower 95%*	*Upper 95%*
Intercept	169.12	54.76	3.08	0.004	56.55	281.70
X Variable 1	0.43	0.17	2.47	0.019	0.07	0.80

γ = Slope = 0.43 > 0

Result : Central Mesaria rainfall (time series) is stationary.

2.10.7 Cumulative Density Function (CDF)

Most widely used CDF distributions in hydrology are:

- Normal
- Log-Normal
- Extreme - Value (Gumble)
- Log-Gumble
- Pearson Type III
- Log-Pearson Type III (Gamma) (Sharifi 2006).

Since the monthly averages of the rainfall (ppt) values were studied in this study, only normal and log-Normal distribution equations were generated since the correlation coefficients results given by Minitab 16® software as presented in Figure 2.7 were fairly good (0.981 and 0.933 respectively).

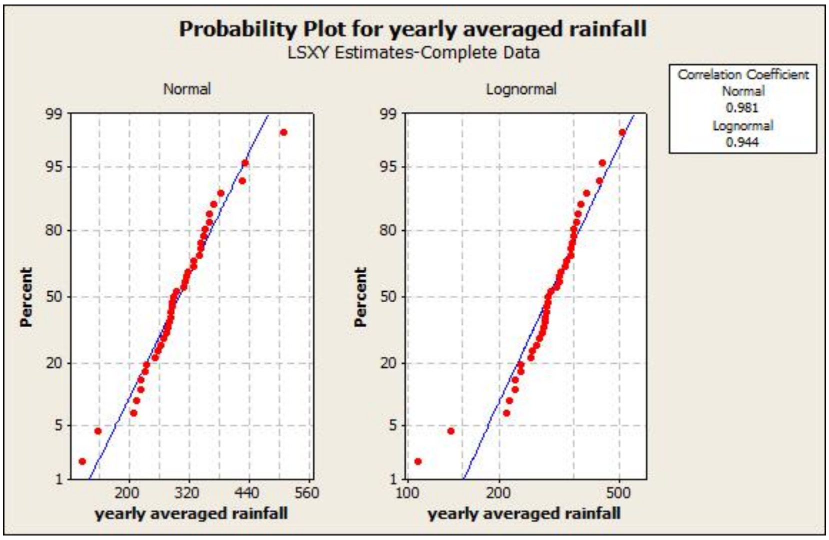


Figure 2.7: Central Mesaria rainfall fit on normal and log-Normal probability distributions

For Central Mesaria generated normal and log-Normal probability distribution equations are:

- Normal $\mathrm{x} = \bar{\mathrm{x}} + \mathrm{s_d}\ \mathrm{z}$ → $\mathrm{x} = 299.9 + 74.9\ \mathrm{z}$

- Log-Normal $\log x = \overline{x_{geo}} + s_{\log x}\, z \qquad \rightarrow \qquad \log x = 2.50 + 0.1\, z$

where ‘z’ is the standard unit, implying the area under the standard curve of that specific probability. Its tabulated form is given in Appendix A.

Result: Comparing the correlation coefficients of the two probability distributions for Central Mesaria rainfall, it is concluded that, Normal distribution has the best fitted curve being greater correlation coefficient value.

2.10.8 Dry or Wet Spell

Knowing the wet and dry years is important in any environmental study the wetness or the dryness can be checked by several methods such as severity index, drought index etc. In this study, an empirical method used which compares the mean rainfall of any regions with the rainfall of any period. If the value is larger than the mean, implying wet and otherwise implying dry.

Below is the results of the above mentioned method performed on the hydrologic year based average rainfall of Central Mesaria from hydrologic years 1975-76 to 2013-14.

Table 2.8: Numerical representation of wet and dry spells of Central Mesaria regions hydrologic year based on average rainfall from hydrologic year 1975-76 to 2013-14

Hydrologic Year	**Rainfall (mm)**	**Wet\ Dry**	**0: Dry 1: Wet**
Mean = 299.9 mm			
1975-1976	351.8	wet	**1**
1976-1977	270.3	dry	**0**
1977-1978	288.4	dry	**0**
1978-1979	350.9	wet	**1**
1979-1980	343.3	wet	**1**
1980-1981	314.1	wet	**1**
1981-1982	224.3	dry	**0**
1982-1983	235.6	dry	**0**
1983-1984	275.2	dry	**0**
1984-1985	295.3	dry	**0**
1985-1986	309.1	wet	**1**
1986-1987	345.4	wet	**1**
1987-1988	369.7	wet	**1**
1988-1989	284.5	dry	**0**
1989-1990	233.9	dry	**0**
1990-1991	138.1	dry	**0**
1991-1992	342.6	wet	**1**
1992-1993	319.9	wet	**1**
1993-1994	279.8	dry	**0**
1994-1995	286.5	dry	**0**
1995-1996	209.6	dry	**0**
Mean = 299.9 mm			
1996-1997	281.5	dry	**0**
1997-1998	252.2	dry	**0**
1998-1999	283.9	dry	**0**
1999-2000	214.9	dry	**0**
2000-2001	385	wet	**1**
2001-2002	427.7	wet	**1**
2002-2003	510.3	wet	**1**
2003-2004	361.7	wet	**1**
2004-2005	331.3	wet	**1**
2005-2006	265.2	dry	**0**
2006-2007	329	wet	**1**
2007-2008	107.5	dry	**0**
2008-2009	257	dry	**0**
2009-2010	433.8	wet	**1**
2010-2011	288.9	dry	**0**
2011-2012	314.6	wet	**1**
2012-2013	360.5	wet	**1**
2013-2014	224.3	dry	**0**

Result: number of wet spells = 18 (47% of the times > the mean), number of dry spells = 21 (53% of the time < the mean). Therefore Central Mesaria is in dry spell during the studied period.

2.10.9 Correlations

Correlation is a measure of the relation between two or more variables. Correlation coefficients can range from -1.00 to +1.00. The autocorrelation function can be defined as:

$$\mathrm{r_k} = \sum_{\mathrm{i=1}}^{\mathrm{n-k}} \frac{(\mathrm{X_i} - \mu)(\mathrm{X_{i+k}} - \mu)}{\sum_{\mathrm{i-1}}^{(\mathrm{n}} (\mathrm{X_i} - \mu)^2} \qquad 2.35$$

in which r_k is the autocorrelation coeffients, k is the interval number of the autocorrelations, n is the total number of data, μ is the mean of the population data and x_i is the i$^{\mathrm{th}}$ period of data.

The value of -1.00 represents a perfect negative correlation whereas a value of +1.00 shows a perfect positive correlation. A value of 0.00 represents a lack of correlation. In this study Minitab 16$^{®}$ software was used for finding correlations.

Chapter 3

TIME SERIES

3.1 Introduction

A collection of organized observations of a quantitative variable taken at successive points in time is called a time series. Time in terms of years, months, days, or hours is a tool that permits one to connect occurrence to a set of common, stable reference points (Schkade 1983).

The phrase ‘time series’ implies a sequence of data points that are typically consisting of successive measurements made over a certain time interval. So, it is in fact an ordered sequence of values of a variable at equally spaced (usually) time intervals. In other words, a time series is a set of statistical data that is usually collected at regular intervals. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, intelligent transport and trajectory forecasting, earthquake prediction, electro-encephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements.

Time series analysis comprises methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data. Note that, many monitoring programs are designed to evaluate these long-term trends. Time series forecasting is the use of a model to predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series.

There are two main goals of time series analysis: (a) identifying the nature of the phenomenon represented by the sequence of observations, and (b) forecasting (predicting future values of the time series variable). Both of these goals require that the pattern of observed time series data is identified and more or less formally described. So the purpose of a time series analysis is to discover the patterns and to predict future values of the time series.

Several types of data analysis available for the time series which are appropriate for different purposes.

- Exploratory analysis which is the clearest way to examine a regular time series manually through a line chart.
- Curve fitting which is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points.
- Function approximation where the time series are matched ‘approximated closely’ to any target function.
- Prediction and forecasting through simple or fully formed statistical models that describes the likely outcome of the time series in the immediate future by the given knowledge of the most recent outcomes.

The fitting of time series models can be an ambitious undertaking. There are many methods of model fitting like:

- Box-Jenkins ARIMA Models,
- Box-Jenkins Multivariate Models,

- Holt-Winters Exponential Smoothing (single, double, triple) Models.

As in most other analyses, in time series analysis, it is assumed that the data consist of a systematic pattern (usually a set of identifiable components) and random noise (error) which usually makes the pattern difficult to identify. One simple method of describing time series is that of classical decomposition. The series can be decomposed into four elements:

- Trend (Tt) — long term movements in the mean;
- Seasonal effects (It) — cyclical fluctuations related to the calendar;
- Cycles (Ct) — other cyclical fluctuations (such as a business cycles);
- Residuals (Et) — other random or systematic fluctuations

Trend represents a general systematic linear or (most often) non-linear component that changes over time and does not repeat or at least does not repeat within the time range captured by that data. Seasonality may have a similar nature of trend but however; it repeats itself in systematic intervals (periodically) over the time. Those two general classes of time series components may coexist in real-life data. Time series analysis techniques involving the filtering out of the noise (residual) make the pattern more salient. In the collection of data taken over time is some form of random variation. Hence, there exist methods for reducing of cancelling the effect due to random variation that forms residuals. The often-used technique is referred as the "smoothing". This technique, when properly applied, reveals more clearly the underlying trend, seasonal and cyclic components. So smoothing is usually done to observe better patterns, since generally smoothing methods eliminates the irregular roughness to see clearer signals.

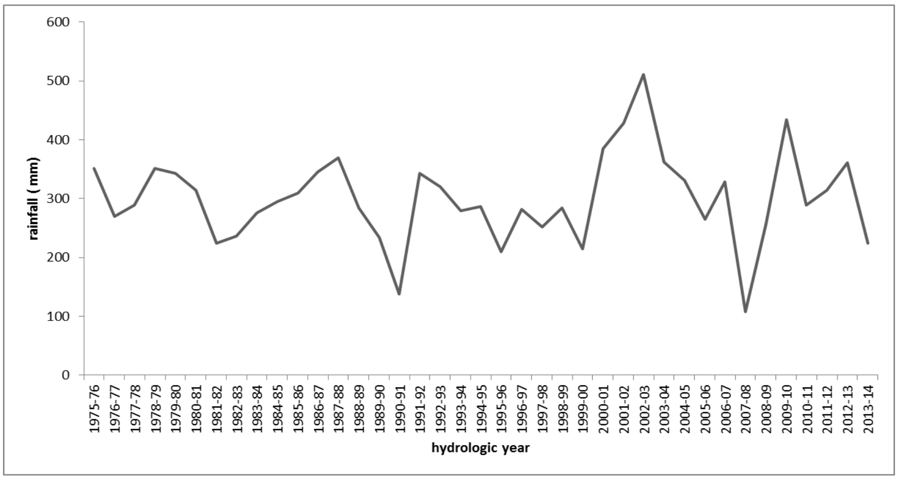


Figure 3.1: Hydrologic year rainfall (time series) of Central Mesaria

Time series models have become popular in recent years since the publication of the book by (Box and Jenkins 1970).

The purpose of a time series analysis is to discover their patterns and to predict future values of the time series.

Since the probability density function and parameters are not enough to find the sample of time series with stochastic process, internal dependencies between sequences must also calculated (Box and Jenkins 1970).

The internal dependencies between sequences of observations are obtained by autocorrelation coefficients. The autocorrelation coefficient can be used to discover non-randomness in data and to detect an appropriate time series model if the data are not random. The autocorrelation function can be obtained by Equation 2.33.

3.2 Trend Analysis

There are no proven automatic techniques to identify trend components in the time series data; however, as long as the trend is monotonous (consistently increasing or decreasing) that part of data analysis is typically not very difficult. The two types of trends that can be statistically analysed are step and monotonic. Monotonic trends are generally gradual changes that are either increasing or decreasing with no reversal of direction. Both step and monotonic trends can be increasing or decreasing, in addition, cycles like seasonal variations can be superimposed on trends. Many monotonous time series data can be adequately approximated by a linear function; if there is a clear monotonous non-linear component, the data first need to be transformed to remove the non-linearity (Serano 2001). Usually a logarithmic, exponential, or (less often) polynomial function can be used. If the time series data contain considerable error, then the first step in the process of trend identification is smoothing. It always involves some form of local averaging of data such that the non-systematic components of individual observations cancel each other out. The most common technique is moving average smoothing which replaces each element of the series by either the simple or weighted average of n surrounding elements, where n is the width of the smoothing window. Even medians can be used instead of means. The main advantage of median as compared to moving average smoothing is that its results are less biased by outliers within the smoothing window. Thus, if there are outliers in the data, median smoothing typically produces smoother or at least more reliable curves than moving average based on the same window width. There are several types of monotonic trend analysis techniques available for use. But not all techniques are appropriate for every data. A trend can be visually examined by plotting the observed data values. However, a statistical test is required to analyse the trend. Note that applying any test may be misleading if seasonal cycles are existing or if the data is not normally distributed and/or if serially correlated so the analyses may even reveal a trend although there is no (Maidment 1993).

3.3 Seasonality Analysis

Seasonal dependency (seasonality) is another general component of the time series pattern. It is formally defined as correlational dependency of order (k) between each i^{th} element of the series; (k) is usually called the lag. If the measurement error is not too large, seasonality can be visually identified in the series as a pattern that repeats every (k) elements. Seasonal patterns of time series can be examined via correlograms. The correlogram (auto-correlogram) displays graphically and numerically the auto-correlation function (ACF), that is, serial correlation coefficients (and their standard errors) for consecutive lags in a specified range of lags (e.g., 1 through 30).

Removing serial dependency has two major reasons first, one can identify the hidden nature of seasonal dependencies in the series and the other reason is to make the series stationary (Serano 2001).

In this study since the forecasting is carried out on yearly data, checking the seasonality is not valid.

3.4 Smoothing Time Series

Smoothing is usually done to help us better see patterns, trends for example, in time series. Generally smoothing methods eliminates the irregular roughness to see a clearer signal.

3.4.1 Smoothing Methods

Traditionally, time series methods have rested heavily on smoothing techniques that try to filter out the effect of the random variation in a time series. Most smoothing methods are based on some simple averaging technique. Like regression analysis; smoothing methods serve to assist in both the clarification of a time series and the forecasting of future values of the series (Schkade 1983 and Serano 2001). Two simple and commonly used smoothing models are:

3.4.1.1 Moving Average

The moving average is a data smoothing method. It spots the trends and leaves out the fluctuations. It is an indicator that shows the average value of an issue over that specific time period. Moving average can be calculated as:

$$Y_i = (2k+1)^{-1} \sum_{j=-k}^{k} X_{i+j} \qquad 3.2$$

where, k is the moving average period (Kottegoda 1980).

3.4.1.2 Exponential Smoothing

Exponential smoothing is used to smoothen the time series and then forecasting it. The exponentially smoothed response value at time period t is denoted by S_t. The smoothing scheme begins by assigning $S_1 = y_1$ at the first period. For the second time period:

$$S_2 = \alpha y_t + (1-\alpha) S_{t-1} \qquad 0 \leq \alpha \leq 1 \qquad 3.3$$

This equation is called the basic equation of exponential smoothing, and the constant α is called the smoothing constant (Kottegoda 1980).

The most important problem when applying exponential smoothing is to find the best smoothing constant α for a particular set of data .Unfortunately , there does not exist a simple formula for finding such a value of α. In general, the more noisy or unstable a time series is, the smaller value of α should be.

3.5 Stationary Time Series

A time series is stationary if it is free of trends, shifts, or periodicity. It means that the statistical parameters of the series such as mean and variance remain constant through time. Otherwise the time series is non-stationary. Generally hydrologic time series defined on an annual time scale are stationary (Maidment 1993).

Therefore a stationary time series should satisfy two conditions:

- Time series data must have constant mean,
- Variance of time series should not change over the time.

In this study Augmented Dickey–Fuller test (ADF) is applied to test the stationarity of the rainfall (time series).

3.6 Forecasting Models

Forecasting model involves the selection of an estimation procedure. A forecast after all, is an estimate of a future outcome of a random process. In this study seven forecasting models were used:

3.6.1 Markov Model

The Markov process considers that the value of an event at one time is correlated with the value of the event at an earlier period. In a first-order Markov process, this correlation exists in two consecutive values of the proceedings. The first order Markov model, which comprises the classic method in synthetic hydrology, declares that the value of a variable y in one time period is dependent on the value of ‘y’ in the preceding time period plus a random component.

$$y_i = d_i + \varepsilon_i \qquad 3.4$$

y_i= rainfall at i^th year,
d_i= deterministic part of i^th year,
ε_i = random part of i^th year.

The values of ε_i are connected with the historical data by certifying that they belong to the same frequency distribution and possess similar statistical properties (mean, deviation, skewness) as the historical series (Gupta, 1989).

A variety of forms and combinations of deterministic and random component are established as different models. Single season (annual) rainfall model of lag 1 is the simplest model which presumes that the amount of the current rainfall is considerably correlated with the previous one value only (Gupta, 1989).

First order Markov Model has been productively applied to many problems. Examples include modelling sequential data using Markov chains, and solving control Problems created in the Markov decision processes (MDP) framework.

If the Markov model's parameters are estimated from data, the standard maximum likelihood estimates consider the first order (single step) transitions only. But for many problems, the first order conditional independence assumptions are not satisfied as a result of the higher order transition probabilities can be poorly approximated by the learned model (Noordin 2010).

Formulation of the Markov model for yearly data (Gupta, 1989):

$$x_i = \bar{x} + r_1(x_{i-1} - \bar{x}) + S\sqrt{(1 - r_1{}^2)}t_i \qquad 3.5$$

where x_i is the rainfall at ith year, $\bar{x}$ is the mean of data, $r_1$is lag one – autocorrelation coefficient, S is the standard deviation of the data, and t_i is the random variate from an appropriate distribution with a mean of zero and variance of unity. For obtaining t_i

the random number should be generated, and in this study Microsoft Excel was used where the inverse error function $erf^{-1}(z)$ as well calculated:

$$\mathrm{erf}^{-1}(\mathrm{z}) = \frac{1}{2}\sqrt{\pi}\left(\mathrm{z} + \frac{\pi}{12}\mathrm{z}^3 + \frac{7\pi^2}{12}\mathrm{z}^5 + \frac{127\pi^3}{40320}\mathrm{z}^7 + \cdots\right) \qquad 3.6$$

Value of z can be obtained from the cumulative distribution function (CDF) of the log-Normal distribution as Figure 3.2 implies.

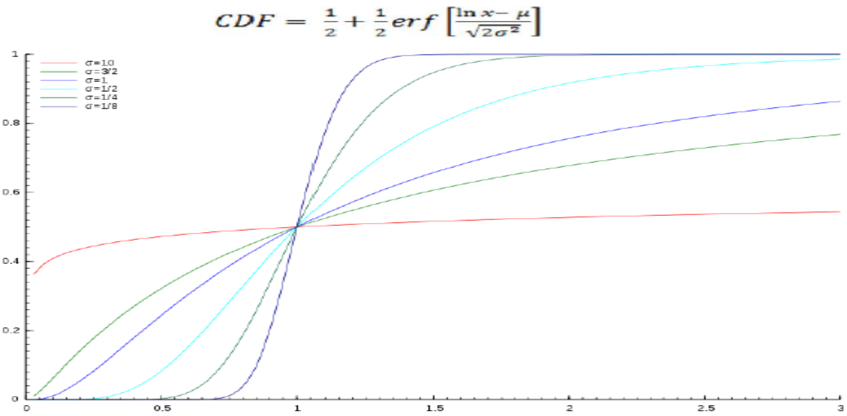


Figure 3.2: Cumulative distribution function of the log-Normal distribution (Gupta, 1989)

The equation of randomness is:

$$\frac{1}{2} + \frac{1}{2}\mathrm{erf}\left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\right] = \mathrm{RAND}() \qquad 3.7$$

As log-Normal random numbers have both mean and standard deviations equal to unity, implies:

$$\mathrm{z} = (\mathrm{RAND}(\ ) - 0.5)2 \qquad 3.8$$

If $\mathrm{erf}(x) = y$, then $\mathrm{erf}^{-1}(y) = x$.

Let $$\frac{\ln x - 1}{\sqrt{2}} = y \qquad 3.9$$

the value of $t = \ln x$. Therefore,

$$\ln x = \left(y\sqrt{2}\right) + 1 \qquad 3.10$$

3.6.1.1 East Mesaria region rainfall as a sample for establishing Markov Model

In order to establish the Markov model for East Mesaria rainfall the mean, autocorrelation coefficient (r1), standard deviation, random number, and other relevant parameters were determined automatically with help of Excel software and the results were tabulated below.

Table 3.1: Markov Model of East Mesaria rainfall

Hydrologic year	Rainfall	mean	r_1	Standard deviation	random	z	erf^{-1}	t_i	Xi (forecasted)
1975-1976	355.4	324.2	0.12	94.8	0.999	0.998	1.327	2.877	556.0
1976-1977	260.2	324.2	0.12	94.8	0.969	0.938	1.170	2.655	601.9
1977-1978	345.8	324.2	0.12	94.8	0.985	0.969	1.249	2.766	617.8
1978-1979	292.3	324.2	0.12	94.8	0.800	0.601	0.595	1.841	532.7
1979-1980	354.0	324.2	0.12	94.8	0.782	0.564	0.550	1.778	516.5
1980-1981	348.3	324.2	0.12	94.8	0.956	0.912	1.110	2.570	589.2
1981-1982	229.0	324.2	0.12	94.8	0.804	0.607	0.603	1.853	530.4
1982-1983	190.1	324.2	0.12	94.8	0.087	-0.826	-0.934	-0.321	318.7
1983-1984	287.7	324.2	0.12	94.8	0.385	-0.230	-0.207	0.708	390.2
1984-1985	372.2	324.2	0.12	94.8	0.202	-0.596	-0.589	0.166	347.8
1985-1986	305.1	324.2	0.12	94.8	0.489	-0.023	-0.020	0.971	418.5
1986-1987	290.3	324.2	0.12	94.8	0.793	0.585	0.576	1.814	506.3
1987-1988	457.0	324.2	0.12	94.8	0.555	0.110	0.098	1.138	453.2
1988-1989	316.9	324.2	0.12	94.8	0.109	-0.782	-0.856	-0.211	319.8
1989-1990	247.5	324.2	0.12	94.8	0.238	-0.523	-0.503	0.289	350.9
1990-1991	161.9	324.2	0.12	94.8	0.694	0.388	0.359	1.507	469.3
1991-1992	434.2	324.2	0.12	94.8	0.950	0.900	1.083	2.532	579.9
1992-1993	392.4	324.2	0.12	94.8	0.711	0.422	0.394	1.557	501.4
1993-1994	291.7	324.2	0.12	94.8	0.785	0.571	0.559	1.790	513.9
1994-1995	284.9	324.2	0.12	94.8	0.407	-0.185	-0.166	0.766	419.0
1995-1996	238.5	324.2	0.12	94.8	0.320	-0.360	-0.331	0.532	385.7
1996-1997	204.7	324.2	0.12	94.8	0.186	-0.627	-0.629	0.111	342.0
1997-1998	265.9	324.2	0.12	94.8	0.551	0.102	0.090	1.128	432.5
1998-1999	271.6	324.2	0.12	94.8	0.604	0.208	0.186	1.263	456.1
1999-2000	257.7	324.2	0.12	94.8	0.303	-0.394	-0.365	0.484	385.6
2000-2001	464.3	324.2	0.12	94.8	0.906	0.813	0.910	2.287	546.8
2001-2002	458.1	324.2	0.12	94.8	0.011	-0.979	-1.273	-0.801	275.5
2002-2003	419.2	324.2	0.12	94.8	0.779	0.558	0.543	1.768	484.7
2003-2004	435.3	324.2	0.12	94.8	0.034	-0.932	-1.156	-0.635	283.7
2004-2005	309.7	324.2	0.12	94.8	0.635	0.270	0.244	1.346	446.0
2005-2006	292.1	324.2	0.12	94.8	0.878	0.757	0.815	2.153	541.4
2006-2007	463.9	324.2	0.12	94.8	0.455	-0.090	-0.080	0.887	433.7
2007-2008	136.4	324.2	0.12	94.8	0.499	-0.001	-0.001	0.998	431.3
2008-2009	300.1	324.2	0.12	94.8	0.457	-0.085	-0.075	0.893	421.1
2009-2010	537.1	324.2	0.12	94.8	0.356	-0.289	-0.262	0.630	395.1
2010-2011	345.0	324.2	0.12	94.8	0.214	-0.571	-0.559	0.209	352.4
2011-2012	466.3	324.2	0.12	94.8	0.116	-0.769	-0.834	-0.180	310.7
2012-2013	366.1	324.2	0.12	94.8	0.365	-0.270	-0.244	0.655	384.2
2013-2014	196.2	324.2	0.12	94.8	0.035	-0.931	-1.153	-0.631	272.0

3.6.2 Auto-Regressive (AR) Model

General expression of auto-regressive model can be defined by:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \qquad 3.11$$

where $\varphi_1, \varphi_2, \ldots \varphi_p$ are auto-regressive coefficients, and ε_t is white noise (residuals) and p is the order of auto-regressive model.

Auto-regressive coefficients ϕ_p can be computed with below matrix form as given in Equation 3.11a.

$$\begin{bmatrix} 1 & r_1 & r_2 & r_3 & r_4 & \ldots\ldots\ldots\ldots r_{p-1} \\ r_1 & 1 & r_1 & r_2 & r_3 & \ldots\ldots\ldots\ldots r_{p-2} \\ r_2 & r_1 & 1 & r_1 & r_2 & \ldots\ldots\ldots r_{p-3} \\ & & & & & \\ r_{p-1} & r_{p-2} & r_{p-3} & r_{p-4} & r_{p-5} & \ldots\ldots\ldots 1 \\ & & & & & \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \\ \phi_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \\ r_p \end{bmatrix} \qquad 3.11a$$

Before applying autoregressive model to time series, order of Autoregressive model (P) should be defined. In this study Akaike information Criterion is used to find the best order of Autoregressive between AR(1),AR(2),AR(3) in order to show best Autoregressive model for deriving suitable synthetic series of data. After modelling series the time series is forecasted by the best Autoregressive model based on the AIC number (Kottegoda 1980 and Maidment 1993).

3.6.2.1 Akaike Information Criterion

The Akaike information criterion (AIC) is a measure of the relative quality of standard models or a given set of data. AIC estimates the quality of each model, relative to each of the other models. Hence, AIC provides a means for model selection (Schmidt 2008). Akaike recommends the following relationship for Autoregressive model:

$$AIC(p) = \min\left(n. ln\sigma_\varepsilon{}^2 + 2(p)\right) \qquad 3.12$$

Where n is the sample size, $\sigma_\varepsilon{}^2$ is the maximum likelihood estimate of the residual variance, p is the order of autoregressive model. The model, which gives the minimum AIC number, is the one to be selected.

3.6.2.2 Steps in Calculating AIC number for Central Mesaria Rainfall

- From the measured and hydrologic yearly averaged rainfall, from 1975-76 to 2003-2004 were used to establish AIC numbers. Note that averaged values from hydrologic years 2004-05 to 2013-14 data will be used to check the prediction.
- Auto-correlation coefficients (r_k) is calculated using Minitab 16®

Table 3.2: Auto-correlation coefficients of Central Mesaria rainfall

Lag (K)	**Auto-correlation Coefficients (r_k)**	**Lag (K)**	**Auto-correlation Coefficients (r_k)**
1	0.39	15	0.06
2	0.05	16	0.11
3	-0.26	17	0.03
4	-0.14	18	-0.03
5	-0.03	19	-0.14
6	-0.08	20	-0.21
7	-0.02	21	-0.12
8	0.05	22	0.04
9	0.18	23	0.11
10	-0.12	24	0.02
11	-0.21	25	0.01
12	-0.32	26	0.00
13	-0.07	27	0.05
14	0.10	28	0.02

- Auto-regressive coefficients of three different orders (AR(1), AR(2), and AR(3)) are used in this study and their coefficients by solving the matrix given in Equation 3.11a were determined through Excel software.
- First order auto-regressive model (AR1) equation is:

$$y_i = 0.39\, y_{i-1} + \varepsilon_i \qquad 3.13$$

- Second order auto-regressive model (AR2) equation is:

$$y_i = 0.43\, y_{i-1} - 0.1\, y_{i-2} + \varepsilon_i \qquad 3.14$$

- Third order auto-regressive model (AR3) equation is :

$$y_i = 0.4\, y_{i-1} + 0.01\, y_{i-2} - 0.28\, y_{i-3} + \varepsilon_i \qquad 3.15$$

- To find the maximum likelihood estimation of the residuals the positive root should be selected.
- For AR(1):

$$y_i{}^2 = (0.39\, y_{i-1} + \varepsilon_i)^2 \rightarrow 1^2 = 0.39^2 * y_{i-1}{}^2 + 2 * 0.39 * y_{i-1} * \sigma_\varepsilon + \sigma_\varepsilon{}^2 \qquad 3.16$$

$$\rightarrow \qquad \sigma_\varepsilon{}^2 = 0.372$$

For AR(2): $\sigma_\varepsilon{}^2 = 0.8$

For AR(3): $\sigma_\varepsilon{}^2 = 0.77$

- AIC numbers are:

For AR(1):	AIC=29*ln(0.372)+2(1) = -26.68
For AR(2):	AIC=29*ln(0.8)+2(2)= -2.47
For AR(3):	AIC=29*ln(0.77)+2(3)= -1.58

- The minimum AIC number which is AR(1) model for this region is selected

3.6.2.3 AIC numbers of all meteorological regions and TRNC

Table 3.3: AIC numbers of all meteorological regions and TRNC

Region	Akaike information criteria(AIC)		
	AR(1)	AR(2)	AR(3)
Central Mesaria	**-26.68***	-2.47	-1.58
East Coast	67.26	**3.12**	5.12
East Mesaria	**-6.34**	2.82	4.21
Karpaz	**-5.21**	2.82	4.82
North Coast	**-3.06**	2.82	4.82
West Mesaria	**0.82**	4.00	6.00
TRNC	-5.21	2.82	**-107.45**

*The bold numbers are representing the relevant AR model of that region i.e being the smallest value among each row.

3.6.2.4 Derivation of the synthetic sequence

After obtaining the suitable model from Akaike information criteria, the derivation of synthetic sequences were obtained by finding the values of normal and independent residual ε_i (Kottegoda 1980). The residual can be defined as if obeys normal distribution by:

$$\varepsilon_i = \mu_\varepsilon + \sigma_\varepsilon Z_i \qquad 3.17$$

where μ_ε is the mean of residuals, σ_ε is the standard deviation of residuals, Z_{1i}, Z_{2i} represent the standard normal random numbers that must be calculated by using the uniform random numbers, η_i varying randomly between 0 and 1. The standard normal random numbers are defined by:

$$Z_{1i} = (-2\ln\eta_{1i})^{1/2}\cos(2\pi\eta_{2i}) \qquad 3.18$$

$$Z_{2i} = (-2\ln\eta_{1i})^{1/2}\sin(2\pi\eta_{2i}) \tag{3.19}$$

3.6.2.5 Derivation of AR(1) Model for Central Mesaria

Through AR(1) model, applying Equations 3.17, 3.18, and 3.19 to hydrologic yearly averaged measured rainfall values of Central Mesaria regions, their relevant synthetic values were generated and tabulated for the period of hydrologic years 1975-76 to 2003-04 below.

Table 3.4: Synthetic data generated by AR(1) model for Central Mesaria

Uniform random number		Standard normal random number		Residual for AR(1) model		y_i values for AR(1) model	Generated synthetic data y_i'	
η_1	0.31	Z_1	-0.38	ε_1	-0.23	0	y'_1	302.95
η_2	0.71	Z_2	-1.48	ε_2	-0.90	-0.90	y'_2	237.16
η_3	0.80	Z_3	-0.60	ε_3	-0.36	-0.71	y'_3	250.87
η_4	0.43	Z_4	0.29	ε_4	0.18	-0.10	y'_4	295.72
η_5	0.83	Z_5	-0.39	ε_5	-0.24	-0.28	y'_5	282.62
η_6	0.36	Z_6	0.47	ε_6	0.28	0.18	y'_6	315.83
η_7	0.84	Z_7	0.29	ε_7	0.18	0.25	y'_7	320.98
η_8	0.83	Z_8	-0.50	ε_8	-0.31	-0.21	y'_8	287.45
η_9	0.18	Z_9	-1.78	ε_9	-1.09	-1.17	y'_9	217.55
η_{10}	0.45	Z_{10}	0.52	ε_{10}	0.32	-0.14	y'_{10}	293.07
η_{11}	0.47	Z_{11}	0.10	ε_{11}	0.06	0.01	y'_{11}	303.47
η_{12}	0.76	Z_{12}	-1.22	ε_{12}	-0.74	-0.74	y'_{12}	248.97
η_{13}	0.99	Z_{13}	-0.10	ε_{13}	-0.06	-0.35	y'_{13}	277.58
η_{14}	0.43	Z_{14}	0.05	ε_{14}	0.03	-0.11	y'_{14}	295.15
η_{15}	0.92	Z_{15}	-0.05	ε_{15}	-0.03	-0.07	y'_{15}	297.59
η_{16}	0.27	Z_{16}	0.41	ε_{16}	0.25	0.22	y'_{16}	319.13
η_{17}	0.33	Z_{17}	0.00	ε_{17}	0.00	0.09	y'_{17}	309.21
η_{18}	0.25	Z_{18}	1.49	ε_{18}	0.91	0.94	y'_{18}	371.72
η_{19}	0.59	Z_{19}	0.77	ε_{19}	0.47	0.83	y'_{19}	363.80
η_{20}	0.88	Z_{20}	-0.69	ε_{20}	-0.42	-0.10	y'_{20}	295.85
η_{21}	0.68	Z_{21}	-0.68	ε_{21}	-0.42	-0.45	y'_{21}	269.80
η_{22}	0.39	Z_{22}	0.55	ε_{22}	0.34	0.16	y'_{22}	314.68
η_{23}	0.97	Z_{23}	0.18	ε_{23}	0.11	0.17	y'_{23}	315.66
η_{24}	0.87	Z_{24}	-0.18	ε_{24}	-0.11	-0.04	y'_{24}	299.68
η_{25}	0.93	Z_{25}	0.07	ε_{25}	0.04	0.02	y'_{25}	304.61
η_{26}	0.22	Z_{26}	0.37	ε_{26}	0.23	0.24	y'_{26}	320.16
η_{27}	0.29	Z_{27}	-1.58	ε_{27}	-0.96	-0.87	y'_{27}	239.50
η_{28}	0.49	Z_{28}	0.06	ε_{28}	0.04	-0.30	y'_{28}	281.24
η_{29}	0.44	Z_{29}	-1.23	ε_{29}	-0.75	-0.87	y'_{29}	239.71

Figure 3.3 shows the fluctuations of the synthetically generated AR(1) data from the measured hydrological yearly averaged rainfall values from 1975-75 to 2003-04.

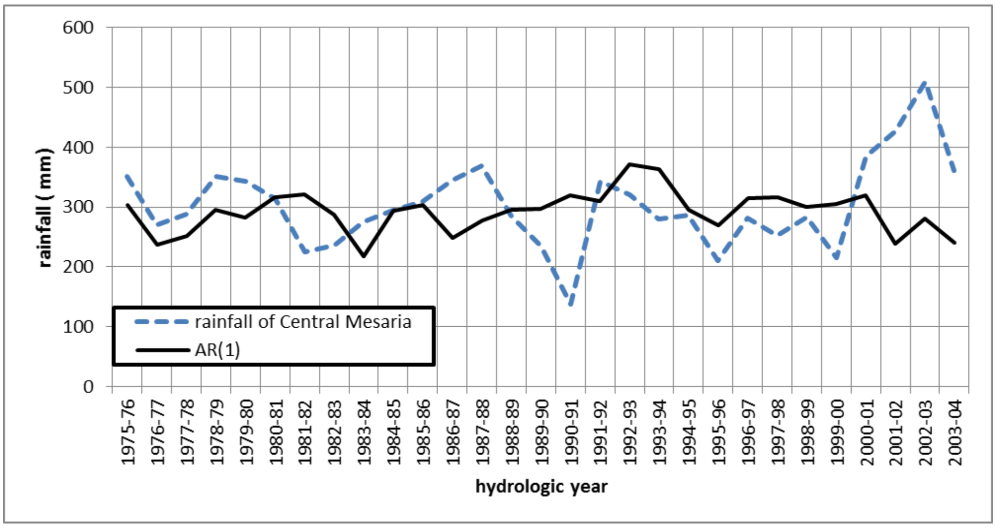


Figure 3.3: Synthetically generated AR(1) model and the respective measured rainfalls of Central Mesaria for the period of hydrologic years 1975-76 to 2003-04

3.6.2.6 Forecasting hydrologic yearly averaged rainfall of Central Mesaria through AR(1) Model

Based on the above discussed methodology, AR(1) model was used to generate the forecasted hydrologic yearly average rainfall of Central Mesaria for the period of to 2004-05 to 2013-14 years. Details are given in Table 3.5.

Table 3.5: Forecasted rainfall by AR(1) for Central Mesaria for the period of hydrologic years 2004-05 to 2013-14.

Uniform random number		**Standard normal random number**		**Residual for AR(1) model**		**y_i values for AR(1) model**	**Forecasted synthetic data y'_i**	
η_{30}	0.11	Z_{30}	0.76	ε_{30}	0.69	0.87	y'_{30}	366.1
η_{31}	0.30	Z_{31}	-1.27	ε_{31}	-1.16	-0.82	y'_{31}	243.0
η_{32}	0.60	Z_{32}	-0.87	ε_{32}	-0.79	-1.11	y'_{32}	221.8
η_{33}	0.04	Z_{33}	-0.46	ε_{33}	-0.42	-0.85	y'_{33}	241.2
η_{34}	0.28	Z_{34}	2.47	ε_{34}	2.24	1.92	y'_{34}	442.8
η_{35}	0.10	Z_{35}	2.16	ε_{35}	1.97	2.71	y'_{35}	500.5
η_{36}	0.01	Z_{36}	0.14	ε_{36}	0.13	1.18	y'_{36}	388.7
η_{37}	0.83	Z_{37}	0.03	ε_{37}	0.02	0.48	y'_{37}	337.9
η_{38}	0.76	Z_{38}	-0.60	ε_{38}	-0.55	-0.36	y'_{38}	276.4
η_{39}	0.30	Z_{39}	1.55	ε_{39}	1.41	1.27	y'_{39}	395.4

The graphical representation of this forecasted data set are shown in Figure 3.4.

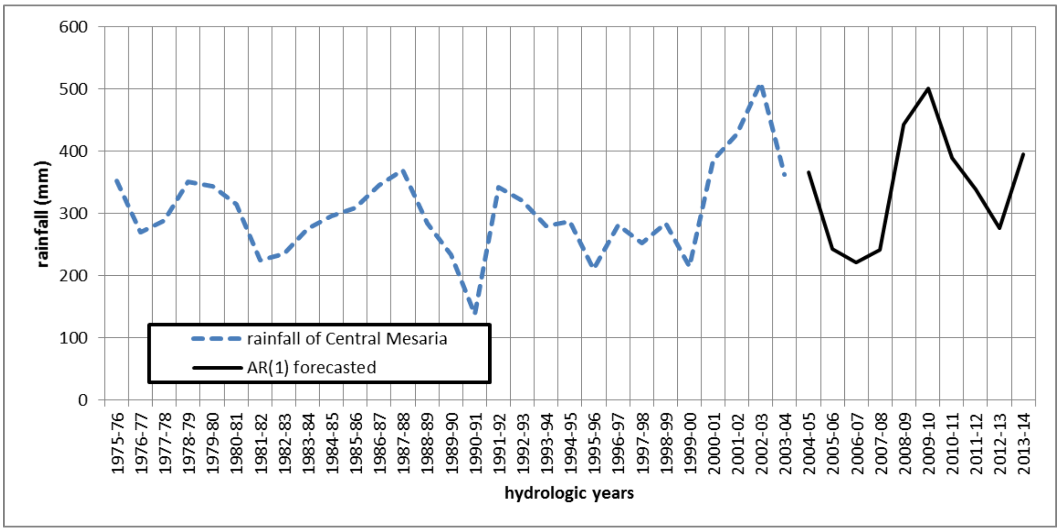


Figure 3.4: Forecasted synthetic data by AR(1) for Central Mesaria for the period of hydrologic years 2003-04 to 2013-14.

3.6.3 Holt-Winters Method

The Holt-Winters seasonal method includes the forecast equation and three smoothing equations — one for the level $\mathbf{l_t}$, one for trend $\mathbf{b_t}$, and one for the seasonal component denoted by $\mathbf{S_t}$, with smoothing parameters α, β and γ.

There are two variations to this method that differ in the nature of the seasonal component. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred (Hyndman 2013).

3.6.3.1 Holt – Winters Additive Method

$$\widehat{y_t} = l_t + hb_t - S_{t-m+h} \tag{3.20}$$

$$l_t = \alpha(y_t - S_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{3.20a}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{3.20b}$$

$$S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)S_{t-m} \tag{3.20c}$$

The error correction form of the smoothing equations is:

$$l_t = l_{t-1} + b_{t-1} + \alpha e_t \qquad 3.21$$

$$b_t = b_{t-1} + \alpha\beta e_t \qquad 3.21a$$

$$S_t = S_{t-m} + \gamma e_t \qquad 3.21b$$

where $$e_t = y_t - (l_{t-1} + b_{t-1} + S_{t-m}) = y_t - \hat{y_t} \qquad 3.21c$$

3.6.3.2 Holt-Winter Multiplicative Method

$$\hat{y_t} = (l_t + hb_t)S_{t-m+h} \qquad 3.22$$

$$l_t = \alpha\left(\frac{y_t}{S_{t-m}}\right) + (1-\alpha)(l_{t-1} + b_{t-1}) \qquad 3.22a$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \qquad 3.22b$$

$$S_t = \gamma(\frac{y_t}{l_{t-1}+b_{t-1}} + (1-\gamma)S_{t-m} \qquad 3.22c$$

and the error correction form of the smoothing equations is:

$$l_t = l_{t-1} + b_{t-1} + \alpha\frac{e_t}{S_{t-m}} \qquad 3.23$$

$$b_t = b_{t-1} + \alpha\beta\frac{e_t}{S_{t-m}} \qquad 3.23a$$

$$S_t = S_t + \gamma\frac{e_t}{(l_{t-1}+b_{t-1})} \qquad 3.23b$$

where $$e_t = y_t - (l_{t-1} + b_{t-1})S_{t-m} = y_t - \hat{y_t} \qquad 3.23c$$

In this study Holt-Winter Multiplicative Model is adopted using Minitab 16® software.

3.6.4 Accuracy Measures of the Forecasted Models

When selecting a forecasting model, or when evaluating any existing model, one has to use measures that summarise the overall accuracy provided by that model. Beside the visual comparison, the suggested models performances are also evaluated using the below mentioned statistical accuracy measures:

- The mean absolute deviation (MAD),

$$\mathrm{MAD} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \qquad 3.24a$$

- The mean square error (MSE),

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \qquad 3.24b$$

- The root mean square error (RMSE), and

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \qquad 3.24c$$

- The mean absolute percentage error (MAPE).

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| * 100 \qquad 3.24d$$

Chapter 4

CALCULATIONS AND RESULTS

4.1 Introduction

In order to study the space-time variability of the rainfall, the monthly rainfall from January1975 to December 2014 of all the meteorologically classified regions of TRNC were gathered from the meteorology office of TRNC. In order to have more useful data, that might be used for different water resources studies; the gathered monthly data for 6 different meteorological regions and for TRNC as a whole were rearranged, based on the hydrological years i.e. from September to August. Hence, the period of this study is from September 1975 to August 2014.

In this chapter, the gathered rainfall were studied in three parts; the first part deals with the basic statistics and probabilistic equations with data quality checks discussed.

The second part mainly concentrates on the application of time series models discussed in previous chapter with forecasting values where the time series analysis was used to understand the occurrence of the random mechanism so as to predict future series based on the past data. For time series model studies, the gathered rainfall values were initially divided into two time series parts,

1. from September 1975 to August 2003, and
2. from September 2004 to August 2014.

The data within the first part was used to train the model and the data set of the second part was used for comparing the error accuracy of the suggested models based on the previously trained data. Once the best fitted model was obtained within the acceptable confidence limits, since the sample size is comparably small (n=39), only 5 years ahead values were predicted based on the most representative model, i.e., up to 2018-19.

In the third part of this chapter, the wetness or the dryness of the months for each region and TRNC (except the three months, June, July, and August since no significant amount of rainfall exists) were examined empirically with respect to the mean of the data (value > mean value implies wetness otherwise dryness).

4.2 Meteorological Region: Central Mesaria

Table 4.1: Total rainfall of Central Mesaria region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic Year	Month												Total
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Agu	
1975-76	0.0	13.1	17.9	118.6	23.4	34.8	38.6	54.8	26.0	1.1	23.5	0.0	351.8
1976-77	5.3	23.0	36.2	40.2	57.8	13.2	48.0	39.2	0.8	0.1	6.5	0.0	270.3
1977-78	22.0	1.3	1.7	76.6	99.7	27.7	40.1	19.3	0.0	0.0	0.0	0.0	288.4
1978-79	0.0	12.4	5.3	96.6	40.2	79.1	48.4	6.6	14.7	46.6	1.0	0.0	350.9
1979-80	0.0	59.2	24.2	90.2	33.8	96.7	31.6	7.6	0.0	0.0	0.0	0.0	343.3
1980-81	0.0	14.2	7.6	40.2	92.6	66.8	46.2	23.7	17.6	5.2	0.0	0.0	314.1
1981-82	0.0	2.4	36.8	31.4	14.9	44.2	62.8	8.0	7.8	10.3	1.0	4.7	224.3
1982-83	2.4	3.4	19.2	18.7	39.0	40.0	34.4	24.4	42.9	11.2	0.0	0.0	235.6
1983-84	0.0	20.8	49.0	32.4	24.8	35.2	25.5	70.5	3.7	0.0	3.6	9.7	275.2
1984-85	0.0	2.8	106.2	42.3	47.9	23.4	31.4	27.7	13.6	0.0	0.0	0.0	295.3
1985-86	13.1	26.8	17.1	49.4	17.5	66.8	19.4	19.5	68.1	11.4	0.0	0.0	309.1
1986-87	1.6	22.0	116.3	42.7	27.1	14.1	97.7	14.2	9.3	0.4	0.0	0.0	345.4
1987-88	0.0	52.2	17.9	79.6	45.3	87.1	68.3	13.8	1.6	0.0	2.7	1.2	369.7
1988-89	5.5	14.6	44.4	87.7	95.2	7.9	22.1	0.0	4.9	2.2	0.0	0.0	284.5
1989-90	1.8	48.2	18.7	27.4	11.3	75.4	30.2	2.4	13.3	0.0	0.0	5.2	233.9
1990-91	0.0	12.7	1.1	16.8	40.0	31.2	30.8	4.0	0.6	0.9	0.0	0.0	138.1
1991-92	0.0	15.9	38.2	118.5	8.9	51.8	9.8	4.8	18.4	32.0	34.1	10.2	342.6
1992-93	0.0	3.7	68.6	80.9	31.4	41.1	46.2	6.4	35.5	6.1	0.0	0.0	319.9
1993-94	0.0	0.0	34.5	9.8	116.3	45.9	43.7	18.0	7.1	0.3	4.2	0.0	279.8
1994-95	4.6	32.8	131.9	26.8	16.4	10.6	10.0	9.6	10.9	0.0	32.8	0.1	286.5
1995-96	0.0	3.1	30.5	5.6	69.2	38.1	29.8	26.9	0.7	5.7	0.0	0.0	209.6
1996-97	0.0	27.3	18.0	49.0	8.0	41.3	49.7	40.1	17.4	30.7	0.0	0.0	281.5
1997-98	12.5	13.0	52.7	53.3	37.7	13.5	28.4	3.3	25.7	12.1	0.0	0.0	252.2
1998-99	4.6	0.6	23.6	74.9	70.2	35.7	24.8	11.5	2.4	27.0	2.9	5.7	283.9
1999-00	2.7	7.3	14.6	9.9	35.3	36.6	29.4	63.4	10.6	0.8	0.0	4.3	214.9
2000-01	20.1	24.4	74.1	109.6	43.6	38.5	13.7	18.6	38.6	0.0	0.0	3.8	385.0
2001-02	0.0	22.6	40.7	117.4	53.9	35.4	30.1	44.4	41.4	15.2	17.8	8.8	427.7
2002-03	6.2	12.1	8.9	155.7	50.8	109.0	83.4	37.5	17.0	28.7	0.2	0.8	510.3
2003-04	1.6	7.2	10.5	68.1	172.5	57.7	0.5	20.0	19.2	4.4	0.0	0.0	361.7
2004-05	0.0	47.3	41.4	64.4	53.9	26.4	18.0	14.6	24.2	39.9	0.0	1.2	331.3
2005-06	5.6	10.8	67.7	5.5	62.0	29.9	44.0	10.6	7.7	2.0	19.4	0.0	265.2
2006-07	10.2	53.8	30.2	5	31.9	78	31.9	19.5	62.5	0.5	1.4	4.1	329.0
2007-08	0	6.6	18.3	28.3	12.3	12.9	6	1.7	19.7	0.7	0.1	0.9	107.5
2008-09	10.1	23.5	8	54.2	35.4	45.5	42	15.2	14.3	0	0	8.8	257.0
2009-10	22.9	30.1	18.1	93.6	102.4	121.7	2.9	13.7	13.3	12.5	0.5	2.1	433.8
2010-11	0	1.7	0.1	40.2	81.4	30.4	24.8	36.4	37.9	35.2	0	0.8	288.9
2011-12	4	5.1	51	35.1	94.9	43.8	17.2	12.4	43.3	0.6	1.6	5.6	314.6
2012-13	0	64.8	48.6	88.1	50.1	15.7	8.6	27.5	56.7	0	0.4	0	360.5
2013-14	0.3	3.3	15.9	57.7	10.2	19.7	22.5	21.2	55.1	14	2.1	2.3	224.3
2014-15	19.2	64.9	21.7	89.1									

Table 4.2: Statistical measures of Central Mesaria rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	C_{Vlog}	C_{Slog}	$\bar{x}_{med}$	C_{dx}	PC_v
299.9	75.9	0.2	0.25	0.1	2.5	0.1	0.035	0.05	-1.2	288.9	0.1	0.54

4.2.1 Empirical determination of minimum required sample size of rainfall for Central Mesaria region based on mean and standard deviation

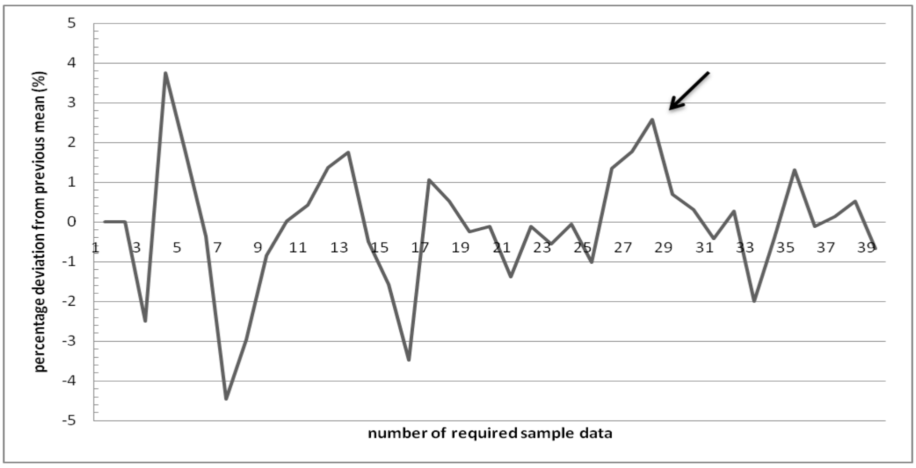


Figure 4.1: Curve showing the required number of sample size for Central Mesaria regions' rainfall, based on the percentage deviations of the mean values

Comment: The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 28.

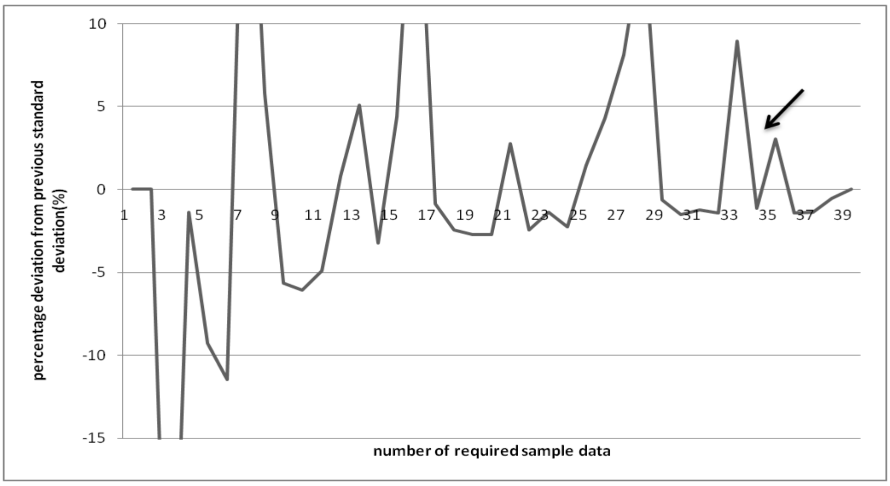


Figure 4.2: Curve showing the required number of sample size for Central Mesaria regions' rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 34.

Hence $n_{min} = 34$ is the required minimum sample size value that satisfies both the mean and the standard deviation approaches for Central Mesaria region rainfall.

Table 4.3: Appropriate rainfall sample size of Central Mesaria for statistic and probabilistic studies.

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
28 < 39 OK	34 < 39 OK

4.2.2 Normality test for Central Mesaria

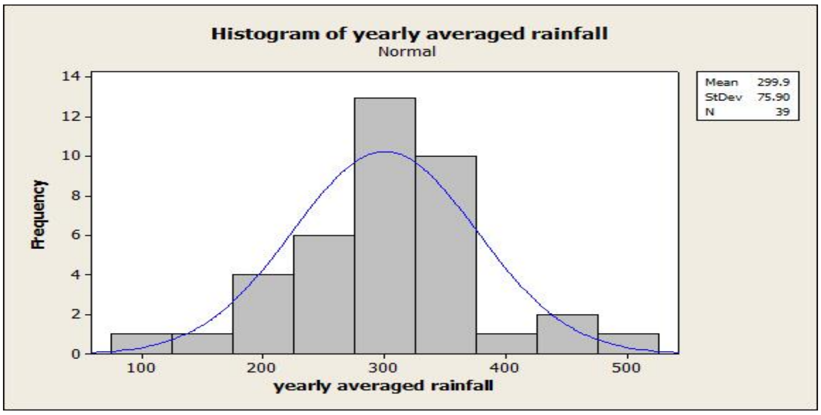


Figure 4.3: Central Mesaria regions’ rainfall histogram

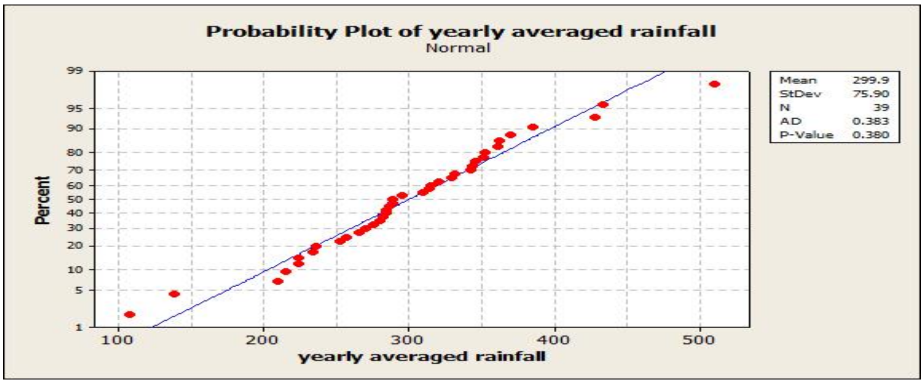


Figure 4.4: Normality test of Central Mesaria rainfall

<u>Result</u>: p-value = 0.38 > 0.05. Therefore, Central Mesaria rainfall is normally distributed.

4.2.3 Homogeneity Test of Central Mesaria

4.2.3.1 Homogeneity test of Central Mesaria rainfall

Four homogeneity tests of time series were done by XLSTAT and the results is given below:

XLSTAT 2015.4.01.20978 - Homogeneity tests - on 8/25/2015 at 5:49:56 PM
Time series: Workbook = homogeneity test.xlsx / Sheet = Sheet1 / Range = Sheet1!$A$1:$A$39 / 38 rows and 1 column
Significance level (%): 5
Maximum time (s): 180
Number of simulations: 10000
Seed (random numbers): 4399059

Summary statistics:

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
351.8	38	0	38	107.5	510.3	298.5	76.437

Pettitt's test (351.8):

K	125.000
t	24
p-value (Two-tailed)	**0.269**
alpha	0.05

The p-value has been computed using 10000 Monte Carlo simulations. Time elapsed: 0s.
99% confidence interval on the p-value:
] 0.258, 0.281 [

Test interpretation:
H0: Data are homogeneous
Ha: There is a date at which there is a change in the data
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 26.92%.

Standard normal homogeneity test (SNHT) (351.8):

T0	3.362
t	24
p-value (Two-tailed)	**0.560**
alpha	0.05

The p-value has been computed using 10000 Monte Carlo simulations. Time elapsed: 0s.
99% confidence interval on the p-value:
] 0.547, 0.573 [

Test interpretation:
H0: Data are homogeneous
Ha: There is a date at which there is a change in the data
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 55.98%.

Buishand's test (351.8):

Q	5.526
t	24
p-value (Two-tailed)	**0.290**
alpha	0.05

The p-value has been computed using 10000 Monte Carlo simulations. Time elapsed: 0s.
99% confidence interval on the p-value:

Test interpretation:
H0: Data are homogeneous
Ha: There is a date at which there is a change in the data
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 29.03%.

---

von Neumann's test (351.8):

N	1.444
p-value (Two-tailed)	**0.038**
alpha	0.05

The p-value has been computed using 10000 Monte Carlo simulations. Time elapsed: 0s.
99% confidence interval on the p-value:
] 0.033, 0.043 [

Test interpretation:
H0: Data are homogeneous
Ha: There is a date at which there is a change in the data
As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.
The risk to reject the null hypothesis H0 while it is true is lower than 3.78%.

Result: Central Mesaria hydrologic yearly averaged rainfall values homogeneity was checked by four different tests and compared through the p-value of confidence interval α=5%:

1- Pettitt test : p-value = 0.27 > 0.05 **Homogenous**
2- SNHT test: p-value= 0.56 > 0.05 **Homogenous**
3- BR test: p-value= 0.29 > 0.05 **Homogenous**
4- VNR test: p-value= 0.04 < 0.05 **Not Homogenous**

4.2.3.2 Homogeneity test of checking the correlation of Central Mesaria rainfall with respect to nearby 5 regions rainfall

Table 4.4: Correlation results of Central Mesaria regions for t-test and F-test with respect to 5 nearby regions

Central Mesaria		East Coast	North Coast	Karpaz	East Mesaria	West Mesaria
mean	299.9	334.7	461.8	449.7	324.2	310.8
Sd	75.9	92.8	117.1	116.8	94.8	88.7
t (Central Mesaria with others)		-1.5608<1.69	-6.2458>1.69	-5.7911>1.69	-1.0775<1.69	-0.5013<1.69
checking t-test		**Acceptable**	**Unacceptable**	**Unacceptable**	**Acceptable**	**Acceptable**
F (Central Mesaria with others)		0.6684<1.72	0.4203<1.72	0.4224<1.72	0.6414<1.72	0.7326<1.72
checking F-test		**Acceptable**	**Acceptable**	**Acceptable**	**Acceptable**	**Acceptable**

Result: Based on t-test and F-test, Central Mesaria region rainfall distribution is correlated with East Coast, East Mesaria, and West Mesaria regions rainfall distributions.

4.2.4 Consistency test of Central Mesaria regions' rainfall with respect to the mean of the nearby 5 regions rainfall using double mass curve

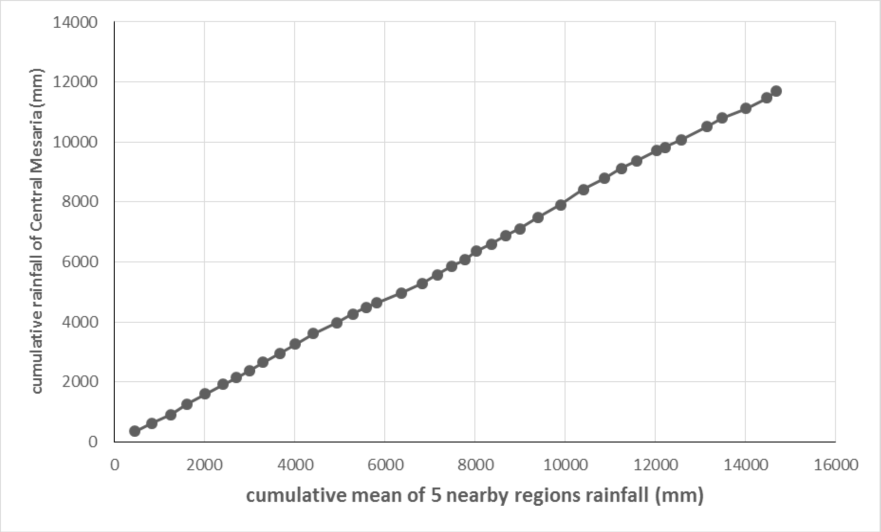


Figure 4.5: Central Mesaria region rainfall consistency check through double mass curve method with respect to nearby 5 other meteorological regions hydrologic yearly average rainfall

Result: Since all the data falls nearly on a straight line with no clear deviation, the rainfall of Central Mesaria region is accepted to be consistent.

4.2.5 Trend (Mann-Kendall and Sen’s Median Slope) Tests of Central Mesaria

The result of XLSTAT is used to perform Mann- Kendall and Sens Median Slope tests for testing the existence of trend and is detailed below:

```
XLSTAT 2015.4.01.20978  - Mann-Kendall trend tests - on 8/26/2015 at 6:39:46 PM
Confidence interval (%): 5
Confidence interval (%)(Sen's slope): 5

Summary statistics:

                                                                              Std.
      Variable        Observations     Minimum      Maximum        Mean     deviation
       351.8               39          107.500      510.300      298.574     76.437

Mann-Kendall trend test / Two-tailed test (351.8):

Kendall's tau                     0.046
S                                32.000
Var(S)                         6326.000
p-value (Two-tailed)              0.697
Alpha (α)                          0.05
The exact p-value could not be computed. An approximation has been used to compute the p-value.

Test interpretation:
H0: There is no trend in the series
H1: There is a trend in the series
As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null
hypothesis H0.
The risk to reject the null hypothesis H0 while it is true is 69.67%.

The continuity correction has been applied.

Ties have been detected in the data and the appropriate corrections have been applied.

Sen's slope:                                        0.521
Confidence interval:                               -0.059,    0.769
```

Result: p-value of Mann-Kendall test is 0.697 > 5%, therefore there is no trend in the rainfall of Central Mesaria although Sens slope is 0.521.

4.2.6 Quality tests table of Central Mesaria

Table 4.5: Quality tests results of Central Mesaria rainfall

Quality Tests	**Parametric or Non-Parametric test checking**
Normality	• p-value = 0.38>0.05. Therefore, the Central Mesaria Rainfall is normally distributed.
Homogeneity	• Based on Pettitt, SNHT, and BR tests the rainfall of Central Mesaria is homogenous (except VNR test). • Based on t-test and F-test, Central Mesaria regions' rainfall distribution is correlated with East Coast, East Mesaria, and West Mesaria regions rainfall hence proving regional homogeneity.
Consistency	• Rainfall of Central Mesaria region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in Central Mesaria rainfall although Sens slope = 0.521 since Mann – Kendall p-value is 0.697 > 0.05
Stationarity	• Central Mesaria rainfall is stationary based on ADF test since slope of regression $\gamma = 0.44 > 0$.

4.2.7 Probability distributions details of Central Mesaria region

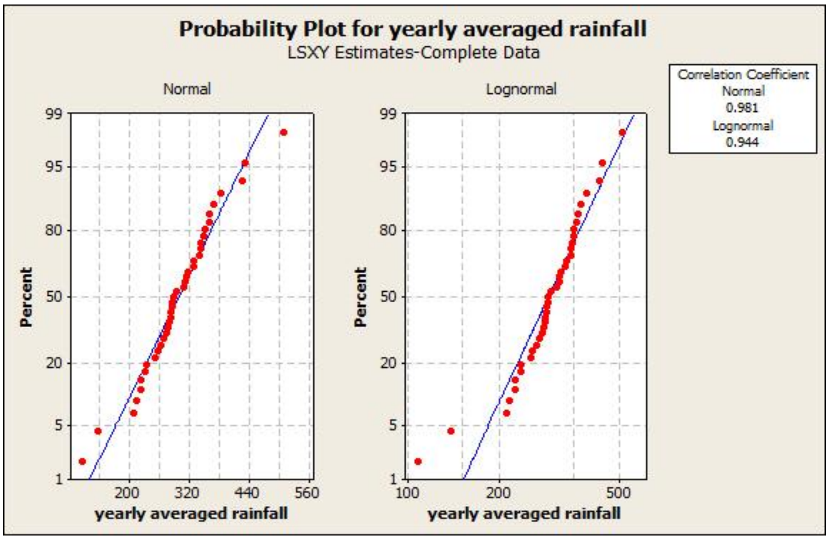


Figure 4.6: Central Mesaria rainfall fit on normal and log-Normal probability distributions

Table 4.6: Equations of the probability distribution functions with their correlation coefficients for Central Mesaria rainfall

Name	Equation	Correlation Coefficient
Normal	x = 299.9 + 74.9 Z	0.981
log-Normal	y = logx = 2.5 + 0.1 Z	0.933

Result: Comparing the correlation coefficients of the two probability distributions for Central Mesaria rainfall, it is concluded that, normal distribution is fitted better than log-Normal distribution being greater correlation coefficient value.

4.2.8 Forecasted values by time series models of Central Mesaria rainfall for the hydrologic years period 2003-04 to 2013-14

4.2.8.1 Markov Model

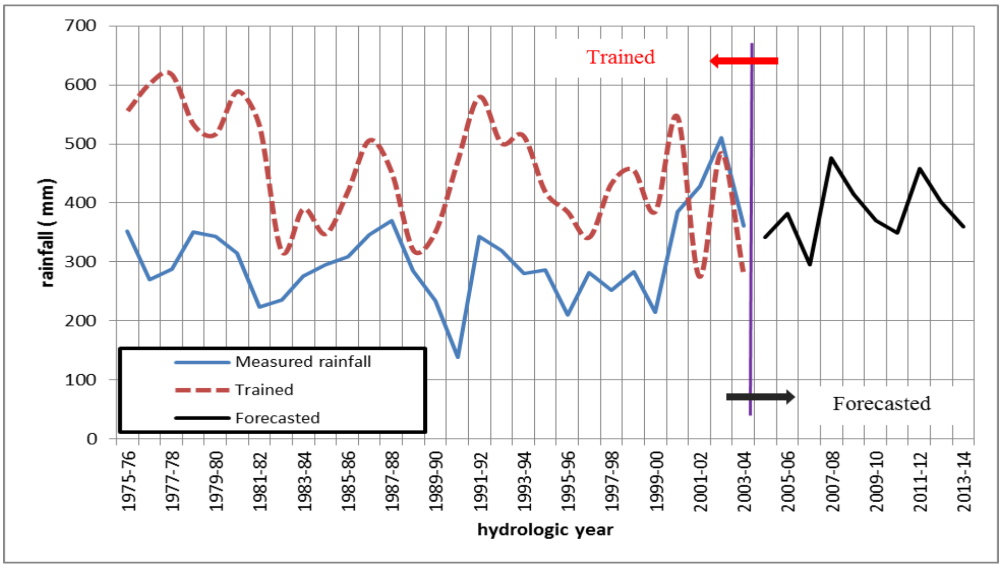


Figure 4.7: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of Central Mesaria

4.2.8.2 Auto-Regressive (AR) Model

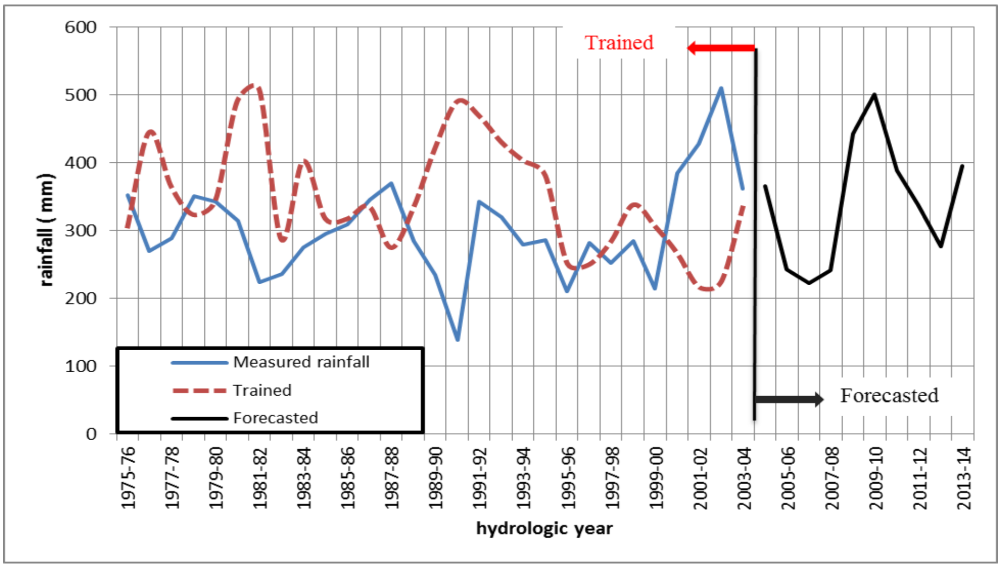


Figure 4.8: Graphical comparison of AR(1) model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of Central Mesaria

4.2.8.3 Holt-Winter Multiplicative Model

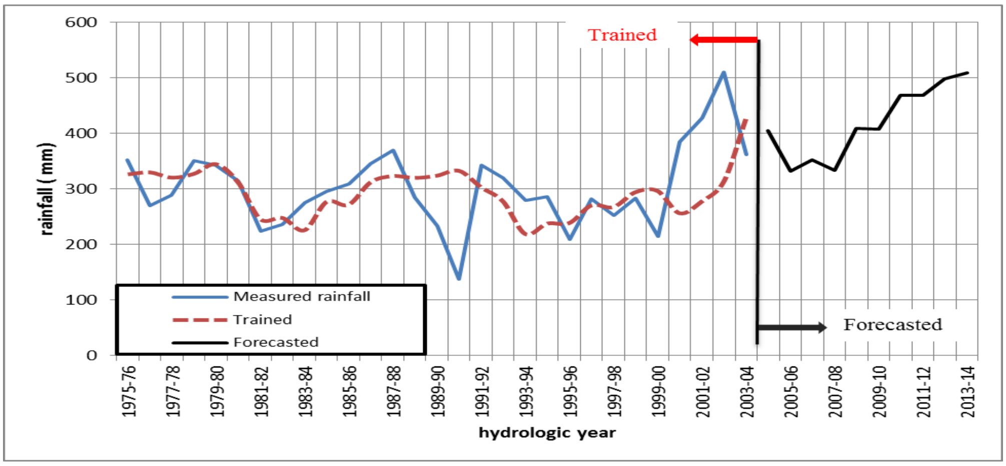


Figure 4.9: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of Central Mesaria

4.2.8.4 Selecting the best fitted time series model for Central Mesaria region

Table 4.7: Accuracy checking of forecasted data of Central Mesaria

Hydrologic Years	**DATA**			
	Measured	**Forecasted by**		
		Markov	**AR(1)**	**Holt-Winter Multiplicative**
2004-05	331.3	340.9	366.1	404.8
2005-06	265.2	381.3	243.0	332.0
2006-07	329.0	295	221.8	352.4
2007-08	107.5	476.2	241.2	333.7
2008-09	257.0	414.4	442.8	409.4
2009-10	433.8	370.4	500.5	407.3
2010-11	288.9	348.4	388.7	468.2
2011-12	314.6	457.5	337.9	468.8
2012-13	360.5	401.8	276.4	498.6
2013-14	224.3	360.1	395.4	509.7
MSE		**19862.8**	**11688.1**	**24434.6**
Ratio w.r.t. Min.		**1.70**	**1.00**	**2.09**
MAPE (%)		**27.4**	**28.4**	**29.9**
Ratio w.r.t. Min.		**1.00**	**1.03**	**1.09**
RMSE		**2235.7**	**1168.8**	**2419.6**
Ratio w.r.t. Min.		**1.91**	**1.00**	**2.07**
MAD		**112.87**	**92.87**	**132.6**
Ratio w.r.t. Min.		**1.22**	**1.00**	**1.43**
Overall		**5.83**	**4.03**	**6.68**

Result : AR(1)model is the best model among the others, having the lowest overall error ratio based on the given 4 standardized error measures. Hence, for Central Mesaria Region, AR(1) model is used to generate (predict) the rainfall for the hydrologic years 2014-15 to 2018-19 which are tabulated below.

4.2.8.5 Prediction of yearly rainfall of Central Mesaria region for hydrologic years 2014-2015 to 2018-2019

Table 4.8: Expected yearly rainfall of Central Mesaria region based on AR(1) model for hydrologic years 2014-15 to 2018-19

Hydrologic Years	Expected yearly rainfall (mm)
2014-2015	269.4
2015-2016	244.6
2016-2017	238.0
2017-2018	217.5
2018-2019	219.6

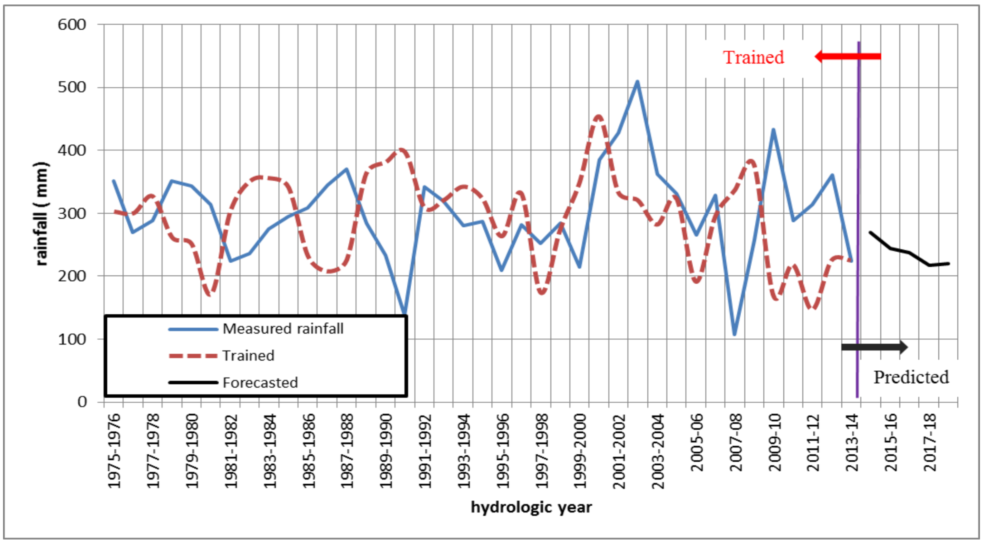


Figure 4.10: Expected (predicted) yearly rainfall of Central Mesaria region based on AR(1) model for hydrologic years 2014-15 to 2018-19

4.2.9 Wet and dry spells of Central Mesaria

Table 4.9: Numerical representation of wet and dry spells of Central Mesaria region

Hydrologic years	Rainfall (mm)	Wet\ Dry	0: Dry 1: Wet
Mean = 299.9 mm			
1975-1976	351.8	wet	1
1976-1977	270.3	dry	0
1977-1978	288.4	dry	0
1978-1979	350.9	wet	1
1979-1980	343.3	wet	1
1980-1981	314.1	wet	1
1981-1982	224.3	dry	0
1982-1983	235.6	dry	0
1983-1984	275.2	dry	0
1984-1985	295.3	dry	0
1985-1986	309.1	wet	1
1986-1987	345.4	wet	1
1987-1988	369.7	wet	1
1988-1989	284.5	dry	0
1989-1990	233.9	dry	0
1990-1991	138.1	dry	0
1991-1992	342.6	wet	1
1992-1993	319.9	wet	1
1993-1994	279.8	dry	0
1994-1995	286.5	dry	0
1995-1996	209.6	dry	0
1996-1997	281.5	dry	0
1997-1998	252.2	dry	0
1998-1999	283.9	dry	0
1999-2000	214.9	dry	0
2000-2001	385	wet	1
2001-2002	427.7	wet	1
2002-2003	510.3	wet	1
2003-2004	361.7	wet	1
2004-2005	331.3	wet	1
2005-2006	265.2	dry	0
2006-2007	329	wet	1
2007-2008	107.5	dry	0
2008-2009	257	dry	0
2009-2010	433.8	wet	1
2010-2011	288.9	dry	0
2011-2012	314.6	wet	1
2012-2013	360.5	wet	1
2013-2014	224.3	dry	0

Result: Number of wet spells = 18 (47 %), number of dry spells = 21 (53 %).

Therefore Central Mesaria region is in dry spell during the studied period.

4.2.10 Study of monthly wet and dry spells of Central Mesaria region for hydrological years 1975-76 to 2013-14

Table 4.10: Monthly wet and dry spell of Central Mesaria

Months	No. of wet spells	No. of dry spells	Conclusion
Sep	13	26	67 % dry
Oct	14	25	64 % dry
Nov	16	23	59 % dry
Dec	17	23	56 % dry
Jan	15	24	62 % dry
Feb	17	22	56 % dry
Mar	19	20	51 % dry
Apr	14	25	64 % dry
May	14	25	64 % dry

Result: Since all the months are in dry spell, Central Mesaria region throughout the year is dry during the studied period.

4.3 Meteorological Region: East Coast

Table 4.11: Total rainfall of East Coast region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic Year	Month												Total
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Agu	
1975-76	0.0	0.5	38.8	100.4	35.8	40.9	44.0	63.5	34.1	0.0	0.0	3.4	361.4
1976-77	22.5	56.6	69.5	92.4	62.8	7.9	42.8	26.2	0.0	0.0	1.5	0.0	382.2
1977-78	4.4	2.7	2.9	95.6	102.2	26.3	22.0	6.8	0.0	0.0	0.0	0.0	262.9
1978-79	0.0	21.2	6.7	151.3	40.3	39.6	41.5	2.1	12.6	2.7	0.0	0.0	318.0
1979-80	5.9	16.8	36.9	127.1	26.9	108.2	33.2	12.8	42.2	0.0	0.0	0.0	410.0
1980-81	1.3	6.7	7.0	46.4	119.8	57.0	37.8	7.2	12.3	45.2	0.0	0.0	340.7
1981-82	0.0	1.8	66.0	31.6	30.8	48.1	22.7	17.1	6.4	0.2	0.0	0.0	224.7
1982-83	2.1	19.7	22.2	27.0	55.0	41.2	48.7	18.7	5.8	9.2	0.0	0.2	249.8
1983-84	1.6	10.9	58.0	23.1	51.7	41.3	50.4	62.7	1.8	0.0	2.2	3.3	307.0
1984-85	0.0	0.7	158.7	70.4	46.4	19.4	24.9	13.0	0.0	3.4	0.0	0.0	336.9
1985-86	0.3	24.8	28.4	85.3	34.9	72.2	6.3	4.2	74.4	5.1	0.0	0.0	335.9
1986-87	0.2	43.2	34.0	59.0	23.1	11.6	125.2	33.9	9.7	0.3	0.0	0.0	340.2
1987-88	0.0	69.9	10.3	166.0	83.4	95.3	85.8	7.4	6.8	9.8	0.0	2.1	536.8
1988-89	0.2	33.3	54.0	89.9	77.5	22.0	35.9	0.0	12.1	5.4	0.0	0.0	330.3
1989-90	0.0	43.0	33.4	29.8	20.2	139.0	29.3	10.6	0.9	0.0	0.0	6.4	312.6
1990-91	0.0	4.1	12.8	9.2	71.1	72.0	45.9	10.8	1.8	0.0	0.0	0.0	227.7
1991-92	0.0	8.6	102.9	277.2	14.0	53.1	18.8	2.7	44.6	13.4	12.4	0.2	547.9
1992-93	0.0	3.8	61.5	125.0	66.8	52.3	53.2	7.9	43.5	25.2	0.0	0.0	439.2
1993-94	0.0	0.0	57.2	10.4	106.1	50.8	55.0	24.5	3.4	2.8	0.0	6.0	316.2
1994-95	0.5	29.0	103.6	52.4	17.7	12.0	7.8	19.3	18.4	0.0	8.0	0.0	268.7
1995-96	0.0	5.8	36.1	5.0	109.0	28.8	45.1	19.3	0.3	0.0	0.0	0.0	249.4
1996-97	0.8	34.3	41.0	39.5	11.1	26.5	31.6	38.6	1.7	0.7	0.0	4.4	230.2
1997-98	33.2	28.7	40.4	63.6	43.8	2.9	34.6	14.0	34.3	0.0	0.0	0.0	295.5
1998-99	0.4	0.0	16.8	85.8	59.2	32.2	20.1	20.7	2.9	10.2	0.0	5.8	254.1
1999-00	17.8	14.8	9.3	23.3	24.7	30.2	44.1	72.2	11.8	0.0	0.0	0.0	248.2
2000-01	16.7	38.4	95.6	109.5	45.5	30.6	4.1	14.9	11.0	0.0	0.0	0.0	366.3
2001-02	1.4	9.4	36.5	180.4	84.5	26.5	20.5	50.9	24.7	0.0	0.0	0.0	434.8
2002-03	16.6	15.2	26.9	114.5	39.2	58.3	80.8	37.0	0.4	13.6	0.0	0.0	402.5
2003-04	0.0	4.4	7.1	93.0	259.7	72.3	0.4	8.9	3.7	7.4	0.0	0.0	456.9
2004-05	0.0	13.3	74.0	104.1	97.2	20.4	13.0	20.2	0.0	31.7	0.0	0.0	373.9
2005-06	6.1	4.9	95.6	6.7	99.9	44.5	23.3	12.7	1.3	0.0	23.1	0.0	318.1
2006-07	6.5	32.8	28.5	16.4	20.4	188	27.5	33.3	56.6	0.4	0	0	410.4
2007-08	0	1.6	20.5	59.1	27.2	18.8	8.4	7.3	19.1	0	0	0	162
2008-09	13.1	12.9	20.5	45.1	52.4	50.5	43.1	18.8	4.2	0	0	0	260.6
2009-10	27.2	12.4	28.8	153.8	69.6	125.4	0.4	9.7	3	3.3	0	0	433.6
2010-11	0.3	12	0	50.3	88.8	37.7	45.7	27.2	12.3	5.6	0	0.2	280.1
2011-12	28.9	16	80.4	65.5	149.4	59.1	20	8.3	83.4	0.9	1.9	0	513.8
2012-13	0	41.6	50.2	111.5	41.3	31.3	3.4	21.6	43.5	0	0	0	344.4
2013-14	8.2	4	0.6	45.4	14.3	15	24.6	12.9	38.3	5.8	0.1	0	169.2
2014-15	5	54.2	29.8	59									

Table 4.12: Statistical measures of East Coast rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	C_{Vlog}	C_{Slog}	$\bar{x}_{med}$	C_{dx}	PC_v
334.7	92.8	0.299	0.277	0.4	2.5	0.1	0.038	0.05	-0.3	330.3	0.211	0.64

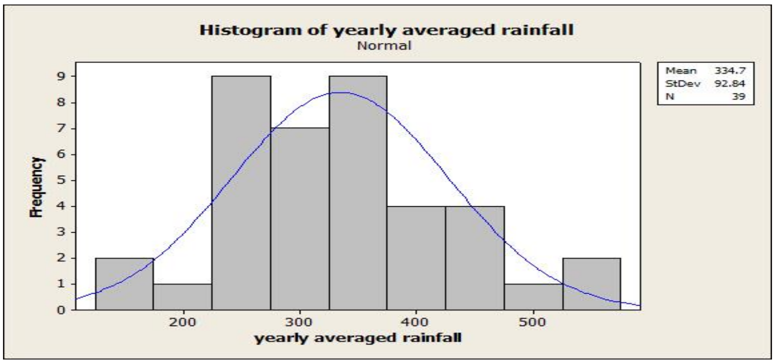


Figure 4.11: East Coast regions' rainfall histogram

4.3.1 Empirical determination of minimum required sample size of rainfall for East Coast region based on mean and standard deviation

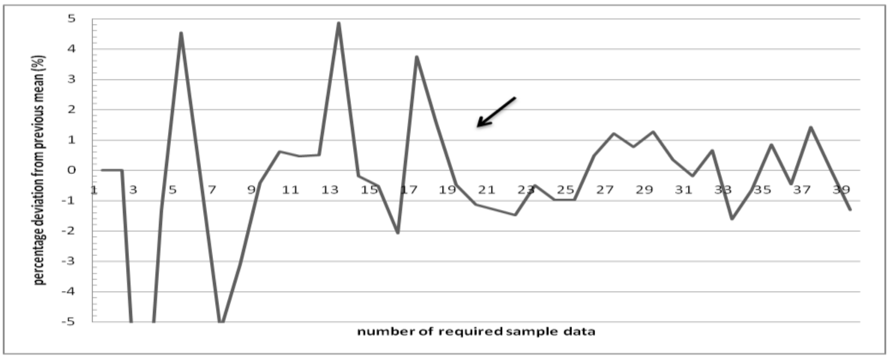


Figure 4.12: Curve showing the required number of sample size for East Coast regions' rainfall, based on the percentage deviations of the mean values

Comment: The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 19.

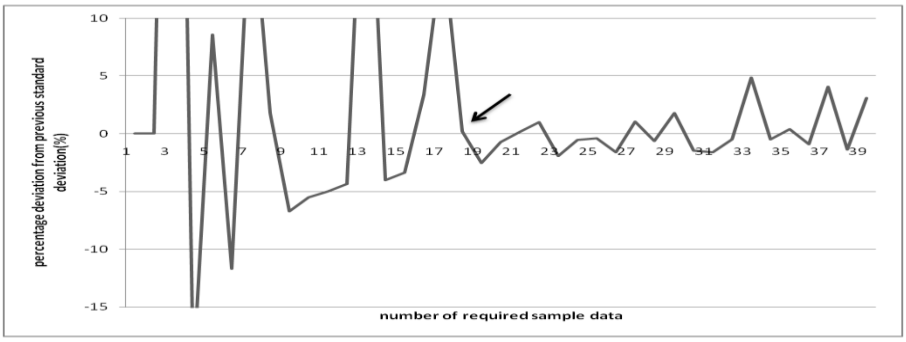


Figure 4.13: Curve showing the required number of sample size for East Coast regions' rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 19. Hence n_{min} = 19 years based on the minimum number of required data.

Table 4.13: Appropriate rainfall sample size of East Coast for statistic and probabilistic studies.

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
19 < 39 OK	19 < 39 OK

4.3.2 Quality Test Table of East Coast

Table 4.14: Quality test results of East Coast rainfall

Quality Tests	Parametric or Non-Parametric test checking
Normality	• p-value = 0.48>0.05. Therefore, the East Coast Rainfall is normally distributed.
Homogeneity	• Based on Pettitt, SNHT, BR, and VNR tests East Coast rainfall time series is Homogenous. • Based on t-test and f-test, East Coast region rainfall distribution is correlated with Central Mesaria, Esat Mesaria,and West Mesaria region rainfall proving homogeneity
Consistency	• Rainfall of East Coast region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in East Coast rainfall because Mann – Kendall p-value = 0.92>0.05 • Sens slope=0.162
Stationarity	• East Coast rainfall is stationary based on ADF test since slope of regression γ = 0.28 > 0.

4.3.3 Probability distributions details of East Coast Region

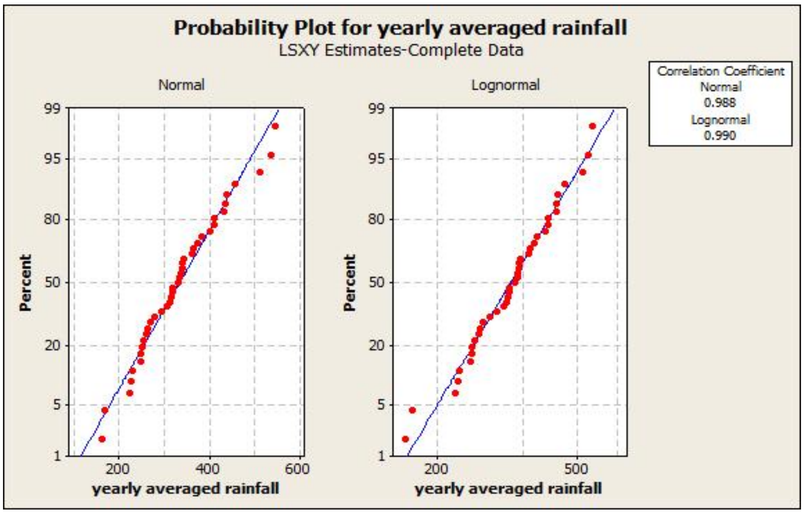


Figure 4.14: East Coast rainfall fit on normal and log-Normal probability distributions

Table 4.15: Equations of the Probability distribution functions with their correlation coefficients for east Coast rainfall

Name	Equation	Correlation Coefficient
Normal	x = 334.7 + 91.6 Z	0.980
log- Normal	y = logx = 2.5 + 0.1 Z	0.990

Result: Comparing the correlation coefficients of the two probability distributions for East Coast rainfall, it is concluded that, log-Normal distribution has the best fitted curve being greater correlation coefficient value.

4.3.4 Forecasted values by time series models of East Coast rainfall for the hydrologic years period 2003-04 to 2013-14

4.3.4.1 Markov Model

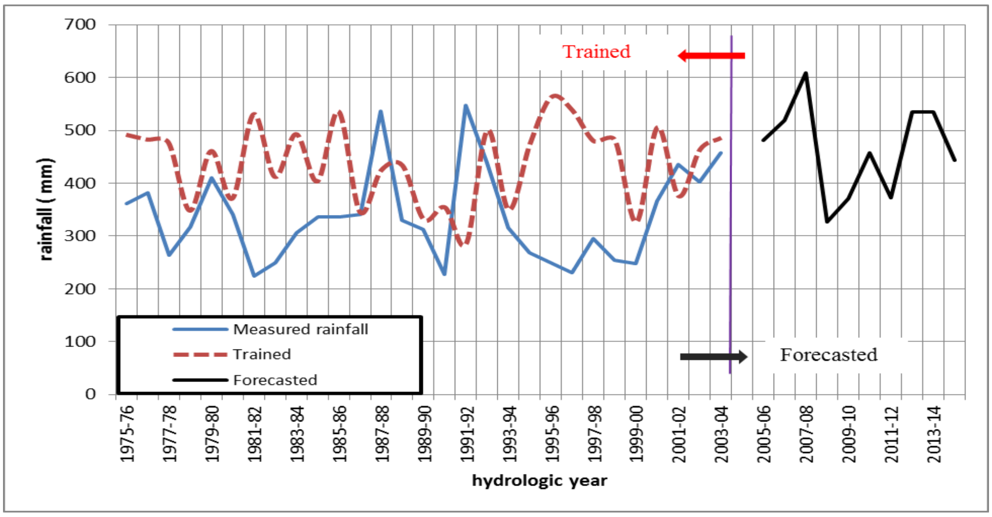


Figure 4.15: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of East Coast

4.3.4.2 Auto-Regressive Model

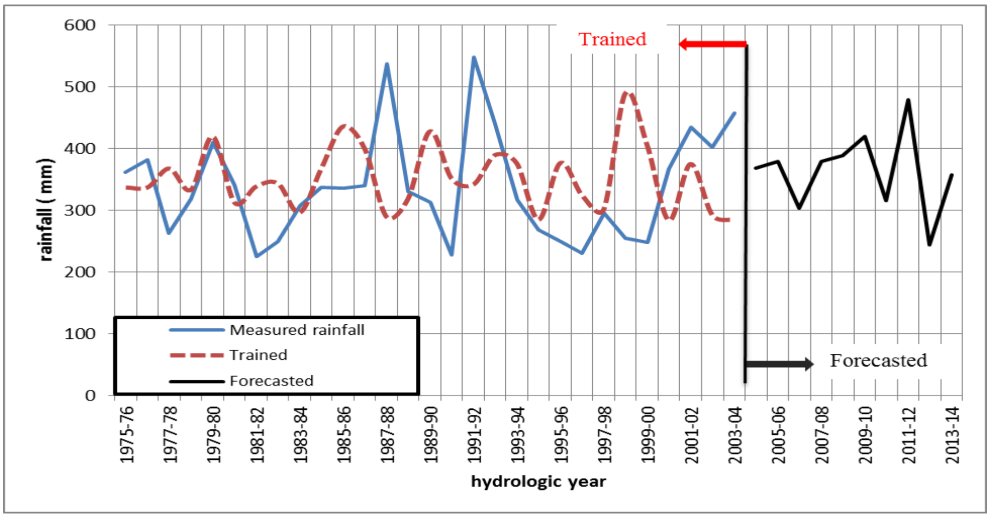


Figure 4.16: Graphical comparison of AR(2) model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of East Coast

4.3.4.3 Holt-Winter Multiplicative Model

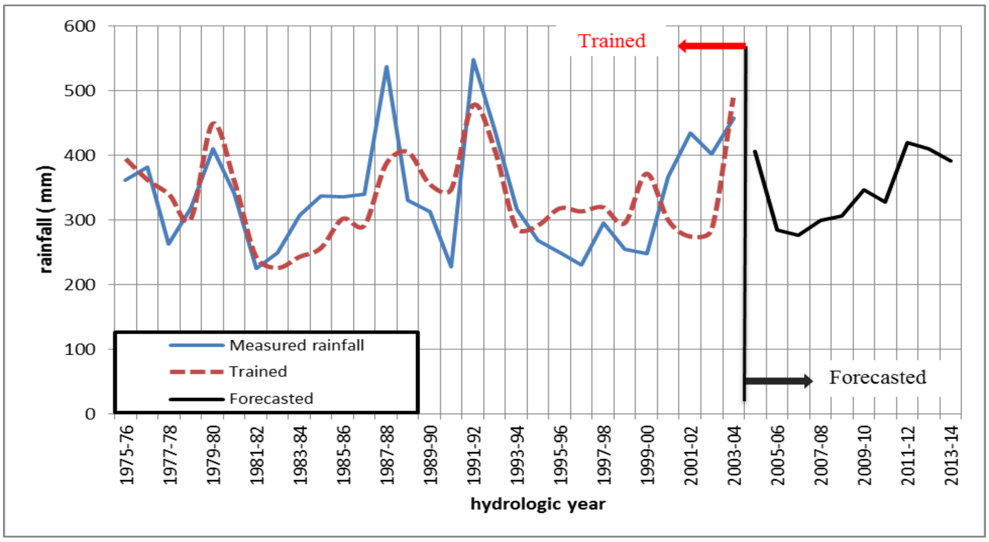


Figure 4.17: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of East Coast

4.3.4.4 Selecting the best fitted time series model for East Coast region

Table 4.16: Accuracy checking of forecasted data of East Coast

Hydrologic Year	**DATA**			
	Measured	**Markov Model**	**AR(2)**	**Holt-Winter model**
2004-05	373.9	481.5	367.8	406.1
2005-06	318.1	519.7	379.1	284.4
2006-07	410.4	609.4	303.8	275.8
2007-08	162.0	326.1	378.5	298.7
2008-09	260.6	370.5	387.7	306.4
2009-10	433.6	458.3	419.8	345.9
2010-11	280.1	372.5	316.0	327.7
2011-12	513.8	534.6	478.4	419.1
2012-13	344.4	534.9	244.4	410.1
2013-14	169.2	444.0	357.0	391.0
MSE		**25224.1**	**12617.1**	**11351.1**
Ratio w.r.t. Min		**2.22**	**1.11**	**1.00**
MAPE (%)		**30.5**	**25.8**	**26.5**
Ratio w.r.t. Min		**1.18**	**1.00**	**1.02**
RMSE		**2522.4**	**1261.7**	**1135.1**
Ratio w.r.t. Min		**2.22**	**1.11**	**1.00**
MAD		**138.5**	**89.0**	**90.1**
Ratio w.r.t. Min		**1.56**	**1.00**	**1.01**
overall		**7.18**	**4.22**	**4.04**

Result: Holt-Winter model is the best model based on the above 4 error measures having the lowest overall error ratio among the other models. Hence, for East Coast region, this model is used to generate the rainfall for the Hydrologic years 2014-15 to 2018-19 and is all tabulated bellow.

4.3.4.5 Prediction of yearly rainfall of East Coast region for hydrologic years 2014-2015 to 2018-19

Table 4.17: Expected yearly rainfall of East Coast region based on Holt-Winter model for hydrologic years 2014-15 to 2018-19

Hydrologic Year	Expected Yearly Total Rainfall (mm)
2014-2015	304.5
2015-2016	441.5
2016-2017	353.1
2017-2018	258.8
2018-2019	274.6

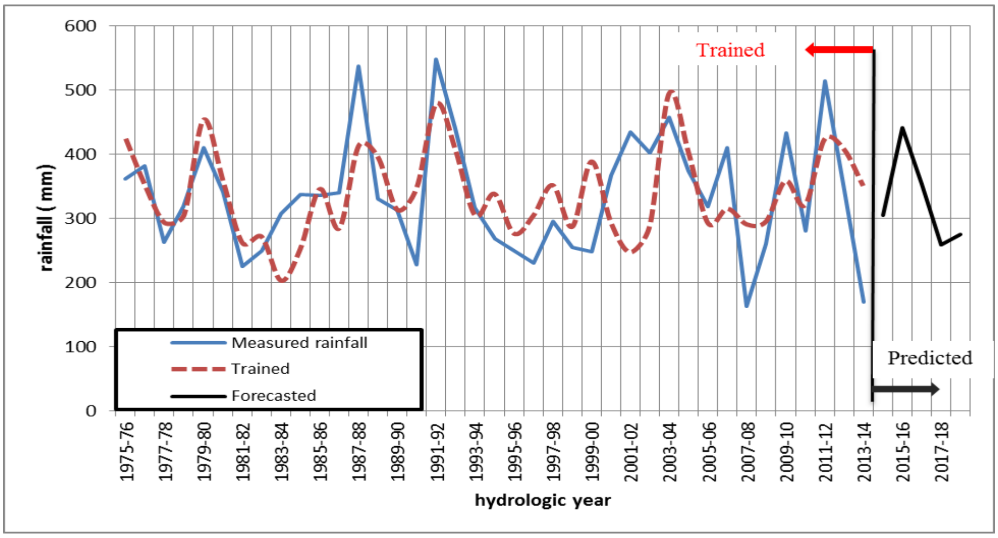


Figure 4.18: Expected (predicted) yearly rainfall of East Coast region based on Holt-Winter model for hydrologic years 2014-15 to 2018-19

4.3.5 Wet and dry spells for East Coast

Table 4.18: Numerical representation of wet and dry spells of East Coast region based on the mean of the data

Hydrologic Year	**Rainfall (mm)**	**Wet\ Dry**	**0: Dry 1: Wet**
Mean = 334.7 mm			
1975-1976	361.4	wet	**1**
1976-1977	382.2	wet	**1**
1977-1978	262.9	dry	**0**
1978-1979	318	dry	**0**
1979-1980	410	wet	**1**
1980-1981	340.7	wet	**1**
1981-1982	224.7	dry	**0**
1982-1983	249.8	dry	**0**
1983-1984	307	dry	**0**
1984-1985	336.9	wet	**1**
1985-1986	335.9	wet	**1**
1986-1987	340.2	wet	**1**
1987-1988	536.8	wet	**1**
1988-1989	330.3	dry	**0**
1989-1990	312.6	dry	**0**
1990-1991	227.7	dry	**0**
1991-1992	547.9	wet	**1**
1992-1993	439.2	wet	**1**
1993-1994	316.2	dry	**0**
1994-1995	268.7	dry	**0**
1995-1996	249.4	dry	**0**
1996-1997	230.2	dry	**0**
1997-1998	295.5	dry	**0**
1998-1999	254.1	dry	**0**
1999-2000	248.2	dry	**0**
2000-2001	366.3	wet	**1**
2001-2002	434.8	wet	**1**
2002-2003	402.5	wet	**1**
2003-2004	456.9	wet	**1**
2004-2005	373.9	wet	**1**
2005-2006	318.1	dry	**0**
2006-2007	410.4	wet	**1**
2007-2008	162	dry	**0**
2008-2009	260.6	dry	**0**
2009-2010	433.6	wet	**1**
2010-2011	280.1	dry	**0**
2011-2012	513.8	wet	**1**
2012-2013	344.4	wet	**1**
2013-2014	169.2	dry	**0**

Result: number of wet spells = 19 (49%), number of Dry spells = 20 (51%). Therefore the East Coast is in dry spell during the studied period.

4.3.6 Study of monthly wet and dry spells of East Coast region for hydrological years 1975-76 to 2013-14

Table 4.19: Monthly wet and dry spell of East Coast

Months	No. of wet spell	No. of dry spell	Conclusion
Sep	12	27	69% dry
Oct	14	25	64% dry
Nov	14	25	64% dry
Dec	18	21	54% dry
Jan	16	23	59% dry
Feb	15	24	62% dry
Mar	18	21	54% dry
Apr	13	26	67% dry
May	13	26	67% dry

Result: Since all the months are in dry spell, East Coast region throughout the year is dry during the studied period.

4.4 Meteorological Region: East Mesaria

Table 4.20: Total rainfal data of East Mesaria region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic Year	Month												Totall
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Agu	
1975-76	0.0	8.7	17.8	136.7	20.4	42.4	35.6	39.5	46.8	2.0	5.5	0.0	355.4
1976-77	0.4	24.3	44.2	54.0	48.9	7.9	43.0	29.6	0.0	0.1	7.8	0.0	260.2
1977-78	13.1	7.3	7.4	113.4	129.3	29.3	37.8	7.8	0.0	0.4	0.0	0.0	345.8
1978-79	0.0	12.1	2.5	82.1	35.4	62.6	34.8	4.8	24.9	29.6	1.4	2.1	292.3
1979-80	0.2	38.0	25.1	110.8	30.9	101.9	28.2	15.2	2.3	0.1	0.0	1.3	354.0
1980-81	0.2	16.4	8.4	42.7	106.8	54.6	44.4	24.7	25.4	24.7	0.0	0.0	348.3
1981-82	0.0	2.7	57.2	30.4	27.4	35.9	39.1	14.6	8.4	10.0	0.1	3.2	229.0
1982-83	4.5	12.7	14.9	15.6	30.7	32.7	30.6	18.0	20.6	9.3	0.0	0.5	190.1
1983-84	0.0	21.1	63.3	13.8	40.8	42.8	33.3	66.5	2.9	0.0	0.2	3.0	287.7
1984-85	0.0	2.2	174.7	60.6	39.0	34.1	31.9	15.3	6.4	8.0	0.0	0.0	372.2
1985-86	10.2	18.6	15.4	56.7	35.6	57.6	16.5	9.5	76.4	8.6	0.0	0.0	305.1
1986-87	1.0	27.6	41.4	36.7	21.4	12.1	119.4	9.6	15.1	0.3	5.2	0.5	290.3
1987-88	0.0	55.6	23.6	132.6	57.3	94.8	66.6	5.5	7.0	9.9	2.2	1.9	457.0
1988-89	4.1	40.1	46.6	92.6	84.1	18.1	20.9	0.0	9.8	0.6	0.0	0.0	316.9
1989-90	0.0	50.7	23.1	24.9	18.3	83.7	25.7	5.6	6.7	0.0	0.0	8.8	247.5
1990-91	0.0	7.2	4.1	10.1	43.2	52.3	36.5	7.4	0.9	0.1	0.0	0.1	161.9
1991-92	0.0	8.2	46.1	209.3	12.8	47.6	15.8	4.7	27.5	40.3	12.3	9.6	434.2
1992-93	0.0	3.6	60.6	126.3	29.8	56.1	52.3	7.3	42.3	14.1	0.0	0.0	392.4
1993-94	0.0	0.5	52.8	6.5	94.4	57.4	56.5	12.3	5.5	0.3	1.7	3.8	291.7
1994-95	3.0	33.4	135.4	25.7	18.0	13.8	9.5	14.9	18.9	0.0	12.3	0.0	284.9
1995-96	0.0	2.0	30.2	5.4	111.7	40.6	24.0	14.2	0.6	9.8	0.0	0.0	238.5
1996-97	0.1	31.3	14.6	50.6	8.2	20.6	29.4	33.1	7.0	6.4	0.2	3.2	204.7
1997-98	28.1	23.3	48.8	49.3	46.2	4.4	21.2	5.7	33.5	5.4	0.0	0.0	265.9
1998-99	0.8	0.0	34.5	78.5	63.8	34.7	14.0	25.5	1.8	17.9	0.0	0.1	271.6
1999-00	2.9	34.2	19.6	10.5	31.8	34.2	40.0	66.2	18.3	0.0	0.0	0.0	257.7
2000-01	13.9	51.1	91.1	150.5	53.0	34.1	7.8	21.2	34.4	0.0	0.0	7.2	464.3
2001-02	0.0	17.4	46.6	156.2	74.1	29.6	26.9	48.0	37.0	1.9	10.3	10.1	458.1
2002-03	8.6	6.4	13.7	120.8	46.6	81.9	91.8	31.5	2.0	15.9	0.0	0.0	419.2
2003-04	0.1	5.5	12.9	95.2	204.7	88.0	2.2	19.1	3.0	4.6	0.0	0.0	435.3
2004-05	0.0	5.8	39.7	76.3	71.0	21.5	16.6	18.1	15.2	45.1	0.0	0.4	309.7
2005-06	16.0	4.5	70.6	7.0	79.8	43.0	37.0	10.2	4.8	3.1	16.1	0.0	292.1
2006-07	2.6	79.3	36.5	6.7	30.8	152.2	37.2	31.9	81.9	0.5	0	4.3	463.9
2007-08	1.8	7	18.8	30.5	20.3	19.8	8.9	13.2	14.4	0	0	1.7	136.4
2008-09	5.3	19.3	20.6	61.9	49.5	40.4	51.4	27.4	18.3	0.7	0	5.3	300.1
2009-10	29.1	29.1	31.6	147.8	94	154.5	9.5	8.1	17.3	9.8	5.8	0.5	537.1
2010-11	1.9	8.5	0.7	67.3	90.6	32.2	23.5	46	36	38.2	0	0.1	345.0
2011-12	24.2	11.1	76.1	57.8	147	38	13	13.1	84.2	0.7	0.2	0.9	466.3
2012-13	0	51.7	42.2	99.3	34.8	26.1	2.7	67.3	40.6	0	0.1	1.3	366.1
2013-14	1	3.3	4	56.9	12.9	10.4	28.1	14.5	57.5	5.5	1.5	0.6	196.2
2014-15	8.8	48.8	24.6	85.1									

Table 4.21: Statistical measures of East Mesaria rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	C_{Vlog}	C_{Slog}	$\bar{x}_{med}$	C_{dx}	PC_v
324.2	94.8	0.253	0.292	0.2	2.5	0.1	0.042	0.05	-0.5	305.1	0.248	0.858

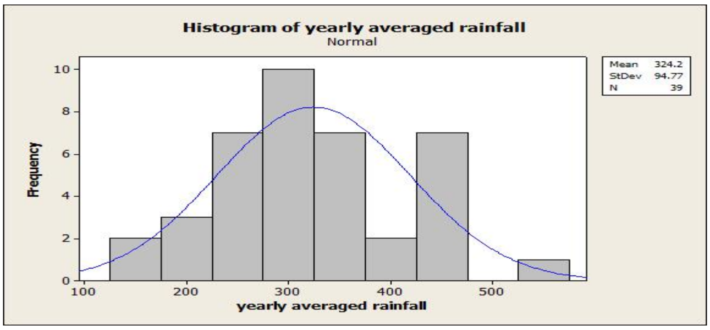


Figure 4.19: East Mesaria regions' rainfall histogram

4.4.1 Empirical determination of minimum required sample size rainfall for East Mesaria region based on mean and standard deviation

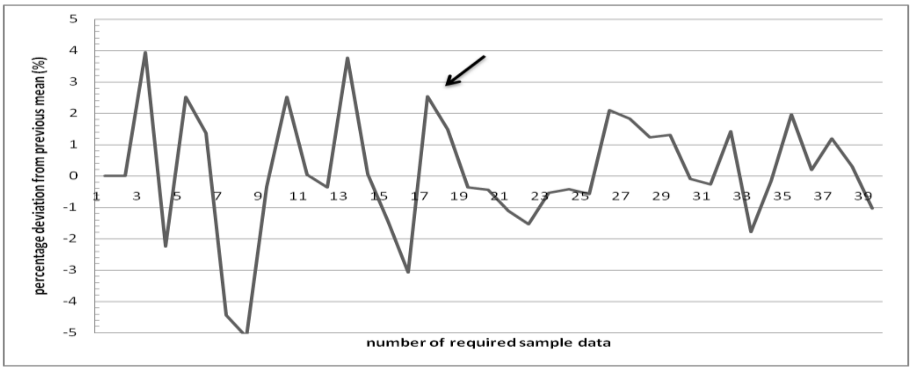


Figure 4.20: Curve showing the required number of sample size for East Mesaria regions' rainfall, based on the percentage deviations of the mean values

Comment: The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 18.

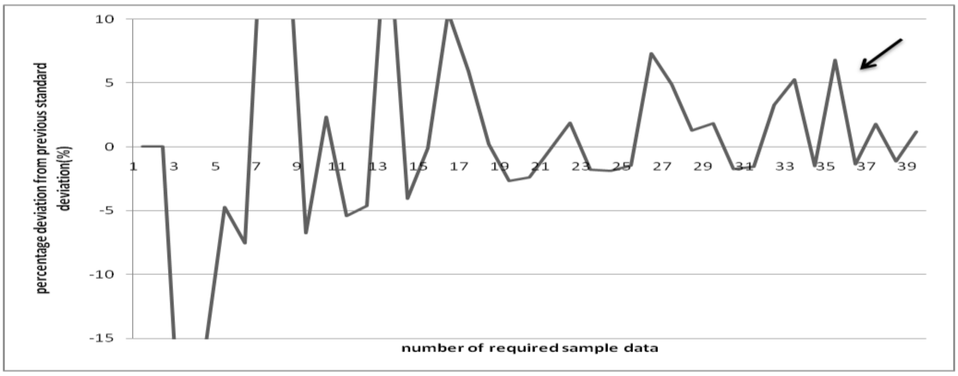


Figure 4.21: Curve showing the required number of sample size for East Mesaria regions' rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 37.

Hence $n_{min} = 37$ is the required minimum number that satisfies both mean and standard deviation for East Mesaria region rainfall.

Table 4.22: Appropriate rainfall sample size of East Mesaria for statistic and probabilistic studies.

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
18 < 39 OK	37 < 39 OK

4.4.2 Quality Checking Tests of East Mesaria

Table 4.23: Quality tests results of East Mesaria rainfall

Quality check Tests	Parametric or Non-Parametric test checking
Normality	• p-value = 0.37>0.05. Therefore, the East Mesaria rainfall is normally distributed.
Homogeneity	• Based on Pettitt, SNHT, BR, and VNR tests East Mesaria rainfall time series is Homogenous. • Based on t-test and F-test, East Mesaria region rainfal distribution is correlated with Central Mesaria,East Coast , and West Mesaria region rainfall distributions.
Consistency	• Rainfall of East Mesaria region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in East Mesaria rainfall because Mann – Kendall p-value = 0.229>0.05 • Sens slope=1.98
Stationarity	• East Mesaria rainfall is stationary based on ADF test since slope of regression $\gamma = 0.34 > 0$.

4.4.3 Probability distributions details of East Mesaria region

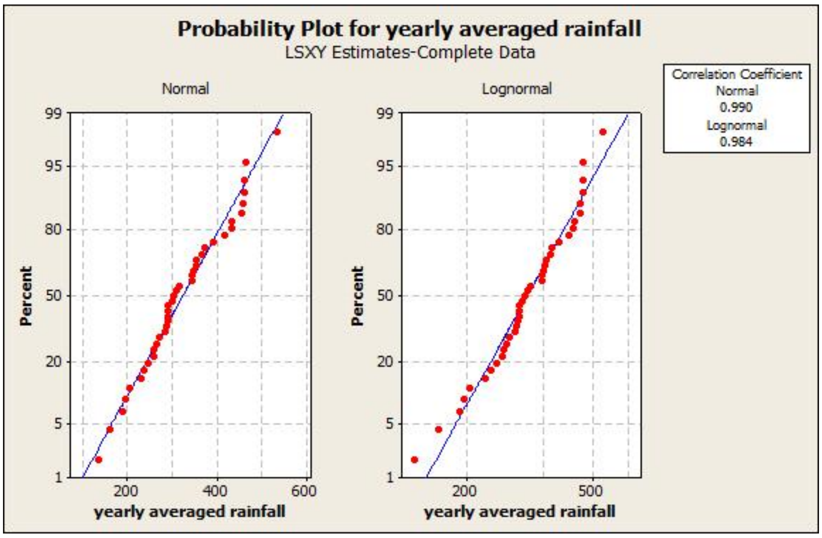


Figure 4.22: East Mesaria rainfall fit on normal and log-Normal probability distributions

Table 4.24: Equations of the probability distribution functions with their correlation coefficients for East Mesaria rainfall

Name	Equation	Correlation Coefficient
Normal	$x = 342.2 + 93.6\ Z$	0.990
log-Normal	$y = \log x = 2.5 + 0.1\ Z$	0.984

Result: Comparing the correlation coefficients of the two probability distributions for East Mesaria rainfall, it is concluded that, normal distribution has the best fitted curve being greater correlation coefficient value.

4.4.4 Forecasted values by time series models of East Mesaria rainfall for the hydrologic years period 2003-04 to 2013-14

4.4.4.1 Markov Model

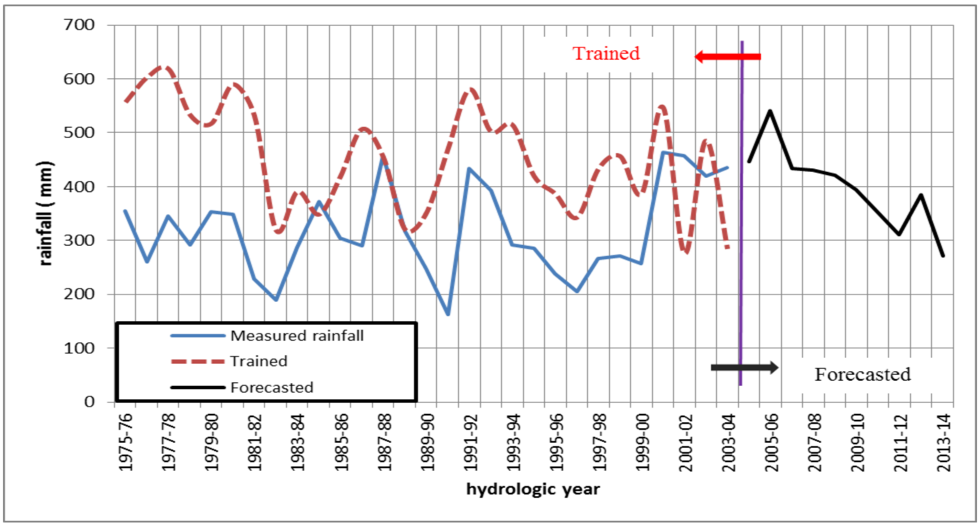


Figure 4.23: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of East Mesaria

4.4.4.2 Auto-Regressive Model

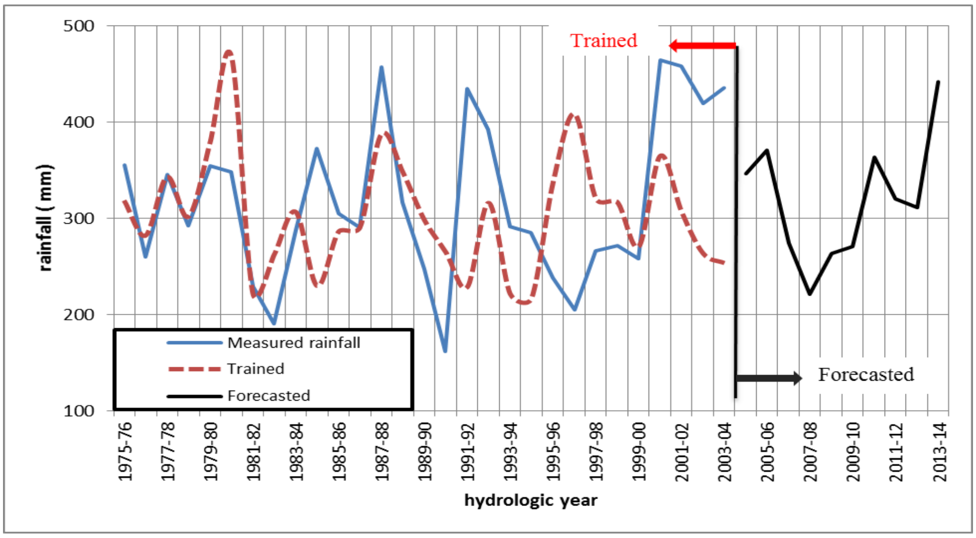


Figure 4.24: Graphical comparison of AR(1) model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of East Mesaria

4.4.4.3 Holt-Winter Multiplicative Model

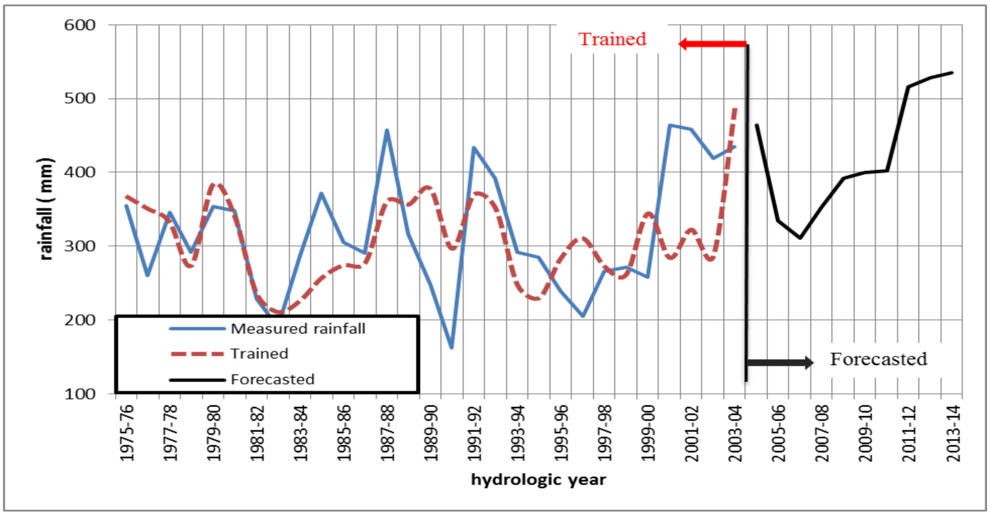


Figure 4.25: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of East Mesaria

4.4.4.4 Selecting the best fitted time series model for East Mesaria

Table 4.25: Accuracy checking of forecasted data of East Mesaria

Hydrologic Year	**DATA**			
	Measured	**Markov model**	**AR(1)**	**Holt-Winter model**
2004-05	309.7	446.0	346.6	464.2
2005-06	292.1	541.4	371.0	334.4
2006-07	463.9	433.7	274.1	311.5
2007-08	136.4	431.3	221.1	352.6
2008-09	300.1	421.1	263.8	392.1
2009-10	537.1	395.1	270.3	400.1
2010-11	345.0	352.4	363.7	402.5
2011-12	466.3	310.7	320.1	516.4
2012-13	366.1	384.2	311.0	528.4
2013-14	196.2	272.0	441.6	535.1
MSE		**23376.8**	**20827.6**	**26987.2**
Ratio w.r.t. Min		1.12	1.00	1.30
MAPE (%)		**30.1**	**37.6**	**33.2**
Ratio w.r.t. Min		1.00	1.25	1.10
RMSE		**2337.7**	**2082.8**	**2698.7**
Ratio w.r.t. Min		1.12	1.00	1.30
MAD		**123.1**	**115.9**	**140.3**
Ratio w.r.t. Min		1.06	1.00	1.21
overall		**4.30**	**4.25**	**4.91**

Result: AR(1) model is the best model based on the above 4 error measures having the lowest overall error ratio among the other models. Hence, for East Mesaria region, this model is used to generate the rainfall for the hydrologic years 2014-15 to 2018-19 and are all tabulated below.

4.4.4.5 Prediction of yearly rainfall of East Mesaria region for hydrologic years 2014-2015 to 2018-19

Table 4.26: Expected yearly rainfall of East Mesaria region based on AR(1) model for hydrologic years 2014-15 to 2018-19

Hydrologic Year	Expected Yearly Total Rainfall (mm)
2014-2015	318.4
2015-2016	338.4
2016-2017	420.3
2017-2018	304.9
2018-2019	330.5

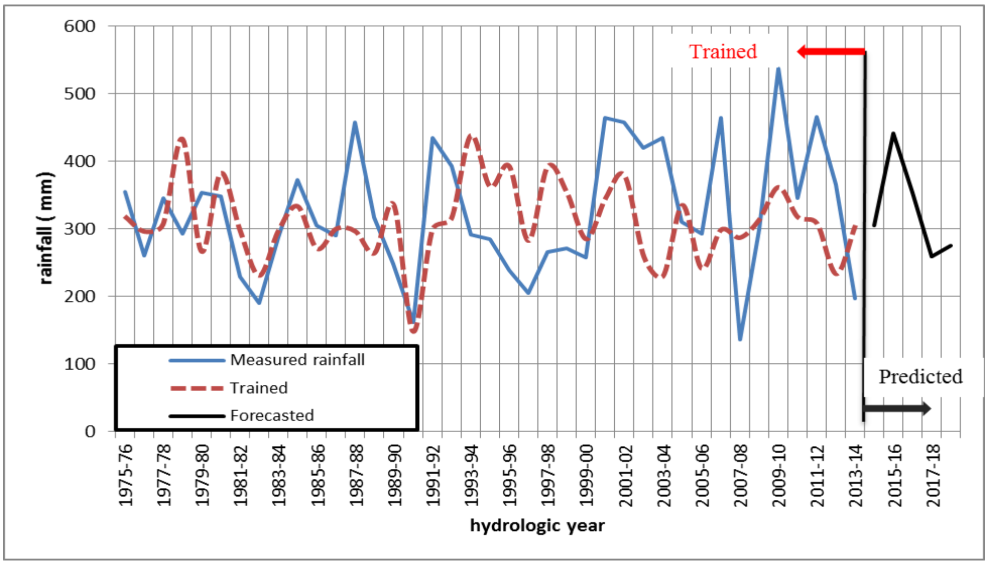


Figure 4.26: Expected (predicted) yearly rainfall of East Mesaria region based on AR(1) model for hydrologic years 2014-15 to 2018-19

4.4.5 Wet and dry spells for East Mesaria

Table 4.27: Numerical representation of wet and dry spells of East Mesaria region

Hydrologic Year	**Rainfall (mm)**	**Wet\ Dry**	**0: Dry 1: Wet**
Mean = 324.2 mm			
1975-1976	355.4	wet	**1**
1976-1977	260.2	dry	**0**
1977-1978	345.8	wet	**1**
1978-1979	292.3	dry	**0**
1979-1980	354	wet	**1**
1980-1981	348.3	wet	**1**
1981-1982	229	dry	**0**
1982-1983	190.1	dry	**0**
1983-1984	287.7	dry	**0**
1984-1985	372.2	wet	**1**
1985-1986	305.1	dry	**0**
1986-1987	290.3	dry	**0**
1987-1988	457	wet	**1**
1988-1989	316.9	dry	**0**
1989-1990	247.5	dry	**0**
1990-1991	161.9	dry	**0**
1991-1992	434.2	wet	**1**
1992-1993	392.4	wet	**1**
1993-1994	291.7	dry	**0**
1994-1995	284.9	dry	**0**
Mean = 324.2 mm			
1995-1996	238.5	dry	**0**
1996-1997	204.7	dry	**0**
1997-1998	265.9	dry	**0**
1998-1999	271.6	dry	**0**
1999-2000	257.7	dry	**0**
2000-2001	464.3	wet	**1**
2001-2002	458.1	wet	**1**
2002-2003	419.2	wet	**1**
2003-2004	435.3	wet	**1**
2004-2005	309.7	dry	**0**
2005-2006	292.1	dry	**0**
2006-2007	463.9	wet	**1**
2007-2008	136.4	dry	**0**
2008-2009	300.1	dry	**0**
2009-2010	537.1	wet	**1**
2010-2011	345	wet	**1**
2011-2012	466.3	wet	**1**
2012-2013	366.1	wet	**1**
2013-2014	196.2	dry	**0**

<u>Result</u>: number of wet spells = 17 (44%) , number of Dry spells = 22 (56%).

Therefore the East Mesaria is in dry spell during the studied period.

4.4.6 Study of monthly wet and dry spells of East Mesaria region for hydrological years 1975-76 to 2013-14

Table 4.28: Monthly wet and dry spell of East Mesaria

Months	No. of wet spell	No. of dry spell	Conclusion
Sep	10	29	74% dry
Oct	15	24	62% dry
Nov	17	22	56% dry
Dec	16	23	59% dry
Jan	14	25	64% dry
Feb	14	25	64% dry
Mar	17	22	56% dry
Apr	13	26	67% dry
May	14	25	64% dry

Result: Since all the months are in dry spell, East Mesaria region throughout the year is dry during the studied period.

4.5 Meteorological Region: Karpaz

Table 4.29: Total rainfall data of Karpaz region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic Year	Month												Totall
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Agu	
1975-76	0.0	1.2	60.4	164.3	57.5	68.4	60.5	92.1	65.3	5.3	0.0	0.0	575.0
1976-77	18.4	34.9	91.8	123.1	144.9	13.5	56.2	40.4	0.0	5.9	7.9	0.0	537.0
1977-78	38.3	16.1	1.7	171.9	170.5	50.0	38.0	14.1	0.0	0.0	0.0	0.0	500.6
1978-79	0.0	29.0	3.3	153.1	78.0	90.6	34.4	27.0	9.0	4.7	0.0	0.0	429.1
1979-80	0.7	53.4	43.1	139.3	71.2	100.0	52.3	14.5	30.3	0.9	0.0	0.4	506.1
1980-81	5.7	15.2	15.3	55.0	167.9	96.8	43.6	22.8	31.4	35.9	0.0	0.0	489.6
1981-82	0.1	7.2	111.9	53.5	43.4	27.7	50.6	17.8	7.1	2.6	0.0	3.3	325.2
1982-83	8.4	41.1	53.8	44.3	48.4	73.3	63.7	21.3	8.9	7.8	0.0	3.6	374.6
1983-84	12.0	38.3	86.5	41.8	43.4	37.4	34.5	67.2	1.1	0.0	1.2	2.3	365.7
1984-85	0.0	4.2	143.5	64.6	96.2	49.4	35.8	9.3	7.0	0.0	0.1	0.0	410.1
1985-86	6.5	40.7	31.5	147.1	35.3	68.5	22.4	5.7	35.0	3.5	0.0	0.0	396.2
1986-87	1.6	136.6	63.4	75.9	53.3	20.0	132.7	20.0	21.5	0.1	1.4	0.0	526.5
1987-88	0.0	53.9	34.9	232.0	79.5	117.6	74.8	7.6	7.2	0.0	0.0	0.0	607.5
1988-89	2.8	34.8	106.6	96.1	88.9	13.8	42.4	0.0	4.5	0.0	0.0	0.0	389.9
1989-90	0.0	67.9	71.1	66.6	49.3	131.5	18.1	12.2	1.1	0.0	0.0	0.3	418.1
1990-91	0.0	35.0	7.1	23.5	89.3	85.2	46.1	3.8	0.0	0.0	0.0	0.0	290.0
1991-92	0.0	51.6	122.9	279.0	87.0	77.2	22.0	5.3	10.9	57.2	0.0	12.9	726.0
1992-93	0.0	4.9	50.1	241.0	72.7	82.6	63.3	8.7	14.5	13.9	0.0	0.0	551.7
1993-94	0.0	0.0	72.5	16.6	99.2	77.3	67.1	26.4	2.8	0.0	3.0	1.4	366.3
1994-95	0.0	68.3	122.0	48.2	21.5	34.4	19.8	31.8	20.8	0.4	9.6	0.0	376.8
1995-96	0.0	6.2	35.9	9.1	144.9	21.4	63.4	19.7	0.3	0.0	0.0	0.0	300.9
1996-97	0.0	66.1	12.8	62.9	11.9	29.5	36.1	29.4	4.2	0.9	0.0	17.8	271.6
1997-98	47.0	55.6	40.2	79.1	104.8	16.0	41.2	16.6	51.7	0.0	0.0	0.0	452.2
1998-99	0.3	0.5	31.0	133.8	87.0	35.3	20.1	25.9	0.3	4.9	0.0	13.3	352.4
1999-00	2.7	37.9	20.2	39.5	58.9	62.4	54.7	59.7	11.5	0.0	0.0	0.0	347.5
2000-01	36.5	38.6	111.4	112.9	41.9	62.3	4.4	14.4	28.9	0.0	0.0	0.0	451.3
2001-02	0.9	23.5	43.3	269.8	147.8	31.7	25.5	42.3	38.8	1.6	6.4	2.3	633.9
2002-03	2.6	11.8	21.3	165.4	43.3	133.0	95.2	11.5	9.0	24.0	0.0	0.0	517.1
2003-04	5.3	12.1	14.0	117.7	339.5	94.7	4.4	2.4	3.5	0.0	0.0	0.0	593.6
2004-05	0.0	17.3	50.1	120.0	164.1	17.4	15.1	28.3	6.9	30.0	0.0	0.0	449.2
2005-06	26.8	12.7	185.2	12.4	133.9	42.1	42.8	16.6	2.1	0.0	3.9	0.0	478.5
2006-07	6.7	69.5	54.9	17.2	28.6	117.8	34.9	8.4	58.7	0.1	0	0	396.8
2007-08	0	8.1	24.5	113	31.7	55.5	9.2	13	26.3	0	0	18.4	299.7
2008-09	14.5	22.6	31	146.1	67.1	120.5	59.6	6	22.4	0	0	0	489.8
2009-10	22.8	50.2	63.2	239.3	71.7	107.9	2.3	8.7	18.6	6.5	0	0	591.2
2010-11	9.3	13.3	0	51.6	111.2	60.4	41	44.4	23.2	13.8	0	0.1	368.3
2011-12	12.5	16.4	125.8	95.4	235.9	81	29.7	26.6	20.5	5.5	1.9	0	651.2
2012-13	0	86.1	130.4	101.8	79.9	54.8	10.9	43.2	21.1	0	0	0	528.2
2013-14	1.2	1.1	3.8	63.7	19.8	18.8	37.3	9.3	36.2	12.5	0	0	203.7
2014-15	34.8	62.4	53.9	61.9									

Table 4.30: Statistical measures of Karpaz rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	C_{Vlog}	C_{Slog}	$\bar{x}_{med}$	C_{dx}	PC_v
449.7	116.8	0.224	0.26	0.2	2.6	0.1	0.036	0.05	-0.5	449.2	0.204	0.654

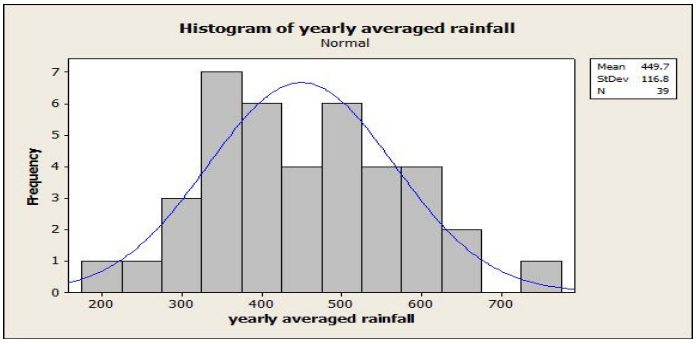


Figure 4.27: Karpaz regions' rainfall histogram

4.5.1 Empirical determination of minimum required sample size of rainfall data for Karpaz region based on mean and standard deviation

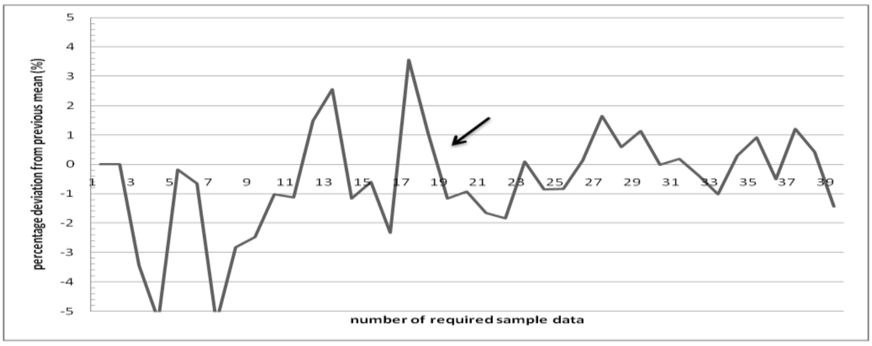


Figure 4.28: Curve showing the required number of sample size for Karpaz regions' rainfall, based on the percentage deviations of the mean values

Comment: The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 19.

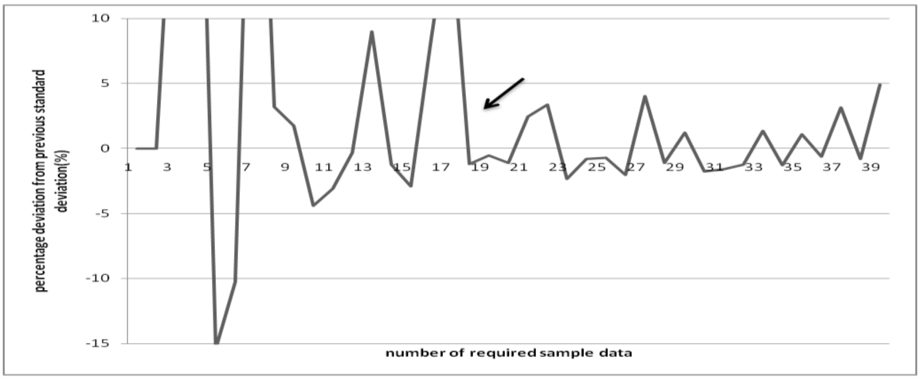


Figure 4.29: Curve showing the required number of sample size for Karpaz regions' rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 19.

Hence $n_{min} = 19$ is the required minimum number that satisfies both mean and standard deviation for Karpaz region rainfall.

Table 4.31: Appropriate rainfall sample size of karpaz for statistic and probabilistic studies

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
19 < 39 OK	19 < 39 OK

4.5.2 Quality Checking Tests of Karpaz

Table 4.32: Quality tests results of Karpaz rainfall

Quality Tests	Parametric or Non-Parametric test checking
Normality	• p-value = 0.89>0.05. Therefore, the Karpaz Rainfall Dta is normally distributed.
Homogeneity	• Based on Pettitt, SNHT, BR, and VNR tests Karpaz rainfall time series is Homogenous. • Based on t-test and F-test, Karpaz region Rainfall distribution is correlated with North Coast and East Coast region Rainfall distributions.
Consistency	• Rainfall of Karpaz region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in Karpaz rainfall because Mann – Kendall p-value = 1.00 >0.05 • Sens slope=-0.1
Stationarity	• Karpaz rainfall is stationary based on ADF test since slope of regression $\gamma = 0.26 > 0$.

4.5.2.1 Probability distributions details of Karpaz region

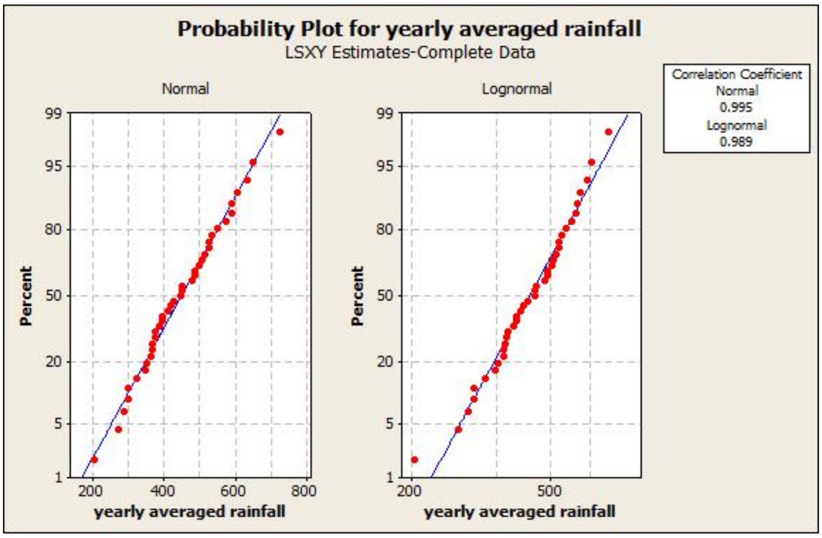


Figure 4.30: Karpaz rainfall fit on normal and log-Normal probability distributions

Table 4.33: Equations of the Probability distribution functions with their correlation coefficients for West Karpaz rainfall

Name	Equation	Correlation Coefficient
Normal	$x = 324.2 + 93.6\ Z$	0.995
log-Normal	$y = \log x = 2.5 + 0.1\ Z$	0.989

Result: Comparing the correlation coefficients of the two probability distributions for Karpaz rainfall, it is concluded that, normal distribution has the best fitted curve being greater correlation coefficient value.

4.5.3 Forecasted values by time series models of Karpaz rainfall for the hydrologic years period 2003-04 to 2013-14

4.5.3.1 Markov Model

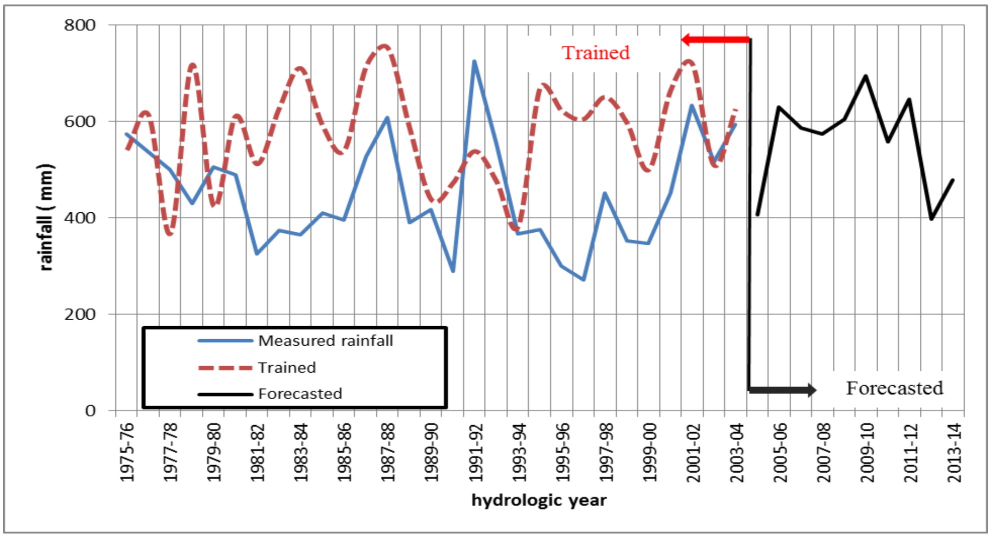


Figure 4.31: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of Karpaz

4.5.3.2 Autoregressive Model

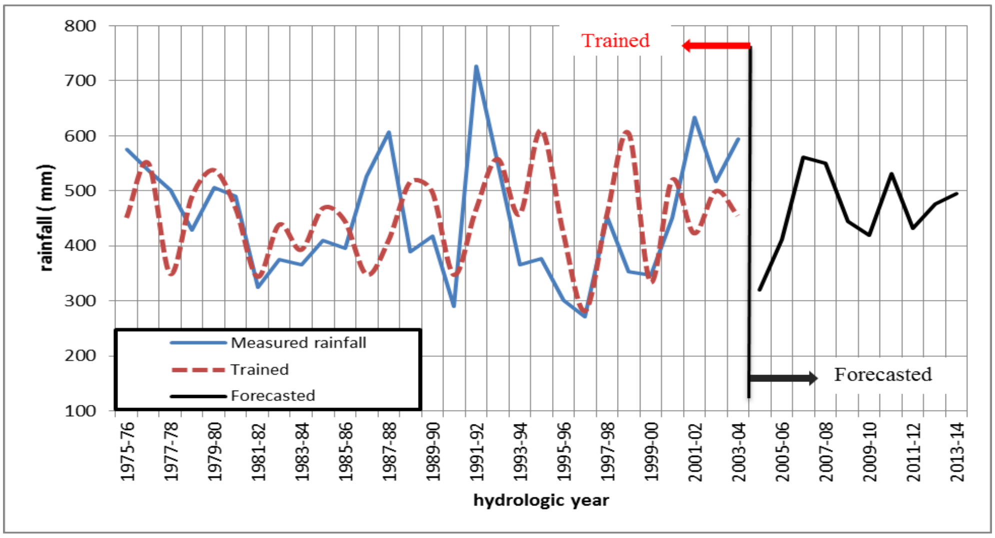


Figure 4.32: Graphical comparison of AR(1) model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of Karpaz

4.5.3.3 Holt-Winter Multiplicative Model

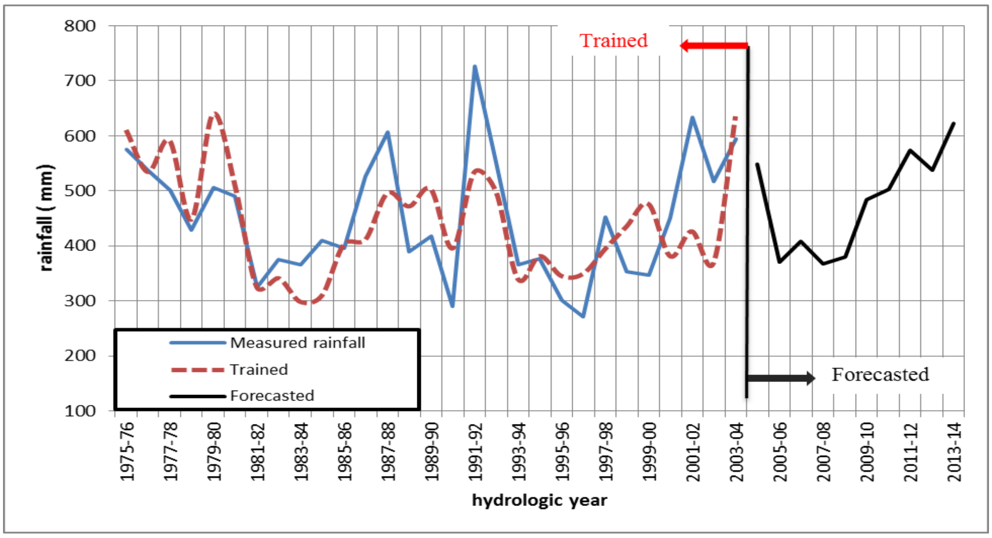


Figure 4.33: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of Karpaz

4.5.3.4 Selecting the best fitted time series model for Karpaz region

Table 4.34: Accuracy checking of forecasted data of Karpaz

Hydrologic year	**DATA**			
	Measured	**Markov model**	**AR (10)**	**Holt-Winter model**
2004-05	449.2	406.6	320.2	548.5
2005-06	478.5	629.3	411.8	370.5
2006-07	396.8	587.6	561.4	409.0
2007-08	299.7	574.1	550.0	367.8
2008-09	489.8	605.6	443.9	380.0
2009-10	591.2	695.2	419.3	484.5
2010-11	368.3	558.7	531.7	502.8
2011-12	651.2	646.0	432.6	573.8
2012-13	528.2	396.9	476.7	536.8
2013-14	203.7	479.4	494.3	622.1
MSE		**29001.5**	**30407.9**	**24896.5**
Ratio w.r.t. Min		**1.16**	**1.22**	**1.00**
MAPE (%)		**27.4**	**33.3**	**22.9**
Ratio w.r.t. Min		**1.20**	**1.46**	**1.00**
RMSE		**2900.2**	**3040.8**	**2489.7**
Ratio w.r.t. Min		**1.17**	**1.22**	**1.00**
MAD		**148.1**	**155.2**	**114.3**
Ratio w.r.t. Min		**1.40**	**1.47**	**1.00**
overall		**3.93**	**5.37**	**4.00**

Result: Holt-Winter model is the best model based on the above 4 error measures having the lowest overall error ratio among the other models. Hence, for Karpaz region, this model is used to generate the rainfall for the hydrologic years 2014-15 to 2018-19 and are all tabulated below.

4.5.3.5 Prediction of yearly rainfall of Karpaz region for hydrologic years 2014-2015 to 2018-19

Table 4.35: Expected yearly rainfall of Karpaz region based on ARIMA(1,0,1) model for hydrologic years 2014-15 to 2023-24

Hydrologic Year	Expected Yearly Total Rainfall (mm)
2014-2015	413.9
2015-2016	601.2
2016-2017	478.8
2017-2018	372.7
2018-2019	356.9

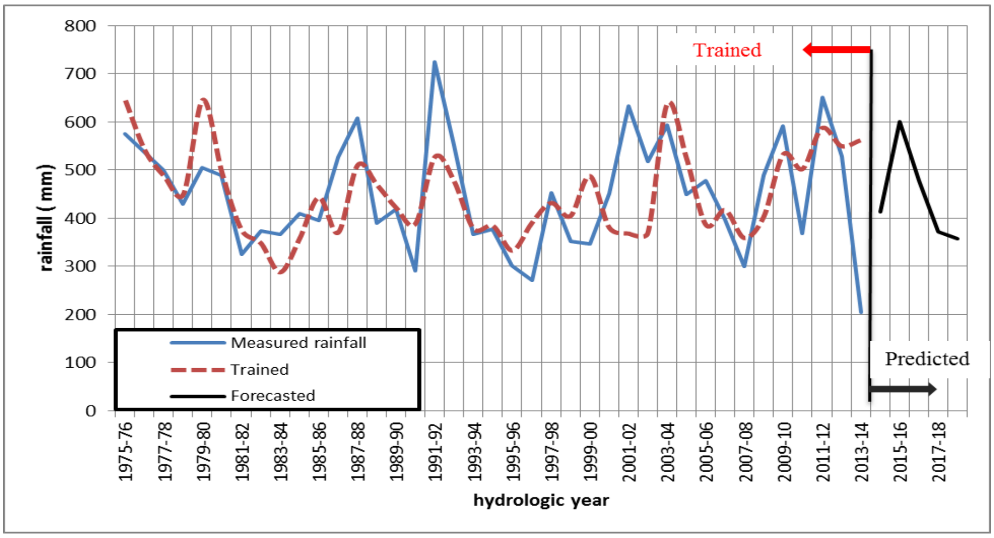


Figure 4.34: Expected (predicted) yearly rainfall of Karpaz region based on Holt-Winter model for hydrologic years 2014-15 to 2023-24

4.5.4 Wet and dry spells for Karpaz

Table 4.36: Numerical representation of wet and dry spells of Karpaz region

Hydrologic Year	Rainfall (mm)	Wet\ Dry	0: Dry 1: Wet
Mean = 449.7 mm			
1975-1976	575	wet	1
1976-1977	537	wet	1
1977-1978	500.6	wet	1
1978-1979	429.1	dry	0
1979-1980	506.1	wet	1
1980-1981	489.6	wet	1
1981-1982	325.2	dry	0
1982-1983	374.6	dry	0
1983-1984	365.7	dry	0
1984-1985	410.1	dry	0
1985-1986	396.2	dry	0
1986-1987	526.5	wet	1
1987-1988	607.5	wet	1
1988-1989	389.9	dry	0
1989-1990	418.1	dry	0
1990-1991	290	dry	0
1991-1992	726	wet	1
1992-1993	551.7	wet	1
1993-1994	366.3	dry	0
1994-1995	376.8	dry	0
1995-1996	300.9	dry	0
1996-1997	271.6	dry	0
1997-1998	452.2	wet	1
1998-1999	352.4	dry	0
1999-2000	347.5	dry	0
2000-2001	451.3	wet	1
2001-2002	633.9	wet	1
2002-2003	517.1	wet	1
2003-2004	593.6	wet	1
2004-2005	449.2	dry	0
2005-2006	478.5	wet	1
2006-2007	396.8	dry	0
2007-2008	299.7	dry	0
2008-2009	489.8	wet	1
2009-2010	591.2	wet	1
2010-2011	368.3	dry	0
2011-2012	651.2	wet	1
2012-2013	528.2	wet	1
2013-2014	203.7	dry	0

Result: number of wet spells = 19 (49%) , number of Dry spells = 20 (51%).

Therefore the Karpaz is in dry spell during the studied period.

4.5.5 Study of monthly wet and dry spells of Karpaz region for hydrological years 1975-76 to 2013-14

Table 4.37: Monthly wet and dry spell of Karpaz

Months	No. of wet spell	No. of dry spell	Conclusion
Sep	11	28	72% dry
Oct	19	20	51% dry
Nov	16	23	59% dry
Dec	18	21	54% dry
Jan	13	26	67% dry
Feb	18	21	54% dry
Mar	18	21	54% dry
Apr	15	24	62% dry
May	17	22	56% dry

Result: Since all the months are in dry spell, Karpaz region throughout the year is dry during the studied period.

4.6 Meteorological Region: North Coast

Table 4.38: Total rainfall of North Coast region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic Year	Month												Totall
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Agu	
1975-76	2.1	4.4	49.9	139.0	56.6	85.2	72.6	89.3	45.5	3.5	3.4	0.0	551.5
1976-77	3.9	48.2	45.6	89.9	106.4	14.7	81.3	30.6	0.1	1.4	0.0	0.0	422.1
1977-78	11.1	4.9	4.4	138.6	227.0	34.2	67.6	20.0	0.0	0.7	0.0	0.0	508.5
1978-79	0.9	27.3	10.1	114.0	95.3	104.3	55.4	14.4	25.3	35.7	0.2	0.0	482.9
1979-80	0.0	34.2	62.1	116.3	61.6	131.0	45.1	16.0	0.8	0.3	0.0	0.0	467.4
1980-81	1.5	36.6	19.9	48.8	137.0	80.9	50.2	17.2	32.1	17.4	0.0	0.0	441.6
1981-82	0.0	5.4	62.6	41.2	51.6	89.6	83.4	15.1	7.9	25.8	0.0	0.5	383.1
1982-83	1.7	11.6	32.2	47.4	58.7	88.7	53.4	27.5	25.1	15.5	0.0	0.0	361.8
1983-84	3.4	24.4	37.2	44.1	48.8	39.0	38.0	76.5	0.0	0.0	1.5	0.5	313.4
1984-85	0.0	4.1	139.7	78.6	173.6	80.3	36.4	11.8	8.3	0.0	0.0	0.0	532.8
1985-86	4.1	28.2	34.1	102.8	44.4	97.2	31.9	8.2	42.4	2.6	0.0	0.0	395.9
1986-87	4.7	43.6	156.8	61.0	66.1	23.3	146.7	15.4	12.9	0.5	1.3	0.8	533.1
1987-88	0.0	66.3	36.8	147.2	70.0	168.6	100.5	11.0	6.4	0.1	2.6	0.1	609.6
1988-89	1.5	44.5	69.7	134.8	113.9	7.5	48.9	0.0	2.9	0.0	0.0	0.0	423.7
1989-90	0.2	69.3	32.3	32.5	25.1	140.2	43.7	7.6	12.0	0.2	0.0	8.1	371.2
1990-91	0.0	8.3	3.8	23.5	95.2	62.0	55.5	13.6	3.2	0.4	0.0	0.0	265.5
1991-92	4.2	13.4	79.6	253.8	44.4	153.8	25.1	5.1	20.1	38.8	2.5	3.7	644.5
1992-93	0.0	0.7	76.5	185.8	71.2	92.7	64.3	7.2	53.7	10.0	0.0	0.0	562.1
1993-94	0.0	1.3	80.8	13.9	136.4	91.4	58.0	29.3	3.2	0.0	2.5	0.2	417.0
1994-95	7.2	42.1	183.3	68.2	25.4	44.8	12.7	13.8	20.0	0.0	19.2	0.0	436.7
1995-96	0.0	7.8	45.1	25.6	108.8	43.9	69.8	39.8	3.2	0.5	0.0	1.1	345.6
1996-97	0.0	56.8	15.9	87.2	20.8	47.4	45.0	43.1	7.1	9.8	0.1	0.1	333.3
1997-98	37.6	20.4	80.0	65.1	65.1	30.3	69.9	9.8	29.4	5.2	0.0	0.0	412.8
1998-99	10.7	0.5	46.9	131.1	135.0	55.6	30.8	24.5	6.3	27.6	3.0	2.3	474.3
1999-00	6.5	20.8	28.4	19.1	73.6	98.9	55.7	82.6	18.1	8.5	0.0	0.2	412.4
2000-01	13.7	58.0	85.6	157.4	50.4	65.4	14.6	20.4	21.8	0.0	0.0	0.9	488.2
2001-02	1.5	23.8	53.0	260.5	104.7	61.5	28.5	42.2	21.6	1.8	6.3	1.5	606.9
2002-03	4.1	9.3	23.4	206.1	82.2	182.3	105.8	14.6	1.3	31.7	0.0	0.2	661.0
2003-04	0.5	15.2	46.1	113.8	196.0	139.2	0.9	10.7	4.3	9.8	0.0	0.0	536.5
2004-05	0.0	16.1	84.2	106.3	94.9	39.4	32.3	31.5	13.4	46.3	0.0	2.9	467.3
2005-06	7.6	17.4	130.1	11.6	148.6	40.6	46.0	12.5	7.3	4.0	18.4	0.0	444.1
2006-07	11.1	95.8	46.9	18.5	41.6	143.5	38	15.6	65	0.3	2.3	0.9	479.5
2007-08	0.1	8.2	39.8	51.9	26.6	42.4	10.2	5.5	10.5	0.3	0.1	0.3	195.9
2008-09	3.5	22.6	16.6	84.2	78.6	95.3	63	11.7	11.6	0	3.1	7.5	397.7
2009-10	34.1	43.8	57.4	173.8	145.1	229.2	15.5	19.7	13.6	11.3	1.3	0	744.8
2010-11	0	17.2	1.4	46.4	132.2	53.9	45	69.1	38.3	10.3	0.1	0.1	414.0
2011-12	12.8	10.7	165.4	72.2	191.3	76.4	30.2	12.4	44.1	1.9	0.4	1.3	619.1
2012-13	0	73.6	125.5	118	109.1	32.2	17.4	46.3	65.3	0	0	0	587.4
2013-14	1.4	8.6	20.4	82.6	22.9	21.6	31.8	10.7	52.4	11.7	0	0	264.1
2014-15	8.7	51.3	45.3	101.1									

Table 4.39: Statistical measures of North Coast rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	C_{Vlog}	C_{Slog}	$\bar{x}_{med}$	C_{dx}	PC_v
461.8	117.1	0.213	0.254	0.1	2.6	0.1	0.034	0.05	-0.7	444.1	0.2	0.67

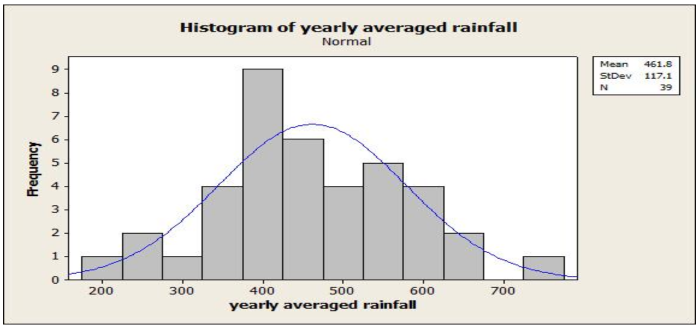


Figure 4.35: North Coast regions’ rainfall histogram

4.6.1 Empirical determination of minimum required sample size of rainfall for North Coast region based on mean and standard deviation

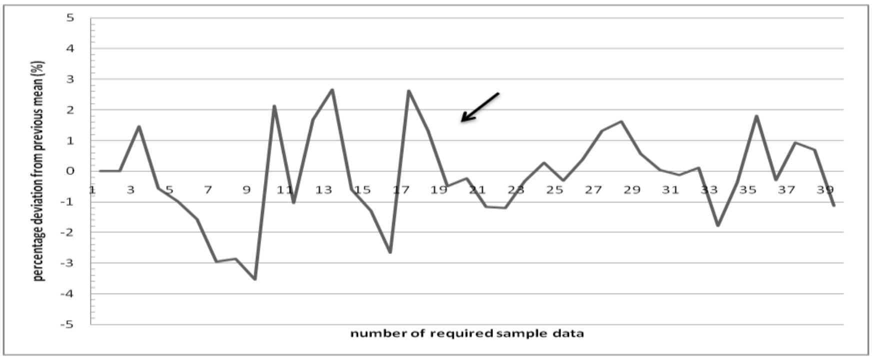


Figure 4.36: Curve showing the required number of sample size for North Coast regions’ rainfall, based on the percentage deviations of the mean values

Comment: The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 19.

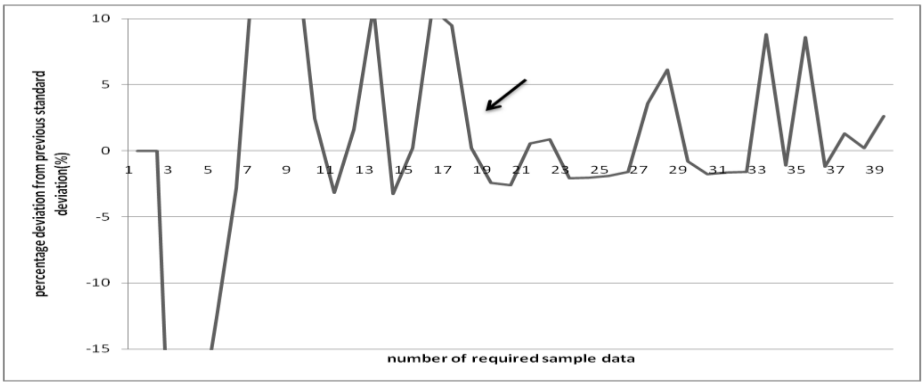


Figure 4.37: Curve showing the required number of sample size for North Coast regions’ rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 37.

Hence n_{min} = 37 is the required minimum number that satisfies both mean and standard deviation for North Coast region rainfall.

Table 4.40: Appropriate rainfall sample size of North Coast for statistic and probabilistic studies

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
19 < 39 OK	37 < 39 OK

4.6.2 Quality Checking Tests of North Coast

Table 4.41: Quality tests results of North Coast rainfall

Quality check Tests	**Parametric or Non-Parametric test checking**
Normality	• p-value = 0.88>0.05. Therefore, the North Rainfall is normally distributed.
Homogeneity	• Based on Pettitt, SNHT, BR, and VNR tests North Coast rainfall time series is Homogenous. • Based on t-test and F-test, North Coast region Rainfall distribution is correlated with Karpaz region Rainfall distributions sets proving Homogeneity
Consistency	• Rainfall of North Coast region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in North Coast rainfall because Mann – Kendall p-value = 0.532>0.05 • Sens slope=1.265
Stationarity	• North Coast rainfall is stationary based on ADF test since slope of regression $\gamma = 0.27 > 0$.

4.6.3 Probability distributions details of North Coast region

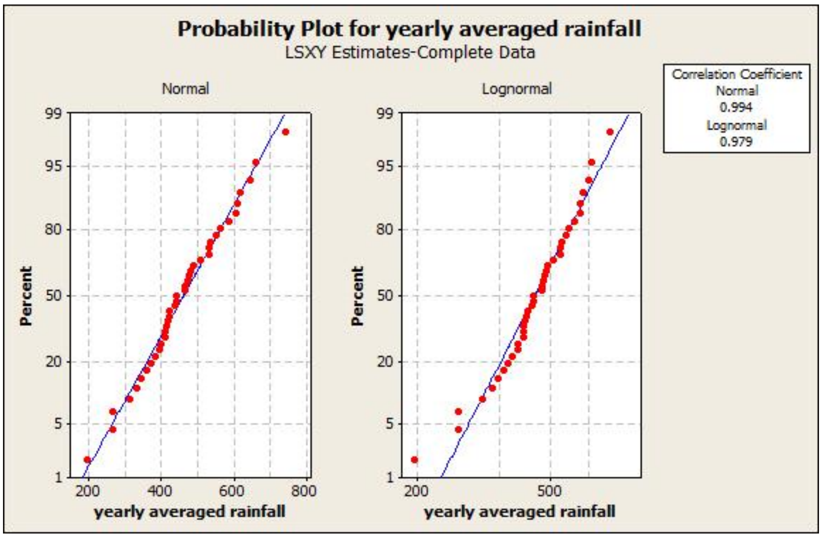


Figure 4.38: North Coast rainfall fit on normal and log-Normal probability distributions

Table 4.42: Equations of the probability distribution functions with their correlation coefficients for North Coast rainfall

Name	Equation	Correlation Coefficient
Normal	$x = 461.8 + 115.6\ Z$	0.994
log-Normal	$y = \log x = 2.6 + 0.1\ Z$	0.979

<u>Result</u>: Comparing the correlation coefficients of the two probability distributions for North Coast rainfall, it is concluded that, normal distribution has the best fitted curve being greater correlation coefficient value.

4.6.4 Forecasted values by time series models of North Coast rainfall for the hydrologic years period 2003-04 to 2013-14

4.6.4.1 Markov Model

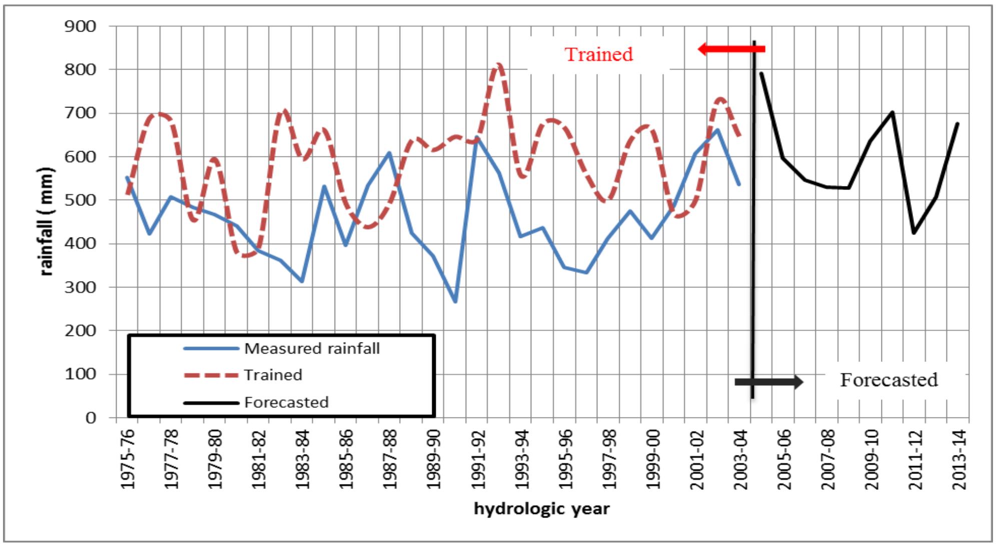


Figure 4.39: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of North Coast

4.6.4.2 Auto- Regressive Model

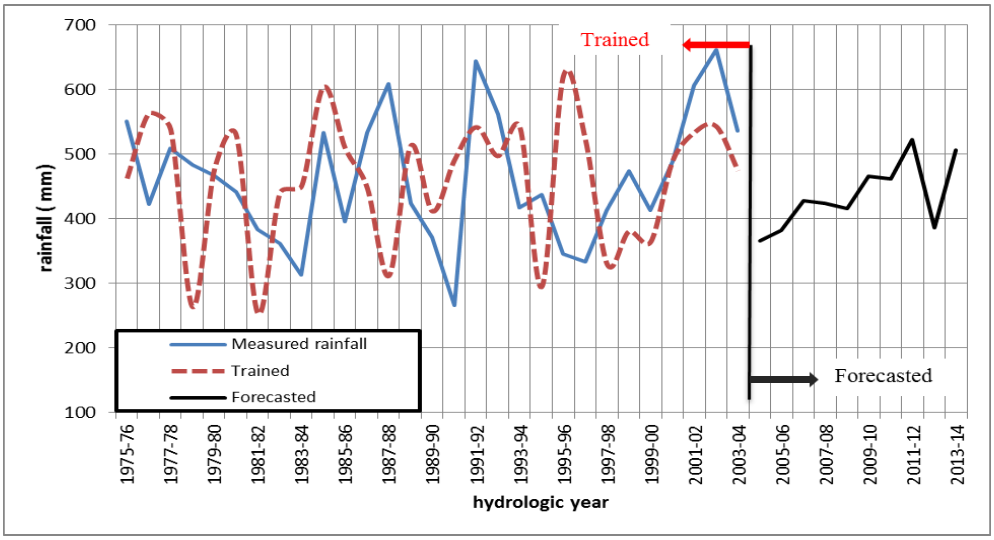


Figure 4.40: Graphical comparison of AR(1) model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of North Coast

4.6.4.3 Holt-Winter Multiplicative Model

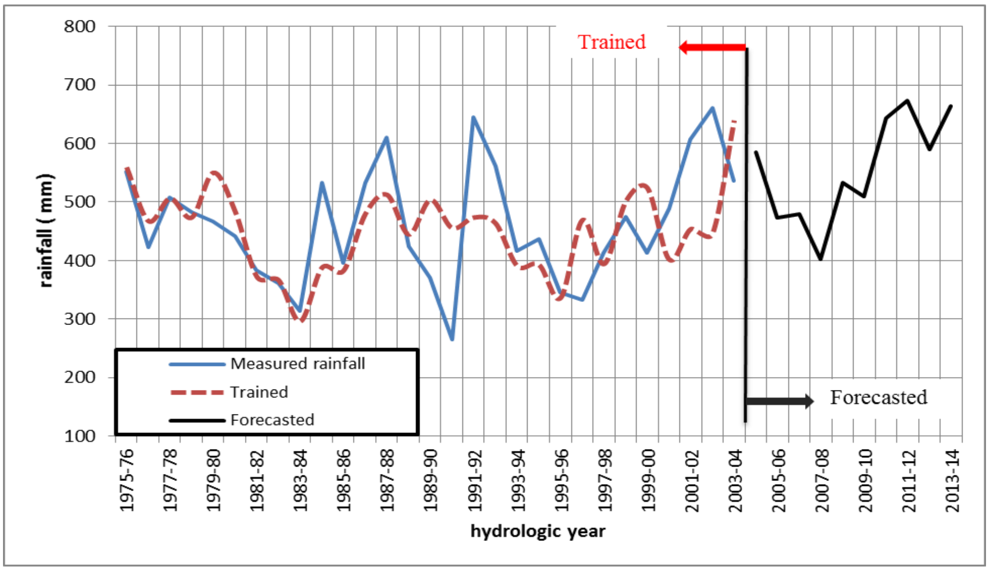


Figure 4.41: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of North Coast

4.6.4.4 Selecting the best fitted time series model for North Coast region

Table 4.43: Accuracy checking of forecasted data of North Coast

Hydrologic Year	**DATA**			
	Measured	**Markov Model**	**AR(1)**	**Winter Model**
2004-05	467.3	791.4	366.1	584.2
2005-06	444.1	595.9	381.4	472.7
2006-07	479.5	546.8	427.7	479.8
2007-08	195.9	529.5	423.9	402.1
2008-09	397.7	528.0	416.1	533.3
2009-10	744.8	634.7	465.4	509.7
2010-11	414.0	702.4	461.1	642.9
2011-12	619.1	425.0	522.5	672.8
2012-13	587.4	508.3	385.4	590.0
2013-14	264.1	675.1	505.8	664.5
MSE		**56903.1**	**25801.4**	**34626.7**
Ratio w.r.t. min.		**2.21**	**1.00**	**1.34**
MAPE (%)		**34.7**	**30.3**	**25.3**
Ratio w.r.t. min.		**1.37**	**1.20**	**1.00**
RMSE		**5690.3**	**2580.1**	**3462.7**
Ratio w.r.t. min.		**2.40**	**1.00**	**1.46**
MAD		**209.0**	**132.9**	**140.8**
Ratio w.r.t. min.		**1.57**	**1.00**	**1.06**
Overall		**7.55**	**4.20**	**4.86**

Result: AR(1) Model is the best model based on the above 4 error measures having the lowest overall error ratio among the other models. Hence, for North Coast region, this model is used to generate the rainfall for the Hydrologic years 2014-15 to 2018-19 and are all tabulated below.

4.6.4.5 Prediction of yearly rainfall of North Coast region for hydrologic years 2014-2015 to 2018-19

Table 4.44: Expected yearly rainfall of North Coast region based on AR(1) model for hydrologic years 2014-15 to 2018-19

Hydrologic Year	Expected Yearly Total Rainfall (mm)
2014-2015	553.0
2015-2016	325.8
2016-2017	371.3
2017-2018	569.6
2018-2019	469.8

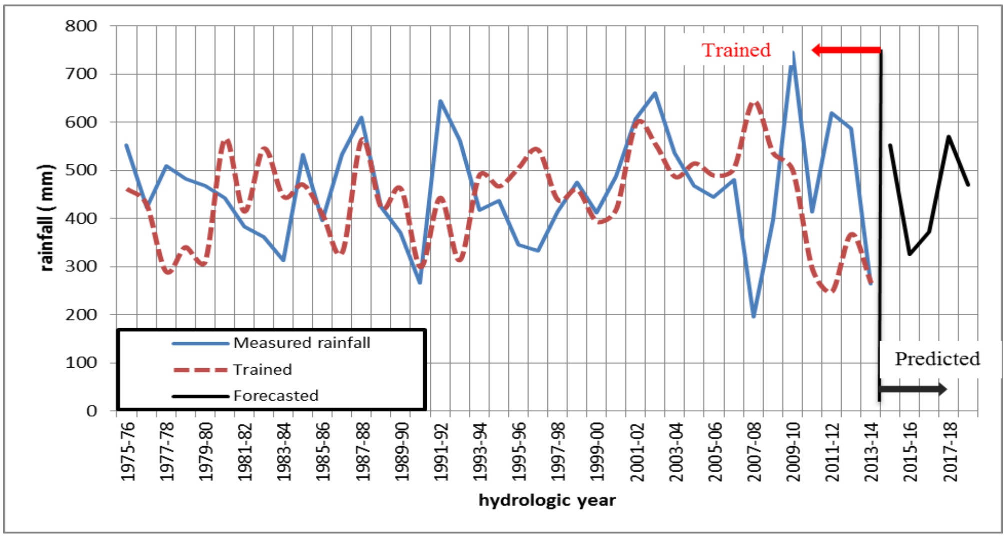


Figure 4.42: Expected (predicted) yearly rainfall of North Coast region based on AR(1) model for hydrologic years 2014-15 to 2018-19

4.6.5 Wet and dry spells for North Coast

Table 4.45: Numerical representation of wet and dry spells of North Coast region based on the mean of the data

Hydrologic Year	Rainfall (mm)	Wet\ Dry	0: Dry 1: Wet
Mean = 461.8 mm			
1975-1976	551.5	wet	**1**
1976-1977	422.1	dry	**0**
1977-1978	508.5	wet	**1**
1978-1979	482.9	wet	**1**
1979-1980	467.4	wet	**1**
1980-1981	441.6	dry	**0**
1981-1982	383.1	dry	**0**
1982-1983	361.8	dry	**0**
1983-1984	313.4	dry	**0**
1984-1985	532.8	wet	**1**
1985-1986	395.9	dry	**0**
1986-1987	533.1	wet	**1**
1987-1988	609.6	wet	**1**
1988-1989	423.7	dry	**0**
1989-1990	371.2	dry	**0**
1990-1991	265.5	dry	**0**
1991-1992	644.5	wet	**1**
1992-1993	562.1	wet	**1**
1993-1994	417	dry	**0**
1994-1995	436.7	dry	**0**

Hydrologic Year	Rainfall (mm)	Wet\ Dry	0: Dry 1: Wet
Mean = 461.8 mm			
1995-1996	345.6	dry	**0**
1996-1997	333.3	dry	**0**
1997-1998	412.8	dry	**0**
1998-1999	474.3	wet	**1**
1999-2000	412.4	dry	**0**
2000-2001	488.2	wet	**1**
2001-2002	606.9	wet	**1**
2002-2003	661	wet	**1**
2003-2004	536.5	wet	**1**
2004-2005	467.3	wet	**1**
2005-2006	444.1	dry	**0**
2006-2007	479.5	wet	**1**
2007-2008	195.9	dry	**0**
2008-2009	397.7	dry	**0**
2009-2010	744.8	wet	**1**
2010-2011	414	dry	**0**
2011-2012	619.1	wet	**1**
2012-2013	587.4	wet	**1**
2013-2014	264.1	dry	**0**

Result: number of wet spells = 19 (49%) , number of Dry spells = 20 (51%).

Therefore the North Coast is in dry spell during the studied period.

4.6.6 Study of monthly wet and dry spells of North Coast region for hydrological years 1975-76 to 2013-14

Table 4.46: Monthly wet and dry spell of North Coast

Months	No. of wet spell	No. of dry spell	Conclusion
Sep	10	29	74% dry
Oct	15	24	62% dry
Nov	15	24	62% dry
Dec	17	22	56% dry
Jan	18	21	54% dry
Feb	19	20	51% dry
Mar	17	22	56% dry
Apr	13	26	67% dry
May	16	23	59% dry

Result: Since all the months are in dry spell, North coast region throughout the year is dry during the studied period.

4.7 Meteorological Region: West Mesaria

Table 4.47: Total rainfall of West Mesaria region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic Year	Month												Total
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Agu	
1975-76	11.8	4.4	23.4	119.8	33.7	50.8	71.2	107.2	55.0	0.0	0.0	0.0	477.3
1976-77	0.0	8.5	44.8	53.8	74.7	12.4	55.7	32.8	0.0	1.0	0.0	0.0	283.7
1977-78	0.9	41.2	2.1	100.8	169.1	57.0	53.1	30.4	0.0	0.0	0.0	0.0	454.6
1978-79	0.0	27.8	10.6	54.1	92.4	30.0	31.6	10.0	8.3	0.0	0.0	0.0	264.8
1979-80	0.0	30.0	41.4	58.5	53.6	64.0	35.2	5.6	0.0	0.0	0.0	0.0	288.3
1980-81	0.0	24.4	13.8	65.6	116.7	67.8	27.2	12.8	17.6	15.3	0.0	0.0	361.2
1981-82	1.0	0.0	51.8	40.6	36.6	58.8	71.8	7.8	14.4	24.7	0.0	0.0	307.5
1982-83	0.0	2.4	19.8	57.4	38.3	75.7	31.4	15.9	18.0	30.8	0.0	0.3	290.0
1983-84	0.0	16.5	28.1	42.1	40.0	17.3	36.6	38.7	0.2	0.0	0.0	0.0	219.5
1984-85	0.0	0.7	60.2	44.5	62.0	60.5	27.9	4.6	1.0	0.0	0.0	0.0	261.4
1985-86	0.0	25.0	15.7	32.1	31.2	50.0	15.9	11.3	37.1	3.3	0.0	0.0	221.6
1986-87	6.3	10.9	98.2	34.0	54.0	19.0	88.7	5.5	2.2	1.4	0.0	0.0	320.2
1987-88	0.0	33.7	28.5	72.0	46.0	114.0	83.2	6.9	1.8	0.0	4.4	0.0	390.5
1988-89	0.0	36.4	42.4	117.7	56.9	9.4	18.6	0.0	0.2	0.0	0.0	0.0	281.6
1989-90	0.0	14.5	28.4	19.4	15.6	67.3	28.4	0.8	3.8	0.7	0.0	0.0	178.9
1990-91	0.0	19.7	10.7	25.8	62.9	40.5	24.3	11.5	0.9	0.0	0.0	0.0	196.3
1991-92	0.0	10.3	33.8	162.8	20.1	104.3	31.4	7.1	8.3	6.3	2.7	1.2	388.3
1992-93	0.0	2.7	59.2	101.9	46.3	65.9	51.9	3.6	41.2	0.0	0.0	0.0	372.7
1993-94	0.0	2.7	32.2	12.4	89.0	54.7	44.4	28.4	1.8	0.0	0.0	0.0	265.6
1994-95	16.9	45.8	91.1	39.7	23.2	14.0	11.5	31.9	6.4	0.0	0.1	0.0	280.6
1995-96	0.8	9.1	36.7	13.5	66.8	51.1	46.0	29.3	0.1	0.0	0.0	14.0	267.4
1996-97	1.5	18.7	6.7	57.3	7.3	28.0	50.5	28.4	10.4	0.3	0.0	0.0	209.1
1997-98	9.9	18.2	61.4	48.7	29.1	13.6	57.4	4.0	4.1	0.0	0.0	0.0	246.4
1998-99	1.1	2.1	33.6	91.3	92.0	33.5	37.1	24.4	0.0	3.8	0.0	0.0	318.9
1999-00	1.6	2.5	18.6	27.4	44.8	52.6	35.6	33.5	5.8	0.0	0.0	0.0	222.4
2000-01	13.4	8.2	58.4	115.6	31.4	56.9	6.4	22.2	4.7	0.0	0.0	0.0	317.2
2001-02	0.0	14.6	20.7	98.3	62.7	34.0	48.6	45.1	4.8	0.2	0.0	0.0	329.0
2002-03	20.7	12.3	38.6	122.3	52.1	167.2	68.2	10.8	0.3	9.8	0.0	0.0	502.3
2003-04	0.0	10.1	37.2	52.7	133.0	54.5	0.1	7.0	2.1	0.2	0.0	0.0	296.9
2004-05	0.1	12.0	66.0	47.4	34.6	48.2	28.8	27.6	8.3	0.7	0.0	0.1	273.8
2005-06	0.0	18.9	57.7	10.1	68.1	27.9	30.6	17.6	4.2	0.0	0.6	0.0	235.7
2006-07	1.6	57.3	31.9	9.4	41.1	84.6	45.6	15.7	99.5	0.6	0.2	0.1	387.6
2007-08	0.1	8.9	41.6	46.9	20.9	29.4	14.7	4	12.8	0	0	0	179.3
2008-09	6.2	16.5	2.8	67.7	79.8	93.7	57.7	15.5	10.5	0	0	0.3	350.7
2009-10	20.2	33.6	35.1	95.9	151.1	120.2	3.7	10.5	15.6	0.2	0	0	486.1
2010-11	0	17.4	0.1	52.4	70	46.2	45.9	37.8	32.9	0	0	0	302.7
2011-12	4.2	7.6	58.2	60.5	145.3	72.3	25.2	12	28.7	10.9	6.4	0	431.3
2012-13	0	80.3	85.9	122.7	55.9	39.4	24.4	23.7	30.7	0	0.1	0	463.1
2013-14	0.2	5.5	25.7	39.7	15	20.4	23.8	7.6	38.9	20.1	0	0	196.9
2014-15	1	43.7	43.1	52.6									

Table 4.48: Statistical measures of West Mesaria rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	C_{Vlog}	C_{Slog}	$\bar{x}_{med}$	C_{dx}	PC_v
310.8	88.7	0.244	0.285	0.6	2.5	0.1	0.039	0.05	0.1	290.0	0.24	0.89

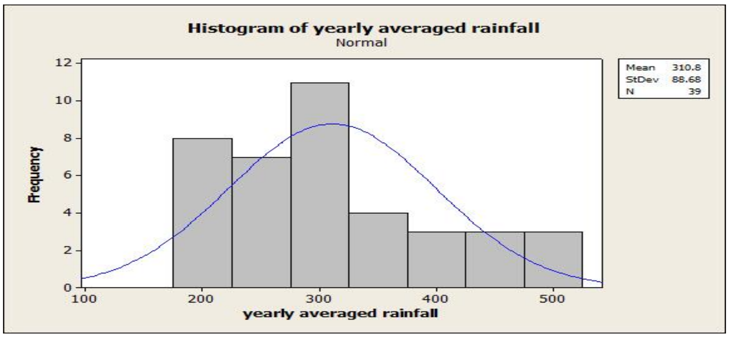


Figure 4.43: West Mesaria regions’ rainfall histogram

4.7.1 Empirical determination of minimum required sample size of rainfall for West Mesaria region based on mean and standard deviation

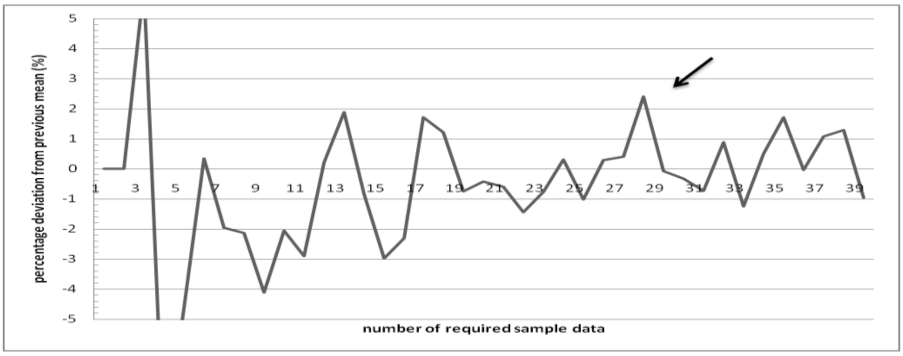


Figure 4.44: Curve showing the required number of sample size for West Mesaria regions’ rainfall, based on the percentage deviations of the mean values

<u>Comment:</u> The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 29.

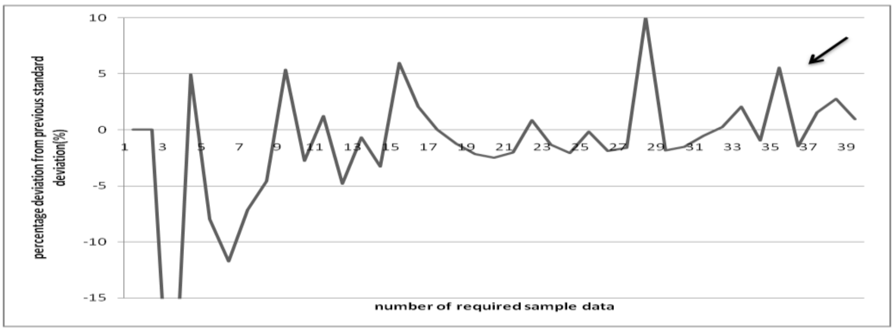


Figure 4.45: Curve showing the required number of sample size for West Mesaria regions’ rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 37.

Table 4.49: Appropriate rainfall sample size of West Mesaria for statistic and probabilistic studies

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
29 < 39 OK	37 < 39 OK

4.7.2 Quality Checking Tests of West Mesaria

Table 4.50: Quality tests results of West Mesaria rainfall

Quality Tests	**Parametric or Non-Parametric Tests Results**
Normality	• p-value = 0.07 > 0.05. Therefore, the West Mesaria data is normally distributed.
Homogeneity	• Based on Pettitt,SNHT,BR,and VNR tests West Mesaria rainfall time series is Homogenous. • Based on t-test and F-test, West Mesaria region data distribution is correlated with Central Mesaria, East Mesaria, and East Coast region distributions; hence the data is homogenous.
Consistency	• Rainfall of West Mesaria region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in West Mesaria rainfall because Mann – Kendall p value = 0.484>0.05 • Sens slope=0.958
Stationarity	• West Mesaria rainfall is stationary based on ADF test since slope of regression $\gamma = 0.08 > 0$.

4.7.3 Probability distributions details of West Mesaria region

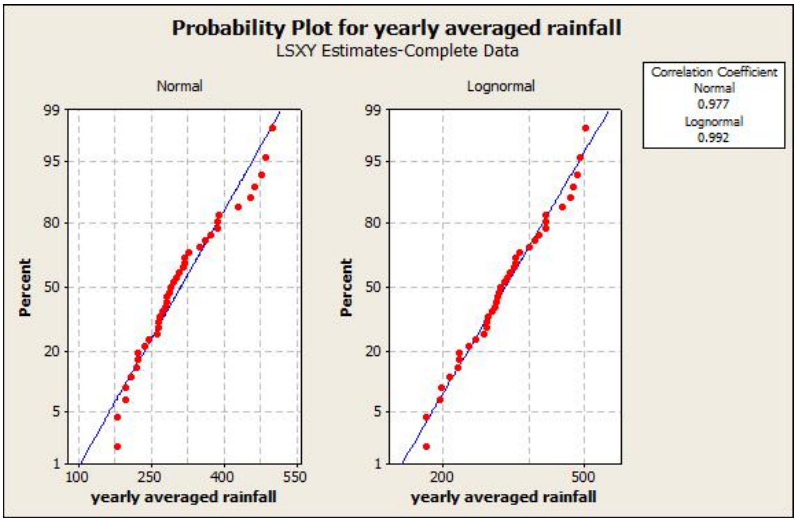


Figure 4.46: West Mesaria rainfall fit on normal and log-Normal probability distributions

Table 4.51: Equations of the Probability distribution functions with their correlation coefficients for West Mesaria rainfall

Name	Equation	Correlation Coefficient
Normal	$x = 310.8 + 87.5\ Z$	0.977
log-Normal	$y = \log x = 2.5 + 0.1\ Z$	0.992

Result: Comparing the correlation coefficients of the two probability distributions for West Mesaria rainfall, it is concluded that, log-Normal distribution has the best fitted curve being greater correlation coefficient value.

4.7.4 Forecasted values by time series models of West Mesaria rainfall for the hydrologic years period 2003-04 to 2013-14

4.7.4.1 Markov Model

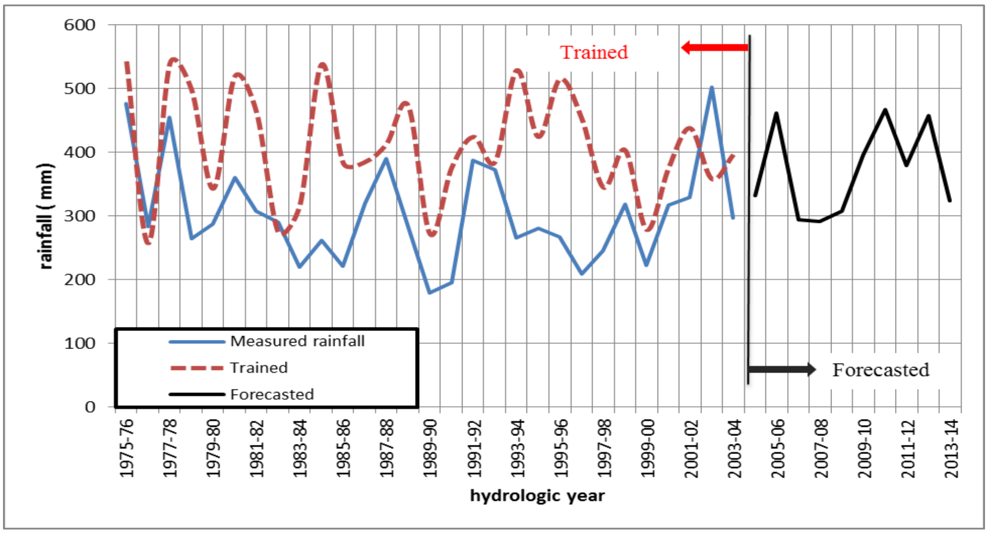


Figure 4.47: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of West Mesaria

4.7.4.2 Auto-Regressive Model

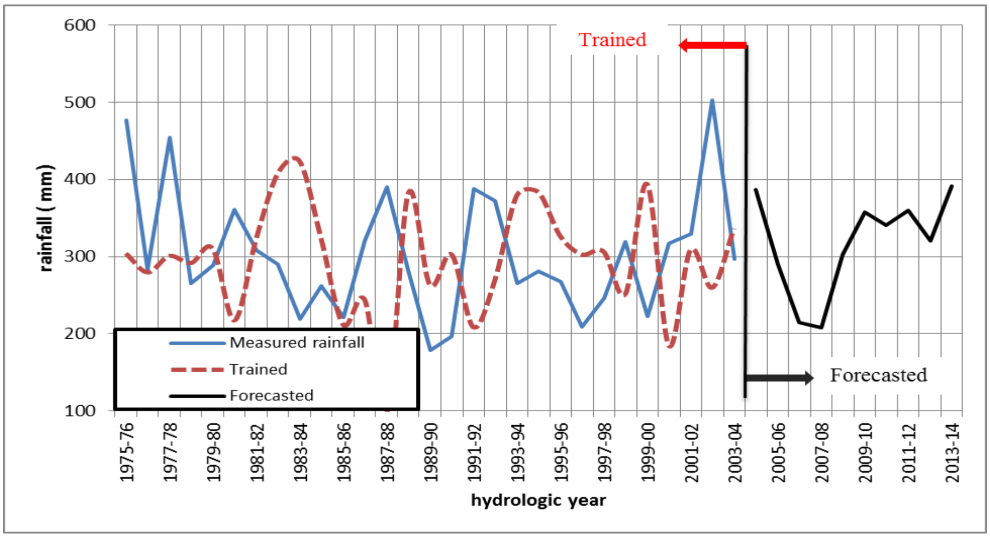


Figure 4.48: Graphical comparison of AR(1) model trained data and hydrologic yearly averaged rainfall (measured) of West Mesaria for the period 1975-76 to 2003-04.

4.7.4.3 Holt-Winter Multiplicative Model

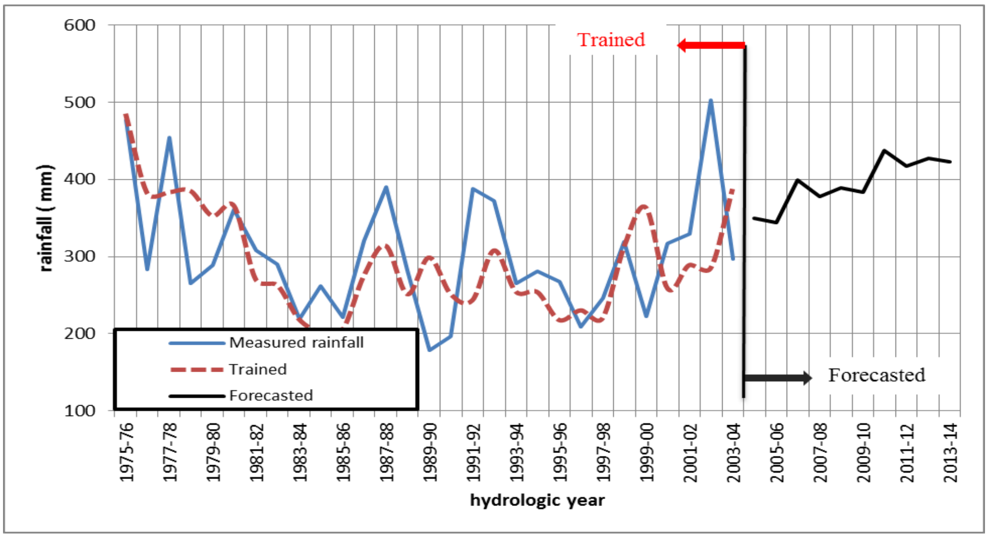


Figure 4.49: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of West Mesaria

4.7.4.4 Selecting the best fitted time series model for West Mesaria region

Table 4.52: Accuracy checking of forecasted data of West Mesaria

Hydrologic Year	**DATA**			
	Measured	**Markov Model**	**AR(1)**	**Holt-Winter model**
2004-2005	273.8	332.0	387.1	349.3
2005-2006	235.7	461.7	290.0	344.1
2006-2007	387.6	294.5	214.3	398.7
2007-2008	179.3	291.1	207.8	377.5
2008-2009	350.7	308.3	302.8	388.6
2009-2010	486.1	395.8	357.0	383.4
2010-2011	302.7	467.0	340.5	438.0
2011-2012	431.3	379.5	359.9	416.8
2012-2013	463.1	457.4	319.9	427.9
2013-2014	196.9	324.2	391.7	422.7
MSE		**13147.5**	**13056.6**	**13958.2**
Ratio w.r.t. min.		**1.01**	**1.00**	**1.07**
MAPE (%)		**26.2**	**32.0**	**24.1**
Ratio w.r.t. min.		**1.09**	**1.33**	**1.00**
RMSE		**1314.8**	**1305.7**	**1395.8**
Ratio w.r.t. min.		**1.01**	**1.00**	**1.07**
MAD		**97.1**	**99.4**	**94.5**
Ratio w.r.t. min.		**1.06**	**1.08**	**1.00**
Overall		**4.17**	**5.13**	**4.14**

Result: Holt-Winter model is the best model based on the above 4 error measures having the lowest overall error ratio among the other models. Hence, for West Mesaria Region, this Model is used to generate the rainfall for the Hydrologic years 2014-15 to 2018-19 and are all tabulated below

4.7.4.5 Prediction of yearly rainfall of West Mesaria region for hydrologic years 2014-2015 to 2018-19

Table 4.53: Expected yearly rainfall of West Mesaria region based on Holt-Winter model for hydrologic years 2014-15 to 2018-19

Hydrologic Year	Expected Yearly Total Rainfall (mm)
2014-2015	383.0
2015-2016	390.3
2016-2017	400.1
2017-2018	322.0
2018-2019	387.2

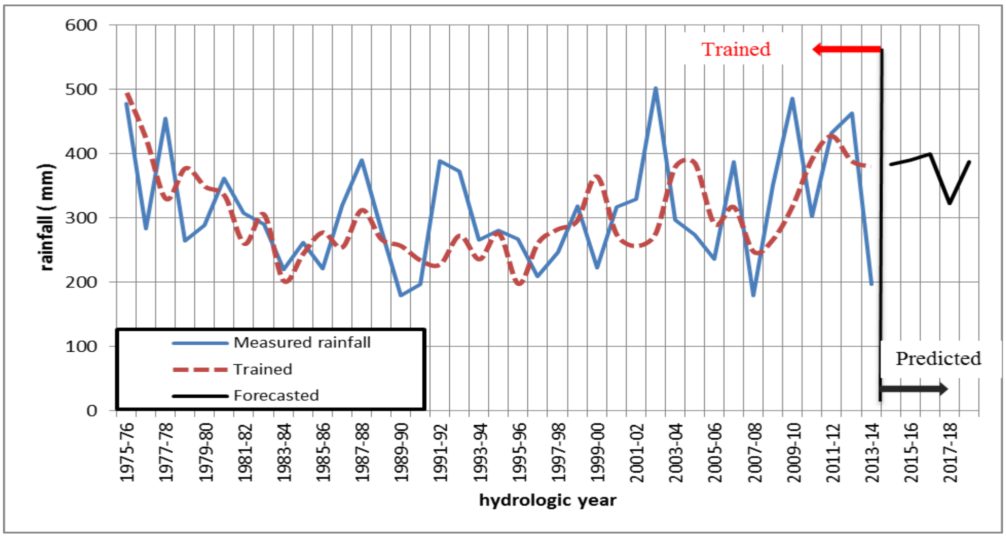


Figure 4.50: Expected (predicted) yearly rainfall of West Mesaria region based on Holt-Winter model for hydrologic years 2014-15 to 2018-19

4.7.5 Wet and dry spells for West Mesaria

Table 4.54: Numerical representation of wet and dry spells of West Mesaria region based on the mean of the data

Hydrologic Year	Rainfall (mm)	Wet\ Dry	0: Dry 1: Wet
Mean = 310.8 mm			
1975-1976	477.3	wet	**1**
1976-1977	283.7	dry	**0**
1977-1978	454.6	wet	**1**
1978-1979	264.8	dry	**0**
1979-1980	288.3	dry	**0**
1980-1981	361.2	wet	**1**
1981-1982	307.5	dry	**0**
1982-1983	290	dry	**0**
1983-1984	219.5	dry	**0**
1984-1985	261.4	dry	**0**
1985-1986	221.6	dry	**0**
1986-1987	320.2	wet	**1**
1987-1988	390.5	wet	**1**
1988-1989	281.6	dry	**0**
1989-1990	178.9	dry	**0**
1990-1991	196.3	dry	**0**
1991-1992	388.3	wet	**1**
1992-1993	372.7	wet	**1**
1993-1994	265.6	dry	**0**
1994-1995	280.6	dry	**0**
1995-1996	267.4	dry	**0**
1996-1997	209.1	dry	**0**
1997-1998	246.4	dry	**0**
1998-1999	318.9	wet	**1**
1999-2000	222.4	dry	**0**
2000-2001	317.2	wet	**1**
2001-2002	329	wet	**1**
2002-2003	502.3	wet	**1**
2003-2004	296.9	dry	**0**
2004-2005	273.8	dry	**0**
2005-2006	235.7	dry	**0**
2006-2007	387.6	wet	**1**
2007-2008	179.3	dry	**0**
2008-2009	350.7	wet	**1**
2009-2010	486.1	wet	**1**
2010-2011	302.7	dry	**0**
2011-2012	431.3	wet	**1**
2012-2013	463.1	wet	**1**
2013-2014	196.9	dry	**0**

Result: Number of Wet spells = 16 (41 %), number of Dry spells = 23 (59 %).

Therefore the West Mesaria is in dry spell during the studied period.

4.7.6 Study of monthly wet and dry spells of West Mesaria region for hydrological years 1975-76 to 2013-14

Table 4.55: Monthly wet and dry spell of West Mesaria

Months	No. of wet spell	No. of dry spell	Conclusion
Sep	9	30	77% dry
Oct	14	15	64% dry
Nov	16	23	59% dry
Dec	14	25	64% dry
Jan	16	23	59% dry
Feb	18	21	54% dry
Mar	16	23	59% dry
Apr	15	24	62% dry
May	12	27	69% dry

Result: Since all the months are in dry spell, West Mesaria region throughout the year is dry during the studied period.

4.8 Meteorological Region: TRNC (General)

Table 4.56: Total rainfall of TRNC region for hydrological years from 1975-76 to 2013-14 in mm

Hydrologic YEAR	MONTH												Totall
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	JUL	AGU	
1975-76	2.2	5.3	37.4	131.9	40.8	60.4	57.5	77.4	45.2	2.5	4.8	0.3	465.7
1976-77	6.6	34.2	52.1	77.3	88.1	11.6	57.5	32.4	0.1	1.5	3.3	0.0	364.7
1977-78	14.5	10.1	4.0	119.5	163.5	35.0	48.8	16.6	0.0	0.3	0.0	0.0	412.3
1978-79	0.3	22.7	7.0	108.7	66.7	75.7	43.4	11.6	17.8	24.2	0.4	0.4	378.9
1979-80	0.6	37.6	44.5	108.0	50.4	107.3	40.8	13.3	8.7	0.2	0.0	0.3	411.7
1980-81	1.5	22.4	13.3	47.2	123.7	73.3	44.1	18.5	25.4	21.2	0.0	0.0	390.6
1981-82	0.0	3.9	63.3	38.8	37.9	57.6	61.1	14.0	7.5	16.6	0.1	1.8	302.6
1982-83	2.9	14.4	28.4	34.6	46.9	65.4	45.2	22.2	19.5	12.1	0.0	0.2	291.8
1983-84	1.4	23.2	50.4	34.9	43.9	37.6	34.7	67.1	1.2	0.0	1.3	2.1	297.8
1984-85	0.0	4.1	145.1	65.4	99.3	54.3	32.7	12.7	6.5	1.6	0.0	0.0	421.7
1985-86	5.4	27.4	25.5	82.7	34.8	72.4	21.3	9.8	54.4	5.6	0.0	0.0	339.3
1986-87	2.8	46.2	95.5	51.8	43.9	17.6	122.7	15.0	12.6	0.5	1.7	0.3	410.6
1987-88	0.0	56.5	27.7	138.6	63.2	120.8	81.6	7.9	5.4	3.0	2.1	0.8	507.6
1988-89	2.5	35.9	61.6	107.4	91.2	12.2	33.6	0.0	5.3	0.9	0.0	0.0	350.6
1989-90	0.3	53.1	33.5	33.2	23.5	108.6	31.2	6.6	7.6	0.1	0.0	5.6	303.3
1990-91	0.0	13.3	5.6	18.6	69.2	56.8	41.9	9.0	1.3	0.2	0.0	0.0	215.9
1991-92	0.8	17.3	69.0	218.8	32.9	90.7	20.6	5.0	20.8	34.6	9.8	7.3	527.6
1992-93	0.0	2.9	64.9	150.1	53.2	69.1	56.4	6.9	31.6	11.0	0.0	0.0	446.1
1993-94	0.0	0.8	58.8	11.6	107.9	67.6	54.9	23.1	4.0	0.3	2.1	1.6	332.7
1994-95	5.5	42.0	138.7	45.3	20.9	25.1	12.1	18.6	16.8	0.1	16.3	0.0	341.4
1995-96	0.1	5.6	36.7	12.6	103.6	38.7	48.3	26.7	1.1	3.0	0.0	2.0	278.4
1996-97	0.3	41.8	13.4	62.6	12.5	34.1	40.6	36.4	8.2	7.8	0.1	3.6	261.4
1997-98	29.3	25.3	58.1	59.8	56.5	16.0	44.2	8.6	30.3	4.4	0.0	0.0	332.5
1998-99	4.2	0.6	34.3	103.3	93.1	40.7	24.7	22.8	2.8	17.9	1.3	3.9	349.6
1999-00	5.0	21.3	20.6	19.8	49.4	59.3	44.9	66.2	14.1	2.6	0.0	0.7	303.9
2000-01	18.1	41.1	86.2	133.3	46.0	50.2	9.6	19.2	24.8	0.0	0.0	10.0	438.5
2001-02	0.6	19.8	43.8	192.0	90.0	40.2	29.7	44.9	28.4	3.3	7.5	4.1	504.3
2002-03	8.2	10.4	21.0	156.6	57.1	130.6	91.2	22.5	4.3	22.6	0.0	0.2	524.7
2003-04	1.0	9.9	24.8	94.4	212.4	93.8	1.5	12.0	5.6	5.0	0.0	0.1	460.5
2004-05	0.0	17.9	60.6	88.2	86.5	30.1	22.3	24.3	12.7	36	0	1.2	379.8
2005-06	10.7	12.2	104.3	9.3	105.1	38.5	39.5	13.0	5.5	2.1	14.4	0.0	354.6
2006-07	7	72.3	39.9	12.4	34.1	128.6	36.6	18	70.8	0.4	0.9	1.8	422.8
2007-08	0	7.3	28.7	52.5	23.1	31.5	9.6	7.6	15.9	0.2	0.3	3.2	179.9
2008-09	7.6	20.5	16.8	78.5	62.2	75.7	54.5	15.7	14.3	0.1	0.9	4.6	351.4
2009-10	27.4	35.5	41.9	154.5	111.8	158.1	7.7	12.8	14.4	8.4	1.7	0.4	574.6
2010-11	1.7	12.5	0.5	51.9	101.5	44.4	37.1	48.4	32.7	18.8	0.1	0.1	349.7
2011-12	14.2	10.8	103.6	64.9	164.7	62.1	23	14.2	50.6	3	1.6	1.4	514.1
2012-13	0	67.5	86	107.3	68.1	32.5	11.8	42.6	46.6	0	0.3	0.3	463.0
2013-14	1.5	4.9	13	62.2	16.7	17.9	28.7	12.5	48.6	11.4	0.6	0.4	218.4
2014-15	12.5	53.6	37	90.9									

Table 4.57: Statistical measures of TRNC rainfall

Parametric										Non-Parametric		
$\bar{x}_{ar}$	S_x	C_{dx}	C_v	C_S	$\bar{x}_{geo}$	S_{logx}	C_{dlogx}	CV_{log}	CS_{log}	$\bar{x}_{med}$	C_{dx}	PC_v
378.8	92.8	0.212	0.245	0	2.6	0.1	0.034	0.04	-0.6	364.7	0.199	0.675

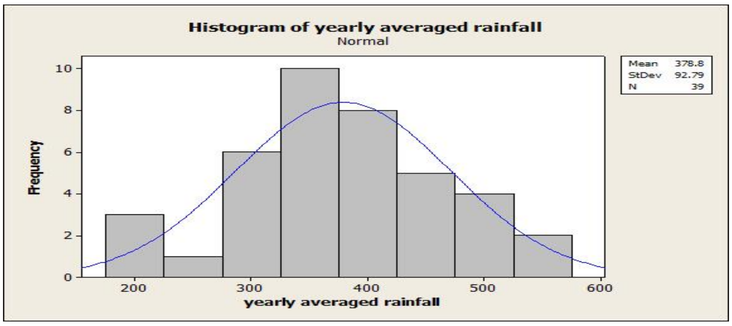


Figure 4.51: TRNC regions' rainfall histogram

4.8.1 Empirical determination of minimum required sample size of rainfall for TRNC based on mean and standard deviation

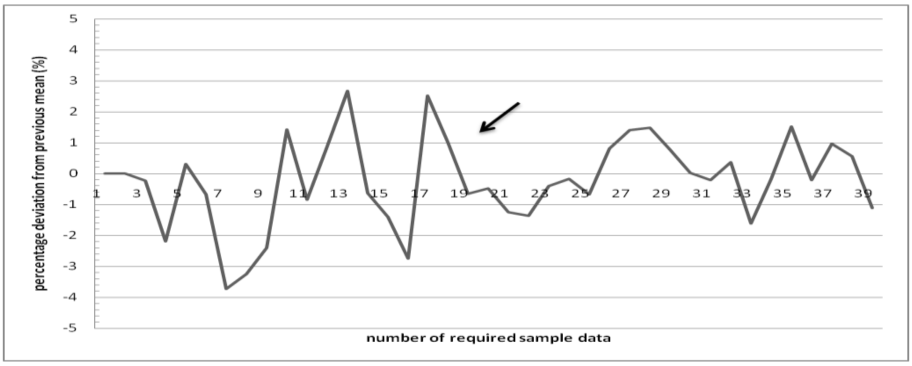


Figure 4.52: Curve showing the required number of sample size for TRNC regions' rainfall, based on the percentage deviations of the mean values

<u>Comment:</u> The curve becomes less than ± 2% error based on the mean values, once the sample size reaches to 19.

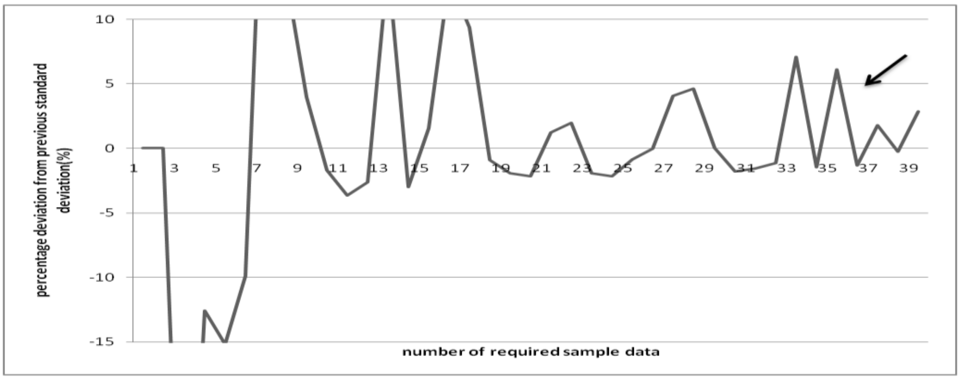


Figure 4.53: Curve showing the required number of sample size for TRNC regions' rainfall, based on the percentage variation of the standard deviations

Comment: The curve becomes less than ± 5% error based on standard deviation once the sample size reaches to 37.

Hence $n_{min} = 37$ is the required minimum number that satisfies both mean and standard deviation for TRNC rainfall.

Table 4.58: Appropriate rainfall sample size of TRNC for statistic and probabilistic studies

Based on Mean (not more than 2% deviation)	Based on Standard Deviation (not more than 5% deviation)
19 < 39 OK	37 < 39 OK

4.8.2 Quality Checking Tests of TRNC

Table 4.59: Quality Tests Results of TRNC rainfall

Quality check Tests	Parametric or Non-Parametric test checking
Normality	• p-value = 0.81>0.05. Therefore, the TRNC rainfall is normally distributed.
Homogeneity	• Based on Pettitt, SNHT, BR, and VNR tests TRNC rainfall time series is Homogenous.
Consistency	• Rainfall of TRNC region is found to be regionally consistent based on double mass curve among nearby 5 regions averaged rainfall.
Trend	• No trend exists in TRNC rainfall because Mann – Kendall p-value = 0.454>0.05 • Sens slope=1.05
Stationarity	• TRNC rainfall is stationary based on ADF test since slope of regression $\gamma = 0.33 > 0$.

4.8.3 Probability distributions details of TRNC region

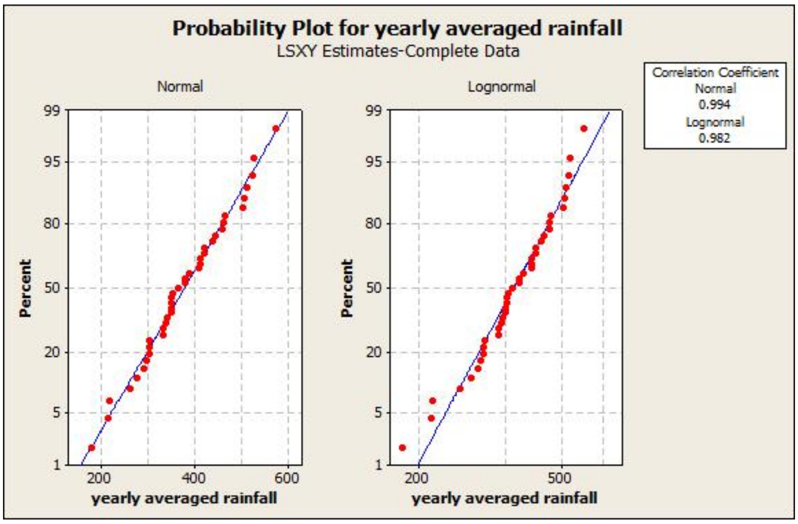


Figure 4.54: TRNC rainfall fit on normal and log-Normal probability distributions

Table 4.60: Equations of the probability distribution functions with their correlation coefficients for TRNC Rainfall

Name	Equation	Correlation Coefficient
Normal	$x = 378.8 + 91.6\ Z$	0.994
log-Normal	$y = \log x = 2.6 + 0.1\ Z$	0.982

Result: Comparing the correlation coefficients of the two probability distributions for TRNC rainfall, it is concluded that, normal distribution has the best fitted curve being greater correlation coefficient value.

4.8.4 Forecasted values by time series models of TRNC rainfall for the hydrologic years period 2003-04 to 2013-14

4.8.4.1 Markov Model

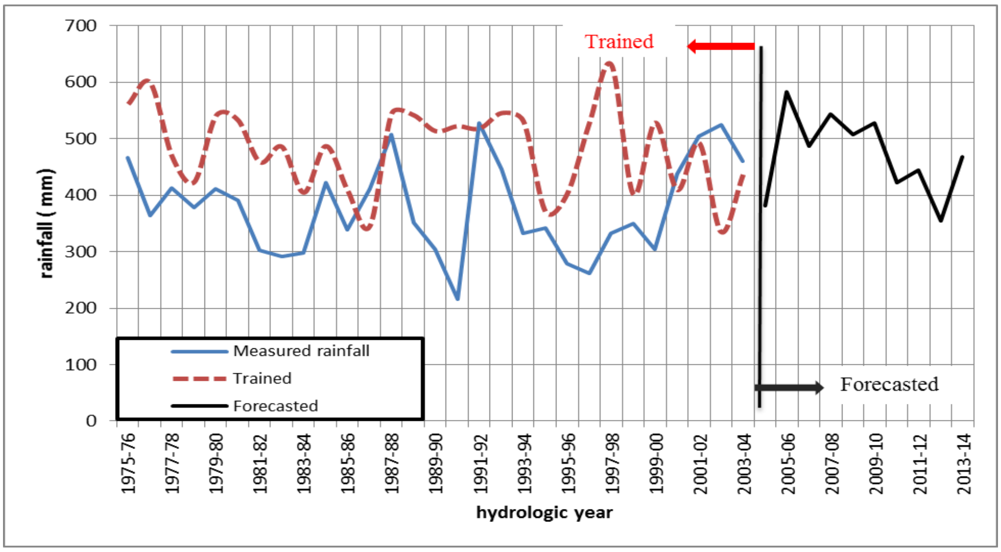


Figure 4.55: Graphical comparison of Markov model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of TRNC

4.8.4.2 Auto-Regressive Model

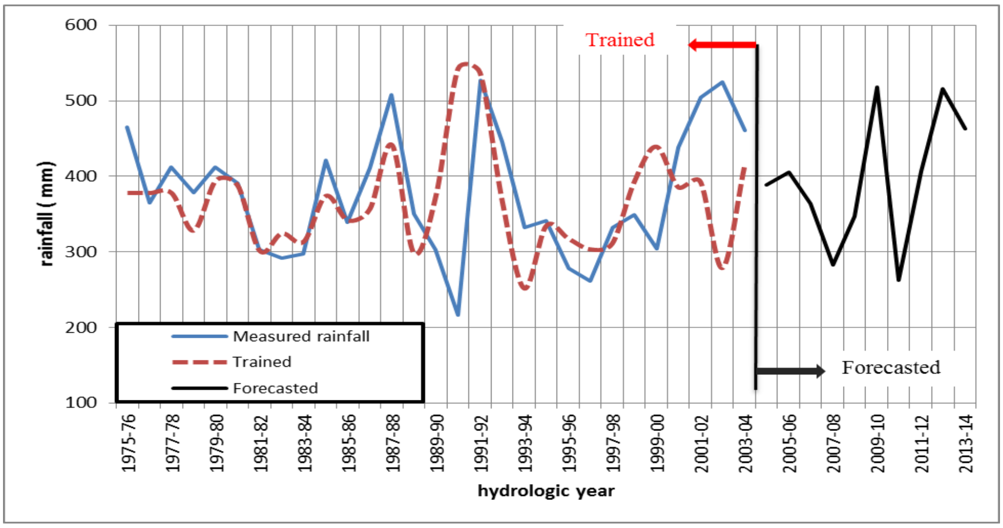


Figure 4.56: Graphical comparison of AR(3) model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of TRNC

4.8.4.3 Holt-Winter Multiplicative Model

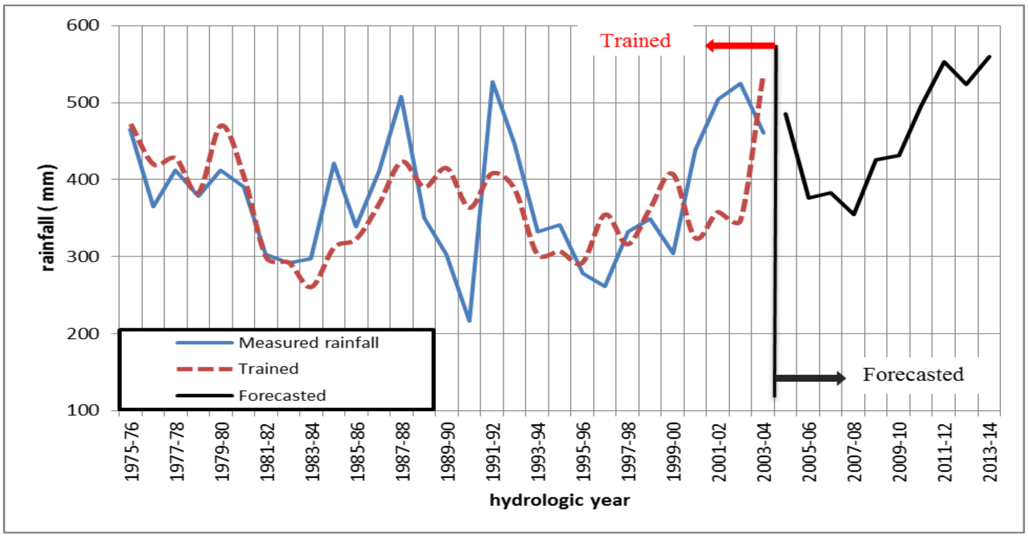


Figure 4.57: Graphical comparison of Holt-Winter Multiplicative method model (trained and forecasted) and the hydrologic yearly averaged rainfall (measured) of TRNC

4.8.4.4 Selecting the best fitted time series model for TRNC region

Table 4.61: Accuracy checking of forecasted data of TRNC

Hydrologic Year	DATA			
	Measured	Markov model	AR(3)	Holt-Winter model
2004-2005	379.8	381.0	388.4	485.9
2005-2006	354.6	582.5	405.9	376.7
2006-2007	422.8	486.2	363.7	383.3
2007-2008	179.9	544.0	283.0	355.1
2008-2009	351.4	506.9	347.0	426.1
2009-2010	574.6	527.8	517.7	431.1
2010-2011	349.7	422.6	262.7	495.2
2011-2012	514.1	443.9	406.3	552.7
2012-2013	463.0	354.9	516.2	523.5
2013-2014	218.4	468.7	462.6	560.0
MSE		29947.1	10175.2	21318.4
Ratio w.r.t. min.		2.94	1.00	2.10
MAPE (%)		27.6	20.3	24.7
Ratio w.r.t. min.		1.36	1.00	1.22
RMSE		2994.7	1017.5	2131.8
Ratio w.r.t. min.		2.94	1.00	2.10
MAD		136.0	77.6	114.7
Ratio w.r.t. min.		1.75	1.00	1.48
Overall		9.00	4.00	6.89

<u>Result:</u> AR(3) Model is the best model based on the above 4 error measures having the lowest overall error ratio among the other models. Hence, for TRNC this model is used to generate the rainfall for the Hydrologic years 2014-15 to 2018-19 and are all tabulated below.

4.8.4.5 Prediction of yearly rainfall of TRNC region for hydrologic years 2014-2015 to 2018-19

Table 4.62: Expected yearly rainfall of TRNC region based on AR(3) model for hydrologic years 2014-15 to 2018-19

Hydrologic Year	**Expected Yearly Total Rainfall (mm)**
2014-2015	281.2
2015-2016	355.7
2016-2017	402.4
2017-2018	357.5
2018-2019	323.6

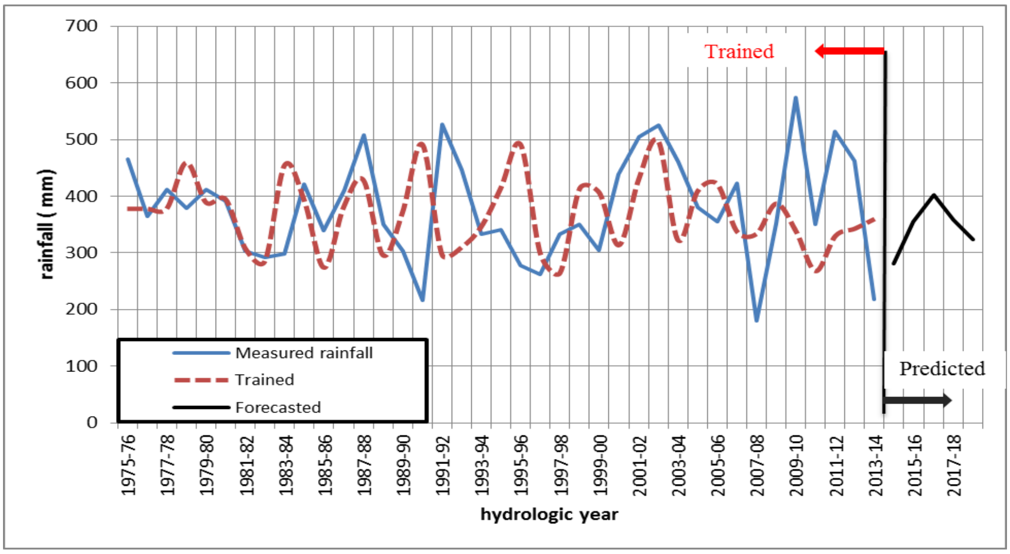


Figure 4.58: Expected (predicted) yearly rainfall of TRNC region based on AR(3) model for hydrologic years 2014-15 to 2018-19

4.8.5 Wet and dry spells for TRNC

Table 4.63: Numerical representation of wet and dry spells of TRNC region based on the mean of the data

Hydrologic Year	Rainfall (mm)	Wet\ Dry	0: Dry 1: Wet
Mean = 378.8 mm			
1975-1976	465.7	wet	**1**
1976-1977	364.7	dry	**0**
1977-1978	412.3	wet	**1**
1978-1979	378.9	wet	**1**
1979-1980	411.7	wet	**1**
1980-1981	390.6	wet	**1**
1981-1982	302.6	dry	**0**
1982-1983	291.8	dry	**0**
1983-1984	297.8	dry	**0**
1984-1985	421.7	wet	**1**
1985-1986	339.3	dry	**0**
1986-1987	410.6	wet	**1**
1987-1988	507.6	wet	**1**
1988-1989	350.6	dry	**0**
1989-1990	303.3	dry	**0**
1990-1991	215.9	dry	**0**
1991-1992	527.6	wet	**1**
1992-1993	446.1	wet	**1**
1993-1994	332.7	dry	**0**
1994-1995	341.4	dry	**0**
Mean = 378.8 mm			
1995-1996	278.4	dry	**0**
1996-1997	261.4	dry	**0**
1997-1998	332.5	dry	**0**
1998-1999	349.6	dry	**0**
1999-2000	303.9	dry	**0**
2000-2001	438.5	wet	**1**
2001-2002	504.3	wet	**1**
2002-2003	524.7	wet	**1**
2003-2004	460.5	wet	**1**
2004-2005	379.8	wet	**1**
2005-2006	354.6	dry	**0**
2006-2007	422.8	wet	**1**
2007-2008	179.9	dry	**0**
2008-2009	351.4	dry	**0**
2009-2010	574.6	wet	**1**
2010-2011	349.7	dry	**0**
2011-2012	514.1	wet	**1**
2012-2013	463	wet	**1**
2013-2014	218.4	dry	**0**

<u>Result</u>: number of wet spells = 19 (49%), number of Dry spells = 20 (51%).

Therefore the TRNC is in dry spell during the studied period.

4.8.6 Study of monthly wet and dry spells of TRNC region for hydrological years 1975-76 to 2013-14

Table 4.64: Monthly wet and dry spell of TRNC

Months	No. of wet spell	No. of dry spell	Conclusion
Sep	13	26	67% dry
Oct	14	25	64% dry
Nov	16	23	59% dry
Dec	17	22	56% dry
Jan	15	24	62% dry
Feb	17	22	56% dry
Mar	19	20	51% dry
Apr	14	25	64% dry
May	14	25	64% dry

Result: Since all the months are in dry spell, TRNC region throughout the year is dry during the studied period.

Chapter 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

- Based on the empirically studied sample size it was observed that, for all the meteorological regions, the used sample size (39) is a sufficient value for any statistic and probabilistic calculations where for each meteorological region different lower values were found.
- Normality, Homogeneity, Consistency, Trend analysis, and Stationarity tests were used as quality test for each meteorological region and TRNC as a whole and found that they are all within the acceptable range of 95 % confidence interval.
- Since yearly average values were used, among the several probability distribution models given in literature, Normal and log-Normal distributions were studied and the best representative for each meteorological region was determined through the best fitting curve approach. Except East Coast and West Mesaria which are log-Normal, the other meteorological regions and TRNC obey fairly well to the Normal distribution model.
- Rainfall of the 6 meteorological regions of North Cyprus and TRNC as a whole were analyzed by three widely used time series models (Markov, Auto-Regressive, and Holt-Winter Multiplicative) and five

successive years of predicting data sets (from hydrologic years 2014-15 to 2018-19) were generated. For this purpose, to determine the best time series model for each regions and for TRNC, the standardized MSE, MAPE, RMSE, and MAD were used. It is worth to add that, while studying the time series models the sample size of 29 (from hydrologic years 1975-76 to 2003-04) is used for training and of size 10 (hydrologic years from 2004-05 to 2013-14) for forecasting although the empirical determined n_{min} suggested the acceptable minimum sizes larger.

- To be able to realize how the meteorological data is varying for a long run, wet and dry spells of yearly and monthly rainfall for each region were as well studied. Interestingly all the data in all the meteorological regions suggest dryness.
- Based on the best representative time series predictions of the five successive years, except North Coast and West Mesaria for hydrologic year 2014-15 will experience lower than the long yearly average rainfall value implying dryness whereas, TRNC will experience higher than its long yearly average during the hydrologic year 2016-17 implying wet period. The other relevant details are given in the following table.
- It is worth to express that for the coming five hydrologic years (2014-15 to 2018-19) West Mesaria will experience total rainfall more than its long years average for all those years whereas Central Mesaria will experience total rainfall less than its long years average for all those years.
- The synopsis of the rainfall parameters and time series models of each meteorological region and TRNC as a whole are tabulated below.

Table 5.1: Synopsis of the rainfall parameters and time series models of each meteorological region and TRNC as a whole

Region	Long years Rainfall Mean (mm)	n_{min}	Quality Check Tests					Representative Probability Distribution		Representative Time Series Model	Dry/ Wet Spell	Below (↓) or Above (↑) Long Years Rainfall Means				
			Normality	Homogeneity	Consistency	Trend	Stationarity	Model	Equation			2014-15	2015-16	2016-17	2017-18	2018-19
Central Mesaria	299.9	34	OK	OK	OK	Reject	OK	Normal	$x = 299.9 + 74.9 Z$	AR(1)	53% dry	↓	↓	↓	↓	↓
East Coast	334.7	19	OK	OK	OK	Reject	OK	log-Normal	$\log x = 2.5 + 0.1 Z$	Holt-Winter Mult.	51% dry	↓	↑	↑	↓	↓
East Mesaria	334.2	37	OK	OK	OK	Reject	OK	Normal	$x = 342.2 + 93.6 Z$	AR(1)	56% dry	↓	↑	↑	↓	↓
Karpaz	449.7	19	OK	OK	OK	Reject	OK	Normal	$x = 324.2 + 93.6 Z$	Holt-Winter Mult.	51% dry	↓	↑	↑	↓	↓
North Coast	461.8	37	OK	OK	OK	Reject	OK	Normal	$x = 461.8 + 115.6 Z$	AR(1)	51% dry	↑	↓	↓	↑	↑
West Mesaria	310.8	37	OK	OK	OK	Reject	OK	log-Normal	$\log x = 2.5 + 0.1 Z$	Holt-Winter Mult.	59% dry	↑	↑	↑	↑	↑
TRNC	378.8	37	OK	-	-	Reject	OK	Normal	$x = 378.8 + 91.6 Z$	AR (3)	51% dry	↓	↓	↑	↓	↓

5.2 Recommendation

- Monthly and yearly rainfall from South Cyprus and other surrounding countries should also be collected and correlated to give more general comments on TRNC general yearly rainfall trend.
- Three different time series models were discussed and used in this study to determine the best representative model of each region for prediction. These models need a set of training data and a set of forecasting data but the empirical determination of n_{min} suggest the acceptable minimum sizes unfortunately except East Coast and Karpaz regions larger than the trained data set size hence, the prediction values generated through those suggested models carry higher risk of deviations. So it is recommended to repeat the similar study at least after 8 hydrologic years later so as to satisfy the minimum sample size of 37 during the training period. It is worth to express that, during prediction, it is assumed that, the global warming and the other unexpected extreme meteorological variations having insignificant effect during the prediction period.
- Now a days, there are other approached like artificial neural networks, wavelet approaches etc... that can be used as time series models for forecasting studies which are highly recommended.
- For dry or wet spell studies, instead of comparing the data with respect to relevant mean value (as was done in this study), for each hydrologic year and for each month, several wetness-dryness bands based on ± standard deviation multiples could be established so as to classify the regional dryness-wetness spells in a detailed manner.

REFERENCES

Alcamo, J., Floerke, M., and Maerker, M. (2007). Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrological Sciences*, 52, 247-275.

Altan, P. and Sen, Z. (2000). Precipitation analysis of TRNC, *ACE 2000*, Eastern Mediterrnean University. Vol. 4. p 1971-1982.

Ashuri, B. and Lu, J. (2010). Time Series Analysis of ENR Construction and Cost Index. *ASCE Journal of Construction Engineering and Management*, 136(11), 1227-1237.

Birpınar, M.E. (Eds.). (2003). Water resources data analysis. *Turkish Water Foundation.*

Box, E.P. and Jenkins, G.M. (1970). Time series analysis: forecasting and control. *Holden-Day.*

Bradley, J.V. (1968). Distribution- free statistical tests. *Prentice-Hall*. ASIN: B0006BVE3K.

Brauner, J.S. (1997). Nonparametric estimation of slope: *Sen's method in environmental pollution.*http://ewr.cee.vt.edu/environmental/sen/sen.html

Bruce, B. and Richard,T.O. (1997). Applied Statistics. *Mcgraw-Hill College*.

Buishand, T.A. (1982). Some methods for testing the homogeneity of rainfall records. *J. Hydrol*., 58: 11--27.

Christensen, J.H. (2007). Regional climate projections. Climate Change. *Cambridge University Press*, Cambridge, 847-940.

Christiane, L. and Bogdan, G. (2010). Meta-learning for time series forecasting and forecast combination. *Elsevier, Neurocomputing.*

Gao, X. and Giorgi, F. (2008). Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model. *Global and Planetary change 62.*

Giorgi, F. (2006). Climate Change Hot-Spots. *Geophysical research letters.* Vol.33,l08707,doi:10.1029/2006GL025734.

Gupta, R. S. (1989). Hydrology and Hydraulic Systems. *Prentice Hall*, pp 343-350.

Kang, H.M. and Yusof, F. (2012). Homogeneity Tests on Daily Raifall Series in Peninsular Malaysia. *Int.J.Contemp.Math.sciences*,Vol.7,2012,n0.1,9-22.

Kimyacı, S. (2004). M.Sc Thesis: *Rainfall, establishment of intensity duration – frequency curves for Cyprus*. Eastern Mediterranean University.

Kostopoulou, E. and Jones, P.D. (2007a). Comprehensive analysis of the climate variability in the eastern Mediterranean, Part I: Map pattern classification, *Int. J. Climatology.*, 27, 1189– 1214.

Kostopoulou, E. and Jones, P.D. (2007b). Comprehensive analysis of the climate variability in the eastern Mediterranean, part II: relationships between atmospheric circulation patterns and surface climatic elements, *Int. J. Climatology.*, 27, 1351–1371.

Kottegoda, N.T. (1980). Stochastic water resources technology. *The Macmillan Press Ltd.*

Kypris, D. C. (1995). Diachronic changes of rainfall and the water resources in Cyprus. Water Resources Management under Drought or Water Shortage Conditions. *A. A. Balkema Publishers.*

Ljung, G. M. and Box, G.E.P. (1983). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297–303.

Maidment, D.R. (Eds.). (1993). Handbook of Hydrology. *McGraw-Hill Inc*.

Meehl, G., Stocker, T., and Collins, W. (2007). The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Inter - governmental Panel on Climate Change. *Cambridge University Press, Cambridge*, p 749–844

Mendenhall, R.B. (1986). Statistics for Management And Economics. *Boston: PWS-KENT.*

Mutreja, K.N. (1990). Applied hydrology. *Tata McGraw-Hill Publishing Company Limited.*

Noordin, J.B. (2010). M.Sc Thesis: *Reservoir storage simulation and forecasting models*. Universiti Teknologi Malaysia.

Pashiardis, S. (2003). Trends of precipitation in Cyprus precipitation analysis for agricultural planning. http://www.fao.org/sd/climagrimed/pdf/ws01_08.pdf

Pettitt, A.N. 1979. A non-parametric approach to the change-point problem. *Journal of Applied Statistics* 28: 126–135

Pinto, J. G., Spangehl, T., Ulbrich, U., and Speth, P. (2006). Assessment of winter cyclone activity in a transient ECHAM4-OPYC3 GHG experiment. *Meteorology*. Z., 15, 279–291.

Rob, J.H. and George, A. (2013). Forecasting principles and practice. *Online open access text book.*

Salvatore, D.(1982). Statistics and econometrics. *McGraw-Hill Publishing Company.*

Schkade, L. (1983). Statistical Analysis for Administative Decisions. *South western publishing co.*

Schmidt, D.F. and Makalic, E. (2008). Model Selection Tutorial #1: *Akaike`s informationCriterion.*http://www.csse.monash.edu.au/Schmitmakalic2008.pdf (latest chack 01sep2015)

Sen, Z. (2003). With the personal discussion of this thesis supervisor. *Istanbul*

Seyhan, E. (1994). Application of statistical methods to hydrology. *Vrije Universiteit, Amsterdam.*

Serano, S.E. (2001). Engineering Uncertainty and Risk Analysis. *Hydro Sience Inc*. 096556438x.

Seyhun, R. and Akıntuğ, B. (2012). Trend Analysis of Rainfall in North Cyprus. *Global Conference on Global Warming* (GCGW - 2012), Istanbul, Turkey.

Sharifi, Y. (2006). M.Sc. Thesis: *Hydro-Climatological Variations and Trend in TRNC.* Eastern Mediterranean University

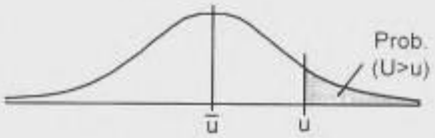
Spiegel, M.R. and Stephens, L.J. (1999). Schaume's Outline series: Statistics. *McGraw-Hill International Editions*. 0070602816.

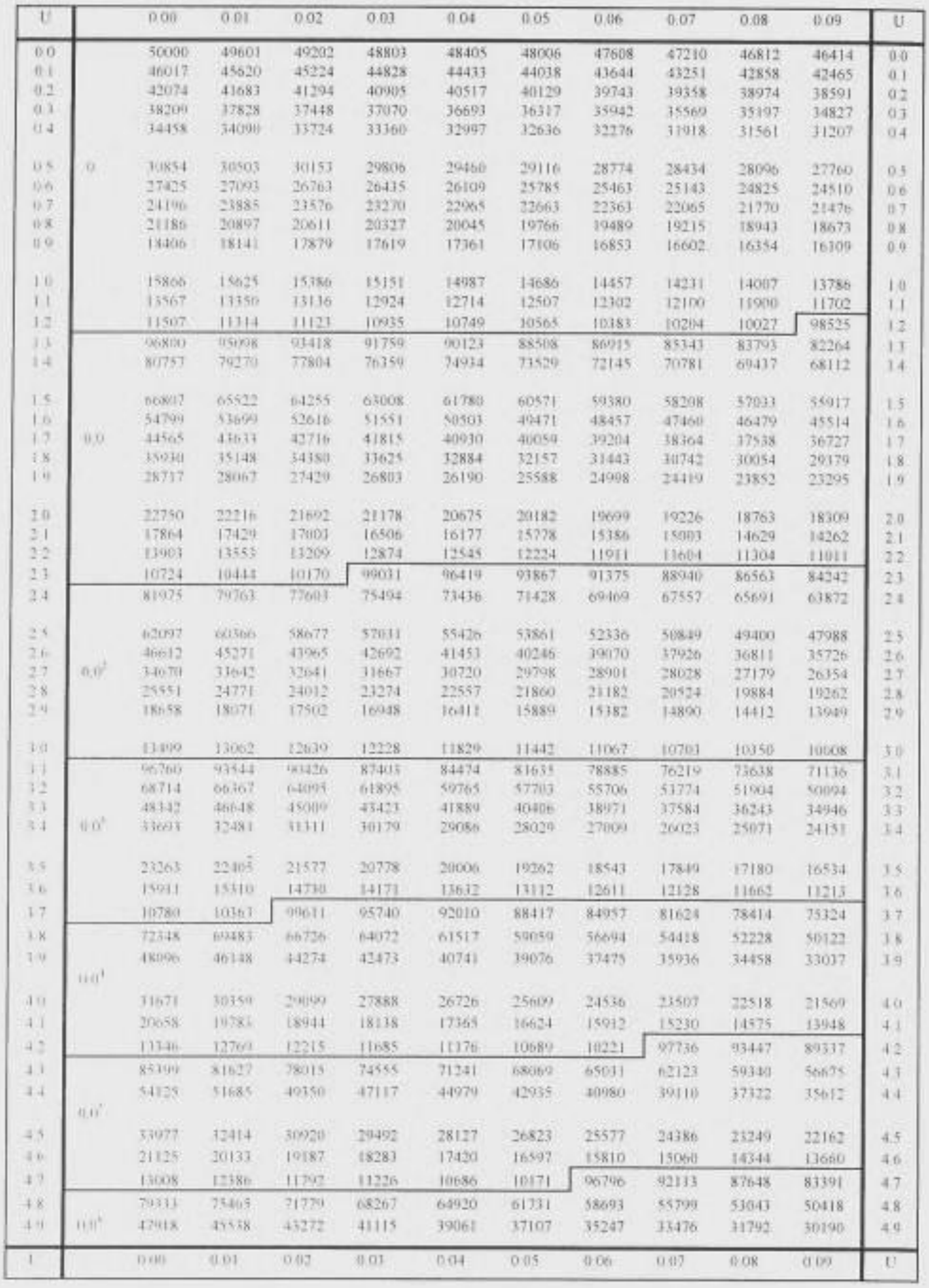
Teelucksingh, S.S. and Watson, P.K. (2002). Apractical introduction to econometric methods, classical and modern. *University of the West Indies press.*

Usul, N. (2001). Engineering hydrology. *METU Press Publishing Company*. Ankara.

APPENDICES

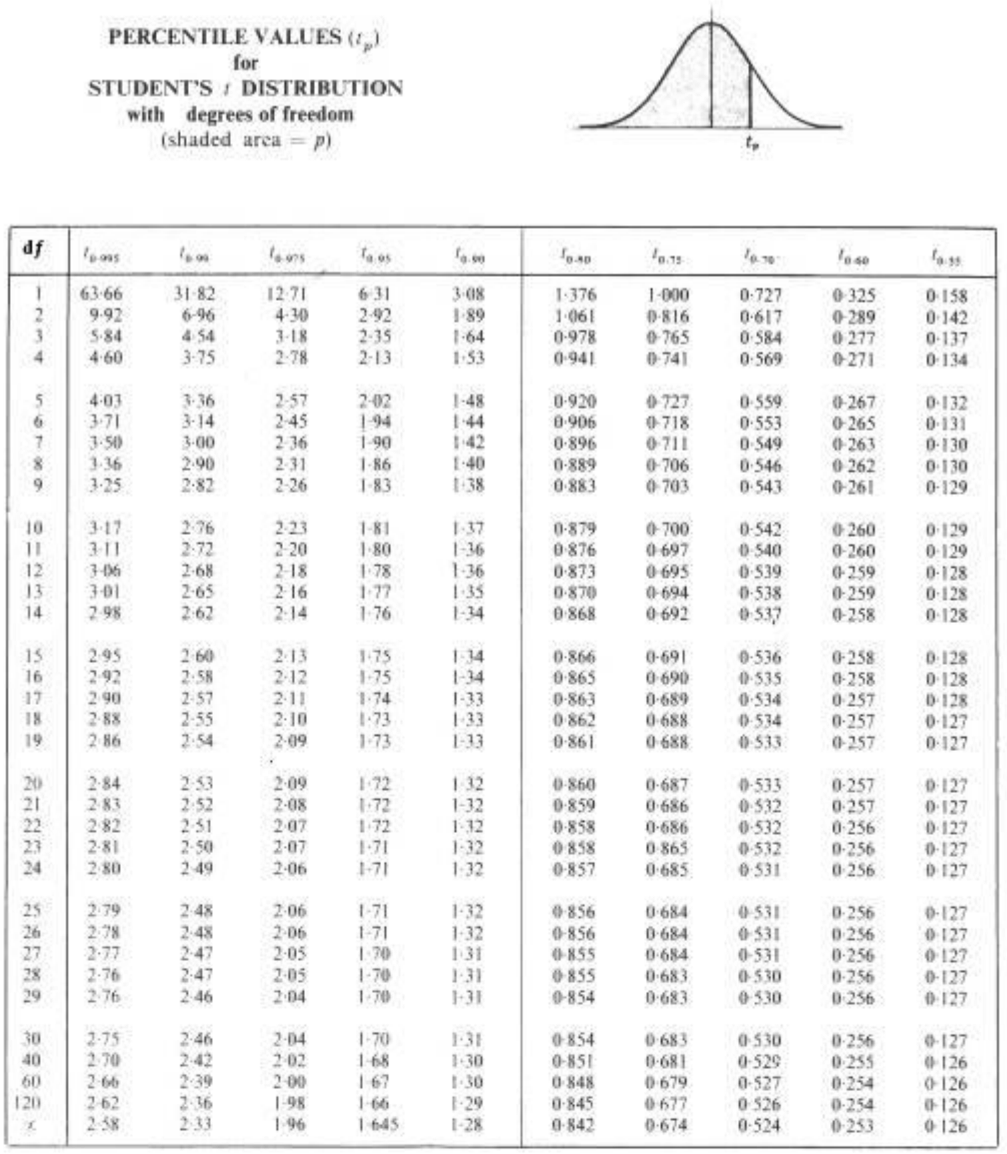
Appendix A: Normal and Log-Normal distributions z-table





U		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	U
0.0		50000	49601	49202	48803	48405	48006	47608	47210	46812	46414	0.0
0.1		46017	45620	45224	44828	44433	44038	43644	43251	42858	42465	0.1
0.2		42074	41683	41294	40905	40517	40129	39743	39358	38974	38591	0.2
0.3		38209	37828	37448	37070	36693	36317	35942	35569	35197	34827	0.3
0.4		34458	34090	33724	33360	32997	32636	32276	31918	31561	31207	0.4
0.5	0	30854	30503	30153	29806	29460	29116	28774	28434	28096	27760	0.5
0.6		27425	27093	26763	26435	26109	25785	25463	25143	24825	24510	0.6
0.7		24196	23885	23576	23270	22965	22663	22363	22065	21770	21476	0.7
0.8		21186	20897	20611	20327	20045	19766	19489	19215	18943	18673	0.8
0.9		18406	18141	17879	17619	17361	17106	16853	16602	16354	16109	0.9
1.0		15866	15625	15386	15151	14987	14686	14457	14231	14007	13786	1.0
1.1		13567	13350	13136	12924	12714	12507	12302	12100	11900	11702	1.1
1.2		11507	11314	11123	10935	10749	10565	10383	10204	10027	98525	1.2
1.3		96800	95098	93418	91759	90123	88508	86915	85343	83793	82264	1.3
1.4		80757	79270	77804	76359	74934	73529	72145	70781	69437	68112	1.4
1.5		66807	65522	64255	63008	61780	60571	59380	58208	57033	55917	1.5
1.6		54799	53699	52616	51551	50503	49471	48457	47460	46479	45514	1.6
1.7	0.0	44565	43633	42716	41815	40930	40059	39204	38364	37538	36727	1.7
1.8		35930	35148	34380	33625	32884	32157	31443	30742	30054	29379	1.8
1.9		28717	28067	27429	26803	26190	25588	24998	24419	23852	23295	1.9
2.0		22750	22216	21692	21178	20675	20182	19699	19226	18763	18309	2.0
2.1		17864	17429	17003	16506	16177	15778	15386	15003	14629	14262	2.1
2.2		13903	13555	13209	12874	12545	12224	11911	11604	11304	11011	2.2
2.3		10724	10444	10170	99031	96419	93867	91375	88940	86563	84242	2.3
2.4		81975	79763	77603	75494	73436	71428	69469	67557	65691	63872	2.4
2.5		62097	60366	58677	57031	55426	53861	52336	50849	49400	47988	2.5
2.6		46612	45271	43965	42692	41453	40246	39070	37926	36811	35726	2.6
2.7	0.0^2	34670	33642	32641	31667	30720	29798	28901	28028	27179	26354	2.7
2.8		25551	24771	24012	23274	22557	21860	21182	20524	19884	19262	2.8
2.9		18658	18071	17502	16948	16411	15889	15382	14890	14412	13949	2.9
3.0		13499	13062	12639	12228	11829	11442	11067	10703	10350	10008	3.0
3.1		96760	93544	90426	87403	84474	81635	78885	76219	73638	71136	3.1
3.2		68714	66367	64095	61895	59765	57703	55706	53774	51904	50094	3.2
3.3		48342	46648	45009	43423	41889	40406	38971	37584	36243	34946	3.3
3.4	0.0^3	33693	32481	31311	30179	29086	28029	27009	26023	25071	24151	3.4
3.5		23263	22405	21577	20778	20006	19262	18543	17849	17180	16534	3.5
3.6		15911	15310	14730	14171	13632	13112	12611	12128	11662	11213	3.6
3.7		10780	10363	99611	95740	92010	88417	84957	81624	78414	75324	3.7
3.8		72348	69483	66726	64072	61517	59059	56694	54418	52228	50122	3.8
3.9		48096	46148	44274	42473	40741	39076	37475	35936	34458	33037	3.9
	0.0^4											
4.0		31671	30359	29099	27888	26726	25609	24536	23507	22518	21569	4.0
4.1		20658	19783	18944	18138	17365	16624	15912	15230	14575	13948	4.1
4.2		13346	12769	12215	11685	11176	10689	10221	97736	93447	89337	4.2
4.3		85399	81627	78015	74555	71241	68069	65031	62123	59340	56675	4.3
4.4		54125	51685	49350	47117	44979	42935	40980	39110	37322	35612	4.4
	0.0^5											
4.5		33977	32414	30920	29492	28127	26823	25577	24386	23249	22162	4.5
4.6		21125	20133	19187	18283	17420	16597	15810	15060	14344	13660	4.6
4.7		13008	12386	11792	11226	10686	10171	96796	92113	87648	83391	4.7
4.8		79333	75465	71779	68267	64920	61731	58693	55799	53043	50418	4.8
4.9	0.0^6	47918	45538	43272	41115	39061	37107	35247	33476	31792	30190	4.9
U		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	U

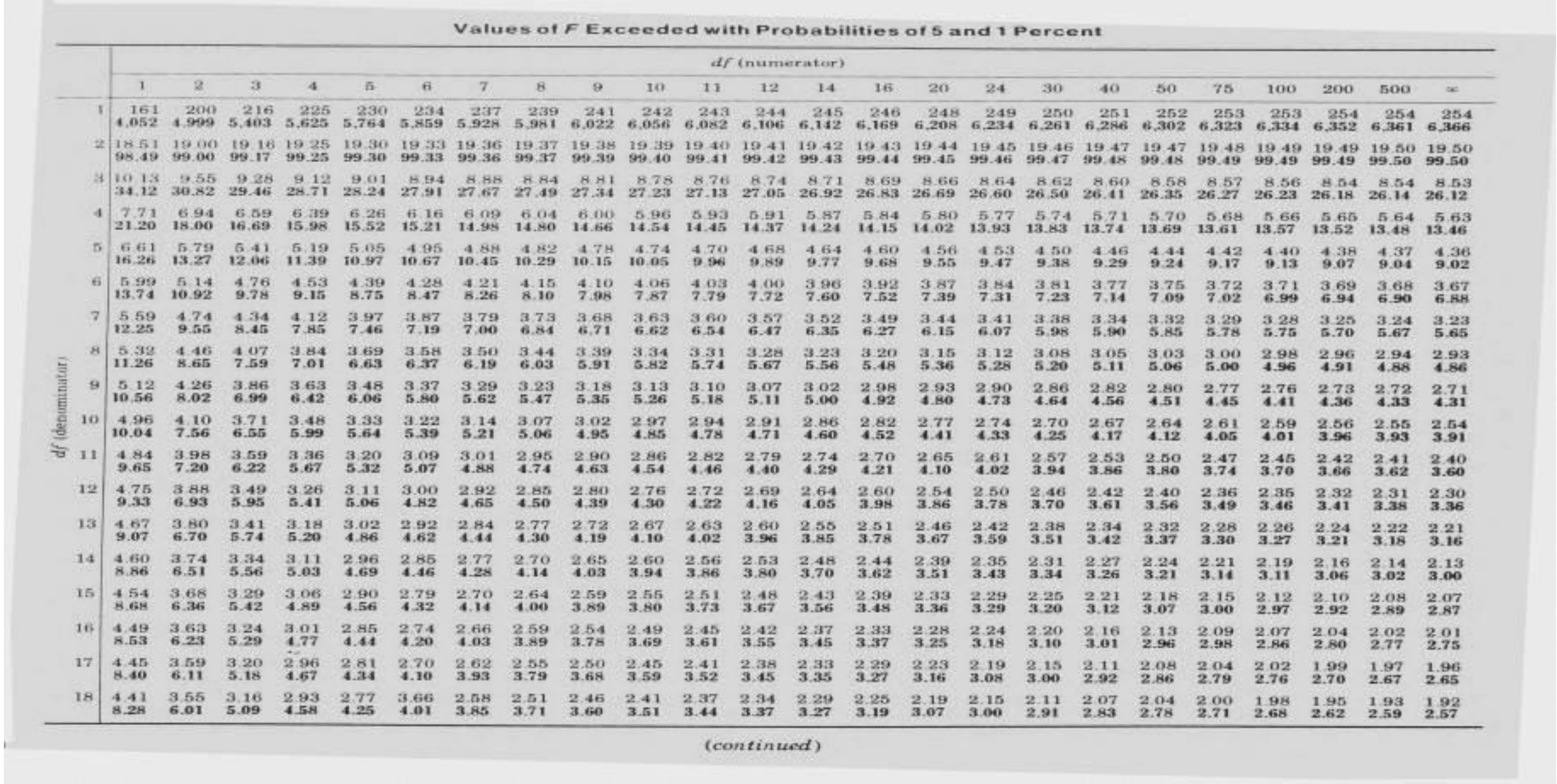
Appendix B: t-test values for different confidence intervals and degree of freedoms



PERCENTILE VALUES (t_p)
for
STUDENT'S t DISTRIBUTION
with degrees of freedom
(shaded area = p)

df	$t_{0.995}$	$t_{0.99}$	$t_{0.975}$	$t_{0.95}$	$t_{0.90}$	$t_{0.80}$	$t_{0.75}$	$t_{0.70}$	$t_{0.60}$	$t_{0.55}$
1	63·66	31·82	12·71	6·31	3·08	1·376	1·000	0·727	0·325	0·158
2	9·92	6·96	4·30	2·92	1·89	1·061	0·816	0·617	0·289	0·142
3	5·84	4·54	3·18	2·35	1·64	0·978	0·765	0·584	0·277	0·137
4	4·60	3·75	2·78	2·13	1·53	0·941	0·741	0·569	0·271	0·134
5	4·03	3·36	2·57	2·02	1·48	0·920	0·727	0·559	0·267	0·132
6	3·71	3·14	2·45	1·94	1·44	0·906	0·718	0·553	0·265	0·131
7	3·50	3·00	2·36	1·90	1·42	0·896	0·711	0·549	0·263	0·130
8	3·36	2·90	2·31	1·86	1·40	0·889	0·706	0·546	0·262	0·130
9	3·25	2·82	2·26	1·83	1·38	0·883	0·703	0·543	0·261	0·129
10	3·17	2·76	2·23	1·81	1·37	0·879	0·700	0·542	0·260	0·129
11	3·11	2·72	2·20	1·80	1·36	0·876	0·697	0·540	0·260	0·129
12	3·06	2·68	2·18	1·78	1·36	0·873	0·695	0·539	0·259	0·128
13	3·01	2·65	2·16	1·77	1·35	0·870	0·694	0·538	0·259	0·128
14	2·98	2·62	2·14	1·76	1·34	0·868	0·692	0·537	0·258	0·128
15	2·95	2·60	2·13	1·75	1·34	0·866	0·691	0·536	0·258	0·128
16	2·92	2·58	2·12	1·75	1·34	0·865	0·690	0·535	0·258	0·128
17	2·90	2·57	2·11	1·74	1·33	0·863	0·689	0·534	0·257	0·128
18	2·88	2·55	2·10	1·73	1·33	0·862	0·688	0·534	0·257	0·127
19	2·86	2·54	2·09	1·73	1·33	0·861	0·688	0·533	0·257	0·127
20	2·84	2·53	2·09	1·72	1·32	0·860	0·687	0·533	0·257	0·127
21	2·83	2·52	2·08	1·72	1·32	0·859	0·686	0·532	0·257	0·127
22	2·82	2·51	2·07	1·72	1·32	0·858	0·686	0·532	0·256	0·127
23	2·81	2·50	2·07	1·71	1·32	0·858	0·865	0·532	0·256	0·127
24	2·80	2·49	2·06	1·71	1·32	0·857	0·685	0·531	0·256	0·127
25	2·79	2·48	2·06	1·71	1·32	0·856	0·684	0·531	0·256	0·127
26	2·78	2·48	2·06	1·71	1·32	0·856	0·684	0·531	0·256	0·127
27	2·77	2·47	2·05	1·70	1·31	0·855	0·684	0·531	0·256	0·127
28	2·76	2·47	2·05	1·70	1·31	0·855	0·683	0·530	0·256	0·127
29	2·76	2·46	2·04	1·70	1·31	0·854	0·683	0·530	0·256	0·127
30	2·75	2·46	2·04	1·70	1·31	0·854	0·683	0·530	0·256	0·127
40	2·70	2·42	2·02	1·68	1·30	0·851	0·681	0·529	0·255	0·126
60	2·66	2·39	2·00	1·67	1·30	0·848	0·679	0·527	0·254	0·126
120	2·62	2·36	1·98	1·66	1·29	0·845	0·677	0·526	0·254	0·126
∞	2·58	2·33	1·96	1·645	1·28	0·842	0·674	0·524	0·253	0·126

Appendix C: F-test Values for different degree of freedoms (df)

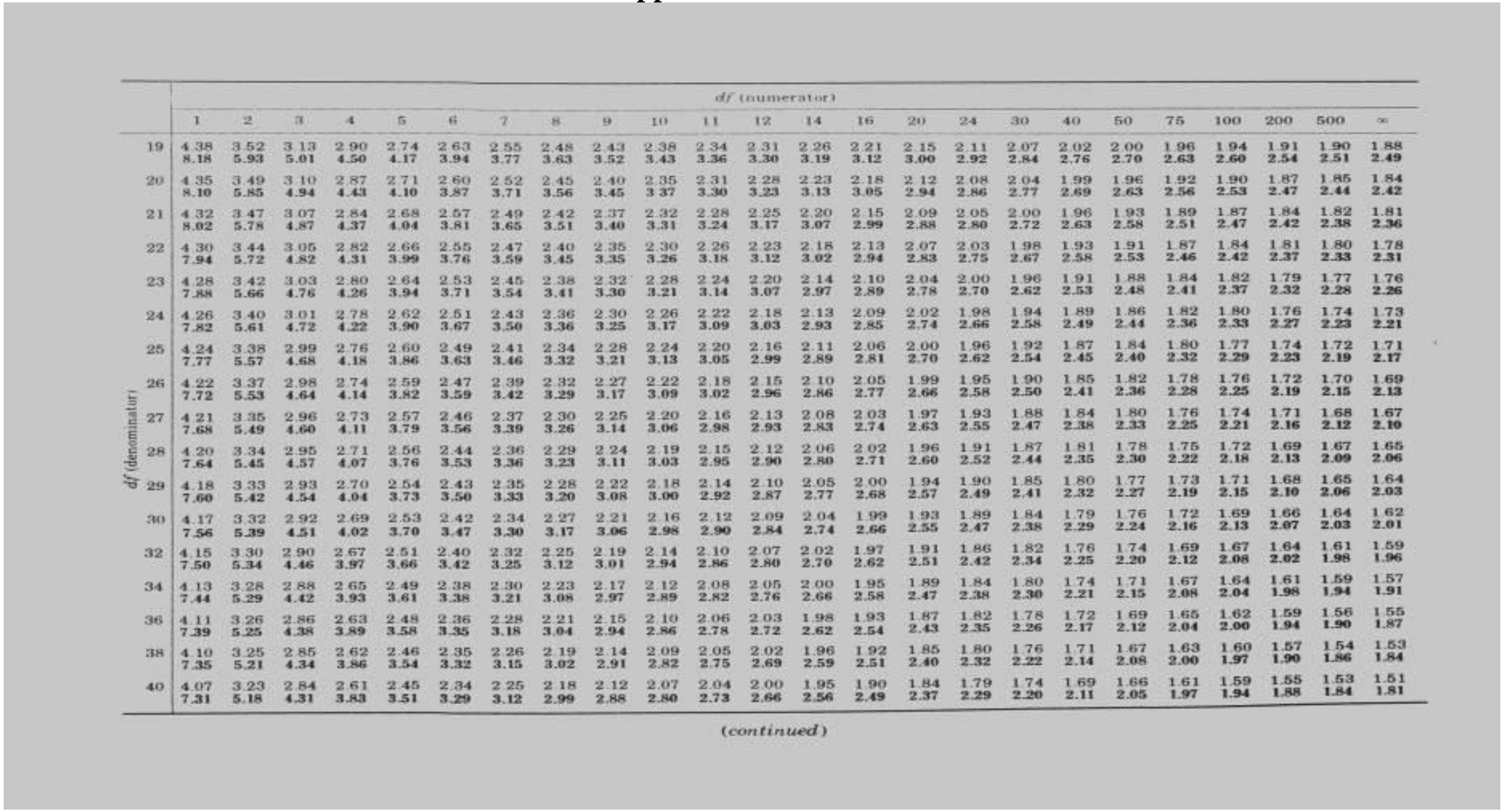


Values of *F* Exceeded with Probabilities of 5 and 1 Percent

df (denominator)	*df* (numerator)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161 4,052	200 4,999	216 5,403	225 5,625	230 5,764	234 5,859	237 5,928	239 5,981	241 6,022	242 6,056	243 6,082	244 6,106	245 6,142	246 6,169	248 6,208	249 6,234	250 6,261	251 6,286	252 6,302	253 6,323	253 6,334	254 6,352	254 6,361	254 6,366
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.36	19.37 99.37	19.38 99.39	19.39 99.40	19.40 99.41	19.41 99.42	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.29	4.78 10.15	4.74 10.05	4.70 9.96	4.68 9.89	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.03 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
7	5.59 12.25	4.74 9.55	4.34 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30	2.72 4.22	2.69 4.16	2.64 4.05	2.60 3.98	2.54 3.86	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.49	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.36
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.77 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.56 3.86	2.53 3.80	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.34	2.27 3.26	2.24 3.21	2.21 3.14	2.19 3.11	2.16 3.06	2.14 3.02	2.13 3.00
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.70 4.14	2.64 4.00	2.59 3.89	2.55 3.80	2.51 3.73	2.48 3.67	2.43 3.56	2.39 3.48	2.33 3.36	2.29 3.29	2.25 3.20	2.21 3.12	2.18 3.07	2.15 3.00	2.12 2.97	2.10 2.92	2.08 2.89	2.07 2.87
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.45 3.61	2.42 3.55	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.18	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.98	2.07 2.86	2.04 2.80	2.02 2.77	2.01 2.75
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.62 3.93	2.55 3.79	2.50 3.68	2.45 3.59	2.41 3.52	2.38 3.45	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.11 2.92	2.08 2.86	2.04 2.79	2.02 2.76	1.99 2.70	1.97 2.67	1.96 2.65
18	4.41 8.28	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	3.66 4.01	2.58 3.85	2.51 3.71	2.46 3.60	2.41 3.51	2.37 3.44	2.34 3.37	2.29 3.27	2.25 3.19	2.19 3.07	2.15 3.00	2.11 2.91	2.07 2.83	2.04 2.78	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59	1.92 2.57

(*continued*)

Appendix C: Continued



df (denominator)	*df* (numerator)																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
19	4.38 **8.18**	3.52 **5.93**	3.13 **5.01**	2.90 **4.50**	2.74 **4.17**	2.63 **3.94**	2.55 **3.77**	2.48 **3.63**	2.43 **3.52**	2.38 **3.43**	2.34 **3.36**	2.31 **3.30**	2.26 **3.19**	2.21 **3.12**	2.15 **3.00**	2.11 **2.92**	2.07 **2.84**	2.02 **2.76**	2.00 **2.70**	1.96 **2.63**	1.94 **2.60**	1.91 **2.54**	1.90 **2.51**	1.88 **2.49**
20	4.35 **8.10**	3.49 **5.85**	3.10 **4.94**	2.87 **4.43**	2.71 **4.10**	2.60 **3.87**	2.52 **3.71**	2.45 **3.56**	2.40 **3.45**	2.35 **3.37**	2.31 **3.30**	2.28 **3.23**	2.23 **3.13**	2.18 **3.05**	2.12 **2.94**	2.08 **2.86**	2.04 **2.77**	1.99 **2.69**	1.96 **2.63**	1.92 **2.56**	1.90 **2.53**	1.87 **2.47**	1.85 **2.44**	1.84 **2.42**
21	4.32 **8.02**	3.47 **5.78**	3.07 **4.87**	2.84 **4.37**	2.68 **4.04**	2.57 **3.81**	2.49 **3.65**	2.42 **3.51**	2.37 **3.40**	2.32 **3.31**	2.28 **3.24**	2.25 **3.17**	2.20 **3.07**	2.15 **2.99**	2.09 **2.88**	2.05 **2.80**	2.00 **2.72**	1.96 **2.63**	1.93 **2.58**	1.89 **2.51**	1.87 **2.47**	1.84 **2.42**	1.82 **2.38**	1.81 **2.36**
22	4.30 **7.94**	3.44 **5.72**	3.05 **4.82**	2.82 **4.31**	2.66 **3.99**	2.55 **3.76**	2.47 **3.59**	2.40 **3.45**	2.35 **3.35**	2.30 **3.26**	2.26 **3.18**	2.23 **3.12**	2.18 **3.02**	2.13 **2.94**	2.07 **2.83**	2.03 **2.75**	1.98 **2.67**	1.93 **2.58**	1.91 **2.53**	1.87 **2.46**	1.84 **2.42**	1.81 **2.37**	1.80 **2.33**	1.78 **2.31**
23	4.28 **7.88**	3.42 **5.66**	3.03 **4.76**	2.80 **4.26**	2.64 **3.94**	2.53 **3.71**	2.45 **3.54**	2.38 **3.41**	2.32 **3.30**	2.28 **3.21**	2.24 **3.14**	2.20 **3.07**	2.14 **2.97**	2.10 **2.89**	2.04 **2.78**	2.00 **2.70**	1.96 **2.62**	1.91 **2.53**	1.88 **2.48**	1.84 **2.41**	1.82 **2.37**	1.79 **2.32**	1.77 **2.28**	1.76 **2.26**
24	4.26 **7.82**	3.40 **5.61**	3.01 **4.72**	2.78 **4.22**	2.62 **3.90**	2.51 **3.67**	2.43 **3.50**	2.36 **3.36**	2.30 **3.25**	2.26 **3.17**	2.22 **3.09**	2.18 **3.03**	2.13 **2.93**	2.09 **2.85**	2.02 **2.74**	1.98 **2.66**	1.94 **2.58**	1.89 **2.49**	1.86 **2.44**	1.82 **2.36**	1.80 **2.33**	1.76 **2.27**	1.74 **2.23**	1.73 **2.21**
25	4.24 **7.77**	3.38 **5.57**	2.99 **4.68**	2.76 **4.18**	2.60 **3.86**	2.49 **3.63**	2.41 **3.46**	2.34 **3.32**	2.28 **3.21**	2.24 **3.13**	2.20 **3.05**	2.16 **2.99**	2.11 **2.89**	2.06 **2.81**	2.00 **2.70**	1.96 **2.62**	1.92 **2.54**	1.87 **2.45**	1.84 **2.40**	1.80 **2.32**	1.77 **2.29**	1.74 **2.23**	1.72 **2.19**	1.71 **2.17**
26	4.22 **7.72**	3.37 **5.53**	2.98 **4.64**	2.74 **4.14**	2.59 **3.82**	2.47 **3.59**	2.39 **3.42**	2.32 **3.29**	2.27 **3.17**	2.22 **3.09**	2.18 **3.02**	2.15 **2.96**	2.10 **2.86**	2.05 **2.77**	1.99 **2.66**	1.95 **2.58**	1.90 **2.50**	1.85 **2.41**	1.82 **2.36**	1.78 **2.28**	1.76 **2.25**	1.72 **2.19**	1.70 **2.15**	1.69 **2.13**
27	4.21 **7.68**	3.35 **5.49**	2.96 **4.60**	2.73 **4.11**	2.57 **3.79**	2.46 **3.56**	2.37 **3.39**	2.30 **3.26**	2.25 **3.14**	2.20 **3.06**	2.16 **2.98**	2.13 **2.93**	2.08 **2.83**	2.03 **2.74**	1.97 **2.63**	1.93 **2.55**	1.88 **2.47**	1.84 **2.38**	1.80 **2.33**	1.76 **2.25**	1.74 **2.21**	1.71 **2.16**	1.68 **2.12**	1.67 **2.10**
28	4.20 **7.64**	3.34 **5.45**	2.95 **4.57**	2.71 **4.07**	2.56 **3.76**	2.44 **3.53**	2.36 **3.36**	2.29 **3.23**	2.24 **3.11**	2.19 **3.03**	2.15 **2.95**	2.12 **2.90**	2.06 **2.80**	2.02 **2.71**	1.96 **2.60**	1.91 **2.52**	1.87 **2.44**	1.81 **2.35**	1.78 **2.30**	1.75 **2.22**	1.72 **2.18**	1.69 **2.13**	1.67 **2.09**	1.65 **2.06**
29	4.18 **7.60**	3.33 **5.42**	2.93 **4.54**	2.70 **4.04**	2.54 **3.73**	2.43 **3.50**	2.35 **3.33**	2.28 **3.20**	2.22 **3.08**	2.18 **3.00**	2.14 **2.92**	2.10 **2.87**	2.05 **2.77**	2.00 **2.68**	1.94 **2.57**	1.90 **2.49**	1.85 **2.41**	1.80 **2.32**	1.77 **2.27**	1.73 **2.19**	1.71 **2.15**	1.68 **2.10**	1.65 **2.06**	1.64 **2.03**
30	4.17 **7.56**	3.32 **5.39**	2.92 **4.51**	2.69 **4.02**	2.53 **3.70**	2.42 **3.47**	2.34 **3.30**	2.27 **3.17**	2.21 **3.06**	2.16 **2.98**	2.12 **2.90**	2.09 **2.84**	2.04 **2.74**	1.99 **2.66**	1.93 **2.55**	1.89 **2.47**	1.84 **2.38**	1.79 **2.29**	1.76 **2.24**	1.72 **2.16**	1.69 **2.13**	1.66 **2.07**	1.64 **2.03**	1.62 **2.01**
32	4.15 **7.50**	3.30 **5.34**	2.90 **4.46**	2.67 **3.97**	2.51 **3.66**	2.40 **3.42**	2.32 **3.25**	2.25 **3.12**	2.19 **3.01**	2.14 **2.94**	2.10 **2.86**	2.07 **2.80**	2.02 **2.70**	1.97 **2.62**	1.91 **2.51**	1.86 **2.42**	1.82 **2.34**	1.76 **2.25**	1.74 **2.20**	1.69 **2.12**	1.67 **2.08**	1.64 **2.02**	1.61 **1.98**	1.59 **1.96**
34	4.13 **7.44**	3.28 **5.29**	2.88 **4.42**	2.65 **3.93**	2.49 **3.61**	2.38 **3.38**	2.30 **3.21**	2.23 **3.08**	2.17 **2.97**	2.12 **2.89**	2.08 **2.82**	2.05 **2.76**	2.00 **2.66**	1.95 **2.58**	1.89 **2.47**	1.84 **2.38**	1.80 **2.30**	1.74 **2.21**	1.71 **2.15**	1.67 **2.08**	1.64 **2.04**	1.61 **1.98**	1.59 **1.94**	1.57 **1.91**
36	4.11 **7.39**	3.26 **5.25**	2.86 **4.38**	2.63 **3.89**	2.48 **3.58**	2.36 **3.35**	2.28 **3.18**	2.21 **3.04**	2.15 **2.94**	2.10 **2.86**	2.06 **2.78**	2.03 **2.72**	1.98 **2.62**	1.93 **2.54**	1.87 **2.43**	1.82 **2.35**	1.78 **2.26**	1.72 **2.17**	1.69 **2.12**	1.65 **2.04**	1.62 **2.00**	1.59 **1.94**	1.56 **1.90**	1.55 **1.87**
38	4.10 **7.35**	3.25 **5.21**	2.85 **4.34**	2.62 **3.86**	2.46 **3.54**	2.35 **3.32**	2.26 **3.15**	2.19 **3.02**	2.14 **2.91**	2.09 **2.82**	2.05 **2.75**	2.02 **2.69**	1.96 **2.59**	1.92 **2.51**	1.85 **2.40**	1.80 **2.32**	1.76 **2.22**	1.71 **2.14**	1.67 **2.08**	1.63 **2.00**	1.60 **1.97**	1.57 **1.90**	1.54 **1.86**	1.53 **1.84**
40	4.07 **7.31**	3.23 **5.18**	2.84 **4.31**	2.61 **3.83**	2.45 **3.51**	2.34 **3.29**	2.25 **3.12**	2.18 **2.99**	2.12 **2.88**	2.07 **2.80**	2.04 **2.73**	2.00 **2.66**	1.95 **2.56**	1.90 **2.49**	1.84 **2.37**	1.79 **2.29**	1.74 **2.20**	1.69 **2.11**	1.66 **2.05**	1.61 **1.97**	1.59 **1.94**	1.55 **1.88**	1.53 **1.84**	1.51 **1.81**

(*continued*)

Appendix C: Continued

	df (numerator)																							
df (denominator)	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
42	4.07 **7.27**	3.22 **5.15**	2.83 **4.29**	2.59 **3.80**	2.44 **3.49**	2.32 **3.26**	2.24 **3.10**	2.17 **2.96**	2.11 **2.86**	2.06 **2.77**	2.02 **2.70**	1.99 **2.64**	1.94 **2.54**	1.89 **2.46**	1.82 **2.35**	1.78 **2.26**	1.73 **2.17**	1.68 **2.08**	1.64 **2.02**	1.60 **1.94**	1.57 **1.91**	1.54 **1.85**	1.51 **1.80**	1.49 **1.78**
44	4.06 **7.24**	3.21 **5.12**	2.82 **4.26**	2.58 **3.78**	2.43 **3.46**	2.31 **3.24**	2.23 **3.07**	2.16 **2.94**	2.10 **2.84**	2.05 **2.75**	2.01 **2.68**	1.98 **2.62**	1.92 **2.52**	1.88 **2.44**	1.81 **2.32**	1.76 **2.24**	1.72 **2.15**	1.66 **2.06**	1.63 **2.00**	1.58 **1.92**	1.56 **1.88**	1.52 **1.82**	1.50 **1.78**	1.48 **1.75**
46	4.05 **7.21**	3.20 **5.10**	2.81 **4.24**	2.57 **3.76**	2.42 **3.44**	2.30 **3.22**	2.22 **3.05**	2.14 **2.92**	2.09 **2.82**	2.04 **2.73**	2.00 **2.66**	1.97 **2.60**	1.91 **2.50**	1.87 **2.42**	1.80 **2.30**	1.75 **2.22**	1.71 **2.13**	1.65 **2.04**	1.62 **1.98**	1.57 **1.90**	1.54 **1.86**	1.51 **1.80**	1.48 **1.76**	1.46 **1.72**
48	4.04 **7.19**	3.19 **5.08**	2.80 **4.22**	2.56 **3.74**	2.41 **3.42**	2.30 **3.20**	2.21 **3.04**	2.14 **2.90**	2.08 **2.80**	2.03 **2.71**	1.99 **2.64**	1.96 **2.58**	1.90 **2.48**	1.86 **2.40**	1.79 **2.28**	1.74 **2.20**	1.70 **2.11**	1.64 **2.02**	1.61 **1.96**	1.56 **1.88**	1.53 **1.84**	1.50 **1.78**	1.47 **1.73**	1.45 **1.70**
50	4.03 **7.17**	3.18 **5.06**	2.79 **4.20**	2.56 **3.72**	2.40 **3.41**	2.29 **3.18**	2.20 **3.02**	2.13 **2.88**	2.07 **2.78**	2.02 **2.70**	1.98 **2.62**	1.95 **2.56**	1.90 **2.46**	1.85 **2.39**	1.78 **2.26**	1.74 **2.18**	1.69 **2.10**	1.63 **2.00**	1.60 **1.94**	1.55 **1.86**	1.52 **1.82**	1.48 **1.76**	1.46 **1.71**	1.44 **1.68**
55	4.02 **7.12**	3.17 **5.01**	2.78 **4.16**	2.54 **3.68**	2.38 **3.37**	2.27 **3.15**	2.18 **2.98**	2.11 **2.85**	2.05 **2.75**	2.00 **2.66**	1.97 **2.59**	1.93 **2.53**	1.88 **2.43**	1.83 **2.35**	1.76 **2.23**	1.72 **2.15**	1.67 **2.06**	1.61 **1.96**	1.58 **1.90**	1.52 **1.82**	1.50 **1.78**	1.46 **1.71**	1.43 **1.66**	1.41 **1.64**
60	4.00 **7.08**	3.15 **4.98**	2.76 **4.13**	2.52 **3.65**	2.37 **3.34**	2.25 **3.12**	2.17 **2.95**	2.10 **2.82**	2.04 **2.72**	1.99 **2.63**	1.95 **2.56**	1.92 **2.50**	1.86 **2.40**	1.81 **2.32**	1.75 **2.20**	1.70 **2.12**	1.65 **2.03**	1.59 **1.93**	1.56 **1.87**	1.50 **1.79**	1.48 **1.74**	1.44 **1.68**	1.41 **1.63**	1.39 **1.60**
65	3.99 **7.04**	3.14 **4.95**	2.75 **4.10**	2.51 **3.62**	2.36 **3.31**	2.24 **3.09**	2.15 **2.93**	2.08 **2.79**	2.02 **2.70**	1.98 **2.61**	1.94 **2.54**	1.90 **2.47**	1.85 **2.37**	1.80 **2.30**	1.73 **2.18**	1.68 **2.09**	1.63 **2.00**	1.57 **1.90**	1.54 **1.84**	1.49 **1.76**	1.46 **1.71**	1.42 **1.64**	1.39 **1.60**	1.37 **1.56**
70	3.98 **7.01**	3.13 **4.92**	2.74 **4.08**	2.50 **3.60**	2.35 **3.29**	2.23 **3.07**	2.14 **2.91**	2.07 **2.77**	2.01 **2.67**	1.97 **2.59**	1.93 **2.51**	1.89 **2.45**	1.84 **2.35**	1.79 **2.28**	1.72 **2.15**	1.67 **2.07**	1.62 **1.98**	1.56 **1.88**	1.53 **1.82**	1.47 **1.74**	1.45 **1.69**	1.40 **1.62**	1.37 **1.56**	1.35 **1.53**
80	3.96 **6.96**	3.11 **4.88**	2.72 **4.04**	2.48 **3.56**	2.33 **3.25**	2.21 **3.04**	2.12 **2.87**	2.05 **2.74**	1.99 **2.64**	1.95 **2.55**	1.91 **2.48**	1.88 **2.41**	1.82 **2.32**	1.77 **2.24**	1.70 **2.11**	1.65 **2.03**	1.60 **1.94**	1.54 **1.84**	1.51 **1.78**	1.45 **1.70**	1.42 **1.65**	1.38 **1.57**	1.35 **1.52**	1.32 **1.49**
100	3.94 **6.90**	3.09 **4.82**	2.70 **3.98**	2.46 **3.51**	2.30 **3.20**	2.19 **2.99**	2.10 **2.82**	2.03 **2.69**	1.97 **2.59**	1.92 **2.51**	1.88 **2.43**	1.85 **2.36**	1.79 **2.26**	1.75 **2.19**	1.68 **2.06**	1.63 **1.98**	1.57 **1.89**	1.51 **1.79**	1.48 **1.73**	1.42 **1.64**	1.39 **1.59**	1.34 **1.51**	1.30 **1.46**	1.28 **1.43**
125	3.92 **6.84**	3.07 **4.78**	2.68 **3.94**	2.44 **3.47**	2.29 **3.17**	2.17 **2.95**	2.08 **2.79**	2.01 **2.65**	1.95 **2.56**	1.90 **2.47**	1.86 **2.40**	1.83 **2.33**	1.77 **2.23**	1.72 **2.15**	1.65 **2.03**	1.60 **1.94**	1.55 **1.85**	1.49 **1.75**	1.45 **1.68**	1.39 **1.59**	1.36 **1.54**	1.31 **1.46**	1.27 **1.40**	1.25 **1.37**
150	3.91 **6.81**	3.06 **4.75**	2.67 **3.91**	2.43 **3.44**	2.27 **3.14**	2.16 **2.92**	2.07 **2.76**	2.00 **2.62**	1.94 **2.53**	1.89 **2.44**	1.85 **2.37**	1.82 **2.30**	1.76 **2.20**	1.71 **2.12**	1.64 **2.00**	1.59 **1.91**	1.54 **1.83**	1.47 **1.72**	1.44 **1.66**	1.37 **1.56**	1.34 **1.51**	1.29 **1.43**	1.25 **1.37**	1.22 **1.33**
200	3.89 **6.76**	3.04 **4.71**	2.65 **3.88**	2.41 **3.41**	2.26 **3.11**	2.14 **2.90**	2.05 **2.73**	1.98 **2.60**	1.92 **2.50**	1.87 **2.41**	1.83 **2.34**	1.80 **2.28**	1.74 **2.17**	1.69 **2.09**	1.62 **1.97**	1.57 **1.88**	1.52 **1.79**	1.45 **1.69**	1.42 **1.62**	1.35 **1.53**	1.32 **1.48**	1.26 **1.39**	1.22 **1.33**	1.19 **1.28**
400	3.86 **6.70**	3.02 **4.66**	2.62 **3.83**	2.39 **3.36**	2.23 **3.06**	2.12 **2.85**	2.03 **2.69**	1.96 **2.55**	1.90 **2.46**	1.85 **2.37**	1.91 **2.29**	1.78 **2.23**	1.72 **2.12**	1.67 **2.04**	1.60 **1.92**	1.54 **1.84**	1.49 **1.74**	1.42 **1.64**	1.38 **1.57**	1.32 **1.47**	1.28 **1.42**	1.22 **1.32**	1.16 **1.24**	1.13 **1.19**
1000	3.85 **6.66**	3.00 **4.62**	2.61 **3.80**	2.38 **3.34**	2.22 **3.04**	2.10 **2.82**	2.02 **2.66**	1.95 **2.53**	1.89 **2.43**	1.84 **2.34**	1.80 **2.26**	1.76 **2.20**	1.70 **2.09**	1.65 **2.01**	1.58 **1.89**	1.53 **1.81**	1.47 **1.71**	1.41 **1.61**	1.36 **1.54**	1.30 **1.44**	1.26 **1.38**	1.19 **1.28**	1.13 **1.19**	1.08 **1.11**
∞	3.84 **6.64**	2.99 **4.60**	2.60 **3.78**	2.37 **3.32**	2.21 **3.02**	2.09 **2.80**	2.01 **2.64**	1.94 **2.51**	1.88 **2.41**	1.83 **2.32**	1.79 **2.24**	1.75 **2.18**	1.69 **2.07**	1.64 **1.99**	1.57 **1.87**	1.52 **1.79**	1.46 **1.69**	1.40 **1.59**	1.35 **1.52**	1.28 **1.41**	1.24 **1.36**	1.17 **1.25**	1.11 **1.15**	1.00 **1.00**