MODELING WEEKLY RAINFALL IN COLOMO CITY

Hondaarachchi Patabendige Thanuja Nilanthi Silva

(168001A)

Thesis submitted in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy

Department of Mathematics

University of Moratuwa
Sri Lanka

April 2020

DECLARATION

ñI declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.ò

ñAlso, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).ò

Signature: Date: 28.05.2020

The above candidate has carried out research for the PhD thesis under my supervision.

Name of the supervisor: Senior Professor T. S. G. Peiris

Signature of the supervisor: Date: 28.05.2020

ABSTRACT

Modeling weekly rainfall has become a demanding assignment due to the complexity of rainfall pattern. Accurate inferences on weekly rainfall prediction facilitate to fill the noticeable gap with respect to the climate monitoring to reduce the climate stress in the country. However, relatively, few measures have been taken to perform the modeling of rainfall in the context of long memory. This study therefore, provides an assessment of such a phenomenon by fitting a novel time series models to weekly rainfall. As the weekly rainfall exhibits the blend features of long memory and time dependence variance, various class of long memory models were fitted by accounting the heteroskedasticity. The best fitted model developed is ARFIMA-GARCH for deseasonalized data. The model was trained using weekly rainfall data from 1990 to 2014 and validated using data from 2015 to 2017 in Colombo city, obtained from the Department of Meteorology, Sri Lanka. The exact maximum likelihood estimation method was utilized to estimate model parameters. For the evaluation of the suitability of the method for parameter estimation, Monte Carlo simulations were carried out with various non seasonally and seasonally fractionally differenced parameter values along with the variance model parameters. The forecasting performance of the five types of long memory models developed was evaluated based on the novel index developed using absolute error for an independent data set in addition to the classical indicators. The rainfall percentiles with the 95% confidence intervals were also developed by exploring temporal variability of weekly rainfall based on parametric approach and bootstrapping approach. It was found that the high likelihood to form extreme rainfall events during beginning of South West Monson (SWM) (30^{th} April to 10^{th} June) and during withdrawal of SWM rainfall (17^{th}-30^{th} September) as well as with the time span from 8^{th} October to 11^{th} November during Second Inter Monsoon (SIM) rainfall. Based on the real coverage probabilities which derived using bootstrap calibration, it was found that there is a discrepancy of the nominal and calculated coverage probabilities of the 95% confidence intervals of rainfall percentiles. The deviation of the normality of the fitted distribution with the small size of sample could be a reason for the such a disparity. The novel long range dependency model is recommended to be used in forecasting weekly rainfall in Colombo city in Sri Lanka since the forecasting performance of the new model is not much diluted with the increase of the forecasting length. The study highlights various challenges for applied statisticians in modeling weekly rainfall.

Keywords: Weekly Rainfall, Long Range Dependency, ARFIMA-GARCH, Forecasting, Coverage Probability, South West Monsoon

DEDICATION

To My Mother, Father, Husband and Lovely Son

ACKNOWLEDGEMENTS

I would like to convey my sincere gratitude to all the eminent people who helped me to complete this research successfully. Firstly, and foremost, I express my heartiest thanks to my supervisor, Prof. T. S. G. Peiris, Senior Professor in Applied Statistics and former Head of the Department of Mathematics, Faculty of Engineering, University of Moratuwa, whose immense knowledge, experience and valuable guidance helped me greatly to complete this research successfully. My sincere appreciation to Prof. S. Samitha, Department of Crop Science, University of Peradeniya who helped me to initiate the research.

I am much obligated to acknowledge gratefully the support extended by the University research grant committee and leave committee, University of Sri Jayewardenepura, Sri Lanka for offering me a research grant (Grant No. ASP/01/RE/HSS/2016/75) and study leave which facilitated the successful completion of this research.

I would like to express my immense thanks to Dr. G. Sanjaya Dissanayake, and Prof. Shelton Peiris, School of Mathematics and Statistics, University of Sydney, Australia for their valuable advice, guidance and inspiration for the successful achievement of my research. I would also like to express my thanks to Prof. Thomas Mathew, University of Maryland Baltimore Country (UMBC), USA for providing his valuable advice and suggestions related to my research during his short visit to the Department of Mathematics, University of Moratuwa.

My heartiest gratitude to the members of my progress review committee, Prof. (Mrs.) N.R. Abeynayake and Dr. (Mrs.) S.C. Mathugama for their valuable suggestions and constructive comments to improve the quality of my research. Also, I would like to thank to Dr. P.M. Edirisinghe, research Coordinator in the Department of Mathematics, University of Moratuwa for giving vital advice and suggestions related to my research.

I would like to express my thanks to Dr. Duminda Kuruppuarachchi, University of Otago, New Zealand for giving great support to me for the successful achievement of my research and my thank also goes to Mr. K. H. M. S. Premalal, a former Director General of the Department of Meteorology, Sri Lanka for giving advice and suggestions to enhance the quality of my research.

I am greatly indebted to Mrs. K. A. D. S. A. Nanayakkara who always encouraged me throughout the period of research by sharing her knowledge and providing her valuable suggestions related to the research. I would like to express my thanks to Mr. J. A. A. U.Weerasena for giving great support to me to success of my research journey and I would like to express my sincere thanks to, Mrs. R. M. K. G. U. Rathnayake and Mrs. P. A. C. P. P. Arachchi, who encouraged me to initialize the study. I would also like to thank all the academic and non-academic members of the Department of Social Statistics, University of Sri Jayewardenapura and all the academic and non-academic members of the Department of Mathematics, University of Moratuwa for their assistance throughout my study.

Last but not the least, I am ever grateful to my mother, father, my husband and my lovely son for their support and encouragement in achieving my goals throughout the period of my research. Their good deeds, indeed, a source of inspiration and a hall mark which gave me motivation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Description
ACF	Auto Correlation Function
AIC	Alkina Information Criterion
ANN	Artificial Neural Network
ARCH	Autoregressive Conditional Heteroskedasticity
ARFIMA	Autoregressive Fractionally Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CV	Coefficient of Variation
FIM	First Inter Monsoon
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAX	Maximum
MFE	Mean Forecast Error
MIN	Minimum
MLE	Maximum Likelihood Estimators
MLR	Multiple Linear Regression
MRE	Mean Relative Error
MSE	Mean Square Error
NEM	North East Monsoon
PACF	Partial Auto Correlation Function
RMSE	Root Mean Square Error
SARFIMA	Seasonal Autoregressive Fractionally Integrated Moving Average
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SD	Standard Deviation
SE	Standard Error
SIM	Second Inter Monsoon

SWM	South West Monsoon
TMIN	Minimum Temperature
VAR	Vector Auto Regression

CHAPTER 1

INTRODUCTION

1.1. Background

Climate change leads to extreme weather conditions which greatly affect the diverse set of human and natural systems in the world. The Intergovernmental Panel on Climate Change (IPCC) defines climate change as "any change in climate over time, whether due to natural variability or as a result of human activity". Observational evidence indicated that the climate change has significantly affected in many countries at different levels which causes a serious threat to sustainable development.

According to the IPCC report in 2014, many key changes arose on various climatic variables, during the period of 1880 to 2012; the average combined land and ocean surface temperature went up by of 0.85^{0}C, mean sea level rose by 19 cm over the period of 1901 to 2010, green gas concentration has increased now higher than ever and the number of heavy precipitation events has increased in many regions. Though the climate variables are components of natural capital, on the other hand, those are the key factors that make severe impact on people wellbeing, economy of the country, environment and social stability. Impacts of climate variability depend on the intensity of the events which affect the global community at different level. Based on the projected changes in the system, IPCC has highlighted alarming trends in changes in global climate and has emphasized the importance of the prediction of climate variables, particularly precipitation in different time scales (IPCC, 2014).

Rainfall, snowfall and other forms of frozen or liquid water falling from clouds are generally known as "Precipitation" (Trenberth, 2005). Solomon et al., (2007) claimed that the number of heavy precipitation events cause to increase likelihood of flooding events in many regions even those where there has been a reduction in total precipitation. According to Dai (2006), the biggest impact on the society may occur due to the changes in precipitation patterns and its variability. Information on key climatic variable predictions allows to various stakeholders to prompt themselves for action in order to reduce adverse impacts and enhance positive effects of climatic variation. Numerous studies have been carried out on climate change and its repercussions, in particularly significant trends in precipitation and temperature (Portmann et al., 2009; Reiter et al., 2012; Shi et al., 2014; Nam et al., 2016). At the local level, particularly in developing regions, there is a need for better information on rainfall patterns and its variability through accurate predictions which help to prepare adaptation in advance.

1.2. Climate Change in Sri Lanka

South Asian countries are frequently vulnerable to extreme whether events and people who live in those regions face a huge challenge to deal with the impact of climate change due to the high population density, poverty and lack of resources to confront climate stress. (Ahmed and Suphachalasai, 2014). Sri Lanka is also a tropical country in South Asian region located at the latitudes of $5^{0}55$ N and 9^{0} 51 N and the longitudes of $79^{0}41$E and $81^{0}53$ E with an area of 65610 square kilometers and the country frequently exposes to erratic weather events.

There is sufficient evidence to claim that climate, in mostly rainfall pattern in Sri Lanka has already changed over the past years (Peiris et al., 2000; Waidyarathne et al., 2006; Manawadu and Fernando, 2008). This change resulted in substantial difference in the atmospheric behavior which inflicts serious consequences on human wellbeing. The mean air temperature of the country increased by 0.016^{0}C during the time span from 1961 to 1990 while mean annual rainfall decreased by 144 mm (Eriyagama et al., 2010). It was revealed that a significant climate change in rainfall and temperature of low country wet intermediate region of Sri Lanka. The percentage reduction in the mean annual rainfall during 1986 to 2001 compared to the period 1932 to 1985 was 9% while the corresponding value for the mean annual maximum temperature was 1.4% (Piyasiri et al., 2004). Manawadu and Fernando (2008) claimed that the number of rainy days in Sri Lanka decreased based on the result of analyzing the spatio-temporal trends in the rainfall using the daily rainfall records at the 22 meteorological stations during the period 1961-2002. Jayawardene et al., (2005) have highlighted that the rate of annual rainfall has increased significantly by

3.15 mm/year in Colombo district while it decreased by 4.87 mm/year and 2.88 mm/year in Nuwara Elliya and Kandy districts respectively.

As in other countries, climate vulnerabilities are expected to be critical in Sri Lanka in the various sectors such as agriculture, fisheries, water, health, urban development, human settlement, economic infrastructure, biodiversity and ecosystem in the country (Mawilmada et al., 2010). Few studies have been carried out to project future climate scenarios with respect to the rainfall in Sri Lanka to assess the impact on agricultural output, economy and water resource of the country. Based on the studies on rainfall projections, Basnayake et al., (2004) highlighted that there is a decreasing trend in mean annual rainfall while De Silva (2006) claimed that the increasing trend in mean annual rainfall. The changes of delay in monsoon onset and an increase in the occurrences of monsoon break period which are caused by enhanced greenhouse emission that could have a substantial impact on decreasing summer precipitation in key areas of South Asia (Ashfaq et al., 2009). The literature on rainfall projections demonstrate the necessity of prediction of rainfall by seasonal basis to obtain accurate rainfall for the country.

1.3. Annual Rainfall Pattern in Sri Lanka

Sri Lanka is the one of the tropical countries in South Asian region that receives rainfall throughout the year. The mean annual rainfall of the direst part of the country is under 900 mm while it is around over 5000 mm in the wettest part of the country *(Source: www.meteo.gov.lk).* Generally, the annual rainfall pattern in many parts of Sri Lanka is bimodal and rainy periods have been classified into four seasons by Domroes (1974).

The four seasons are:

1. First Inter Monsoon (FIM) from March to April
2. South West Monsoon (SWM) from May to September
3. Second Inter Monsoon (SIM) from October to November
4. North East Monsoon (NEM) from December to February

The rainfall of the country is strongly governed by the seasonal varying monsoon system. With the mean annual rainfall 1861mm, 60% of rainfall is received during SWM and SIM while 14%, and 26% rainfall is received during FIM and NEM respectively (Premalal, 2013). The southwestern part of the country receives rainfall at any time of the day during southwest monsoon seasons and the amount of the rainfall varies from 100mm to over 3000mm. During the period of NEM, the dry and the cold wind blowing from the Indian land occurs result in cool but dry weather over many parts of the island and the rainfall amount varies from 177mm to 1281mm.

FIM rains mostly spread in South-Western region and the rainfall amount varies from 250mm to 700mm. It is particularly observed that the thunderstorm type rain during the afternoon or evening. SIM enriches from wide spread rain with strong winds that sometimes leads to floods or landslides. Rainfall in SIM furnishes balance distribution over Sri Lanka and the rainfall amount varies from 750mm to 1200mm *(Source: www.meteo.gov.lk).* Various studies have been conducted on the onset of four rainy seasons and the length of spells in the four seasons (Ramesh et al., 1996; Peiris et al., 2000; Omotosho et al., 2000; Goswami and Gouda, 2007). However, the result of those studies are highly varied.

The large-scale climatic drivers also contribute considerably to rainfall variability in the Sri Lanka. The two main large-scale climatic drivers that influence the rainfall pattern of Sri Lanka are, the Southern Oscillation (SO) and the Indian Ocean Dipole (Zubair, 2002; Zubair et al., 2003; Zubair and Ropelewski, 2006). El Nińo and La Nińa are generated due to the changes in the winds, atmospheric pressure and sea water in the Pacific Ocean. El Nińo and La Nińa are the opposite phases such that El Nińo leads to wetter conditions during October to December and drier conditions during January to March and July to August on average (Zubair et al., 2008). Those two climate drivers are extreme weather conditions such that El Nińo being the warm extreme and La Nińa the cold extreme and those do not change with the regularity of the seasons such as winter and summer, however, they might be recur on average about every three or four years (Permalal, 2013). Malmgren et al., (2003) have claimed that the no change was observed NEM rainfall with respect to the El Nińo Southern Oscillation (ENSO). Also, they reported that the no similar influence from ENSO is seen for the two seasons NEM and FIM while there were differences in SWM rainfall pattern with respect to the ENSO climate divers. Furthermore, it is important to note that rainfall is the main erratic variable in tropical countries like Sri Lanka.

1.4. Impact of Unpredicted Rainfall

Rainfall is the one of the most important climatic variables in planning and decision making in the agricultural sector, particularly in those regions where livelihood of the people concerned depend on rain fed agriculture. According to Jayewardene et al., (2005), the 22% of the agricultural exports and 75% of the industrial exports use electricity which 62% is generated through hydropower.

Rice cultivation plays an imperative role in Sri Lanka as most of the other countries in Asia. It is projected that by 2050, the majority of paddy growing areas in Sri Lanka will be faced to water related issues particularly during Maha season and as a result, paddy cultivation in those regions will become low down (De Silva et al., 2007). Furthermore, Amarsinghe et al., (2015) projected that the irrigation requirement increased by 10-17% based on the late onset of the rainfall. According to the prediction of climate change for the 2050, De Silva et al., (2007) claimed that the paddy irrigation water requirement will be increased by 23% due to the average rainfall decreased by 17% during the wet season. Paddy farming output falling by 20%-30% in the next 20 to 30 years due to the erratic weather conditions will result in such a negative impact on agriculture employment in this section which has a 35% of the working population in Sri Lanka (Baba, 2010).

Declining rainfall is a serious threat to the tea industry and it is estimated that the tea yield would be reduced by 30-80 kg/Ha with a reduction in monthly rainfall by 100mm (Wijeratne et al., 2007). It was found that the variability of coconut production mainly depends on the two key factors, changes in monsoon rainfall and increases in maximum air temperature. Based on the result of the projection under six different climate scenarios, Peiris et al., (2004) claimed that the coconut production would not be sufficient to cater the local demand after 2040 when other external factors are non-limiting. According to Fernando et al., (2007), the loss of income to the economy with respect to coconut production was between US$ 32 million to US$ 73 due to low rainfall which is caused by climate change.

Changes in rainfall pattern cause to increase the likelihood of short and long run crop production decline which leads to a food insecurity in the country. The shifts in the monsoonal rainfall pattern due to global warming alarming to South Asian with respect to the food production and Sri Lanka is predicted to be one of the countries that faces risk of food insecurity in the Asian Pacific region (De Zoysa and Inoue, 2014). Thus, the change of pattern and the quantity of rainfall most of the time create a serious threat to the sustainable development of countries at different levels.

Sri Lanka is also under stress to face such climate changes which result in extreme weather conditions, particularly in rainfall events. Rainfall in the Wet Zone is mostly, intensive resulting flooding, landslides, soil erosion, damage to properties and infrastructures. Lo and Koralegedera (2015) claimed that the cities including Colombo in Sri Lanka would be seriously faced to water related issues due to changes in rainfall patterns, urbanization and installation of complex infrastructure. Furthermore, they reported that the very heavy rainfall may occur in future in the Colombo city. Sri Lanka has witnessed a number of extreme rainfall events in South Western region during the south west monsoon season. The most recently, a flood event in Sri Lanka was reported in May 2016 and rainfall varied between 74.7mm to 137.7mm. Sri Lanka was hit by severe tropical storms that caused widespread flooding and landslides in 22 districts of the country (OCHA Report, 2016). Each year the government of Sri Lanka spends huge amount to reconstruct and renovate the infrastructures which damage caused by floods in the wet zone.

Changing in trends of rainfall can potentially increase the transmission of mosquito vector born disease, such as Dengue in many districts in the country (Pathirana et al., 2009). According to Najim et al., (2012), the coastal and marine resources in South

Asian region including Sri Lanka have been affected seriously due to the changes of climate events. Also, the total climate change cost in South Asia based on the economic findings using the integrated assessment model, will increase in the long term (Ahmed and Suphachalasai, 2014). Some studies have been carried out to highlight the rainfall impact on the country based on the past data, but none of the studies were reported to predict rainfall in the short-term basis.

The studies conducted so far with respect to the rainfall in Sri Lanka indicates that the properties of rainfall including rainfall trend, amount and intensity have changed (Malmgren et al., 2003; Jayawardena et al., 2005; Waidyaratne and Peiris, 2006; Wickramagamage, 2010; Mathugama and Peiris, 2011 & 2012) but very few attempts were made to predict the amount of rainfall either on short-term or a long term basis. Though many authors have used different models to predict rainfall on annual, seasonal or monthly basis, either on agro-ecological or district basis (Soltani et al., 2007; Kaushik and Singh, 2008; Nirmala and Sundaram, 2010; Ghalhari et al., 2015) there are many drawbacks with respect to the statistical and non statistical aspects. Nevertheless, extremely very few studies have been reported in forecasting weekly rainfall (Burt and Weerasinghe, 2014).

1.5. Motivation to the Study

Sri Lanka is an agricultural country and its main energy source is hydropower. Time or quantity variations in rainfall could have directly affected on agricultural output which would be cased to sever damage to the Gross National Product in the country. Prior knowledge of short-range raining behavior will help Sri Lankan farmers to take advantage of rainfall by having proper water management practices which cause to maximize the crop harvest and minimize the human hardship during erratic rainfalls. Due to the fact that Sri Lanka is situated in a tropical location, the high variability of weather conditions can be formed which causes unexpected heavy rain, floods, lighting, landslides and high winds. Moreover, Sri Lankan economy is directly linked with aviation and shipping which mainly depend on exports; tea, rubber, coconut, minor agricultural products, apparels and tourism (Rathnayake et al., 2011). Thus, the advance knowledge of short range, rainfall is valuable to many fields.

Furthermore, advance knowledge of short-term rainfall amount will assist early prevention and control as well as future preparation regarding public health issues. Information on rainfall could be used in decision making with respect to the spread of diseases and pets (Dhiman et al., 2010). Information on short range forecasting in rainfall is utilized in the construction field which contributes to the development of the country. Most of the time, people who engaged in the construction field need short range forecast information to plan their activities to get maximum benefits by reducing the unnecessary cost which arises due to unexpected rainy events. Rainfall has a strong influence on traffic and sewer system in urban areas in the country. Therefore, prior knowledge of short-range rainfall is very essential to make effective decisions and planning to prevent future obstacles.

Forecasting rainfall can also be used for climate monitoring, detecting of droughts and bad weather conditions. Warning systems specially for floods may require a quantitative rainfall forecast to increase the lead time for warning. Frequent floods and landslides are already causing extensive damages to our infrastructure in the region and threaten to urban development. Hidayah et al., (2011) highlighted that the importance of simulation of continuous rainfall in hydrological research, especially for flood estimations.

Thus, it is clear that the changes of the pattern and quantity of rainfall has a considerable impact on various sectors and human wellbeing in the country at different levels. Lack of accurate knowledge about the occurrences and the amount of rainfall has significantly affected the growth of the country directly or indirectly. Many stakeholders need to make early actions to reduce risk of climate change with accuracy prediction of climatic events, especially rainfall events. Therefore, prior knowledge is very essential to make effective decisions and planning policies to prevent future obstacles.

It has been observed that though the mean annual rainfall pattern is normal, the unusual pattern of rainfall with a short period may disturb many activities in the country such as agricultural, drinking water supply, construction, commercial and social stability during the recent past years in Sri Lanka.

Various prediction methods have been developed on annual and monthly time scales but less attention was given for weekly rainfall prediction. Nevertheless, rainfall forecasting on a weekly basis is very essential to the government, businessmen, people in the industrial sectors, to increase the productivity, maximize economic benefits and minimize losses. Also, it will be useful to policy makers to implement new policies which will help to develop the country. Sri Lanka needs to address climate change adaptation to ensure the economic development with carefully investigated information on rainfall pattern and its variability which result in from the predictions of the best fitted rainfall models in various regions.

1.6. Objectives of the Study

In view of the above explanation in details, the objectives of this study are:

- To study the temporal variability of weekly rainfall in SWM and SIM
- To identify extreme weekly rainfall events during the period of SWM and SIM
- To identify the impact of exogenous variables; temperature, relative humidity and vapor pressure on weekly rainfall
- To develop a novel model to forecast weekly rainfall in Colombo city
- To validate the model

1.7. Chapter Outline

Organization structure of the theses is presented in this section as follows. Chapter 2 will provide the comprehensive literature review on rainfall studies. Various types of models such as Box and Jenkins models, Artificial Neural Network models, Regression models, Hybrid models and long range dependency models are discussed in detail by different time scales in this chapter. Data description of the study and the theories used for modeling the rainfall in different context are given in Chapter 3.

The result of the explanatory data analysis of the rainfall at different time scales along with the impact of exogenous variables for modeling are discussed in Chapter 4. Temporal variability of SWM and SIM are explained based on parametric and bootstrapping methods in the Chapter 5. Modeling via classical time series approaches are discussed in Chapter 6. The development of the novel models in detail are discussed in Chapter 7. Chapter 8 will provide conclusions and recommendations along with suggestions for further researchers. Finally, Chapter 9 provides list of publications generated by this study.

CHAPTER 2

LITERATURE REVIEW

Rainfall is one of the most complex and difficult elements of the hydrological cycle to understand and model due to the its high variability in both space and time (French et al., 1992). However, due to the importance of rainfall, over the past decades, several models have been developed to predict the rainfall with different degrees of accuracy. In this Chapter, a critical evaluation of past studies on modeling rainfall amount is carried out with emphasis on forecasting models based on the various time scales.

2.1. Prediction of Rainfall Using ARIMA/SARIMA

Box and Jenkins (1976) time series approach have been extensively used to model and forecast total rainfall in various time scales: annual, seasonal, monthly and weekly basis by various authors. The common used models are ARIMA (Auto Regressive Integrated Moving Average) and SARIMA (Seasonal ARIMA).

2.1.1. Prediction of Annual Rainfall

Ogallo (1986) employed ARMA (3,1) to the areal annual rainfall of two homogenous regions in East Africa using rainfall records from 1922 to 1980. In order to determine the annual rainfall anomalies, composite indices which were developed through empirical orthogonal analysis. The model could be accounted for only 50% of the total observed variability. However, Ogallo suggested that the model accuracy would be improved by considering seasonal variation. Partheepan et al., (2005) developed ARIMA model to forecast the annual rainfall of Batticaloa district in Sri Lanka using rainfall data from 1900 to 2003. The data for 100 years were used for the model development and remaining 4 years were used for model validation. Parameters of the model are significance at 0.05 level of significance and the correlation coefficient between actual and fitted values of the validation period was as 0.82 ($p<0.05$). A study conducted to predict the annual rainfall of Chattisgarh State in India by Chakraborty et al., (2010) and they developed AR (1) for annual rainfall. The model was tested for 2006. The model AR (1) was decided based on the AIC value MFE, MAE, MRE and RMSE. However, the time span that used for model building was very low and also no attempt has been made to test the significance of the parameters of the model.

2.1.2. Prediction of Seasonal Rainfall

ARIMA models have been developed to model the pre monsoon rainfall (March, April, May) for six stations of the Western region in India using data from 1949 to 2009 by Narayanan et al., (2013). Based on the predicted result for the time span between 2010 to 2030, it was concluded that there would be considerable rise in the pre monsoon rainfall over the northwest part of the country. Ghalhari et al., (2015) suggested SARIMA (2,0,0) (5,1,0), SARIMA (1,0,1) (5,1,0) and SARIMA (0,0,1) (5,1,0) models to seasonal rainfall for the three stations in the South of Kerman province in Iran. The seasonal rainfall (Winter and Falls) data from 1963 to 2008 were used for the model development and the model validity was done using data for a period of 5 years (2009-2013). Even though the study claimed the existence of correlation between actual and forecasted for the independent series but it failed to give the significance levels to get the real idea about the overall accuracies of the models. It should also be noted that the prediction for 20 years ahead from such a model is not statistically sound.

2.1.3. Prediction of Monthly Rainfall

Delleur and Karvas (1978) suggested ARMA (1,1) model to the square root transformation of monthly basin average rainfall series over 15 basins located in the Midwestern United State. The record lengths varied from 492 to 684 months and the basin average rainfall were obtained by the Thiessien Polygon method. However, no attempt has been made to give overall model accuracies. Soltani et al., (2007) also developed an ARIMA model using monthly rainfall data of 28 main cities of Iran during the period of 1970 to 2000. Though they tested the parameters and model assumptions, no attempt has been made to validate the model for an independent data set. Kaushik and Singh (2008) developed SARIMA (3,0,2) $(2,0,1)_{12}$ model for prediction of rainfall on monthly scales during the period of 1994-2006. The study was conducted at Mirzapur, Uttar Pradesh in India. The study did not check the model validity using an independent data set, nevertheless, the authors claim that the accuracy of predictions made by the model was fairly less. Another drawback of this model is that replacement of missing values by zero.

Kingdom of Saudi Arabia, Momani (2009) developed seasonal ARIMA $(1,0,0)(0,1,1)_{12}$ model using rainfall records from 1922 to 1999 to forecast monthly rainfall 10 years ahead. Significance of the model parameters as well as diagnostics of error were carried out. However, it compared the actual and forecasted rainfall based on the time series plot only. According to the plot, the model was not able to represent the peak values of the rainfall. A similar study was carried out by Ali (2013) to forecast monthly rainfall of Baghdad International airport station in Iraq during the period of 1980 to 2012. It should be noted that both studies have not made any attempt to find overall model accuracy. SARIMA $(0,1,1)(0,1,1)_{12}$ has been developed for forecasting monthly rainfall of Tamil Nadu in India by using monthly rainfall data from 1871 to 2006 (Nirmala and Sundaram, 2010). The parameters were estimated and tested for statistical significance, but the accuracy of the model has not been done. Nevertheless, authors suggested that the accuracy of the model could be improved by adding more input parameters such as El Nińo Southern Oscillation (ENSO) and Land surface temperature. A similar type of model has been suggested by the same authors to predict monthly rainfall in Tamil Nadu using data time span from 1950-2008.

Gerretsadikan and Sharma (2011) suggested SARIMA $(0,0,1)(1,1,4)_{12}$ to study monthly rainfall of Mekele station of Ethiopia using rainfall records from the period 1975 to 2009. They claimed that the model was adequately fitted to the historical data and there was no violation of assumptions in relation to model adequacy. However, parameters were not tested for the significance. In the case study of Abadeh region in Iran, Shamsnia et al., (2011) modeled the monthly average precipitation 1989 to 2009 using seasonal ARIMA $(0,0,1)(1,1,1)_{12}$ model. However, in this study as well, no attempt was made to work out the overall model accuracy and parameters were not tested for the significance. Using data from 2001 to 2013 in Central Java region in Indonesia, Nugroho and Simanjuntak (2014) fitted ARIMA (6,0,3) for monthly rainfall. Generally, it is not recommended to use higher order AR models. Nevertheless, the authors have not justified why they selected higher order models.

Yusof and Kane (2012) applied two time series analysis techniques namely, seasonal ARIMA and the state space model based on exponential smoothing modeling methods for forecasting monthly rainfall of two weather stations in Malaysia using monthly rainfall records from 1968 to 2003. The study developed SARIMA (1,1,2) $(1,1,1)_{12}$ and SARIMA (4,0,2) $(1,0,1)_{12}$ as the best fitted models for the two regions and parameters of the seasonal ARIMA were tested for the significance. Though they claimed that the exponential smoothing state space models have been adequately fitted to the data, it was not given the model accuracies. Mahsin et al., (2012) developed seasonal ARIMA (0,0,1) $(0,1,1)_{12}$ model to monthly rainfall in Dhaka Station in Bangladesh using the rainfall data over the period from 1981 to 2010. Aziz et al., (2013) developed SARIMA (0,0,0) $(2,1,0)_{12}$ model for monthly rainfall predictions using time span from 1974 to 2010.

In another study in the Eastern region of Ghana by Ampaw et al., (2013) developed SARIMA (0,0,1) $(2,1,1)_{12}$ model for monthly rainfall. However, the detail of the diagnostic test statistics did not provide. However, they claimed that the difference between actual and predicted rainfall varied between -6.92 mm to 13.75mm. A study done in the Shouguag city, China by Wang et al., (2013) formed SARIMA (2,0,2) $(1,1,1)_{12}$ using monthly data for the period of 1996 to 2009. However, in this study too, no attempt has been made to test the significance of the parameters of the model. The percentage error varied within 20%, except for the January, May, September and December. It also claimed the reason for the high relative error of January and December. This was due to fact that the model is not sensitive. Selvaraj et al., (2013) employed ARIMA (0,0,12) model to forecast monthly rainfall in Tamilnaudu, India. The training data set made from January 2001 to December 2012. No comments are made on diagnostics tests and forecasting.

Patrick et al., (2014) used monthly rainfall (1970-2009) to develop ARIMA (5,1,1) model to Kinshasa city of Congo in Africa. These authors have not made attempts to test the parameters of the model for the significance. Wang et al., (2014) developed ARIMA models by accounting both the inter annual and inter monthly variation in forecast of monthly rainfall. Preliminary clustering analysis was performed for the monthly rainfall data on Lanzhou precipitation station in Lanzhou, China from 1951 to 2000 and the characteristics of each cluster, namely minimum, maximum and truncated mean rainfall of each cluster were fitted with separate ARIMA models.

Etuk and Mohamed (2014) developed SARIMA (0,0,0) $(0,1,1)_{12}$ to predict monthly rainfall in Gezira irrigation scheme in Sudan. They have checked the model diagnostic but no attempt had been made to evaluate forecasting accuracy. Dastorani et al., (2014) also developed different order of ARIMA and SARIMA models for nine stations in North Khorasan, Iran. Babazadeh and Shamsnia (2014) recommended SARIMA (0,0,0) $(2,1,0)_{12}$ to monthly rainfall in area Shiraz in Iran. It should be noted again that none of the above three studies made any attempt to test the parameters for significance and the percentage errors.

SARIMA (0,0,1) $(1,1,1)_{12}$ model was built to forecast monthly rainfall using rainfall data from 1980 to 2006 of Sylhet station, Bangladesh (Bari et al., 2015). The data (1980 to 2006) were used as the training set and data from 2007 to 2010 were used as validation set. No comparison has been done between actual and predicted. Eni and Adeyeye (2015) forecasted monthly rainfall in 2013 using SARIMA (1,1,1) $(0,1,1)_{12}$. The study was done in the Warri town of Nigeria using the past rainfall records from 2003 to 2012. Using pair wise t distribution, they claimed that the differences between the actual and observed were not significant at 0.05 level of significance.

Chonge et al., (2015) modeled SARIMA (0,0,0) $(0,1,2)_{12}$ for data (1977-2014) in Gishu Country, Kenya and forecasted for two years ahead. In this study, Mean absolute percentage error (7.78%) was taken as a statistical indicator to judge the model. It should be noted that MAPE does not provide any sense of the magnitudes of the error for each point. SARIMA (1,1,5) $(1,1,2)_{12}$ was identified as the best fitted model to forecast monthly rainfall of the Urmia lake catchment area of India by Alimirzaie et al., (2015). In this study, 42 years monthly rainfall data (1968-2010) were used to train the model but no comparison was done between actual and forecast value even for the trained data sets.

Savoe (2015) made an attempt to predict long term precipitation over the Ghanaian segment of Ghana using data from 1967 to 2000. SARIMA (2,1,1) $(1,1,1)_{12}$ was proposed as the best fitted model for the average monthly rainfall. The authors have predicted for 41 years ahead (2007 to 2047). It should be pointed that prediction of such a long period is not statistically valid using ARIMA models. Furthermore, as in most of the past studies, the significance of the model parameters as well as comparison of actual and forecast values were not done.

A recent study by Mohamed and Ibra (2016) have developed multiplicative seasonal autoregressive integrated moving average (MSARIMA) to forecast monthly rainfall of Nyala station in Sudan using rainfall data from 1971 to 2010. Those models have been selected based on the RMSE and MAE. However, these indicators are not suitable to justify the accuracy of a model. Zafor et al., (2016) also developed eight different seasonal ARIMA models to predict monthly rainfall in different locations in Sylhet district, Bangladesh. Models were developed using rainfall records from 2001 to 2012 while the performance of the models have been validated using data in 2011 only. This is not sufficient and further it is not an independent data set.

2.1.4. Prediction of Weekly Rainfall

Unlike monthly rainfall, not much studies have been carried out to forecast weekly rainfall under ARIMA/SARIMA environment. Zakaria et al., (2012) modeled weekly rainfall data from four rainfall stations in the North West of Iraq for the period 1990-2011 using seasonal ARIMA approach. They considered only 30 rainy weeks every year for this study. The four models suggested are ARIMA (3,0,2) $(2,1,1)_{30}$, ARIMA (1,0,1) $(1,1,3)_{30}$, ARIMA (1,1,2) $(3,0,1)_{30}$ and ARIMA (1,1,1) $(0,0,1)_{30}$. They have considered that there is a seasonal pattern with the length of 30 weeks and they have highlighted about the complexity of representing seasonal periods to the models. The performance of the model was evaluated by using the same data, but not for an independent data set. Though, the forecasted values were obtained up to 2006, the predicted values were not compared with the actual values.

Popale and Govantiwar (2014) developed seasonal ARIMA (1,1,1) $(1,0,1)_{52}$ model for forecasting weekly rainfall using 31 years (1982-2011) data of Rahuri region in India. The length of the seasonality has been assumed as 52 without any statistical justification. In this case too, model parameters were not tested for the significance and the forecast values were not compared with actual values.

2.2. Use of Artificial Neural Network for Modeling Rainfalls

In section 2.1 it was shown that ARIMA (either seasonal or no seasonal) models have achieved success in their own linear domains. However, rainfall is a result of many complex atmospheric parameters which cannot easily be determined on the assumption linearity among variables within the same series. Thus, some authors have used Artificial Neural Network (ANN) approach to model rainfall to address the problems belonging to nonlinear forecasting.

Kumarasiri and Sonnadara (2006) developed an ANN model for the annual rainfall in the Colombo city in Sri Lanka. Rainfall of the past ten years were used as input vector and the network was trained from 1869 to 1973 using feed forward back propagation algorithm and it was tested using the time span from 1974 to 2003. The authors claimed that the proposed model has been successful in forecasting the annual rainfall amount one year ahead. However, the proposed model should not be able to forecast beyond two years ahead, as accuracy becomes exceedingly low. Nanda et al., (2013) proposed ARIMA (1,1,1) and three different ANN namely, multilayer perceptron (MLP), legendre polynomial equation (LPE), functional link ANN (FLANN) to estimate yearly rainfall. They found that the model FLANN yield better prediction in comparison to other models in forecasting yearly rainfall. Nirmala (2015) employed ANN to predict annual rainfall of Tamilnadu in India using rainfall records for 136 years (1871-2006). Out of those data, 100 years (1871-1970) were used for training the network using three algorithms, namely gradient descent (GD) algorithm, scaled conjugate gradient (SCG) algorithm and radical basis function (RBF) algorithm. The RBF was selected as the best out of the three algorithms discussed above due to lower mean absolute percentage error. However, the accuracy of the models were not tested for individual forecast values.

Krishnankutty (2006) applied two approaches, namely ANN and multiple linear regression (MLR) for forecasting the southwest monsoon rainfall of the 14 districts of Kerala in India. Fourteen separate feed forward ANN and MLR models were developed for distinct districts in Kerala. The data for 51 years (1941-1991) have been used for training the network using back propagation learning algorithm while the data for a period 1992-2004 were used for testing purpose. The correlation coefficient between the forecasted and actual values for the district area-weighted model was 0.95. The author claimed that the ANN models outperformed MLR models based on the visual observation of plots.

Kumar et al., (2007) developed an ANN in forecasting regional rainfall of Orissa State in India by accounting information on large scale climate tele connections namely, El Ninó Southern Oscillation (ENSO), Equatorial Zonal Wind Index (EQWIN), Ocean-Land Temperature Contrast (OLTC). The summer monsoon seasonal rainfall during the four months period, from June to September were considered for this study. Genetic Optimizer algorithm was used to optimize the feed forward back propagation network architecture. The result revealed that the correlation coefficient of forecasted and actual was 0.8951. But no comparison was carried out for individual points. Many studies were carried out to evaluate the seasonal rainfall in different regions by using ANN (Mekanik and Imteaz, 2012; Gupta et al., 2013; Golabi et al., 2013; Rasel et al., 2015).

Two authors; Kumarasiri and Sonnadara (2006) employed ANN for the purpose of predicting monthly rainfall in the Colombo city in Sri Lanka. The monthly rainfall was classified into 06 categories based on the rainfall depths since the large error have been occurred when using actual depth of rainfall as input. The multilayered feed forward network was trained by using the back-propagation algorithm for 50 years from 1949 to 1998 and testing was done for 05 years from 1999 to 2003. The authors claimed that the model have been reasonably successful in forecasting monthly rainfall a month ahead.

In 2008, Mar and Naing applied three-layer feed forward neural network on monthly rainfall data for the period of 1970 to 2006 for the purpose of forecasting monthly rainfall in Myanmar. Based on the low RMSE (9.881) of the model, the authors claimed that ANN has a good ability to forecast monthly rainfall of Myanmar. Khodashenas et al., (2010) used ANN model to predict monthly precipitation of Mashhad synoptic station in Iran using rainfall data from 1958 to 2008. The proposed model consists of 4 hidden neurons with one out put neuron which could predict monthly rainfall with a high accuracy. The correlation coefficient of predicted values and actual vales of rainfall was reported as 0.84 ($P<0.05$).

Vamsidhar et al., (2010) developed a multilayered feed forward neural network model for forecasting monthly rainfall in India using rainfall data for the period 1901-2000. Pressure, humidity and dew point were considered as the input of the network. The model with 7 hidden neurons was selected as the best model among the tested models and the model accuracy has been calculated (94.28%) using the formula, Accuracy = 100-MSE. Mekanika et al., (2011) made an attempt to develop long term rainfall prediction model (12 months in advance) using ANN to forecast monthly rainfall for the West mountainous region in Iran. Three ANN models were formed based on the Levemberg-Marquardt algorithm with different inputs using monthly rainfall from 1977 to 2002. The model was tested for the year 2003.

Deshpande (2012) employed four distinct ANN models, namely multilayer perceptron neural network (MLP), jordon elmann neural network (JENN), Self organized feature map (SOFM), recurrent neural network (RNN) for predicting monthly rainfall of Maharashtra State in India. Terzi and Cevik (2012) applied ANN to forecast monthly rainfall in Isparta and compared with result of multiple linear regression model. ANN model was selected as the best model for monthly rainfall estimation in the study region based on the correlation coefficient and RMSE.

A feed forward neural network model have been employed for forecasting monthly rainfall of Mirzapur district, Uttar Pradesh in India by Kumar and Yadav (2013). Ten climatic variables monthly temperature (average, diurnal, minimum and maximum), evaporation (potential and reference crop), relative humidity, clouds cover and frequency (ground frost and wet days) data from 1995 to 2002 have been used as model inputs. The authors claimed that the agreement between predicted and observed had been good and the determination of coefficient was reported as (R^2 =0.8563).

Gupta et al., (2014) developed ANN model for forecasting monthly rainfall of Madhya Pradesh in India. Normalized monthly rainfall data of 10 years from 2000 to 2010 were used as input variables to the multi layered feed forward neural network. In this study, 75% of data were used for training, 15% of data were used for validating and 10% data were used for the testing. The model was trained using back propagation algorithm. The correlation value of the actual and predicted was 0.9360 (P value <0.05). Many researchers all over the world made attempts to model monthly rainfall using different ANN with increasing degree of accuracy (Alhashimi, 2014; Dubey, 2015; Mesgari et al., 2015).

Luk et al., (2001) applied three different ANN, namely multilayer feed forwards neural network (MLFNN), partial recurrent neural networks (PRNN) and time delay neural networks (TDNN) on rainfall amounts which were taken during 15 minutes intervals from 1991 to 1996 of an urban catchment in Western Sydney, Australia for the purpose of predicting rainfall. They selected the eight ANN based on NMSE and the authors claimed that the all ANN models could make reasonable forecasts accuracies for one-time step ahead (15 minutes). Based on the NMSE of the validation period, it was reported that the TDNN has more accuracy in predicting rainfall in comparison to MLFNN and PRNN. A study was conducted to forecast hourly rainfall of Bangkok in Thailand by Hung et al., (2009) using ANN model. Eight different ANN models were formulated using different input such as humidity, air pressure, wet bulb temperature, rainfall intensities and cloudiness using hourly data from 1977 to 1999. Two simple multilayered ANNs were trained by using sigmoid activation function while six generalized feed forward ANN were trained by using sigmoid and hyperbolic tangent transformation algorithms. The data of the year 2003 were used for the testing purpose. Furthermore, sensitivity analysis was performed to rank the input contribution and beside the rainfall itself, the most important input was the wet bulb rainfall temperature in forecasting rainfall. The correlation coefficients between actual and predicted for 1h, 2h and 3h were 0.99,0.92 and 0.84 respectively.

Charaniya and Dudual (2013) proposed two distinct ANN models namely, generalized feed forward (GFNN) and focused time lag delay (FTLNN) neural networks for daily rainfall predictions on the basis of preceding events of rainfall data. This study has been carried out at Nagpur region in the Central part of India using rainfall daily rainfall records for 30 years (1977-2006). The result indicated that FTLNN model made better forecast accuracy than the GFNN for learning a temporal pattern which gave least normalized Mean Squared Error. Similar studies were carried out for modeling daily rainfall using ANN approach in various countries (Weerasinghe et al., 2010; Amesh and Negaresh, 2013; Omidvar, 2015).

2.3. Use of Multiple Linear Regression for Rainfall Forecasting

Krishnankutty (2006) applied Multiple Linear Regressing (MLR) for forecasting the southwest monsoon rainfall of the 14 districts of Kerala in India using rainfall as independent variable and compared those models with the corresponding feed forward ANN and concluded that the ANN models ANN models outperformed MLR models. Kannan et al., (2010) developed MLR to predict the summer monsoon (September to November) rainfall using monthly rainfall data during summer monsoon of previous year in Tamil Nadu, India and claimed that accuracy of predicting value is low. The study did not provide the corresponding R^2 value and furthermore, errors of the model were not tested for white noise and nothing has been mentioned about the significance of the model parameters.

Rasel et al., (2015) applied two statistical approaches; MLR and ANN to forecast seasonal rainfall in South Australia using spring rainfall as dependent variable. Single and combined lagged large-scale remote climate drivers such as, EI Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD) and Southern Annular Mode (SAM) were considered as potential predictor variables on long term spring rainfall. The study revealed that lagged Dipole Model Index (DMI)-SAM have been more effective on spring rainfall predictability than other combination of climate drivers and using those predictor variables, ANN model could increase the model correlation up to 87% ($p < 0.05$) whereas, the rainfall predictability of MLR was 52% ($p < 0.05$).

Terzi and Cevik (2012) applied MLR technique for monthly rainfall in Isparta and compared it with three-layer feed forward ANN model. Based on the R^2 value and RMSE, they concluded that ANN is more superior than MLR. Alhashimi (2014) also developed MLR model to monthly rainfall (1970 - 2008) in Iran taking air mean temperature, relative humidity and wind speed as independent variables without validation of diagnostic tests related to MLR. They also claimed that ANN model is superior than MLR model in forecasting monthly rainfall based on the correlation coefficient and RMSE values. It should be mentioned that in ANN models, parameters are not tested for statistical significance, though the authors claimed that the ANN models are superior than other models.

Armesh and Negaresh (2013) employed MLR model and ANN model to forecast the maximum daily rainfall of Saravan in Iran considering various meteorological variables and climate indices from 1986 to 2010 as potential predictors. They found that the variables; monthly maximum and minimum relative humidity and climate indicators have made a significant effect on the maximum daily rainfall in Saravan. This regression model was fitted for the purpose of model comparison with radical basis function (RBF) NN and multilayered feed forward back propagation (MLFBP) NN for forecasting daily maximum rainfall. The correlation coefficient of the models RBF, MLFBP and MLR were 0.95.0.95 and 0.89 respectively. Thus, based on the correlation coefficient, RMSE and MAE, authors claimed that the ANN model considerably yield better forecasting result. However, no attempt was made to test the model assumptions and parameters of the regression model for significance.

2.4. Hybrid Models for Rainfall Forecasting

In the previous section three types of modeling approaches were discussed independently. Each approach has advantages as well as drawbacks. Therefore, in recent years researchers in all over the world have proposed hybrid models which combining two or more different types of models to forecast the rainfall at different time scales for improving the forecast accuracy.

Yu and Yu (2012) proposed modular radical basis function neural network (M-RBF-NN) coupled with the singular spectrum analysis (SSA) and partial least square (PLS) regression for forecasting monthly rainfall of Liuzhou in China. Monthly rainfall data from January 1949 to December 2006 were used to train the model. Initially, the technique SSA was applied to the rainfall series to the purpose of removing of trends and to reform the new time series. Next, a triple phase non linear M-RBF-NN model was utilized for rainfall forecasting by linking different activation functions. Then the result in a suitable number of RBF-NN predictors were selected using the partial least square technology. Finally, the model was assembled by the least squares support vector regression (LS-SVR). Based on the absolute relative error (ABRE), root mean square error (RMSE) and Pearson relation coefficient (PRC), authors claimed that the M-RBF-NN has the highest accuracy. However, it should be pointed out such indicators are not recommended to judge a model for forecasting accuracy.

Mahalakshmi et al., (2014) developed a hybrid model with a combination of ARIMA and ANN to predict monthly rainfall in Tamil Nadu, India using data from 1950 to 2012. The seasonal ARIMA $(0,1,1) \times (0,0,1)_{12}$ model was fitted to the historical data and the residual derived from the ARIMA model was fitted by ANN. For the ANN they used a data set 756, out of which, 700 data were used to train the network using

feed forward back propagation algorithm. The authors revealed that the better forecast result could be obtained using a hybrid model than ARIMA or ANN alone.
Patel and Parekh (2014) employed a hybrid model by combining the ANN and Adaptive Neuro Fuzzy Inference System (ANFIS) to forecast monthly monsoon rainfall for Gandhinagar station in India. Various membership functions were utilized to derived eight different models and they considered climate parameters as input variables for all models. It was found that the hybrid model with seven membership functions with three inputs, namely relative humidity, temperature and wind speed produced the best performance to forecast the rainfall in this area. It should be pointed out that unlike statistical models there is no justification why 8 models were considered and also no significance is considered when variables and their transformation are included into the model.

Abbot and Marohasy (2014) used ANN model with genetic optimization to find the lagged relationships among temperature, atmospheric pressure, climate indices to study the monthly rainfall of three geographical distinct regions in Queensland. Meteorological data, including rainfall from 1893 to 2012 were used for this study and result was compared with Predictive Ocean Atmospheric Model for Australia (POAMA) which is the General Circulation Model currently used to produce the official seasonal rainfall forecast. The result indicated that the forecasts using ANN for three areas were superior compared to forecasts from the best available general circulation model (POAMA).

A survey which consists of several artificial intelligence models that have been used to forecast the rainfall was conducted by Pallavi and Singh (2016). The ANN, combined model of support vector machine (SVM) & Fuzzy logic method and NN-Fuzzy method were models considered for this study. Based on the survey, the author concluded that the hybrid combination of ANN and Fuzzy logic; adaptive neuron fuzzy inference system (ANFIS) was the best approach for rainfall forecasting.

A hybrid model combining the self organizing map (SOM) and multilayer perceptron neural network (MLPN) have been developed to forecast the typhoon rainfall from

July to October in Taiwan by Lin and Chang Wu (2009). In this study, hourly rainfall data of 10 rain gauges of Tanshui river basin was used. Firstly, they analyzed input data using SOM technique that was used to decompose the input into distinct clusters. Then, MLPN was carried out for each cluster. The Proposed model was applied and compared with the conventional ANN model result. It was found that the forecasting power of the proposed model outperform conventional ANN.

Yusof et al., (2013) developed a hybrid ARIMA (2,1,2)-GARCH (1,1) and ARIMA (3,1,1)-GARCH (1,1) to model daily rainfall of Ipoh and Alorestar in Malaysia respectively using data for the period from 1968 to 2003. Residuals in both models were found to be white noise as well as no ARCH effects. Though they claimed that the models fit the daily rainfall data set at two locations. well, no attempt was made to find the accuracy rate of the predictions.

Two hybrid models have been developed to forecast daily precipitation of two locations in Iran by Teimoorzadeh et al., (2015). Firstly, single genetic programming (GEP) and ANN were applied on daily rainfall data for 9 years (2000-2008). Then, two hybrid models, namely wavelet genetic programming (WGEP) and wavelet ANN (WANN) were developed using same data set due to low accuracy of first models. It was found that the forecast accuracy was significantly increased using the hybrid model and also forecast accuracy using WANN outperform using WGEP. A study was carried out to forecast daily precipitation using hybrid models, namely wavelet-artificial neural network (WANN) at Verayneh station, Iran and the result was compared with adaptive neuro fuzzy inference system (ANFIS) by Solgi et al., (2014). Wavelet transformation was applied to the daily precipitation data and original time series were decomposed to multiple sub time series which could be applied as input data for artificial neural network. Different structures in ANFIS were applied to the same dataset for the purpose of comparison. Based on the result, the best model was the WANN which had less error than ANFIS along with the high correlation coefficient (r =0.95).

Faulina and Suhartono (2013) applied different techniques to forecast daily rainfall of six area in Indonesia by using rainfall records from 1996 to 2012. This study focused to develop both individual and hybrid models, namely Adaptive Neuro Fuzzy Inference System (ANFIS), ARIMA, ARIMAX, ARIMA-ANFIS and ARIMAX-ANFIS. Traingular, Gaussian and Gbell functions were used as membership functions of ANFIS and best model was selected based on the RMSE. It was found that the individual ARIMA model yields a more accurate forecast than other complex models. Also, the authors mentioned that complicated models do not always yield better forecast than the simple one. Furthermore, it should be added that complicated models do not recommend the inferences with some confidence and consequently most of the complicated models are subjective.

2.5. Impact of Other Climatic Variables on Rainfall

Various studies have been carried out to find the dynamic relationship between annual or monthly rainfall and other external variables using vector auto regression (VAR) models (Adenomon et al., 2013) or regression models (Malmgren et al., 2003). Kumar et al., (2007) found that seasonal rainfall of Orissa State in India has significant association with large scale climate tele connections namely, El Nińo Southern Oscillation (ENSO), Equatorial Zonal Wind Index (EQWIN), and Ocean-Land Temperature Contrast (OLTC). Tularam (2010) found a relationship between ENSO and rainfall in South East Queensland. Mekanik and Imteaz (2012) claimed that the there is a significant relationship between large climatic drivers; ENSO, Indian Ocean Dipole (IOD) and Southern Annular Mode (SAM) with spring rainfall of Victoria in Australia. A similar study was carried out by Rasel et al., (2015) claimed that the significant relationship between large scales climatic drivers such as ENSO, IOD and SAM with seasonal rainfall in South Australia.

The findings of the above studies provided evidences to the significant relationship between the large climatic drivers with the seasonal rainfall. Also, it is noted that the few studies in literature reported the significant relationship between rainfall and other external factors as temperature, relative humidity, vapor pressure etc.

2.6. Long Memory Models

In recent past, time series models with long memory features became very popular among researchers in many fields in particularly in financial time series modeling. Features of a autoregressive fractionally integrated moving average (ARFIMA) long memory model was initially introduced by Granger and Joyeux (1980) and Hosking (1981). It was an extension of the traditional ARMA process with a fractional differencing parameter. The model defined as ARFIMA (p,d,q) allows the parameter “d” to take fractional values for differencing. There is a fundamental change in the correlation structure of the ARFIMA model, when compared with the correlation structure of the conventional ARIMA model (See Chen et al., 1994). According to Granger and Joyeux (1980), the slowly decaying autocorrelation exhibited in long range dependency or long memory models differ from stationary ARIMA models that decay exponentially. This is the primary detection for the development of long memory models. However, development of such models is not an easy task as there are many unsolved problems in this area of research.

2.6.1. Estimating of Fractional d of ARFIMA Models

Many researchers proposed different methods to estimate the fractional differencing parameter. Gewek and Porter-Hudak (1983) proposed a method for estimating the long memory differencing parameters based on a simple linear regression of the log periodogram. An approximate maximum likelihood method for parameter "d" was proposed by Fox and Taqqu (1986). Fundamental properties of the ARFIMA family and the estimation of the model parameters were discussed by Andel (1986). An exact maximum likelihood estimation method for differencing parameter was introduced by Sowell (1992). Chen et al., (1994) developed a regression type estimator of ‘'d" using lag window spectral density estimators. A method based on the smoothed periodogram for estimating of ARFIMA parameters was proposed by Resien (1994). Comparison study assessments were done by Cheung and Diebold (1994) on maximum likelihood estimators for fractionally differenced parameters using two types of maximum likelihood (ML) estimators in the form of frequency-domain ML and exact domain ML of time series processes with an unknown mean.

2.6.2. Use of Long Memory Models

Montanari et al., (1997) proposed fractional differenced ARIMA models for daily and monthly inflows of lake Maggiors in Italy and monthly rainfall in Genoa. They applied an approximation in the spectral domain of the Gaussian maximum likelihood function called Whittle estimators for the model parameter estimation. However, authors claimed that though the fractional ARIMA improved the modeling of inflows with long range persistence the analysis of monthly rainfall shows that the absence of long-term effects. Further, the authors suggested that in fractionally difference method with a seasonal component would improve the capability of representing both long and short memory persistence of the hydrological time series.

2.6.3. Use of Gegenbauer ARMA Models

Due to the practical success of the ARFIMA model, a more generalized fractionally differenced long memory time series model called the Gegenbauer ARMA(GARMA) was probed in detail by Gray et al., (1989). This type of long memory class illustrates multiple unbounded spectral peaks away from the zero. Chung (1996) extended the work in introducing a grid-based parameter estimation procedure of an elementary GARMA process. A concise summary of fractionally differenced Gegenbauer processes with long memory was provided in Dissanayake (2016). An extensive review of fractionally differenced Gegenbauer processes with long memory carried out by Dissanayake et al., (2018).

2.6.4. Use of Seasonal Autoregressive Fractionally Integrated Moving Average (SARFIMA) Models

Though the ARFIMA model was able to capture the long-range dependency, it does not take into account the seasonal variation patterns present in some real data series particularly in rainfall series. The SARFIMA (Porter-Hudak, 1990) is a natural extension of the ARFIMA process with an additional seasonal filter. The model consists of long memory dependency features with periodic behavior in terms of the data. SARFIMA model was utilized for forecasting of the monthly IBM product revenue in Ray (1993). Peiris and Singh (1996) suggested a convenient method to calculate predictors for seasonal and non seasonal fractional parameters of long memory models under certain conditions. The work done by Bisognin and Lopes (2009) described number of properties of seasonally fractional ARMA process in detail. SARFIMA model was applied to forecast Iraqi oil production and model parameters were estimated using conditional sum of squares by Mostafaei and Sakhabakhsh (2011). Additionally, Reisen et al., (2014) proposed a semi parametric approach to estimate two seasonal fractional parameters in a SARFIMA model and the performance was evaluated through a Monte Carlo experiment.

However, extremely few attempts have been made to study the rainfall behavior in context of long memory. A study done by Yaya and Fashae (2014) made an attempt to fit SARFIMA models for rainfall data in six rainfall zones of Nigeria but they claimed that could not develop significant SARFIMA models which the seasonal behavior with the long-range dependency of the real data. However, these types of models can be developed to tackle to model complex time series such as weekly rainfall.

2.6.5. Models for Capture Heteroskedasticity

There has been growing interest in modeling time series in many disciplines such as finance, economics, environmental science, hydrology etc. having heteroscedasticity property. The heteroscedasticty in time series is generally handled using autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) models (Engle, 1982; Bollerslev, 1986). Various authors (Ling and Li, 1997; Henry, 2001; Jenson, 2005; Sena et al., 2006) developed ARFIMA-GARCH models for different applications, but less attention was given for short-term or long-term prediction of the series and thus those work has less importance from practical point of view.

Kane and Yusof (2013) employed GARCH (1,1) model to the residual of the ARFIMA that fitted for the daily rainfall data from 1975 to 2008 in Malaysia. They estimated the fractional differencing parameter using the method proposed by the two researchers Gewek and Porter-Hudak (1983) and claimed that the by adding the

GARCH specification for the ARFIMA error helps to capture the serial correlation in the squared residual. Reisen et al., (2014) proposed SARFIMA-GARCH model for modeling and forecasting daily average PM_{10} (Airborne Ambient Particulate Matter) concentration. Semi parametric procedure suggested by Reisen et al., (2014a) was used to estimate the fractional differencing parameter in along with time dependence condition error variance. The authors claimed that some features of the data such as seasonality, long memory and volatility were able to capture by the proposed model.

2.6.6. GARMA Class of Models with Heteroskedasticity

Fresh interest in the econometric community infused into the process the introduction of a GARMA class of models with heteroskedasticity by Dissanayake and Peiris (2012). It was followed by the casting of the process driven by Gaussian white noise in state space by Dissanayake et al., (2016a) to establish a parameter estimation based optimal lag order validated by predictive accuracy. A similar experiment in which the process was driven by GARCH errors (instead of Gaussian white noise) was presented by Dissanayake et al., (2014) with the validation of parameter estimation based optimal lag order done through log likelihood measures.

2.7. Summary of the Chapter 2

Many researchers have attempted to predict rainfall at different time scales. The type of models used are MLR, ARIMA, SARIMA, ANN, Hybrid and long memory. Of those, ARIMA, SARIMA, and ANN were found to be more popular in modeling rainfall irrespective of time scales. However, there were various drawback in such models with respect to the statistical aspect as well as non-statistical aspects. Almost all models were not tested for an independent data set. Though some studies provided correlation coefficient of predicted and observed, most of those failed to give p-value which need to get overall judgment about the model. Thus, most of those models are not recommended to use. All authors have claimed that prediction of weekly rainfall, in particularly in tropical countries is more difficult than prediction of annual, seasonal or monthly due to various noisy structure.

Nevertheless, very few studies were reported to model weekly rainfall but no proper prediction was carried out.

Almost all authors claimed that ANN models are better than MLR. However, the main drawback of this methodology is that these model parameters are not tested statistically. Many researchers developed hybrid models coupled with artificial neural network for the purpose of forecasting rainfall. However, some of them claimed that more complicated model not always give better forecast in comparison to the simple ones. Another drawback of hybrid models or ANN model is that the results depend on the methodology used to estimates parameters and no study has been claimed that their results are invariant of the methodology used.

The long memory models with a fractional differencing parameter have been popular among the researchers in modeling complicated time series data. Most of the researchers applied long memory models for the financial time series. Generally, rainfall series in tropical countries are also complex as financial series. There has been a still noticeable gap modeling persistent rainfall in view of long memory. It is very important to develop a novel model to forecast weekly rainfall series since very less attention have been given to model the rainfall at weekly basis. Nevertheless, this extensive literature review certainly provides immense information on the direction of developing novel model for weekly rainfall in Colombo city.

CHAPTER 3

RESEARCH METHODOLOGY

In order to enrich the understanding of characteristics of the weekly rainfall series, many techniques are applied and those methods are described in detail in this chapter. Initially, the study site and the data description of the weekly rainfall along with the exogenous climatic variables are discussed.

3.1. Study Site

Sri Lanka is a tropical country in South Asian region and the Colombo city is the commercial capital of Sri Lanka, situated with latitudes 6^0 55′ N and Longitude 79^0 51′ E and is chosen as the study site. Colombo meteorological location is the main station of the department of Meteorology in Sri Lanka. Many meteorological variables including daily rainfall data have been recorded without missing values by the Colombo station since 1870. The corresponding study site is presented by the Figure 3.1.

3.2. Data Description

Daily rainfall data and the climatic variables: minimum and maximum temperature, relative humidity (AM & PM), minimum and maximum vapor pressure were obtained at daily basis from 1960 to 2017 in Colombo city from the Department of Meteorology, Sri Lanka.

The daily rainfall (mm) data has been converted into weekly rainfall by dividing a year into 52 weeks such that week 1 corresponds to 1-7 January, week 2 corresponds to 8-14 January week 3 refer as 15-21 January and so on. The corresponding weeks are presented in Table 3.1. In order to make homogenous period irrespective the years. February 29^{th} wasn't taken into account when making 52 weeks.

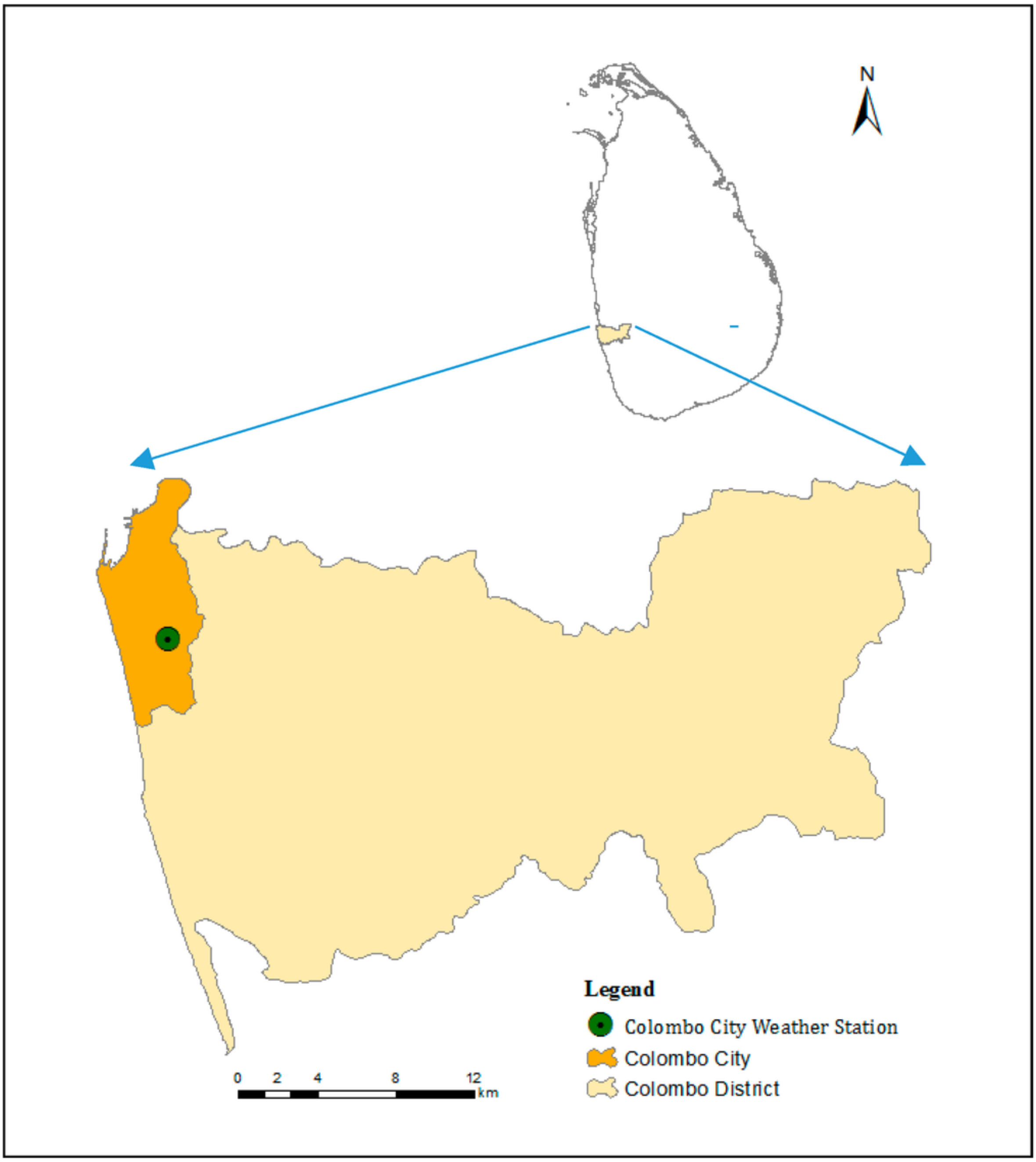

Figure 3.1: The city of Colombo is in the Colombo district

Table 3.1: Standard weeks in a year

Week	Date	Week	Date
1	January 01 -07	27	July 02-08
2	January 08 -14	28	July 09-15
3	January 15 -21	29	July 16-22
4	January 22 -28	30	July 23-29
5	January 29 -February 04	31	July 30-August 05
6	February 05 -11	32	August 06-12
7	February 12 -18	33	August 13-19
8	February 19-25	34	August 20-26
9	February 26 -March 04	35	August 27-September 02
10	March 05-11	36	September 02-09
11	March 12-18	37	September 10-16
12	March 19-25	38	September 17-23
13	March 26 -April 01	39	September 24-30
14	April 02-08	40	October 01- 07
15	April 09-15	41	October 08- 14
16	April 16-22	42	October 15- 21
17	April 23-29	43	October 22- 28
18	April 30- May 06	44	October 29- November 04
19	May 07- 13	45	November 05 -11
20	May 14-20	46	November 12-18
21	May 21-27	47	November 19-25
22	May 28-June 03	48	November 26-December 02
23	June 04-10	49	December 03-09
24	June 11-17	50	December 10-16
25	June 18-24	51	December 17-23
26	June 25-July 01	52	December 24-31

3.3. Analysis of the Weekly Rainfall Percentiles for SWM

The city Colombo is located in the Western part of the country which directly receives rainfall from SWM. The rainfall percentiles analysis was utilized on weekly rainfall series during SWM to pursue and underline the temporal fluctuations during the time span from 1960 to 2015. The weeks 18-39 were pertaining to the SWM is presented by Table 3.2.

Table 3.2: The weeks pertaining to the SWM

Weeks	Date	Weeks	Date
18	April 30-May 06	29	July 16-22
19	May 07-13	30	July 23-29
20	May 14-20	31	July 30-August 05
21	May 21-27	32	August 06-12
22	May 28-June 03	33	August 13-19
23	June 04-10	34	August 20-26
24	June 11-17	35	August 27-September 02
25	June 18-24	36	September 03-09
26	June 25- July 01	37	September 10-16
27	July 02-08	38	September 17-23
28	July 09-15	39	September 24-30

In addition to standard weeks, running totals of weekly rainfall were also considered for the analysis of weekly rainfall percentiles. The running totals of weekly rainfall were obtained during SWM period is presented by Table 3.3. It is calculated total of 148 running weekly totals of weeks which belongs to SWM.

Table 3.3: The running totals of weeks pertaining to the SWM

Running totals of weeks during SWM	Date
Week 1	April 30-May 06
Week 2	1-7 May
Week 3	2-8 May
Week 4	3-9 May
....	....
....	....
Week 148	24-30 September

Before analyzing rainfall percentiles, the trend analysis was carried out and tested the linear and quadratic trend pattern in weekly rainfall during time period from 1960 to 2015. Also, the weekly rainfall series pertaining to the SWM checked for the randomness using auto correlation plots. Furthermore, the normal probability plots of each weekly rainfall series which belong to SWM were obtained to test the normality. The weekly rainfall percentile analysis was done in context of confidence intervals using two approaches namely, parametric and bootstrapping. Under the parametric approach, estimates were made by fitting the probability distributions for weekly series in SWM and those details are presented from the section 3.5.

3.4. Analysis of the Weekly Rainfall Percentiles for SIM

The SIM furnishes considerable shower to the city Colombo during the October to November. This analysis mainly focused to assess the temporal variability of the weekly rainfall during the SIM. The weeks 40-48 were belongs to the SIM is presented by Table 3.4.

Table 3.4: The weeks pertaining to the SIM

Weeks	**Date**	**Weeks**	**Date**
40	October 01- 07	45	November 05 -11
41	October 08- 14	46	November 12-18
42	October 15- 21	47	November 19-25
43	October 22- 28	48	November 26-December 02
44	October 29- November 04		

The running totals of weekly rainfall were also considered as SWM for the analysis of weekly rainfall percentiles. Those running totals of weekly rainfall were utilized to study the variability of the rainfall and those are presented by Table 3.5. It is calculated total of 57 running weekly totals of weeks which pertaining to SIM.

Table 3.5: The running totals of weeks pertaining to the SIM

Running totals of Weeks during SIM	Date
Week 1	1-7 October
Week 2	2-8 October
Week 3	3-9 October
Week 4	4-10 October
....	....
....	....
Week 57	26th November to 2nd December

Linear and quadratic trend patterns in weekly rainfall during the SIM was tested before carrying out the analysis of rainfall percentile. Moreover, randomness of the series was assessed using auto correlation plots. Two approaches: parametric and bootstrapping were applied to underline the rainfall percentiles along with the 95% confidence intervals.

3.5. The Best Fitted Statistical Distribution for Weekly Rainfall

Many known probability distributions; Log normal, Exponential, Gamma, Weibull, Largest Extreme Value, Smallest Extreme Value, Logistic, Log Logistic along with the different forms of some distributions such as 3-parameter Gamma, 2-parameter Exponential, 3-parameter Log Logistic and 3-parameter Weibull distributions were utilized to fit probability distribution for the weekly rainfall series and two test Anderson-Darling and Kolmogorov-Smirnov test were used to identify the best fitted probability distributions.

Weekly rainfall percentiles at 50, 60, 70, 80 and 90 and the corresponding 95% confidence intervals were calculated using best fitted distribution for the standard weeks. It is also fitted many probability distributions mentioned above for the running weekly totals of 148 weeks in SWM and 57 weeks in SIM and the same procedure was carried out to select the best fitted probability distributions and calculated the five percentiles and the corresponding 95% confidence intervals for the running weekly totals.

3.6. The Use of Bootstrapping Approach

This is a non parametric distribution free method. Resampling with replacement procedure was utilized to create the number of samples with the same size based on the original sample. Here, large number of statistics made based on the large number of repeated samples created. The 95% confidence intervals for the weekly rainfall percentiles were developed using percentile bootstrap approach.

In order to identify the time period which, form the extreme rainfall events, the rainfall percentiles along with the 95% confidence intervals which made using two approaches were utilized and those statistics were used to enrich the understanding of the weekly rainfall variation during the SWM and SIM. Furthermore, the analysis of the running weekly total with their 95% confidence intervals used to confirm the result of the analysis which was made based on the two approaches.

3.7. Coverage Probability for Weekly Rainfall Percentiles

The coverage probability of the confidence intervals is one of the measurements that can be used to test the accuracy of the confidence interval bands. To compute the accurate level of confidence intervals for weekly rainfall percentile, the parametric bootstrapping approach was applied based on the real coverage probability which derived from the bootstrapping calibration. Under the parametric approach all the inferences including confidence interval bands made based on the distribution which we selected as best fitted either skewed on symmetric. This attempt has been made to test the inferences as confidence interval bands created by using skewed distribution.
Here, one weekly rainfall series were considered as population and considered the 95% confidence intervals made based on the best fitted distribution. A simulation was carried out to calculate the coverage probability of the 95% confidence interval for the rainfall percentiles at the small sample size. It is important to noted that the 2000 random samples were generated using parametric bootstrapping approach. It was compared the calculated coverage probability with the nominal coverage probability. Based on the result, it was highlighted that the corresponding accurate confidence level for the percentiles to achieve the real coverage probability as 95%.

3.8. Modeling Weekly Rainfall Using Classical Models

Before moving to novel approach, it is better to model the rainfall using conventional approach since many reasons. The presence of complex models do not give better result always and to overcome much difficulties in modeling the weekly rainfall, conventional method is initially used.

In this study, an effort is made to model the weekly rainfall series from 1990 to 2014 using Seasonal Autoregressive Integrated Moving Average (SARIMA) model by accounting the correlation structure of the series. Since the heteroskedasticity presence of the residuals derived from the above best fitted model a variance model called Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model developed to capture the stochastic volatility. Model assumption were also tested simultaneously, to assess the accuracy of the fitted model. In addition to above models, an ARIMA-GARCH model was fitted to the deseasonalized weekly rainfall series to improve the model accuracy and the forecasting performance.

Furthermore, the study is moved to model the weekly rainfall with exogenous variables to identify effect of the exogenous variables on weekly rainfall. Vector Auto Regressive (VAR) model was utilized to find the affect the nine variables such as minimum, maximum and average of three variables temperature, relative humidity and vapor pressure on weekly rainfall. Also, the Granger causality test was applied to test one time series of the exogenous variable is useful in forecasting rainfall.

3.8.1. Stationary Series

A stochastic process $\{Y_t\}$ is said to be a stationary if for arbitrary points t_1, t_2, $t_3 ... t_n$, the joint distribution of the random variable $\{Y_{t1}, Yt_2, Yt_3 ... Y_{tn}\}$and $\{Yt_{1+h}, Yt_{2+h} ... Ytn_{+h}\}$ are the same. Observed series was tested for the stationary using Argument Dickey Fuller Test (ADF). This was developed by Dickey and Fuller (1979). This is used to test whether a unit root is present in an autoregressive model.

3.8.2. ARIMA Modeling

Box-Jenkings Auto-regressive Moving Average (ARMA) is a one of the most popular techniques used for rainfall forecasting. An autoregressive model of order p is typically classified as AR (p) and a moving average model with q terms is known as an MA (q). A model that consists of p autoregressive terms and q moving average terms is called ARMA (p, q). Usually, the original time series employs a lag operator B to define the ARMA (p, q) and model may be written as

$$\varphi(B)y_t = \theta(B)\varepsilon_t$$

B is the backward shift operator defined as $B^k y_t = y_{t-k}$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3 - .... - \varphi_p B^p$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + ... + \theta_q B^q$$

The series should be a stationary to model ARMA (p, q). If it is non stationary, the series should be transformed into a stationary series by getting differencing d. It is defined as ARIMA(p,d,q) and simply it can be written by using a back shift operator,

$$\varphi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t$$

3.8.3. SARIMA Modeling

If a time series exhibit a periodic behavior within certain time intervals then the series are said to be a seasonal time series. Those series can be model using seasonal ARIMA model and can be denoted by SARIMA(p,d,q)× $(P,D,Q)_s$. The formula can be formed as;

$$\varphi(B)\Phi(B^s)(1 - B)^d (1 - B^s)^D y_t = \theta(B)\Theta(B^s)\varepsilon_t$$
$$\Phi(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \Phi_3 B^{3s} - .... - \Phi_{Ps} B^{Ps}$$
$$\Theta(B) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \Theta_3 B^{3s} + ... + \Theta_Q B^{Qs}$$

Where $\varphi(B), \theta(B), \Phi(B^s)$ and $\Theta(B^s)$ are polynomial of order p, q, P and Q respectively. p and P are the order of non seasonal and seasonal autoregressive and q and Q are the order of non seasonal and seasonal moving averages. Also, d and D are the number of non seasonal and seasonal differences and s is the length of season.

3.8.4. Concept of ARCH/GARCH Modeling

The Autoregressive Conditional Heteroskedasticity (ARCH) model was first introduced by Engle (1982) for modeling the time dependent conditional variance. This model was generalized by Bollerslev (1986) called as Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. Time dependent variance or conditional variance commonly called as heteroskedasticity cannot be captured by the ARIMA/SARIMA models. Thus, GARCH models are utilized to capture the conditional variance existed from the residuals derived which from the ARIMA/SARIMA models. The GARCH (p, q) model can be written as;

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \alpha_3\varepsilon_{t-3}^2 + ... + \alpha_r\varepsilon_{t-r}^2 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2 + ... + \beta_s\sigma_{t-s}^2$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r}\alpha_i\,\varepsilon_{t-i}^2 + \sum_{i=1}^{s}\beta_i\sigma_{t-i}^2$$

where $\varepsilon_t = \sigma_t\, e_t$, $e_t \sim iiDN(0,1)$ and $V(\varepsilon_t) = \sigma_t^2$

Here, ε_t indicates the uncorrelated residuals of the SARIMA model that have time dependent variance while e_t is a random variable i.i.d with mean zero and variance 1. Thus, in SARIMA-GARCH model, the conditional mean is described by the SARIMA while conditional variance is described by GARCH model.

3.8.5. Testing for the Serial Correlation

The existence of serial autocorrelation violates the standard assumption in ARIMA/SARIMA models. Estimates and forecast values are no longer efficient and the estimates are biased and inconsistence when ignoring the serial correlations. Breusch-Godfrey serial correlation Lagrange Multiplier (LM) test is used to check the serial correlation of a given series up to specific lag.

3.8.6. Testing for the ARCH Effect

Time dependent variance cannot be tested using ACF of the residuals. Thus, squared residuals from the mean model is used to identified the heteroskedasticity of the residuals and this is known as the ARCH effect. If the residuals exist an ARCH effect, the Lagrange multiplier test is used. Initially, estimate the mean equation:

$$R_t = c + u_t$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + ... + \gamma_q u_{t-q}^2$$

The corresponding hypothesis are;

$$H_0 = \gamma_0 = \gamma_1 = \gamma_2 = ... = \gamma_q = 0 \text{ Vs } \quad H_1 = \text{at least one } \gamma_i \neq 0$$

The test statistic is defined as TR^2 (The multiplication of the number of observation and the coefficient of the multiple correlation) and under H_0, this statistic follows the chi squared distribution with q degree of freedom (Engle, 1982).

3.8.7. VAR Modeling

Vector Autoregressive Model (VAR) are used for multivariate time series and its structure is a linear combination of past lags of the same variable along with the past lags of the other variables. This type of models examines the dynamic relationship among the interrelated variables. The $\{Y_t\}$ is a VAR process order 1 [VAR (1)] and for k=2 (k denoted as number of exogenous variables) model can be written as;

$$y_{1t} = \varphi_{10} + \varphi_{11} y_{1,t-1} + \varphi_{12} y_{2,t-1} + \varepsilon_{1t}$$

$$y_{2t} = \varphi_{20} + \varphi_{21} y_{1,t-1} + \varphi_{22} y_{2,t-1} + \varepsilon_{2t}$$

3.8.8. Granger Causality Test

This test is used to find the direction of the relationship among set of time series. The causality is away to investigate two variables in a time series. If the variable X is necessary to forecast the variable Y, then X is said to Granger Cause Y.
The corresponding null hypothesis is,

H_o: X does not Granger Cause Y Vs H_1: X Granger Cause Y

3.9. Modeling Weekly Rainfall Using Novel Approach

Modeling rainfall becomes a demanding assignment since the complexity of rainfall pattern has changed day by day. It is noted that the rainfall in Sri Lanka shows the erratic variation. Thus, it cannot expect high forecasting accuracy by modeling rainfall using conventional approach. Accordingly, we have to move to a new technique to address this issue. Relatively, few measures have been taken to perform the modeling of rainfall in the context of long memory. This study provides an assessment of such a phenomenon by fitting an appropriate time series model by counting the long memory features. The long-range dependency model is allowed to take fractional values for the differencing. According to the Granger and Joyeux (1980), the fractional differencing is the infinite filter that corresponding to the expansion of $(1\text{-}B)^d$, where B is the backwards shift operator while d is the fractional differing parameter. However, according to the Hosking (1981), the fractional differencing operator can be defined as an infinite binomial series expansion in power of the backward shift operator. There are several parameters estimation methods for the long memory parameter d were proposed by the researchers. The spectrum based semi parametric whittle estimation method, regression method, wavelet-based method, approximately maximum-likelihood methods based on truncations of the infinite autoregressive expansion of the process and the truncation of the infinite moving average expansion of the process and exact maximum likelihood method with the Cholesky decomposition and with the Durbin-Levinson algorithm are some model parameter estimation methods for long memory models.

A long-range dependency model is proposed to fit weekly rainfall data to explore characteristics of persistence through an unbounded spectral density. Since the weekly rainfall exhibited the persistence, initially, autoregressive fractionally integrated moving average (ARFIMA) model is fitted. The exact maximum-likelihood method with Durbin-Levinson algorithm was utilized to estimate the long memory parameter of the model and this was not tested for the previous rainfall studies. However, a Monte Carlo simulation was carried out with different fractionally differing parameters to measure the suitability of the method for parameter estimation. Best fitted model is chosen based on the minimum of the mean absolute error.

Careful examination of the data exhibits periodic fluctuations as an additional feature. Since, the rainfall series exhibit periodic variations and persistence, a seasonal autoregressive fractionally integrated moving average (SARFIMA) model is fitted to weekly rainfall series. Here also used MLE method for the parameter estimation. Same as above, Monte Carlo simulation was done with different seasonal and non seasonal fractionally differencing parameters to measure the aptness of the method for parameter estimation.

In addition to the observed series, the deseasonalized series also considered and fitted long memory model for the purpose of the improve the forecasting accuracy. Since the heteroskedasticity existence in the all above models, variance models are developed to the above long-range dependency models to gain good modeling accuracy.

The number of long range dependency models (5); ARFIMA, ARFIMA for the deseasonalized data, ARFIMA-GARCH, ARFIMA-GARCH for the deseasonalized and adjusted SARFIMA-GARCH model are developed for the weekly rainfall series and selected best fitted model to describe the features of the weekly rainfall by accounting forecasting performance of the next year.

3.9.1. The Discrepancy Between Short and Long Memory Series

Let assume that the process $\{Y_t\}$is a stationary time series with autocorrelation $\rho(k) = corr(Y_t, Y_{t+k})$ and the normalized spectral density function is $f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\rho_k e^{-i\omega\omega}$; $-\pi < \omega < \pi$ where ω is the Fourier frequency then can be identified the following differences between short memory and long memory series.

Table 3.6: The difference between short and long memory series

Short Memory Series	Long Memory Series
ρ_k is exponentially decay	ρ_k is hyperbolically decay
$\rho_k \sim r^k$ for $\lvert r \rvert < 1$	$\rho_k \sim k^{-d}$ for $d > 0$
$\sum \rho_k < \infty$	$\sum \lvert \rho_k \rvert = \infty$
$\lim_{\omega \to 0} f(\omega)$ exsist and bounded	$\lim_{\omega \to 0} f(\omega)$ not exsist or unbounded

If the shape of the auto correlation function in between exponentially and hyperbolically called as intermediate memory series.

CHAPTER 4

EXPLORATORY DATA ANALYSIS

Analysis of pattern of weekly rainfall would enhance the management of water resources which enables us to face the impact of climate change. A detailed explanatory data analysis was carried out to explore the features of rainfall in this chapter. The rainfall characteristics are discussed at different time scales such as annual, seasonal, monthly and weekly. Also, features of the exogenous variables is described in this chapter.

4.1. Descriptive Analysis of Annual Rainfall

The summary statistics of fifty-six years of annual rainfall data are presented in Table 4.1. Also Figure 4.1 depicts the annual rainfall trend in Colombo city during the study period.

Table 4.1: The summary statistics of annual rainfall total (in mm) for the period of 56 years (1960-2015)

Number of Years	Mean	SD	Median	Minimum	Maximum	CV (%)
56	2402.2	456.1	2395.4	1456.6 (1986)	3934.5 (1963)	18.99

During the period of 1960 to 2015, the annual rainfall of Colombo city was varied from 1456.6 mm to 3934.5 mm. The mean annual rainfall over the 56 year period was 2402.2 mm with a coefficient of variance of 19% confirms that the less variation in annual rainfall. The minimum annual rainfall amount was recorded in 1986 while maximum rainfall was reported in 1963.

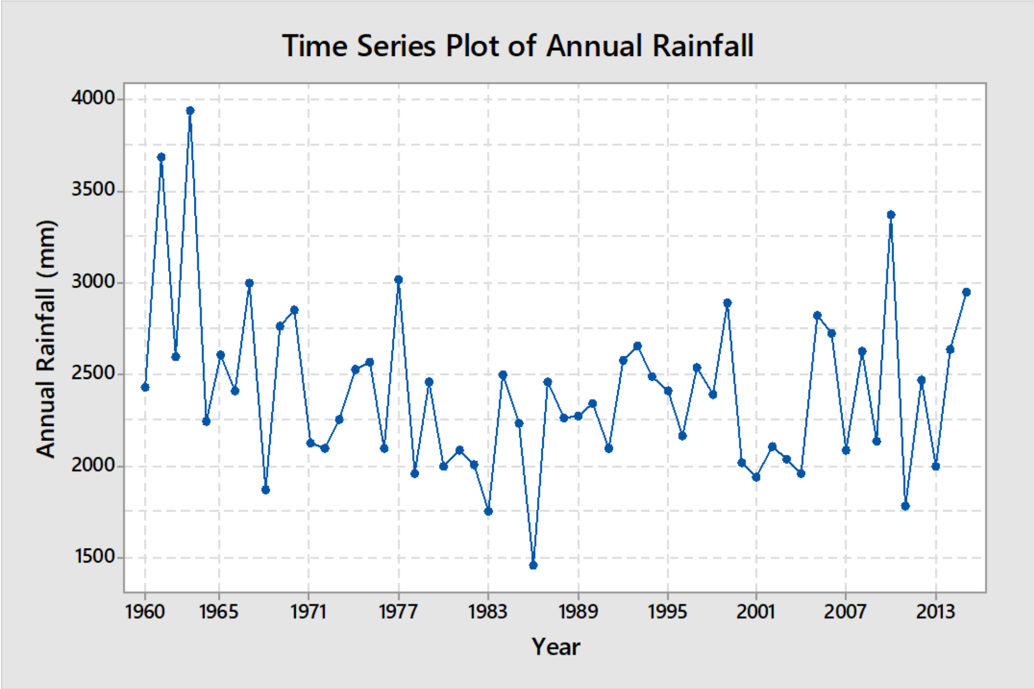

Figure 4.1: Annual rainfall in Colombo city in mm (1960-2015)

Figure 4.1 illustrates much slight decreasing rainfall pattern during the 56 years period. However, the variation in annual rainfall pattern like to be much consistent during the study period. All the years except 1986 enriched from the rainfall with more than 1750mm amount in the Colombo city.

4.2. Descriptive Analysis of Seasonal Rainfall

The rainfall patterns of the country are predominantly governed by the seasonally varying monsoon system. Rainy periods of the country mainly have been classified into four seasons. Two monsoon periods and two inter monsoon periods. The Colombo city is located in the Western part of the country, Sri Lanka. Due to the geographical location, the city of Colombo is influenced by erratic rainfall mainly during two seasons namely, SWM and SIM. The summary statistics of seasonal rainfall are presented in Table 4.2 and the seasonal rainfall behavior was graphically presented from the Figure 4.2 and Figure 4.3.

Table 4.2: The summary statistics of seasonal rainfall total (in mm) from 1960 to 2015

Season	Mean	SD	CV(%)	Median	Min	Max	Mean Intensity
First Inter Monsoon (FIM) *(March to April)*	374.8	158.5	42.3	336.9	101.8 (2004)	736.8 (1961)	6.2
South West Monsoon (SWM) *(May to September)*	1024.1	256.3	25.1	992.8	509.8 (1986)	1737.8 (1963)	6.7
Second Inter Monsoon (SIM) *(October to November)*	695.4	226.7	32.6	684.1	221.9 (1986)	1264.4 (2005)	11.4
North East Monsoon (NEM) *(December to February)*	302.1	155.6	51.5	295.2	58.4 (1981)	631.7 (2014)	3.4

Parenthesis indicates the year of which minimum or maximum occurred

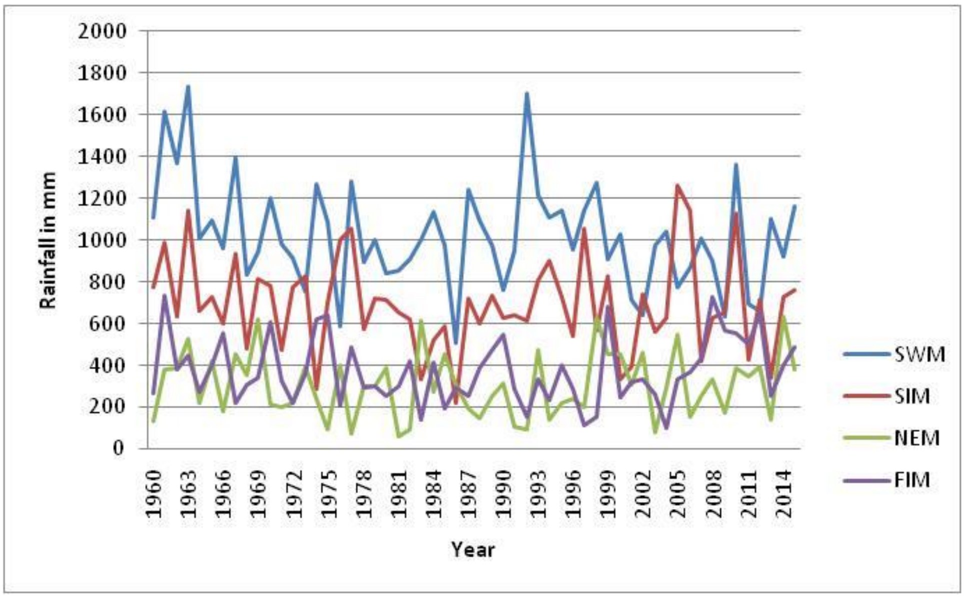

Figure 4.2: Seasonal rainfall during the time span from the 1960 to 2015

According to the Figure 4.2, the highest rainfall amount was received to the city Colombo was reported during the SWM from May to September. Based on the statistics of the Table 4.2, the mean seasonal rainfall of SWM was accounted as 1024.1mm. The maximum seasonal rainfall during the SWM was reported in 1963 with a 1737.8 mm. Since the heavy shower particularly beginning of the SWM can be occurred during this season, sometimes leads to occur the floods events in the city. The season SWM enriched from the rainfall with more than 509 mm. The second highest rainfall amount was received in the city during the SIM from October to November. The highest and lowest rainfall amount during the SIM were reported in years 1986 and 2005 respectively. The SIM showed the highest intensity rainfall which causes to floods and landslides. The lowest rainfall amount was received for the Colombo city during the NEM. The mean seasonal rainfall was 302.1 mm during this season. It is noticed that the considerable rainfall amount were received for the city Colombo during the NEM in the year 2014.

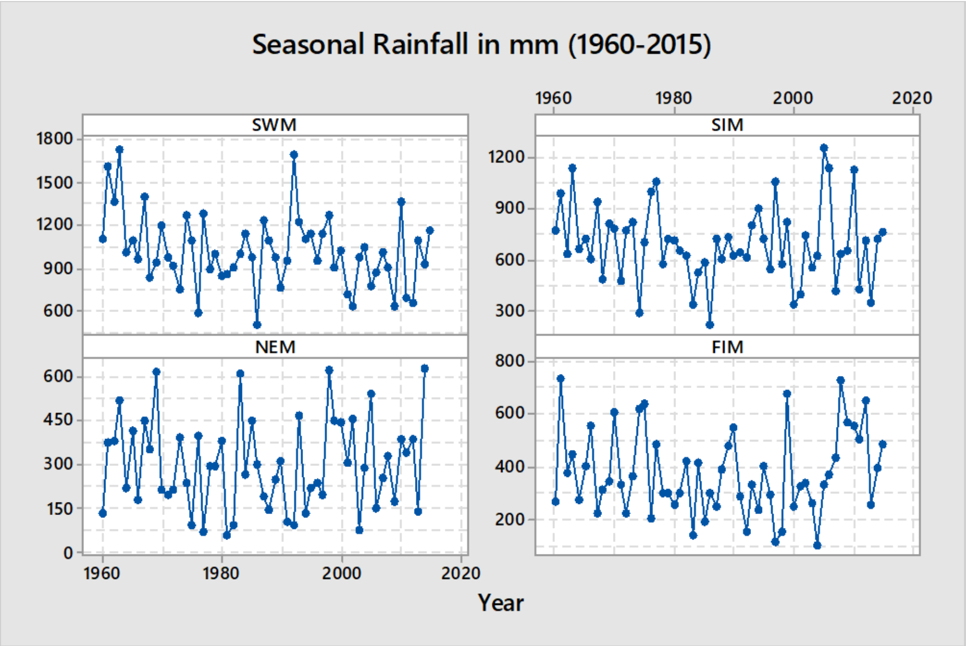

Figure 4.3: Distribution of rainfall for the four seasons from 1960 to 2015

The highest coefficient of the variance is marked as 51.5% for the period of NEM indicated that the much variation in seasonal rainfall than the other three for the city Colombo. According the Figure 4.3, it can be seen that the much peaks of the seasonal rainfall in NEM. However, those peak rainfall values not give much effect on the city due to those were in low range.

4.3. Descriptive Analysis of Monthly Rainfall

The pattern of the monthly rainfall is more beneficial for the many field in the country. To examine the monthly rainfall characteristics, the summary statistics of the monthly rainfall during the time span from 1960 to 2015 is obtained and presented in Table 4.3.

Table 4.3: The summary statistics of monthly rainfall total (in mm) for the period of 56 years (1960-2015)

Month	Mean	SD	CV(%)	Median	Minimum	Maximum
January	70.68	58.78	83.2	68.4	0.0*	211.5 (1984)
February	72.90	61.38	84.2	64.8	0.0 **	221.8 (1963)
March	119.4	78.40	65.6	105.4	0.2 (1996)	356.8 (1977)
April	255.4	134.6	52.7	231.7	70.9 (2004)	617.9 (1999)
May	346.2	173.4	50.1	320.6	83.5 (2014)	750.7 (1977)
June	193.0	93.20	48.3	183.6	63.8 (1986)	602.3 (1992)
July	132.3	101.6	76.8	109.1	10.5 (1986)	482.2 (1998)
August	114.3	91.30	79.9	91.6	2.5 (2001)	435.2 (1962)
September	238.3	138.9	58.3	209.8	29.3 (1976)	631.4 (2015)
October	363.0	169.8	46.8	358.1	94.8 (1983)	871.2 (1977)
November	332.5	164.0	49.3	295.1	52 (2000)	971.5 (2010)
December	163.8	109.5	66.8	156.5	6.6 (2003)	476.5 (2014)

Parenthesis indicates the year of which minimum or maximum occurred

*(1974,1983,1977), ** (1972,1976,1980,1987,1998)

According to the above table, the two months, January and February showed less rainfall than the other months. There was not any rainfall during the months January and February in several years (1974,1983 and 1977 for January and 1972, 1976, 1980, 1987, 1998 for February). The months, April, May, September, October and November are enriched from rainfall while out of the those, May, October and November give heavy shower to the city. The coefficient of variation in monthly rainfall marked noticeably high values than the seasonal and annual. The highest variation of coefficients reported during months January and February which illustrated the not much consistence rainfall. The highest monthly shower in the city recorded in the month November in 2010 which was 971.5 mm during the study period.

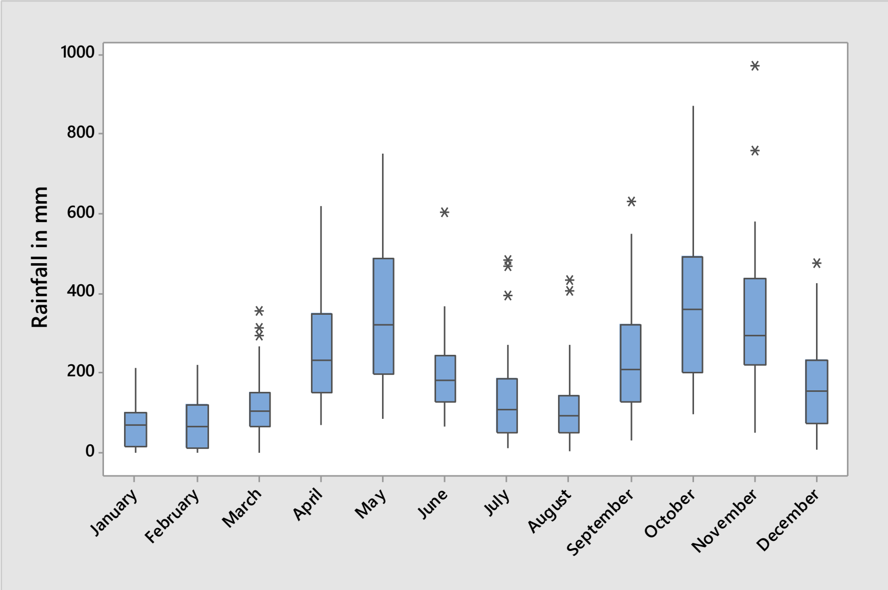

Figure 4.4: Box plot of the monthly rainfall

According to the box plot of the monthly rainfall, it can be clearly seen that an extreme rainfall event in the month November. Some several months also marked extreme rainfall events by which were badly affected on the city. The city of Colombo was enriched by the rainfall significantly during the three months May, October and November.

4.4. Explanatory Analysis of Weekly Rainfall

Descriptive analysis of weekly rainfall with respect to the four seasons are carried out separately.

4.4.1. Descriptive Analysis of Weekly Rainfall for SWM

In order to examine the pattern and behavior of the weekly rainfall the descriptive statistics were obtained and those were presented by seasons. The corresponding summary statistics of the weekly rainfall pertaining to the SWM is presented in Table 4.4 and the box plot of the weekly rainfall in SWM is depicted from the Figure 4.5.

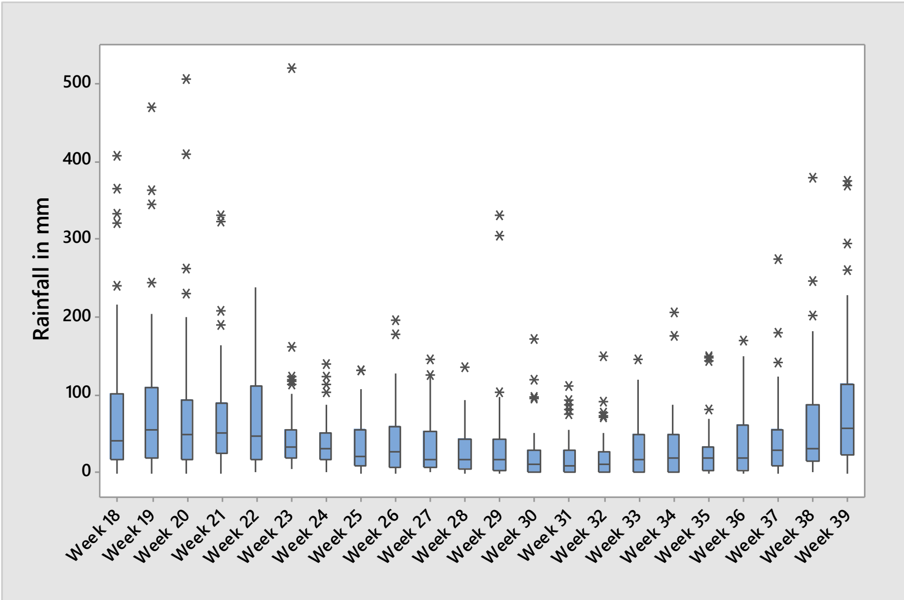

Figure 4.5: Box plot of the weekly rainfall pertaining to the SWM

Table 4.4: The summary statistics of weekly rainfall total pertaining to SWM (week 18-39) from 1960 to 2015

Week No and Period:	Mean	SD	CV(%)	Median	Min	Max
18 (April 30- May 06)	81.6	96.1	117.8	42.0	0.0	407.7
19 (May 07- 13)	86.2	96.4	111.9	55.6	0.0	470.3
20 (May 14-20)	75.5	96.7	128.2	48.5	0.0	506.8
21 (May 21-27)	69.5	69.2	99.5	50.7	0.0	331.8
22 (May 28-June 03)	69.0	61.8	89.6	48.0	0.5	239.1
23 (June 04-10)	52.5	72.7	138.4	33.1	5.1	519.8
24 (June 11-17)	40.6	32.0	79.0	32.0	2.0	141.0
25 (June 18-24)	35.8	35.5	99.3	20.8	0.0	132.8
26 (June 25-July 01)	39.5	41.9	106.2	28.2	0.0	196.4
27 (July 02-08)	34.2	36.3	106.0	17.6	0.1	146.3
28 (July 09-15)	29.0	32.3	111.2	16.2	0.0	135.2
29 (July 16-22)	37.2	60.9	163.6	17.9	0.0	331.6
30 (July 23-29)	22.7	32.3	141.9	11.9	0.0	173 .0
31 (July 30-August 05)	19.5	26.5	135.9	8.6	0.0	111.6
32 (August 06-12)	21.5	28.7	133.7	11.6	0.0	150.4
33 (August 13-19)	30.8	35.0	113.7	17.9	0.0	146.6
34 (August 20-26)	29.4	40.2	136.6	20.1	0.0	205.9
35 (August 27-September 02)	27.7	35.4	128.1	19.4	0.0	150.4
36 (September 03-09)	36.1	43.6	120.7	19.3	0.0	170.1
37 (September 10-16)	44.3	52.7	119.0	28.8	0.0	275.6
38 (September 17-23)	62.1	74.6	120.1	32.1	0.3	379.9
39 (September 24-30)	86.6	91.1	105.2	56.5	0.0	376.4

Figure 4.5 illustrates that the distributions of the weekly rainfall data during SWM and all those distributions are positive skewed. Also, this depicts that there is a significant high variation in weekly rainfall in the week 18-23 while much lower variation in weeks 24 and 25. The lowest (19.5mm) and highest (86.6mm) mean weekly rainfall were reported in the 31st week and 39th week respectively. It is noted that mean weekly rainfall in SWM gradually decreases from 19th week to 31st week and then the pattern has changed to increase. An almost similar pattern was observed in median weekly rainfall also. Maximum weekly rainfall was recorded in the 23rd week (519.8mm) in 1992. The weeks 23, 24, 27 and 38 received rainfall continuously over the 56 year. It is also noted that the 82% of the weeks in SWM marked more than 100% coefficient of variation which indicates the high variation in weekly rainfall.

4.4.2. Descriptive Analysis of Weekly Rainfall for SIM

The summary statistics of the weekly rainfall during the period of SIM is presented in Table 4.5 and the box plot of the distribution of the weekly rainfall in SIM is illustrated from the Figure 4.6 respectively.

Table 4.5: The summary statistics of weekly rainfall total pertaining to SIM (week 40- 48) from 1960 to 2015

Week No and Period:	Mean	SD	CV(%)	Median	Min	Max
40 (October 01- 07)	51.9	52.1	100.3	35.1	0.2	237.2
41 (October 08- 14)	81.7	89.4	109.3	46.4	0.0	370.1
42 (October 15- 21)	98.6	93.4	94.7	71.1	0.0	413.7
43 (October 22- 28)	89.7	74.2	82.8	73.5	0.0	362.4
44 (October 29- November 04)	102.9	78.1	76.0	82.9	0.0	337.0
45 (November 05 -11)	91.1	84.9	93.1	73.7	0.0	464.0
46 (November 12-18)	76.8	76.6	99.7	53.4	0.0	347.3
47 (November 19-25)	62.4	64.9	104.1	53.8	0.0	388.5
48 (November 26-December 02)	55.2	53.7	97.2	32.0	0.0	232.1

Mean weekly rainfall of SIM varies from 51.9 mm to 102.90 mm. The lowest and highest mean week rainfall was reported the 40th week and the 44th week respectively. The highest weekly rainfall in SIM was reported in 2010. It can be seen in a similar pattern of weekly rainfall in mean and median in SIM. Also noted that in all the weeks in SIM higher mean rainfall than median rainfall. The week 40^{th} received rainfall continuously over the period of 1960 to 2015 and minimum of the rainfall of this week reported as 0.2mm. Almost mean and median weekly rainfall during SIM is much higher than the weekly rainfall in SWM. It can be seen that the high variability in weekly rainfall at the middle of the seasons. However, coefficient of variance values indicates that the considerable much low fluctuation in the weekly rainfall pertaining to SIM than SWM.

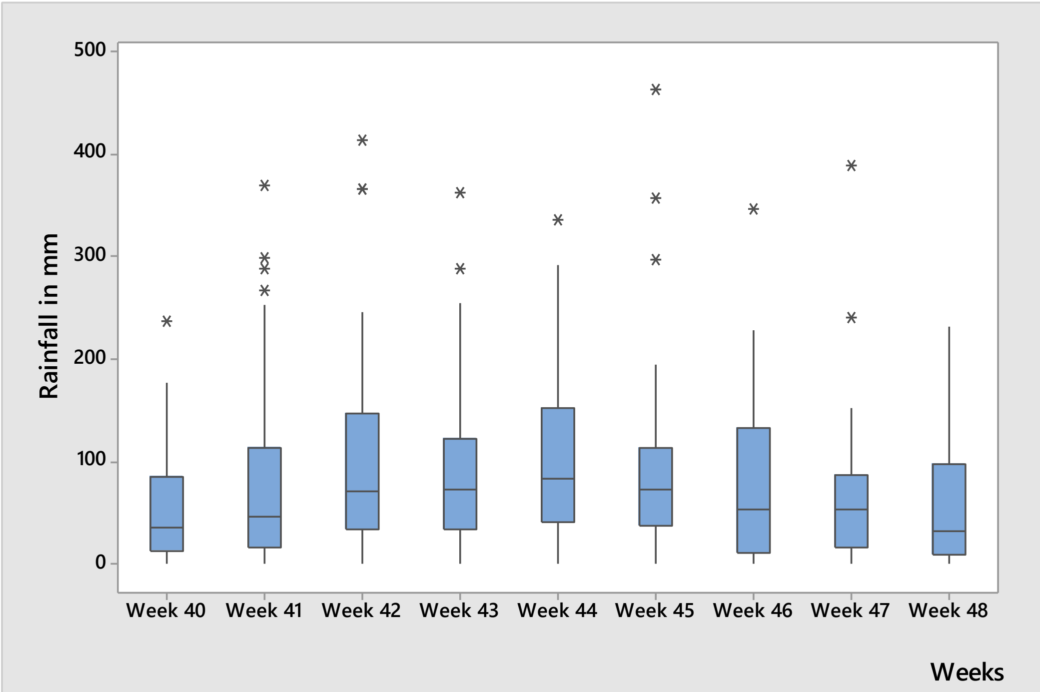

Figure 4.6: Box plot of the weekly rainfall pertaining to the SIM

Figure 4.6 depicts the distributions of the weekly rainfall data during SIM. All the weekly rainfall, pertaining to the SIM were positive skewed with longer tail to the right. Beginning and the withdrawal of the season shows much low rainfall than the middle. Many extreme weekly rainfall events can be seen at the weeks 41-47.

4.4.3. Descriptive Analysis of Weekly Rainfall for FIM

Table 4.6 gives the summary statistics of the weekly rainfall during the FIM period. Also, distribution of the weekly rainfall in this period is presented from the Figure 4.7.

Table 4.6: The summary statistics of weekly rainfall total pertaining to FIM (Week 10-17) for the period of 56 years.

Week No and period:	Mean	SD	CV(%)	Median	Min	Max
10 (March 05-11)	22.3	27.0	121.0	12.9	0.0	113.2
11 (March 12-18)	29.1	36.7	126.0	14.9	0.0	141.5
12 (March 19-25)	26.0	38.3	147.4	13.4	0.0	220.9
13 (March 26 -April 01)	31.5	28.9	91.8	24.9	0.0	129.4
14 (April 02-08)	50.1	52.9	105.5	34.2	0.0	227.2
15 (April 09-15)	49.7	47.3	95.2	38.4	0.0	261.4
16 (April 16-22)	72.0	67.6	93.9	56.8	0.0	336.3
17 (April 23-29)	71.6	63.3	88.4	55.6	0.2	280.6

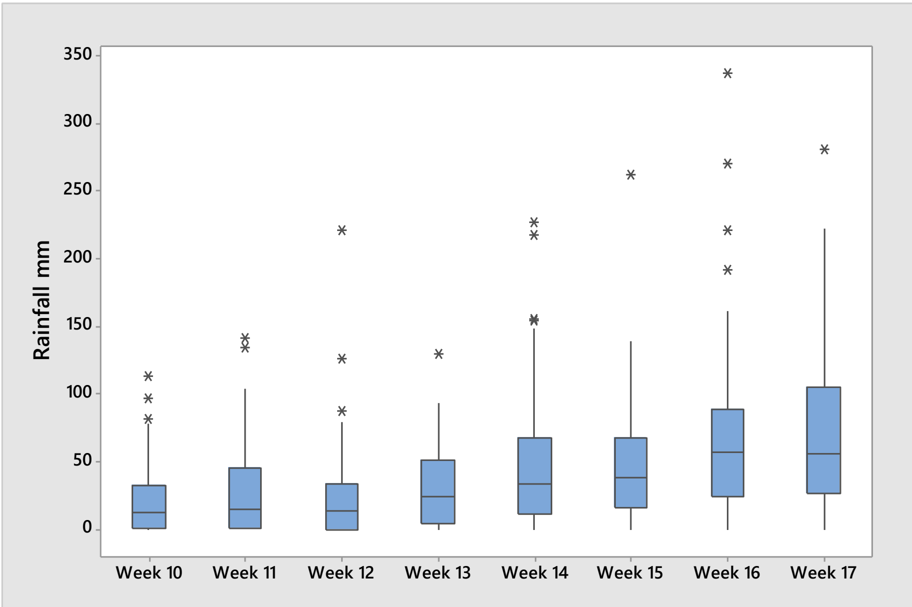

Figure 4.7: Box plot of the weekly rainfall pertaining to the FIM

According to the Figure 4.7, there is an increasing trend of weekly rainfall since beginning of the season. Positive skewed distribution can be clearly identified in weekly rainfall in all the weeks pertaining to the FIM. The maximum weekly rainfall in FIM was recorded in 1999 (336.3mm) at the 16^{th} week. It is noted that the week 17^{th} received continuous rainfall over the 56 years. The high mean rainfall was observed at the weeks 16-17. The high coefficient of variance implies that the much heavy variation in weekly rainfall during the FIM. However, the considerable low shower can be identified during the FIM than the SWM and the SIM in Colombo city.

4.4.4. Descriptive Analysis of Weekly Rainfall for NEM

The summary statistics of the weekly rainfall in NEM is presented in Table 4.7.

Table 4.7: The summary statistics of weekly rainfall total pertaining to NEM (week 49-52 and week 1-9) for the period of 56 years (1960-2015)

Week No:	Mean	SD	CV(%)	Median	Min	Max
49 (December 03-09)	47.1	55.0	116.9	30.9	0.0	259.2 (2010)
50 (December 10-16)	39.3	40.5	102.9	27.6	0.0	172.7 (2015)
51 (December 17-23)	30.2	32.0	105.9	17.7	0.0	135.7 (1965)
52 (December 24-31)	29.9	46.1	154.4	10.7	0.0	268.7 (1969)
1 (January 01 -07)	16.1	26.1	162.9	4.4	0.0	121.4 (1986)
2 (January 08 -14)	24.0	36.0	150.2	4.3	0.0	136.5 (1969)
3 (January 15 -21)	11.3	22.8	201.3	0.1	0.0	108.3 (1999)
4 (January 22 -28)	12.2	20.3	166.2	1.1	0.0	70.8 (2001)
5 (January 29 -February 04)	14.5	24.0	165.2	1.2	0.0	110.3 (1990)
6 (February 05 -11)	21.6	31.9	147.9	2.8	0.0	135.1 (1984)
7 (February 12 -18)	13.6	23.2	170.8	0.7	0.0	103.6 (2012)
8 (February 19-25)	20.4	30.6	150.1	9.2	0.0	154.0(1964)
9 (February 26 -March 04)	25.4	34.0	133.8	11.1	0.0	127.0 (1974)

According to the Table 4.7, maximum weekly rainfall was marked in 1969 (268.7 mm) during the week 52 in NEM. The mean weekly rainfall during this season showed much low amount than the SWM and SIM. Most of the mean weekly rainfall was less than 20 mm. It is noted that the much considerable rainfall in weeks 49-52 than the other weeks in the NEM. Also, all the weeks illustrated the more than 100% coefficient of variation indicated that the very high variation in weekly rainfall during this season. It is noticed that the coefficient of the variation in week 3 was 201.3% implies that the much heavy variation in weekly rainfall during this week. This is the week which have largest coefficient of variation among the 52 weeks.

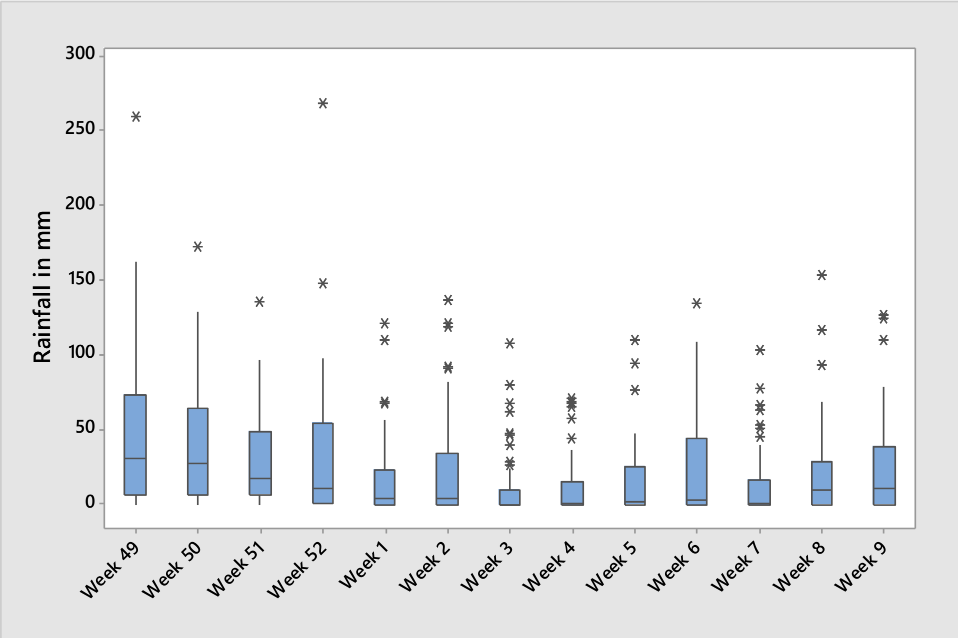

Figure 4.8: Box plot of the weekly rainfall pertaining to the NEM

Figure 4.8 depicts the distribution of weekly rainfall in NEM. As we expect, all weekly rainfall distribution positive skewed while many extreme rainfall amounts can be seen in weekly rainfall in NEM. It can be clearly seen that the much heavy shower in weekly rainfall at the beginning of the season. As above mentioned, Figure 4.8 also showed the high variability in weekly rainfall during the week3 than the others. However, though it has high variation it gives much low rainfall to the city.

4.5. Descriptive Analysis of the Weekly Temperature

Temperature is considered as one of the exogenous climatic variables in modeling weekly rainfall. To enrich the understanding of the behavior of temperature a descriptive analysis was carried out for minimum, maximum and mean temperature during the period from 1990 to 2014.

4.5.1. Minimum Weekly Temperature

In order to examine the temporal variability of the minimum weekly temperature, the time series plot and the summary statistics were obtained and presented in Figure 4.9 and Table 4.8 respectively.

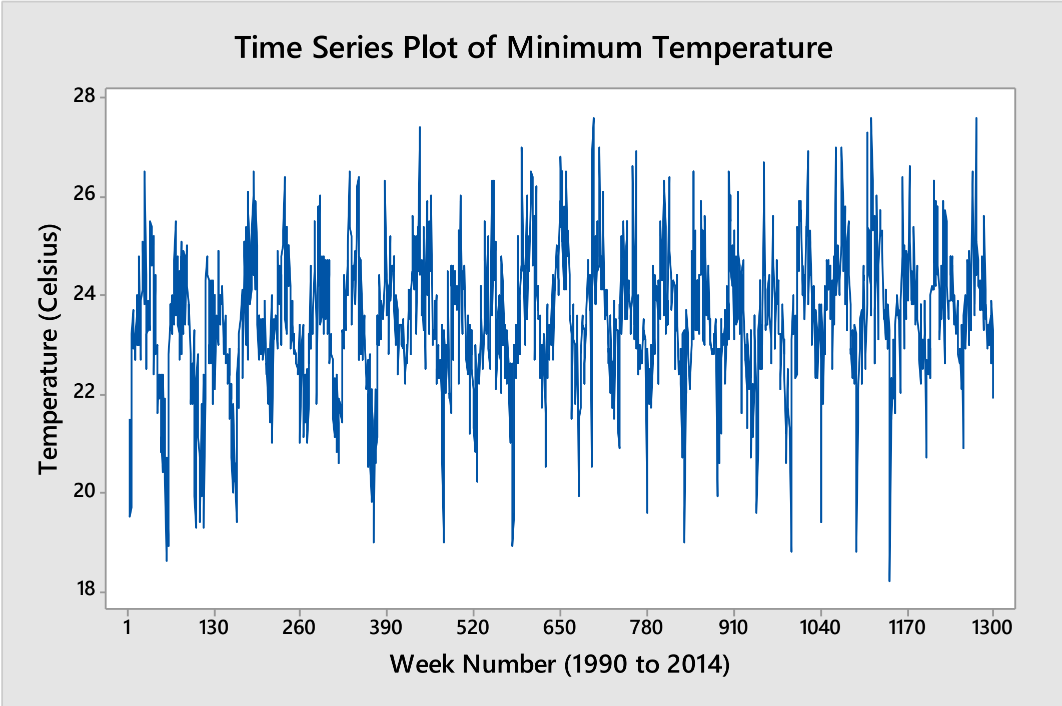

Figure 4.9: The time series plot of the minimum weekly temperature

Table 4.8: The summary statistics of the minimum weekly temperature

Mean	Median	Minimum	Maximum	CV(%)	Skewness
23.4	23.5	18.2	27.6	6.0	-0.29

The Figure 4.9 depicts a random pattern in minimum temperature and based on the Table 4.8, the lowest minimum weekly temperature in Colombo city was recorded as 18.2 ^{0}C during the period of 1990 to 2014. The coefficient of variance (6%) provided evidence to low fluctuation in minimum weekly rainfall while this has a moderate negative skewed distribution. Moreover, the mean minimum weekly temperature was considered to describe the pattern of the minimum temperature over the considered time span.

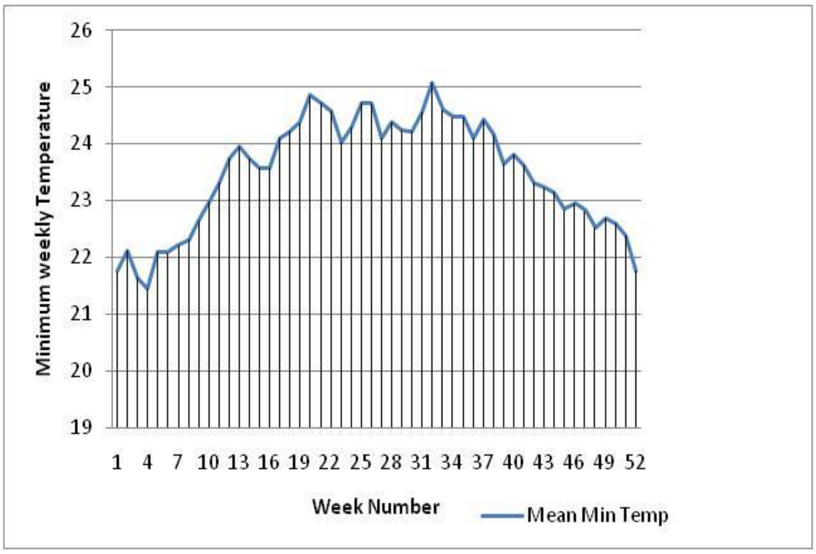

Figure 4.10: Mean minimum weekly temperature

Beginning of the year, the mean minimum weekly temperature indicated much lower value while it was gradually increased till 13th week. It can be clearly seen that the after 31st week, a steady decline until end of the year.

4.5.2. Maximum Weekly Temperature

The Table 4.9 and Figure 4.11 depicts the descriptive statistic of the maximum weekly temperature and the corresponding maximum mean weekly temperature over the 56 years.

Table 4.9: The descriptive statistics of the maximum weekly temperature

Mean	Median	Maximum	Minimum	CV(%)	Skewness
31.8	31.7	36.1	29.6	3.47	0.74

The maximum weekly temperature varies from 29.60C to 36.1^{0}C while it can be identified low fluctuations in maximum weekly temperature (coefficient of variance 3.47%). This is also skewed with a tail to positive.

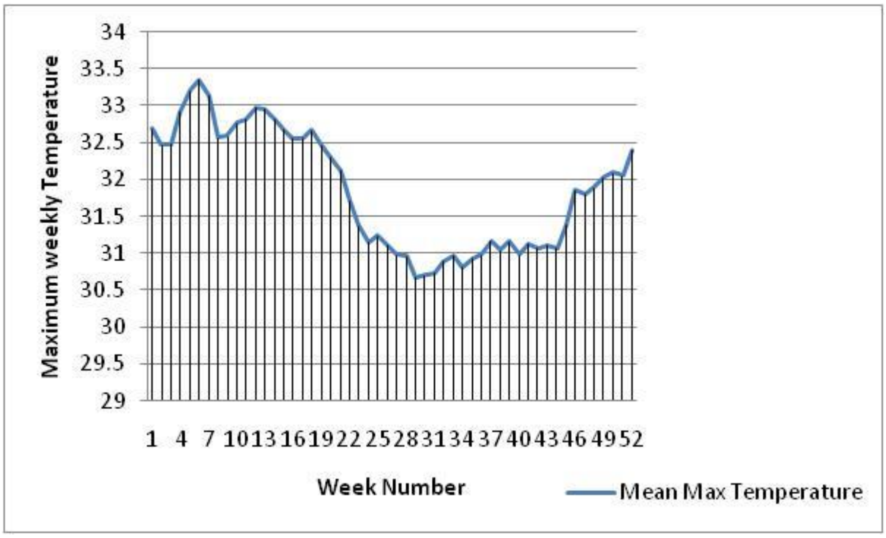

Figure 4.11: Mean maximum weekly temperature

The Figure 4.11 depicts high mean maximum weekly temperature at the beginning of the year and the pattern has changed to decline from the weeks 16th to 30th. An increasing trend pattern is depicted from the week 30th to until end of the year.

4.5.3. Average Weekly Temperature

The summary statistics of mean weekly temperature were presented in Table 4.10.

Table 4.10: The summary statistics of the average weekly temperature

Mean	Median	Minimum	Maximum	CV(%)	Skewness
27.7	27.6	24.4	30.4	2.84	0.04

The lowest average weekly temperature in Colombo city was reported as 24.4 ^{0}C while the highest was recorded as 30.4^{0}C during the study period. The Table 4.10 shows that the low coefficient of variation than the minimum weekly temperature and this indicated that the much consistency pattern. This has slight positive skewed distribution.

4.6. Descriptive Analysis of the Relative Humidity

Relative humidity is a ratio, express in percent, of the amount of atmospheric moisture present relative to the amount that would be presented if the air were saturated at the given temperature. This is a function of the both moisture content and the temperature. To assess the relationship between relative humidity and rainfall three measurements such as the minimum, maximum and average of relative humidity were considered and those characteristics were described in the following sections.

4.6.1. Minimum Weekly Relative Humidity

The some features of the minimum weekly relative humidity (MinRH) is described using the time series plot (Figure 4.12) along with the summary statistics table of the MinRH. (Table 4.11).

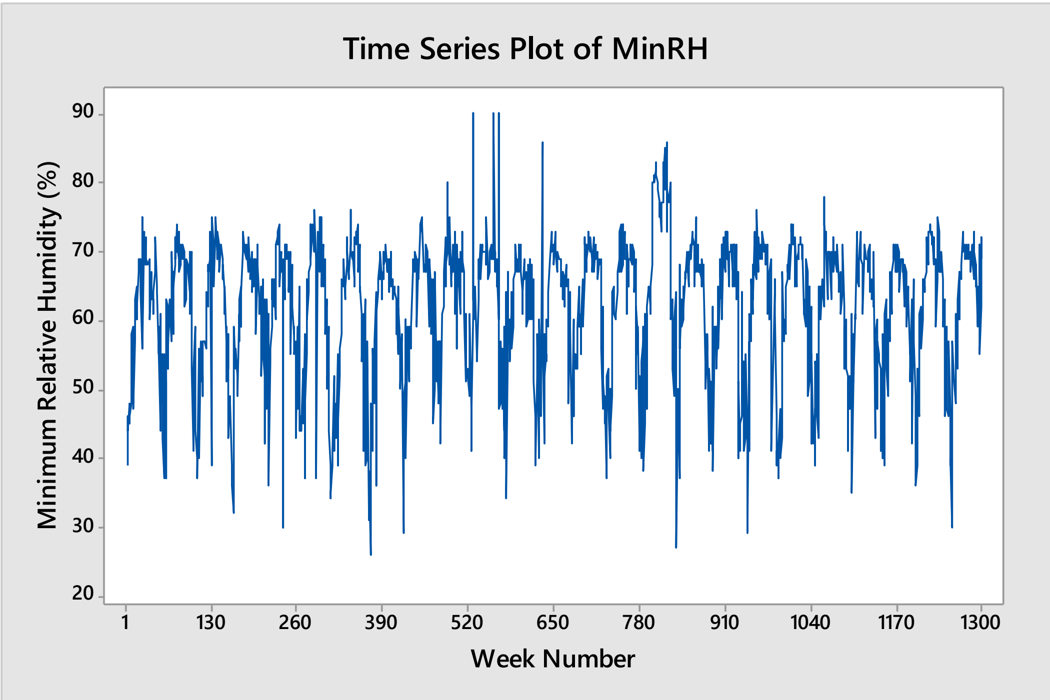

Figure 4.12: The time series plot of the minimum weekly relative humidity

The Figure 4.12 showed the most of the MinRH in between 40 and 70 while some of them were above to 80 and below to 30.

Table 4.11: The descriptive statistics of the minimum weekly relative humidity

Mean	Median	Minimum	Maximum	CV(%)	Skewness
61.87	65.0	26.0	90.0	16.37	-0.82

The MinRH was reported as 26% while maximum was 90% during the considered time span. This is slightly negative skewed distribution while it is noted that the much considerable fluctuation (CV=16.37%). To explain the variation of the minimum relative humidity with respect to the weeks in a year, the mean minimum weekly relative humidity was taken and those are depicted by Figure 4.13.

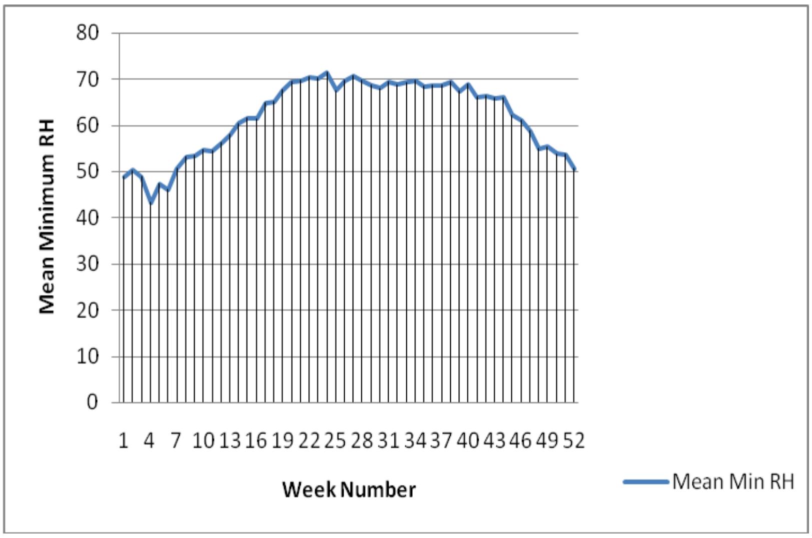

Figure 4.13: Mean minimum weekly relative humidity

According to the Figure 4.13, there was a slight decline in mean minimum weekly relative humidity at the beginning of the year and after 4th week it can be clearly identified the increasing trend until the 25th week. Also, it depicts steady pattern in between the 26-46 weeks and there was a gradual decrease until end of the year from 46th week.

4.6.2. Maximum Weekly Relative Humidity

The summary statistics of the maximum weekly relative humidity (MaxRH) was taken to observe features of the MaxRH and result is presented in Table 4.12.

Table 4.12: The descriptive statistics of the maximum weekly relative humidity

Mean	Median	Minimum	Maximum	CV(%)	Skewness
95.4	81.2	74.0	100.0	2.9	-1.98

The minimum of the maximum RH was reported as 74% while maximum was 100%. This showed the negative skewed distribution with longer tail to the left. However, this showed the lowest fluctuation than the minimum and average relative humidity. To explain the variation of the maximum relative humidity with respect to the weeks in a year, the mean maximum weekly relative humidity was taken and those are depicted by Figure 4.14.

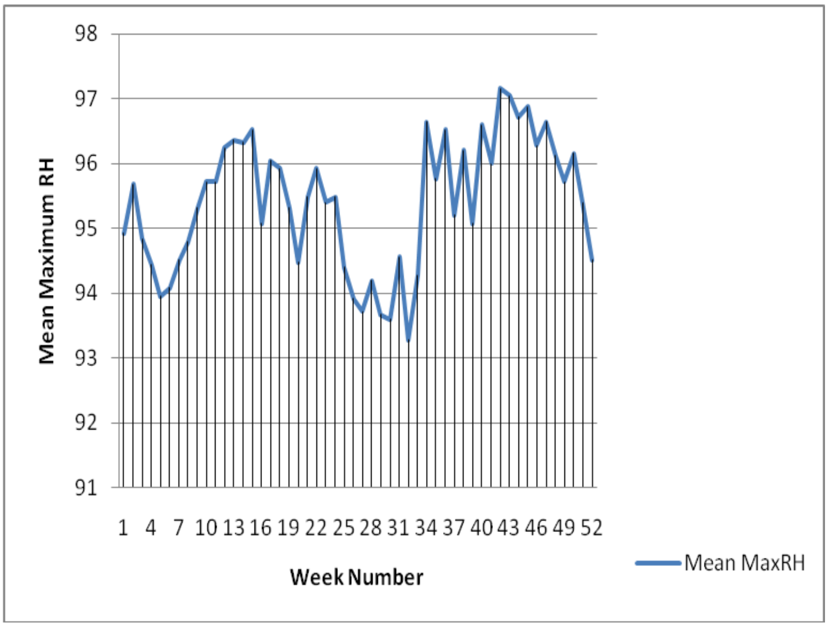

Figure 4.14: Mean maximum weekly relative humidity

According to the above Figure 4.14, there is a decreasing trend in mean maximum relative humidity from 2nd to 5th week. However, after the 5th week, it showed an increasing trend up to 17th week and again displayed the decreasing trend up to the week 32 and changed the pattern again to increasing trend. After that it can be identified decreasing trend till end of the year. Though the trend pattern has been changed, the amount of the change is considerable low.

4.6.3. Average Weekly Relative Humidity

In order to examine the temporal variability of the average weekly relative humidity (AvgRH), time series plot and summary statistics of AvgRH were obtained and result is presented in Figure 4.15 and Table 4.13 respectively.

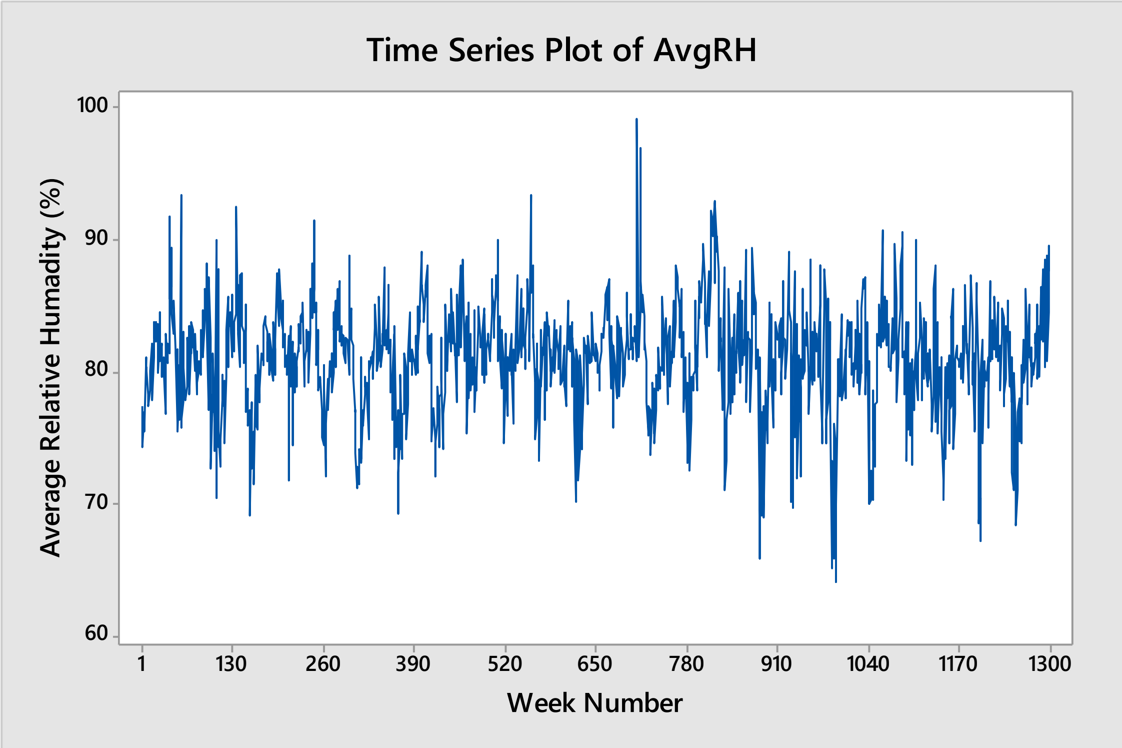

Figure 4.15: The time series plot of the average weekly relative humidity

Based on the above Figure 4.15, there is much low fluctuation in Avg RH than MinRH. Most of the AvgRH in between 75%-85%.

Table 4.13: The descriptive statistics of the average weekly relative humidity

Mean	Median	Minimum	Maximum	CV(%)	Skewness
81.0	81.2	64.0	99.1	5.0	-0.33

The Table 4.13 showed that the minimum average relative humidity as 64% while maximum was 99.1. The coefficient of variation indicated that the much consistent variation compared with the minimum weekly relative humidity. This also showed negative skewed distribution.

4.7. Descriptive Analysis of the Vapor Pressure

Vapor pressure is the amount of pressure of water vapor in the air and it measured by milibars. The maximum amount of moisture that can be in the air is called saturation vapor pressure for a given temperature. The relative humidity gets 100% at the saturation vapor pressure. The characteristics of the minimum, maximum and mean weekly vapor pressure is discussed in detail in the following sections.

4.7.1. Minimum Weekly Vapor Pressure

In order to examine the temporal variability of the minimum weekly vapor pressure (MinVapPres), the time series plot and the summary statistics were obtained and those are presented in Figure 4.16 and Table 4.14 respectively.

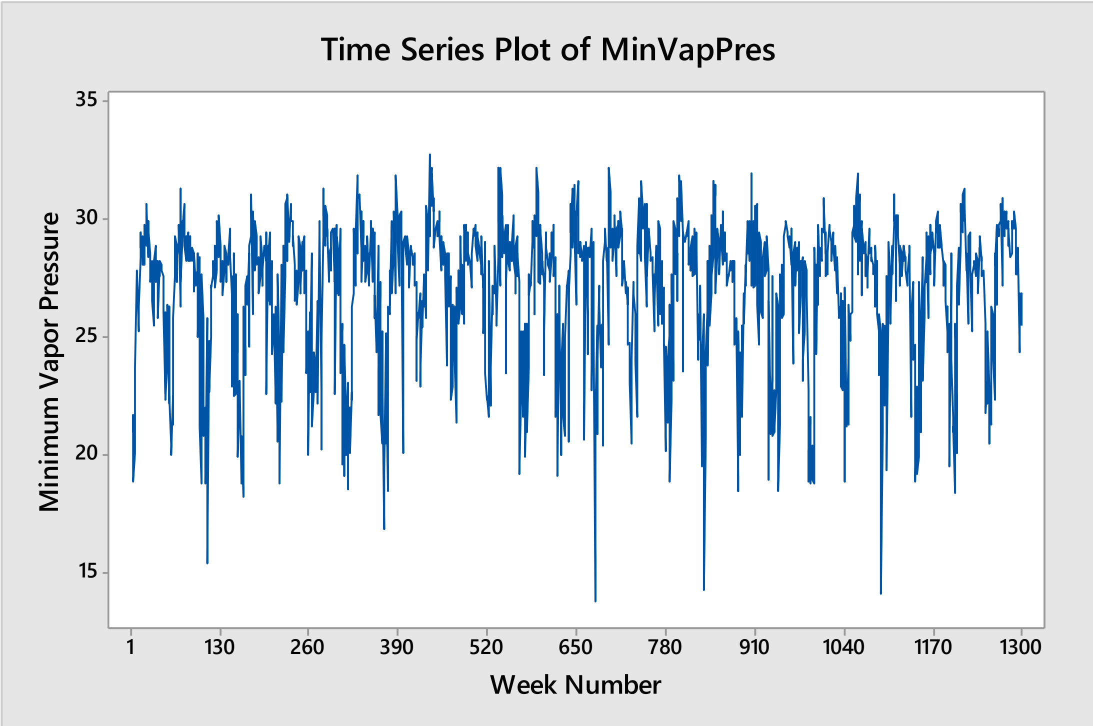

Figure 4.16: The time series plot of the minimum weekly vapor pressure

The Figure 4.16 indicates that the most of the minimum vapor pressure in between 20-34 milibar. Also, it is noted that some weeks' vapor pressure is considerable low from the others.

Table 4.14: The descriptive statistics of the minimum weekly vapor pressure

Mean	Median	Minimum	Maximum	CV(%)	Skewness
27.0	27.9	13.8	32.8	11.1	-1.16

The minimum and the maximum of the minimum vapor pressure were reported as 13.8 milibar and 32.8 milibar respectively. This also showed a negative skewed distribution with longer tail to the left. To explain the variation of the minimum vapor pressure with respect to the weeks in a year, the mean minimum weekly vapor pressure was taken and those are depicted by Figure 4.17.

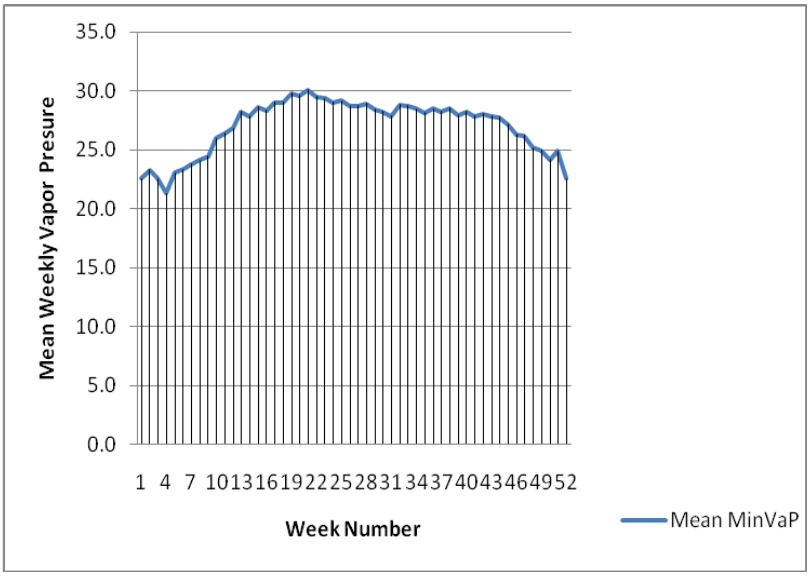

Figure 4.17: Mean minimum weekly vapor Pressure

According to the Figure 4.17, the mean weekly minimum vapor pressure showed increasing trend pattern from week 4 to week 22 and thereafter it depicted the much stable pattern up to week 44. Also, it showed the decreasing trend at the end of the year.

4.7.2. Maximum Weekly Vapor Pressure

In order to examine the temporal variability of the maximum weekly vapor pressure (MaXVapPres) the time series plot and the summary statistics of the variable were obtained and result are presented in Figure 4.18 and Table 4.15 respectively.

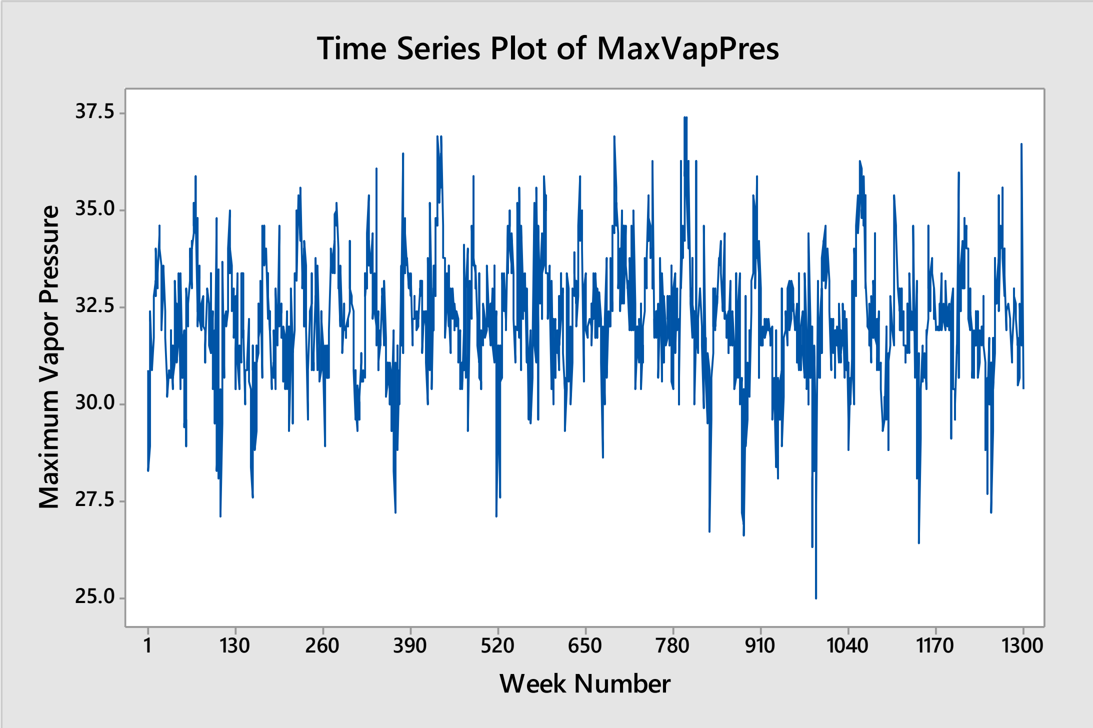

Figure 4.18: The time series plot of the maximum weekly vapor pressure

The Figure 4.18 indicates that the most of the maximum vapor pressure in between 30-34 milibar. Also, it is noted that some weeks' vapor pressure is considerable low from the others.

Table 4.15: The descriptive statistics of the maximum weekly vapor pressure

Mean	Median	Minimum	Maximum	CV(%)	Skewness
32.2	32.2	25.0	37.4	5.0	-0.14

The maximum vapor pressure was varied from 25.0 milibar to 37.4 milibar. It is seems to be much consistent pattern in maximum weekly vapor pressure when compared with the minimum and average vapor pressure. This is a slightly negative skewed distribution. To explain the variation of the maximum vapor pressure with respect to the weeks in a year, the mean maximum weekly vapor pressure was taken and those are depicted by Figure 4.19.

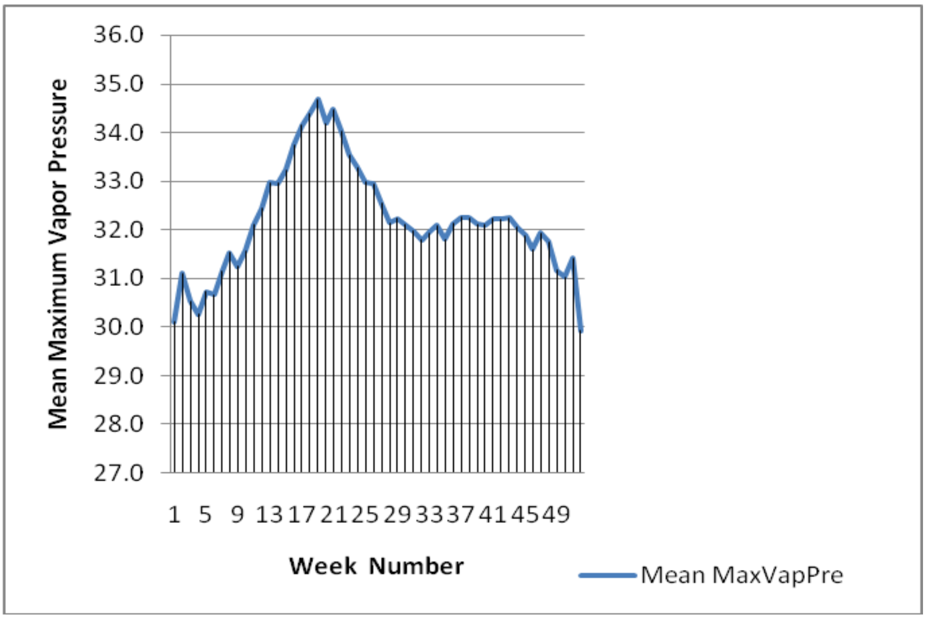

Figure 4.19: Mean maximum weekly vapor Pressure

There is a increasing trend at the beginning of the year and after the week 21 it changed to the decreasing trend pattern.

4.7.3. Average Weekly Vapor Pressure

In order to examine the temporal variability of the average weekly vapor pressure the summary statistics of the variable was obtained and result is presented in Table 4.16.

Table 4.16: The descriptive statistics of the average weekly vapor pressure

Mean	Median	Minimum	Maximum	CV(%)	Skewness
29.8	30.1	22.2	34.5	6.5	-0.91

The average weekly vapor pressure varied from 22.2 milibar to 31.5 milibar. The variation of the series considerable low (CV= 6.5%) compared with the minimum vapor pressure.

4.8. Summary of the Chapter 4

There was a slight decreasing pattern in the annual rainfall in Colombo city. SWM and SIM bring more rainfall to the city than FIM and NEM. During 1960 to 2015, the total rainfall of SWM varied from minimum of 509.8mm (1986) to maximum of 1737.8 mm (1963) with the mean of 1024.1mm and standard deviation of 256.3mm. Also, it is noted that the rainfall with high intensity in SIM. However, rainfall variability is much higher during FIM and NEM than that in other two seasons.

The weekly rainfall by seasons exhibited an increasing trend in weekly rainfall at the beginning as well as withdrawal of the SWM. The coefficient of variation is more than 100% in 82% of weeks in SWM that provided evidence to high fluctuation in weekly rainfall during the SWM. There is a much possibility to form the floods during the beginning and the withdrawal of the SWM since heavy shower along with the much variation in weekly rainfall. In contrast, low rainfall amount was received during the beginning and the end of the SIM. There is much consistence in weekly rainfall was observed during the season SIM than SWM. Though an increasing trend of weekly rainfall can be seen at the beginning of the season FIM, it provided considerable low shower through the season to the city.

CHAPTER 5

COVERAGE PROBABILITY FOR WEEKLY RAINFALL PERCENTILES CONFIDENCE LIMITS

Modeling rainfall percentile is one of the successful techniques that can be used to describe the temporal variability of the rainfall and to highlight the its behavior. However, before modeling the rainfall, it is better to study the temporal variability in weekly rainfall in different aspects. Therefore, the main focus of this chapter is to model weekly rainfall percentile in the context of confidence intervals and study the coverage probability for weekly rainfall percentile confidence intervals.

5.1. Trend Estimation

The parametric trend analysis was carried out for all weeks separately and tested the linear and quadratic trend pattern during the time span from 1960 to 2015. The plots of time series demonstrate the sense about the trend pattern of the weekly rainfall. The time series plots for the randomly selected weeks in SWM and SIM are shown in Figure 5.1 and Figure 5.2 respectively. Also, the remaining plots were shown in Appendix 1.

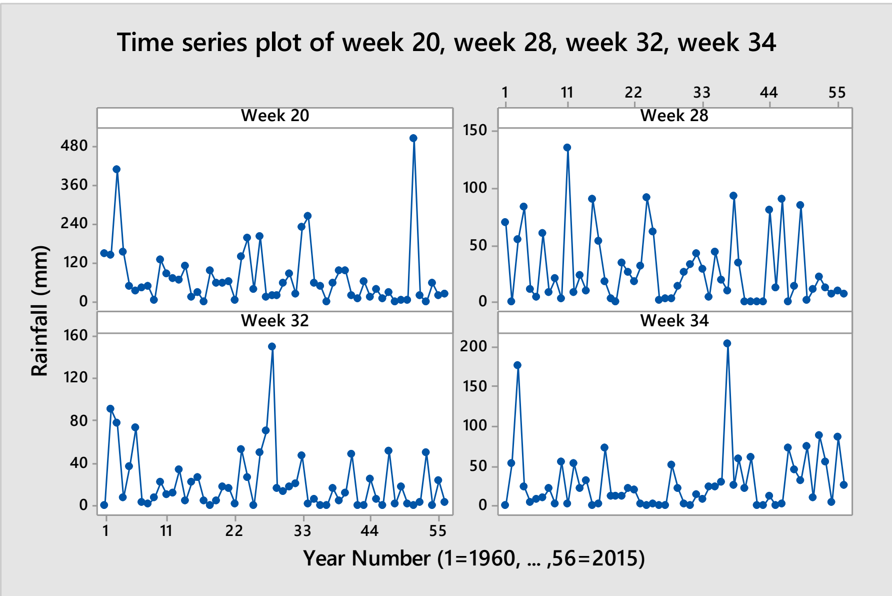

Figure 5.1: The time series plots of the weekly rainfall of the selected weeks (20, 28, 32 and 34) in SWM

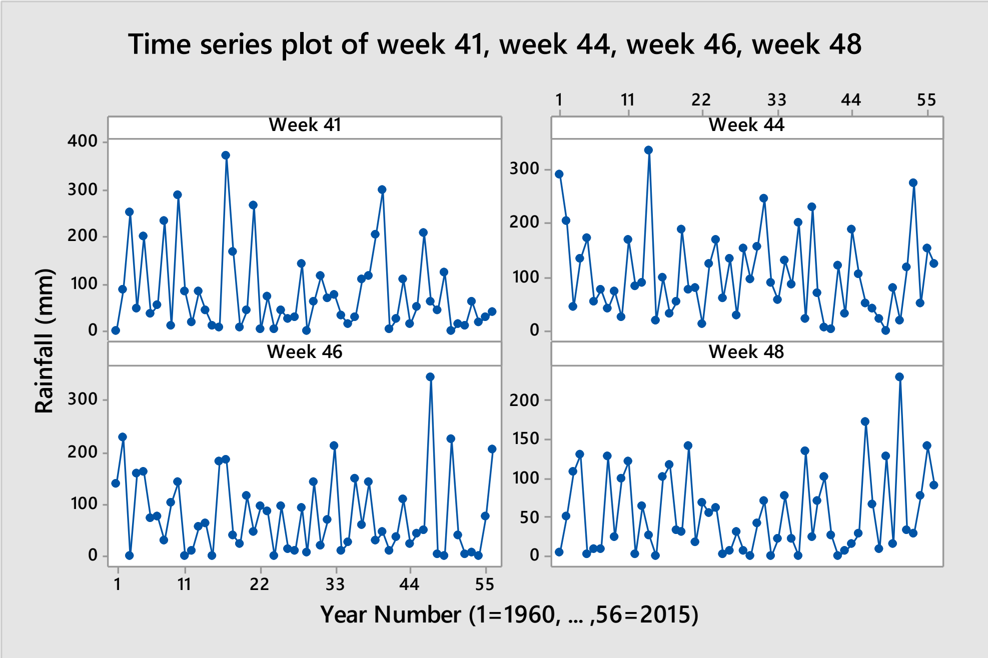

Figure 5.2: The time series plots of the weekly rainfall of the selected weeks (41, 44, 46 and 48) in SIM

The Figure 5.1 and Figure 5.2 do not provide any sense about the presence of trend. The rest of the time series plots also demonstrate the same pattern (Appendix 1). However, to avoid the conclusion subjectively the hypothesis tests were carried out for the parameters of the linear and quadratic trend. The result of the test of parameters in linear trend is shown in Table 5.1.

Table 5.1: The coefficients (slope) of the linear trend along with the P-values

Week	Slope and P-value	Week	Slope and P-value	Week	Slope and P-value	Week	Slope and P-value
1	0.156 (0.4777)	14	0.126 (0.776)	27	-0.394 (0.191)	40	0.478 (0.271)
2	0.223 (0.460)	15	0.360 (0.361)	28	-0.355 (0.186)	41	-1.050 (0.157)
3	0.149 (0.436)	16	0.012 (0.984)	29	-0.369 (0.469)	42	-0.729 (0.350)
4	-0.026 (0.881)	17	-0.240 (0.651)	30	-0.283 (0.293)	43	-0.305 (0.623)
5	0.353 (0.076)	18	-0.596 (0.459)	31	-0.067 (0.762)	44	-0.615 (0.346)
6	0.300 (0.259)	19	-1.831 (0.120)	32	-0.357 (0.134)	45	0.351 (0.622)
7	0.263 (0.171)	20	-1.052 (0.191)	33	0.121 (0.679)	46	-0.427 (0.505)
8	-0.724 (0.103)	21	-0.730 (0.204)	34	0.277 (0.410)	47	0.693 (0.200)
9	-0.157 (0.580)	22	0.798 (0.119)	35	0.206 (0.488)	48	0.392 (0.382)
10	0.193 (0.392)	23	0.090 (0.882)	36	0.632 (0.079)	49	0.577 (0.207)
11	0.133 (0.665)	24	0.411 (0.122)	37	-0.140 (0.752)	50	0.141 (0.678)
12	-0.500 (0.115)	25	-0.035 (0.906)	38	-0.389 (0.533)	51	-0.426 (0.108)
13	0.281 (0.243)	26	-0.447 (0.200)	39	-0.132 (0.863)	52	-0.199 (0.606)

It can be concluded that no significant linear trend is presented in any week since none of coefficients are significantly different from zero since the corresponding p-values are not less than 5%. Furthermore, no quadratic trend also presence in weekly rainfall within the same time period. The some of the weeks' result is presented in Table 5.2 and the rest weeks also showed the similar result.

Table 5.2: The coefficients of linear and quadratic along with the P-values of the randomly selected weeks

Weeks	Linear Coefficient (P-Value)	Quadratic Coefficient (P-Value)
11	-1.89 (0.127)	0.0356 (0.093)
23	3.66 (0.139)	-0.0626 (0.137)
29	-2.40 (0.248)	0.0357 (0.312)
32	-0.353 (0.715)	-0.0001 (0.997)
46	-3.57 (0.172)	0.0551 (0.214)
52	-1.80 (0.255)	0.0280 (0.296)

5.2. Weekly Rainfall Percentiles

Rainfall percentiles are employed in designing of water related structures in many fields. Sound awareness about the rainfall pattern is vital to mitigate the various issues derived from heavy rainfall and long dry spell existence due to climate change. Based on the result of the explanatory data analysis, it is found that the city Colombo was enriched from the heavy shower during the South West Monsoon (SWM) and Second Inter Monsoon (SIM). Furthermore, it is explored that the heavy intensity rain events over the SIM. Thus, SWM and the SIM are the two rainy seasons which can have much possibility to form extreme rainfall events. Fifty six year weekly rainfall series varied from 1960 to 2015 were considered for this analysis and weeks 18-39 were pertaining to the SWM while weeks 40-48 belong to SIM. Here, the series week 18 was made by taking into consideration of the 18^{th} week (April 30-May 06) rainfall in every year. The remaining weekly series also made in the same manner.

5.3. The 95% Confidence Intervals for the Weekly Rainfall Percentiles using Parametric Approach

Most of the researchers made attempt to make inferences about the rainfall amount by using point estimates derived from the different theoretical probability distributions for rainfall percentiles. Sharma and Singh (2010) used the Generalized Extreme Value distribution, Gamma and Log Pearson distributions for the maximum weekly rainfall in the monsoon period at the Pantnagar region in India to study the temporal variability of maximum weekly rainfall. According to the review of rainfall percentiles carried out by Sharda and Das (2005), the Weibull distribution is more likely fitted for describing weekly rainfall at Dehradunin India. Also, they used the probability distribution models for computing minimum assured amount of rainfall at different probability levels. Beta and Weibull distributions were fitted for the weekly rainfall during the monsoon and non monsoon periods, respectively, and those best fit distributions are employed for computing minimum assured amount of rainfall at different probability levels for the Command area by Mishra et al., (2013). Moreover, many researchers have fitted theoretically probability distributions for the rainfall data at different timescales mainly monthly, seasonally and annually for the purpose of making inferences about the rainfall using point estimates.(Varathan et al., 2010; Singh et al., 2012; Alghazali and Alawadi, 2014; Mayooran and Laheetharan, 2014; Ghosh et al., 2016).

However, extremely few studies were reported in Sri Lanka with respect to the rainfall variation at weekly scale. Waidyarathne et al., (2006) analyzed weekly rainfall data to investigate the change of the onset of FIM rain in coconut growing agro ecological regions in Sri Lanka. However, it might be more risky depending on a single value formed from probability distributions to mitigate the circumstances which would be existed due to climate change. Confidence interval is one of the most popular techniques that can be used to measure the uncertainty. Based on the literature, there is no study has been conducted for weekly rainfall quantities in context of the parametric confidence interval approach.

5.3.1. Distribution of Weekly Rainfall

Many probability distributions were fitted to the weekly rainfall series pertaining to the SWM and SIM and five rainfall percentiles; P_{50}, P_{60}, P_{70}, P_{80} and P_{90} along with the 95% confidence intervals were made based on the best fitted distribution. To get the sound knowledge of the distributions of the weekly rainfall initially, histograms were obtained and some of those are presented from the Figure 5.3 and Figure 5.4 respectively. Moreover, the remaining plots were illustrated in Appendix -1.

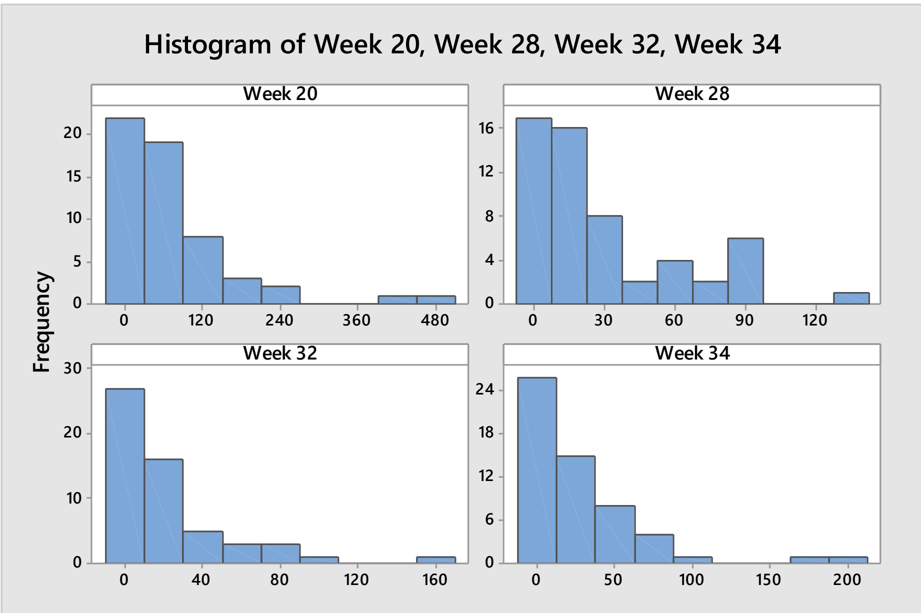

Figure 5.3: Histogram of the total weekly rainfall for week numbers: week 20, 28, 32 and 34 in SWM

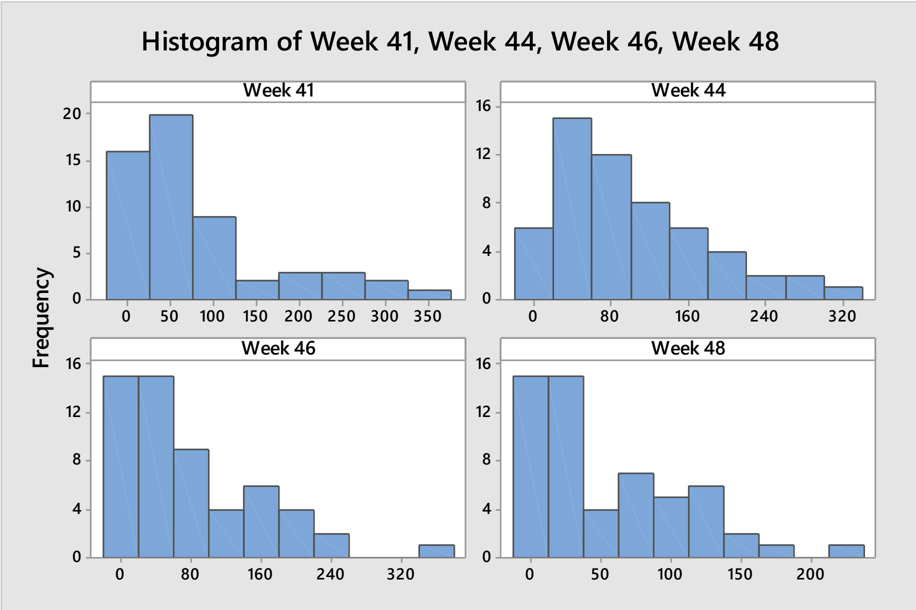

Figure 5.4: Histogram of the total weekly rainfall for week numbers: week 41,44,46 and 48 in SIM

Histogram of dataset provides clear evidence that the distributions of the weekly rainfall which pertaining to the SWM as well as SIM are skewed with longer tail to the right. Four randomly selected weeks 20, 28, 32 and 34 which belong to SWM are depicted in Figure 5.3 and weeks 41, 44, 46 and 48 in SIM are presented in Figure 5.4. An almost similar pattern was observed in remaining data series also.

5.3.2. Randomness of the Weekly Series

The randomness of the weekly rainfall series were checked by using the autocorrelation plots. The auto correlation plots of the randomly selected four weeks which two weeks belongs to SWM and rest pertains to the SIM are presented from the Figure 5.5, Figure 5.6, Figure 5.7 and Figure 5.8 respectively.

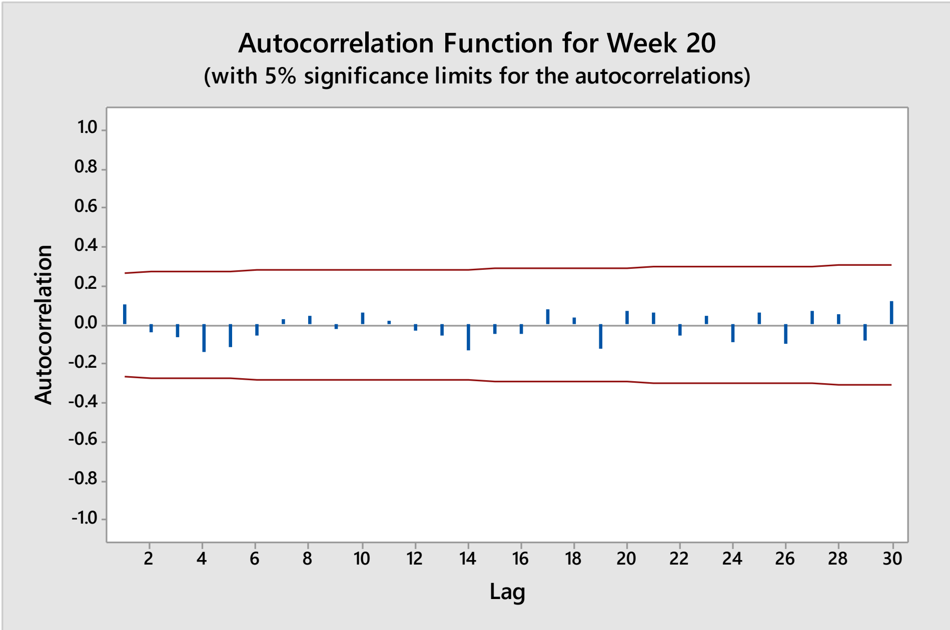

Figure 5.5: The auto correlation plot of the week 20 belongs to SWM

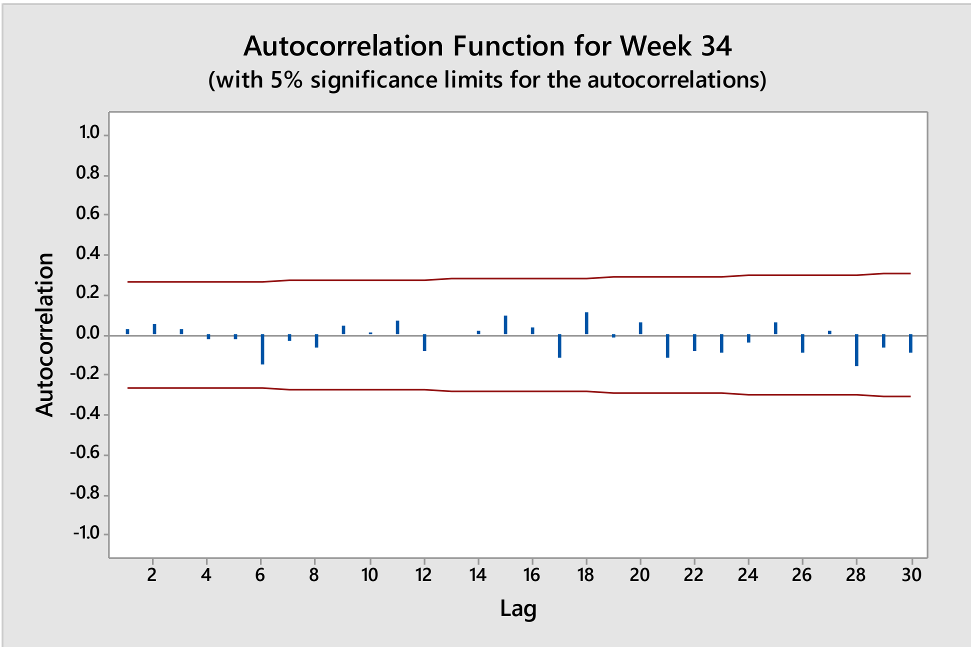

Figure 5.6: The auto correlation plot of the week 34 belongs to SWM

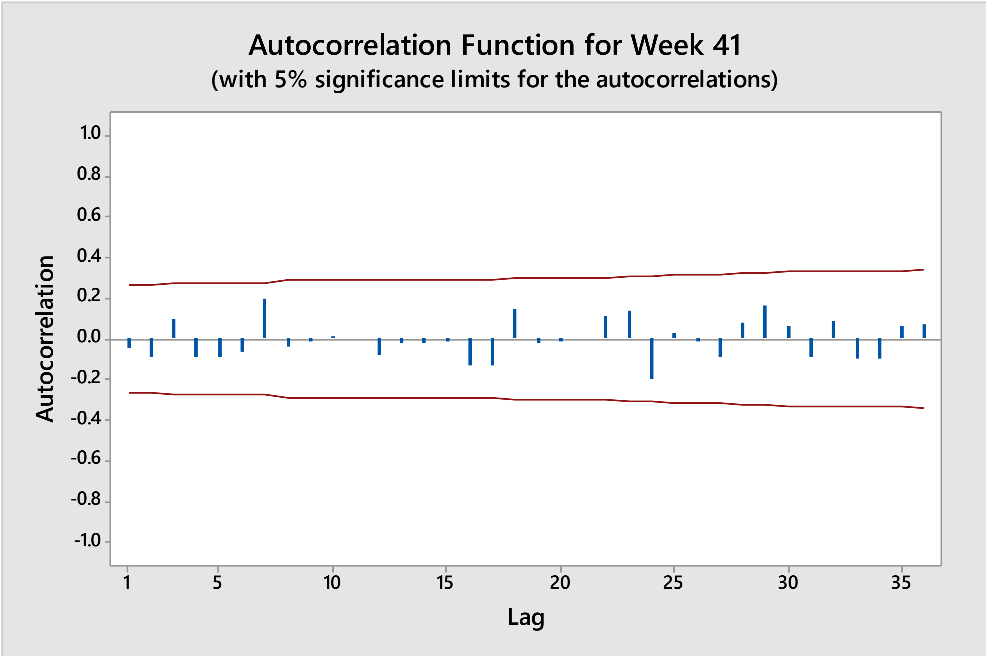

Figure 5.7: The auto correlation plot of the week 41 pertains to SIM

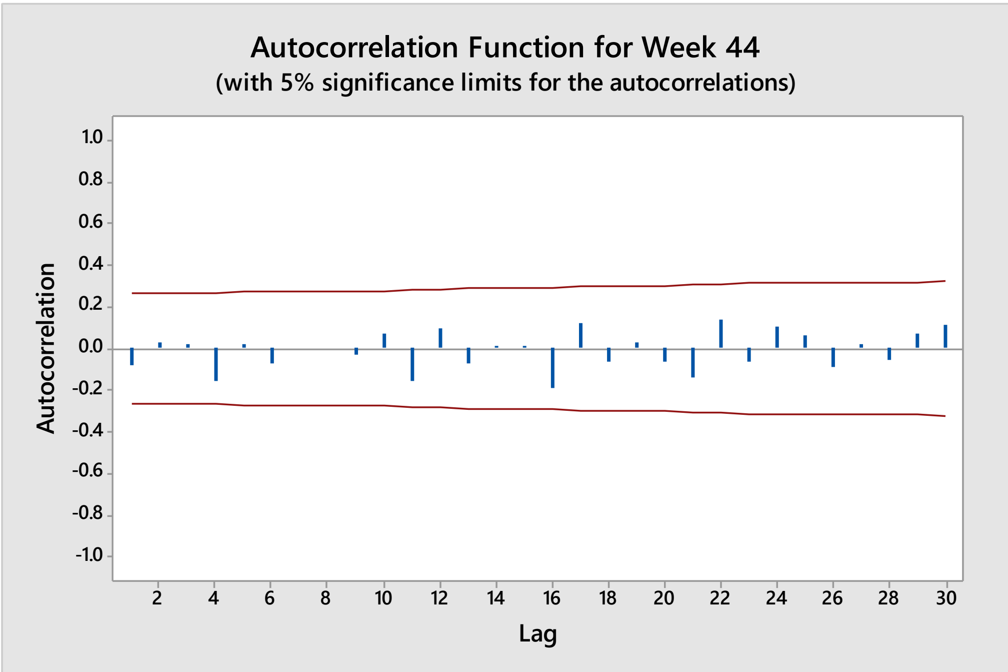

Figure 5.8: The auto correlation plot of the week 44 pertains to SIM

According to the above figures that the corresponding weekly rainfall series for the weeks 20, 34, 41 and 44 are in random manner. Also, similar patterns of autocorrelation were observed reaming weekly series (Appendix 2). Thus, it can be concluded weekly rainfall series of all the weeks 18-48 have a random pattern.

5.3.3. Normality of Weekly Rainfall Series

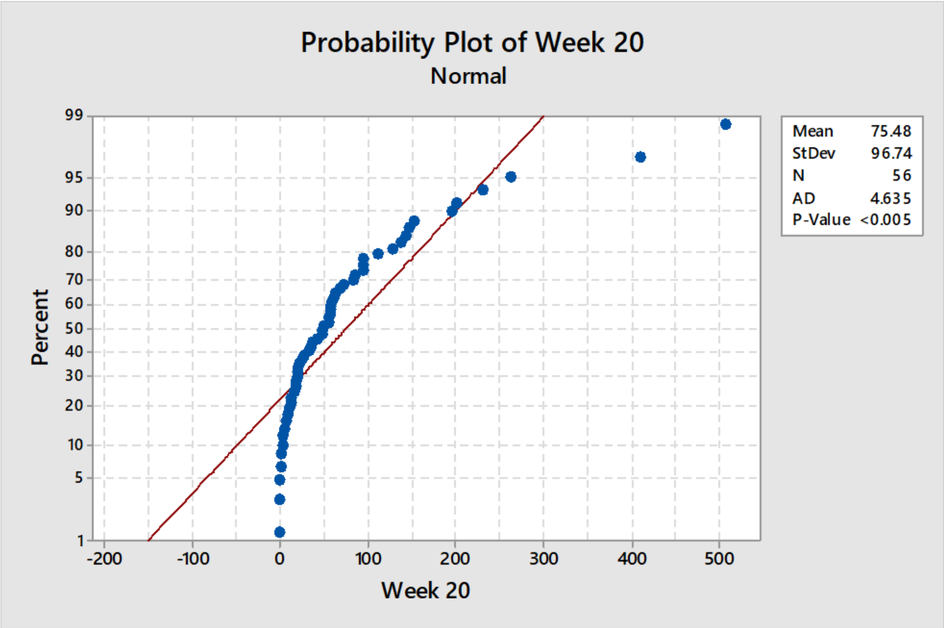

Figure 5.9: The normal probability plot of the week 20 in SWM

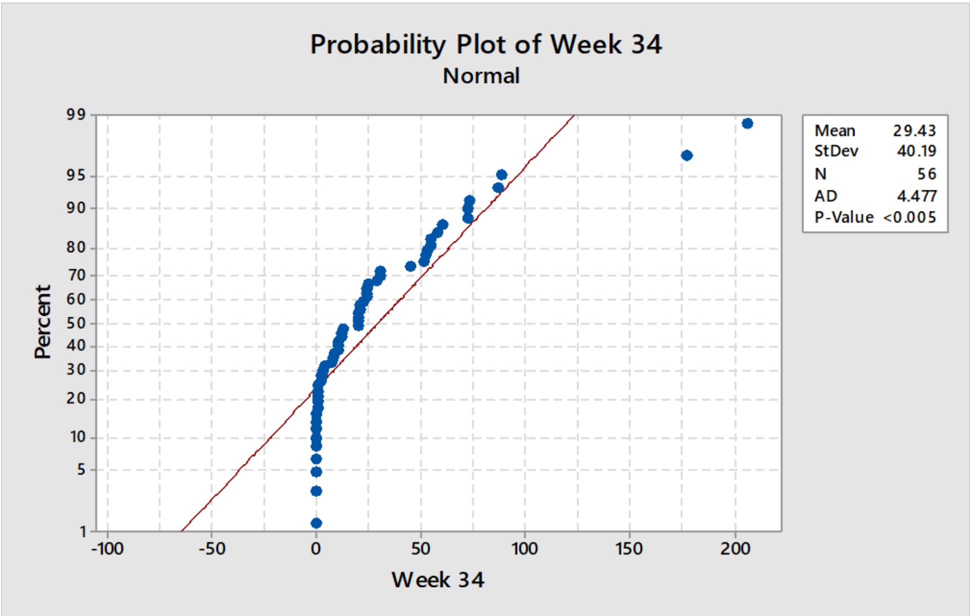

Figure 5.10: The normal probability plot of the week 34 in SWM

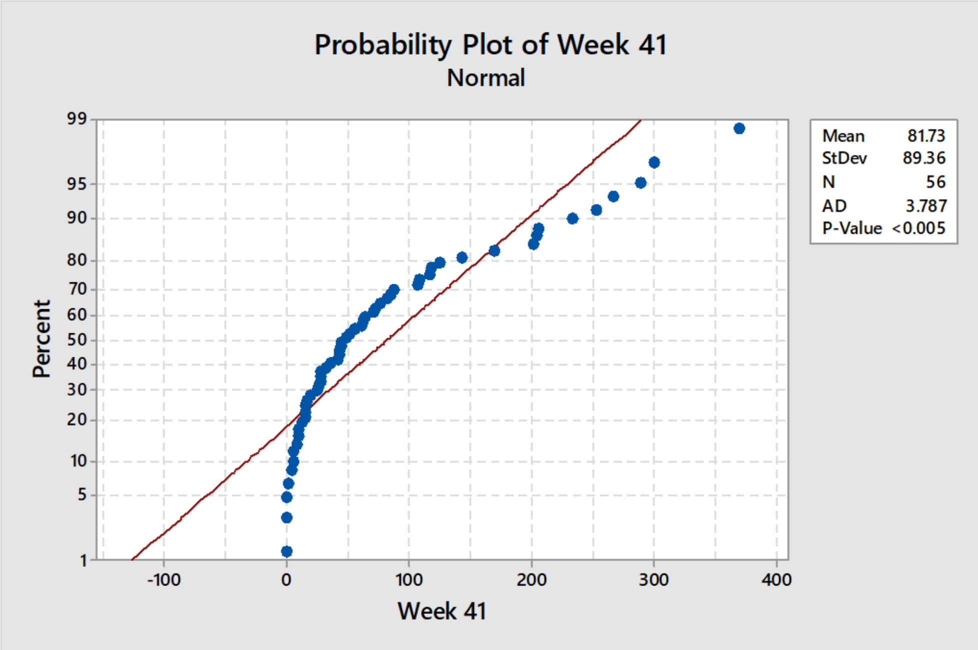

Figure 5.11: The normal probability plot of the week 41 in SIM

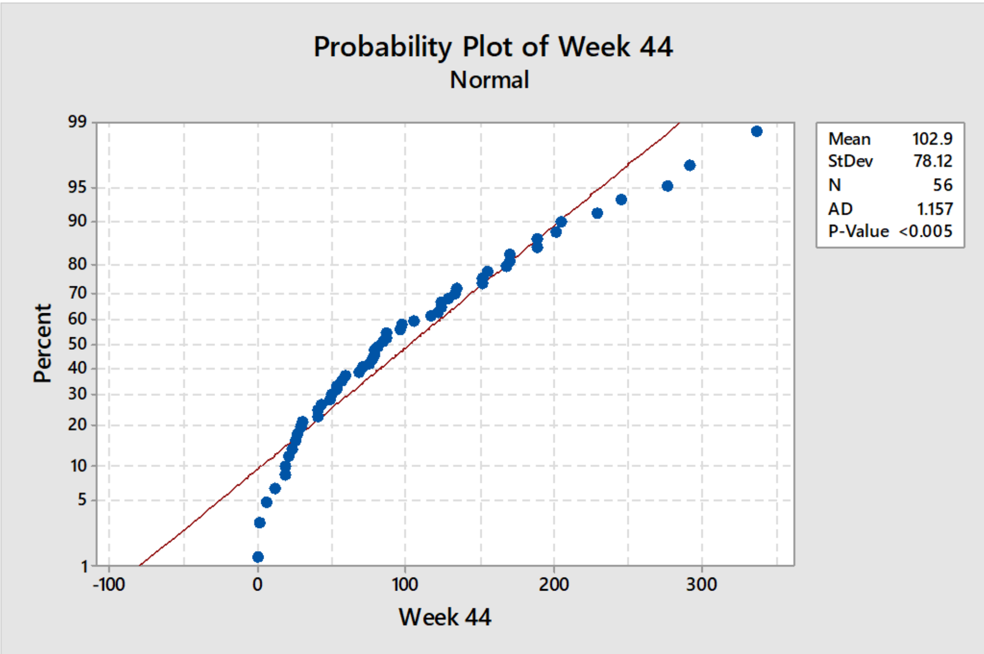

Figure 5.12: The normal probability plot of the week 44 in SIM

The Anderson Darling test confirmed that the distributions of weekly rainfall significantly different from the normal distribution. Furthermore, all graphs provide the evidence to reject the null hypothesis that the series is followed normal distribution. A similar result was obtained in remaining data series too. This implies that none of the weekly rainfall series belongs to SWM as well as SIM followed normal distribution.

5.3.4. Common Distributions for Weekly Rainfall Totals

The following known distributions: Log normal, Exponential, Gamma, Weibull, Largest Extreme Value, Smallest Extreme Value, Logistic, Log Logistic along with the different forms of some distributions such as 3-parameter Gamma, 2-parameter Exponential, 3-parameter Log Logistic and 3-parameter Weibull distributions were utilized to select best fitted distribution for the weekly rainfall in 18 to 48. The Table 5.3 represents some of the selected probability distributions with their probability density functions. The distribution parameters were estimated using maximum likelihood approach.

Table 5.3: The Probability density functions

Distribution	**Probability Density Function**	**Parameters**
Lognormal	$f(x)=\frac{1}{\sigma x\sqrt{2\pi}}\ \exp\left(-\frac{(\ln x-\mu)}{2\sigma^2}\right)$	μ - Location Parameter, σ -Scale Parameter $\mu \geq 0, \sigma > 0, X \geq 0$
Exponential	$f(x)=\frac{1}{\alpha}\ \exp\left(-\frac{x}{\alpha}\right)$	α -Scale Parameter $\alpha > 0$
2 Parameter Exponential	$f(x)=\frac{1}{\alpha}\ \exp\left(-\frac{(x-\lambda)}{\alpha}\right)$	α -Scale Parameter, λ -Threshold parameter $\alpha > 0, \lambda < X$
Largest Extreme Value	$f(x)=\frac{1}{\sigma}\exp\left[\left(\frac{x-\mu}{\sigma}\right)\right]\exp\left\{-\exp\left(\frac{(x-\mu)}{\sigma}\right)\right\}$	μ - Location Parameter, σ -Scale Parameter $\mu \geq 0, \sigma > 0, X \geq 0$

Table 5.3: (Continued...)

Weibull	$f(x)=\frac{\beta}{\alpha^{\beta}}x^{\beta-1}\ \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)$	α -Scale Parameter, β -Shape Parameter $\alpha > 0,\ \beta > 0,\ X \geq 0$
3-Patarameter Weibull	$f(x)=\frac{\beta}{\alpha^{\beta}}(x-\lambda)^{\beta-1}\ \exp\left(-\left(\frac{x-\lambda}{\alpha}\right)^{\beta}\right)$	α-Scale Parameter, β-Shape Parameter,λ-Threshold parameter $\alpha > 0,\ \beta > 0,\ \lambda < X$

The Anderson Darling and Kolmogorov-Smirnov test were used to identify the best fitted probability distributions for weekly rainfall series. The Table 5.4 is presented the best fitted probability distributions for the weekly rainfall series parting to the SWM and the corresponding maximum likelihood estimates along with the two test statistics called Anderson Darling test statistics (AD) and Kolmogorov-Smirnov test statistic (KS).

5.3.4.1. Properties of the Best Fitted Models for Weeks in SWM

Table 5.4: The best fitted statistical models and maximum likelihood estimates for weekly rainfall during SWM

Week No.	Best Fitted Distribution	AD	KS	Estimated Parameters (MLE)
18	3 - Parameter Weibull	0.317 (0.501)	0.0782 (0.884)	α = 77.061, β = 0.878, λ = - 0.838
19	3 - Parameter Weibull	0.131 (0.520)	0.0526 (0.996)	α = 82.249, β = 0.888, λ = - 0.935
20	3 - Parameter Weibull	0.247 (0.510)	0.0684 (0.956)	α = 67.331, β = 0.804, λ = - 0.508
21	3 - Parameter Weibull	0.362 (0.461)	0.1027 (0.596)	A = 73.570, β = 1.086, λ = - 1.752
22	Exponential	0.457 (0.540)	0.0857 (0.773)	α= 68.989
23	Lognormal	0.319 (0.526)	0.0700 (0.928)	μ = 3.518, σ = 0.912
24	Weibull	0.291 (0.257)	0.0691 (0.934)	α = 43.645, β = 1.267
25	3 - Parameter Weibull	0.498 (0.222)	0.0752 (0.910)	α = 34.182, β = 0.884, λ = - 0.383
26	2- Parameter Exponential	0.912 (0.103)	0.1099 (0.110)	α = 40.204, λ = - 0.718
27	3 - Parameter Weibull	0.275 (0.521)	0.073 (0.926)	α= 32.535, β= 0.887, λ = - 0.269
28	3 - Parameter Weibull	0.531 (0.186)	0.069 (0.952)	α= 24.822, β= 0.741, λ = - 0.131
29	2- Parameter Exponential	0.873 (0.107)	0.1813 (0.182)	α = 37.875, λ = - 0.676
30	3 - Parameter Weibull	0.596 (0.210)	0.1066 (0.548)	α= 16.711, β= 0.626, λ = - 0.038
31	2- Parameter Exponential	0.841 (0.126)	0.1823 (0.232)	α = 19.853, λ = - 0.355
32	3 - Parameter Weibull	0.617 (0.113)	0.1002 (0.627)	α= 15.263, β= 0.602, λ = - 0.029
33	3 - Parameter Weibull	0.445 (0.531)	0.094 (0.706)	α= 23.975, β= 0.651, λ = - 0.067

Table 5.4: (Continued...)

34	2- Parameter Exponential	0.694 (0.101)	0.1193 (0.194)	α =29.964, λ = - 0.535
35	3 - Parameter Weibull	0.607 (0.120)	0.1186 (0.410)	α= 22.408, β= 0.698, λ = - 0.089
36	3 - Parameter Weibull	0.544 (0.328)	0.0888 (0.770)	α= 26.012, β=0.602, λ = - 0.049
37	3 - Parameer Weibull	0.246 (0.531)	0.0662 (0.967)	α = 40.709, β= 0.838, λ = - 0.366
38	3 - Parameter Weibull	0.438 (0.315)	0.0979 (0.656)	α= 57.303, β= 0.855, λ = - 0.261
39	3 - Parameter Weibull	0.397 (0.394)	0.0679 (0.958)	α= 81.654, β= 0.863, λ = - 0.831

* The value in parenthesis represents the corresponding P value

It can be seen that the most of the weeks (15 out of 22) belong to the SWM were well fitted with the 3 parameter Weibull distribution. However, weeks 22-24, Exponential, Lognormal and Weibull distributions were found to be most appropriate distributions. Two parameter Exponential distributions were most probable distribution for the weeks 26, 29, 31 and 34. Moreover, 68% of the running weekly totals are well fitted with the 3 parameter Weibull distribution while 22% are fitted with the two parameter Exponential distribution and the remaining are well fitted with the Exponential, Largest Extreme Value, Weibull and Lognormal distributions.

5.3.4.2. Properties of the Best Fitted Models for Weeks in SIM

The best fitted probability distribution and the corresponding test statistics during the SIM are presented in Table 5.5.

Table 5.5: The best fitted statistical models and maximum likelihood estimates for weekly rainfall during SIM

Week No.	**Best Fitted Distribution**	**AD (Pvalue)**	**KS (Pvalue)**	**Estimated Parameters (MLE)**
40	3 - Parameter Weibull	0.372 (0.443)	0.0536 (0.725)	α = 49.918, β = 0.901, λ = - 0.403
41	3 - Parameter Weibull	0.181 (0.520)	0.0156 (0.973)	α = 77.341, β = 0.875, λ = - 0.832
42	3 - Parameter Weibull	0.360 (0.465)	0.0002 (0.999)	α = 104.811, β = 1.097, λ = -2.583
43	3 - Parameter Weibull	0.245 (0.510)	0.0179 (0.965)	α= 95.553, β = 1.189, λ = - 0.671
44	3 - Parameter Weibull	0.137 (0.540)	0.0057 (0.891)	α = 113.014, β = 1.309, λ = - 1.717
45	Largest Extreme Value	0.574 (0.144)	0.0730 (0.926)	μ = 58.397, σ =51.092
46	2- Parameter Exponential	0.842 (0.126)	0.1180 (0.417)	α = 78.205, λ = - 1.397
47	2- Parameter Exponential	0.645 (0.322)	0.1107 (0.499)	α = 63.506, λ = - 1.134
48	2- Parameter Exponential	0.692 (0.103)	0.1034 (0.587)	α = 56.235, λ = - 1.005

* The value in parenthesis represent the corresponding pvalues

Most of the weeks belong to SIM were well fitted with the 3 parameter Weibull distribution while 2-parameter Exponential and Largest extreme value distributions were found to be most appropriate distributions for remaining. Furthermore, 65% of the running weekly totals are well fitted with the 3 parameter Weibull distribution while 30% are fitted with the two parameter Exponential distribution and the remaining are well fitted with the Largest Extreme Value distributions.

5.3.5. Confidence Intervals for Weekly Rainfall in SWM

The formulas used for the percentile and its variance calculation based on the probability distribution is also shown in Table 5.6. Furthermore, Table 5.7 depicts the formulas that were employed for the confidence bands of percentiles.

Table 5.6: The formulas used for percentiles and variance estimates

Distribution	**Percentiles** $\left(\hat{X}_p\right)$	**Variance of Percentile Var**$\left(\hat{X}_p\right)$
Lognormal	$\hat{\mu} + z_p \hat{\sigma}$	$Var(\hat{\mu}) + {z_p}^2 Var(\hat{\sigma}) + 2Z_p Cov(\hat{\mu}, \hat{\sigma})$
Exponential	$-\ln(1-p)\hat{\alpha}$	$[-\ln(1-p)]^2 Var(\hat{\alpha})$
2 Parameter Exponential	$\hat{\lambda} + [-\ln(1-p)\hat{\alpha}]$	$Var(\hat{\lambda}) + [-\ln(1-p)]^2 Var(\hat{\alpha}) + 2[-\ln(1-p)]Cov(\hat{\lambda}, \hat{\alpha})$
Largest Extremes Value	$\hat{\mu} + z_p \hat{\sigma}$	$Var(\hat{\mu}) + {z_p}^2 Var(\hat{\sigma}) + 2Z_p Cov(\hat{\mu}, \hat{\sigma})$
Weibull	$\hat{\alpha}\,[-\ln(1-p)]^{1/\hat{\beta}}$	$\frac{\hat{X}_p^2}{\hat{\alpha}^2} Var(\hat{\alpha}) + \frac{\hat{X}_p^2}{\hat{\beta}^4} {z_p}^2 Var(\hat{\beta}) - 2Z_p \frac{\hat{X}_p^2}{\hat{\alpha}\,\hat{\beta}^2} Cov(\hat{\alpha}, \hat{\beta})$
3-Parameter Weibull	$\hat{\lambda} + \hat{\alpha}\,[-\ln(1-p)]^{1/\hat{\beta}}$	$Var(\hat{\lambda}) + \hat{\omega}^2 Var(\hat{\alpha}) + \frac{\hat{\alpha}^2}{\hat{\beta}^4} \hat{\omega} Z_p^2 Var(\hat{\beta}) - 2\frac{\hat{\alpha}}{\hat{\beta}^2} \hat{\omega}^2 {z_p}^2 Cov(\hat{\alpha}, \hat{\beta}) + 2\hat{\omega} Cov(\hat{\alpha}, \hat{\lambda})$ $-2\frac{\hat{\alpha}}{\hat{\beta}^2} Z_p \hat{\omega}\; Cov(\hat{\beta}, \hat{\lambda})$

Table 5.7: The formulas used for confidence intervals for percentiles

Distribution	**Confidence Bands**
Lognormal Exponential Weibull	$\left\{\exp\left[\ln(\hat{X}_p) - Z_{\alpha/2} \frac{\sqrt{Var(\hat{X}_p)}}{\hat{X}_p}\right], \exp\left[\ln(\hat{X}_p) + Z_{\alpha/2} \frac{\sqrt{Var(\hat{X}_p)}}{\hat{X}_p}\right]\right\}$
2-Parameter Exponential 3- Parameter Weibull	If $\lambda < 0$ $\left\{\hat{X}_p - Z_{\alpha/2}\sqrt{Var(\hat{X}_p},\quad \hat{X}_p + Z_{\alpha/2}\sqrt{Var(\hat{X}_p}\right\}$ $\lambda > 0$ $\left\{\hat{X}_p - Z_{\alpha/2}\sqrt{\frac{Var(\hat{X}_p)}{\hat{X}_p}},\quad \hat{X}_p + Z_{\alpha/2}\sqrt{\frac{Var(\hat{X}_p)}{\hat{X}_p}}\right\}$
Largest Extreme Value	$\left\{\hat{X}_p - Z_{\alpha/2}\sqrt{Var(\hat{X}_p},\quad \hat{X}_p + Z_{\alpha/2}\sqrt{Var(\hat{X}_p}\right\}$

Weekly rainfall percentiles and the corresponding 95% confidence intervals which calculated using above formulas are presented in Table 5.8. Those intervals were made for the weekly rainfall percentiles at 50, 60, 70, 80 and 90 based on the probability distributions which were selected as best fitted for corresponding weeks.

Table 5.8: The percentiles of the weekly rainfall and the corresponding 95% confidence intervals during SWM in the city Colombo

Week Number	**PERCENTILES**				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
18	49.9 (32.1, 67.7)	68.9 (46.5, 91.4)	94.4 (65.4, 123.3)	131.7 (92.1, 171.3)	198.5 (136.3, 260.7)
19	53.5 (34.6, 72.4)	73.6 (49.9, 97.3)	100.45 (70.0, 130.9)	139.7 (98.1, 181.2)	209.6 (144.2, 274.9)
20	42.2 (25.9, 58.5)	59.9 (38.7, 81.0)	84.3 (56.2, 112.4)	121.2 (81.5, 160.9)	189.5 (124.5, 254.6)
21	50.8 (35.9, 65.6)	66.1 (48.5, 83.8)	85.5 (64.1, 107.0)	112.3 (84.8, 139.8)	156.8 (116.8, 196.8)
22	47.8 (36.8, 62.1)	63.2 (48.6, 82.1)	83.1 (63.9, 107.9)	111.0 (85.4, 144.3)	158.9 (122.3, 206.4)
23	33.72 (26.6, 42.8)	42.5 (33.3, 54.1)	54.4 (42.1, 70.2)	72.6 (55.0, 95.9)	108.4 (78.5, 149.8)
24	32.7 (25.6, 41.7)	40.7 (32.6, 50.9)	50.5 (41.0, 62.3)	63.5 (51.7, 78.2)	84.3 (67.8, 104.8)
25	22.2 (14.3, 30.1)	30.6 (20.7, 40.4)	41.8 (29.1, 54.5)	58.2 (40.8, 75.6)	87.4 (59.8, 115.1)
26	27.2 (19.9, 34.4)	36.1 (26.5, 45.8)	47.7 (35.0, 60.4)	64.0 (47.0, 80.9)	91.9 (67.6, 116.1)
27	21.3 (13.8, 28.7)	29.2 (19.8, 38.6)	39.8 (27.8, 51.9)	55.4 (38.9, 71.8)	83.0 (57.1, 108.9)
28	15.0 (8.7, 21.3)	21.9 (13.6, 30.3)	31.8 (20.3, 43.2)	47.0 (30.3, 63.8)	76.4 (47.6, 105.2)
29	25.5 (18.7, 32.5)	34.0 (24.9, 43.1)	44.9 (33.0, 56.9)	60.3 (44.3, 76.2)	86.5 (63.7, 109.4)
30	9.3 (4.7, 13.8)	14.5 (8.0, 21.0)	22.5 (12.9, 32.0)	35.7 (20.7, 50.7)	63.4 (35.1, 91.6)
31	13.4 (9.8, 17.0)	17.8 (13.1, 22.6)	23.6 (17.3, 29.8)	31.6 (23.2, 40.0)	45.4 (33.4, 57.3)
32	8.3 (4.0, 12.5)	13.2 (7.0, 19.3)	20.8 (11.6, 29.9)	33.6 (18.9, 48.3)	61.0 (32.7, 89.2)
33	13.6 (7.1, 20.0)	20.9 (11.9, 29.9)	31.8 (18.8, 44.8)	49.8 (29.6, 69.9)	86.4 (49.1, 123.6)

Table 5.8: (Continued...)

34	20.2 (14.8, 25.7)	26.9 (19.7, 34.1)	35.5 (26.1, 45.0)	47.7 (35.1, 60.3)	68.5 (50.4, 86.5)
35	13.2 (7.3, 19.0)	19.7 (11.7, 27.6)	29.2 (18.0, 40.3)	44.2 (27.6, 60.9)	74.0 (44.4, 103.5)
36	14.1 (6.9, 21.3)	22.5 (12.0, 32.9)	35.4 (19.8, 51.0)	57.3 (32.3, 82.3)	103.9 (155.6, 152.2)
37	25.9 (16.3, 35.6)	36.3 (24.0, 48.7)	50.4 (34.3, 66.6)	71.5 (49.0, 93.9)	109.8 (73.6, 146.0)
38	37.1 (23.6, 50.5)	51.5 (34.4, 68.6)	70.9 (48.7, 93.1)	99.7 (69.1, 130.3)	151.7 (103.1, 200.3)
39	52.6 (33.6, 71.6)	73.0 (48.9, 97.0)	100.4 (69.3, 131.6)	140.9 (97.8, 184.0)	213.7 (144.6, 282.8)

The result indicated that there was much heavy rainfall at the begins of the SWM. Also, Weeks 18-23 marked considerable rainfall with high variability. It is noted that 90^{th} percentiles of weeks 18-23 vary between 108.4mm to 209.6 mm which bring a greater amount of rainfall to this region. According to the table 5.8, there is a 90% chance to have 209.6 mm maximum rainfall, during the 19^{th} week and this value can be varied between 144.2 mm and 274.9 mm at 95% confidence level. However, a clear decreasing pattern of weekly rainfall can be identified after the 23^{rd} week.

The weeks 31 and 32 marked lower rainfall amount than others during SWM. After 35^{th} week, again it can be seen an increasing trend of weekly rainfall till the end of the season. The week 39 records the highest rainfall amount in the SWM. The median rainfall of the 39^{th} week series was 52.6 mm while the 70^{th} percentile of this week marked more than 100 mm rainfall amount which is a large quantity for the area. Week 38 also brings much heavy rainfall with noticeable variation in this season for this region.

The rainfall percentiles and corresponding 95% confidence intervals for running totals of weekly rainfall were also constructed during the SWM in Colombo. Figure 5.13 represented only 90^{th} percentile of running total and its 95% confidence bands. It also depicts the high rainfall variation with the arrival of SWM. Also, Figure.5.13 illustrates the much heavy rainfall due to the withdrawal of the SWM. Based on the result of the running total of the weekly rainfall, it can be further confirmed that there was heavy rainfall with great variation during the period of weeks 18-23 (30th April to 10th June) and weeks 38-39 (17th-30th of September).

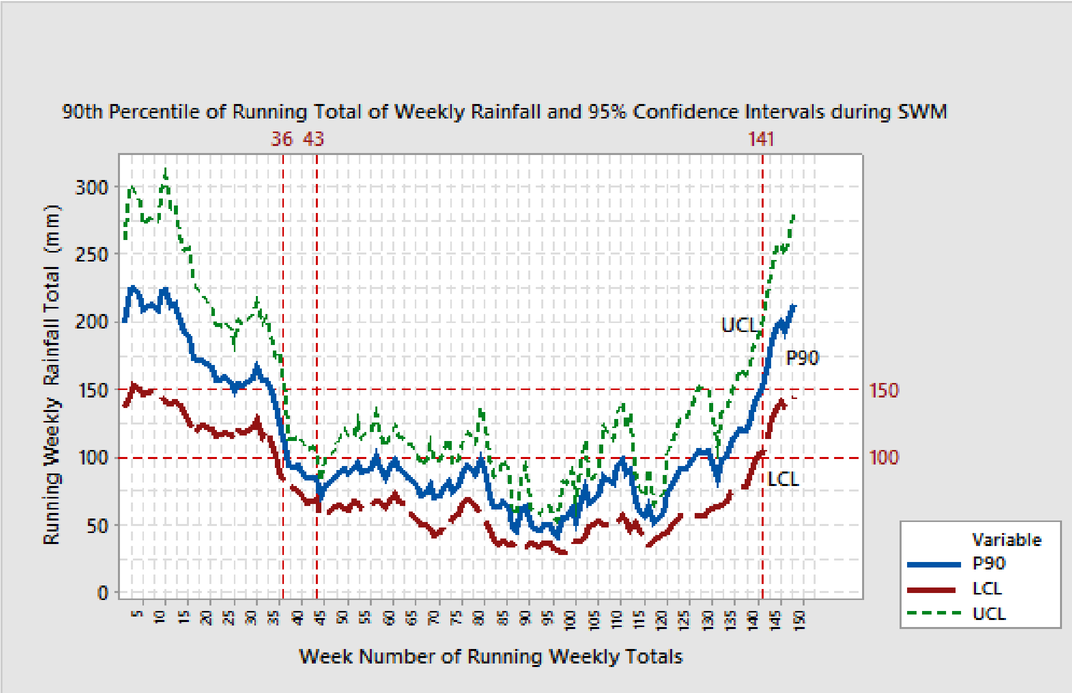

Figure 5.13: The 90th percentiles of running total of weekly rainfall and 95% confidence intervals during SWM in Colombo

Based on the analysis of past extreme rainfall events in Colombo area during SWM, it can be identified that the many floods occurred in the months May and June. Most recently (on 15 May 2016) Sri Lanka was hit by a severe tropical storm that caused heavy flooding in Colombo. Furthermore, floods occurred in Colombo in the past years; 1975, 1989, 1992, 2008 from May to June period (Jegarascsingam, 1998).

5.3.6. Confidence Intervals for Weekly Rainfall in SIM

Weekly rainfall percentiles which pertaining to the SIM with the corresponding 95% confidence intervals are presented in Table 5.9. Those intervals also were made for the weekly rainfall percentiles at 50, 60, 70, 80 and 90 based on the probability distributions which were selected as best fitted for corresponding weeks.

Table 5.9: The percentiles of the weekly rainfall and the corresponding 95% confidence intervals during SIM in the city Colombo

Week	**PERCENTILES**				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
40	32.8 (21.5, 44.2)	44.9 (30.8, 59.0)	60.9 (42.8, 79.0)	84.2 (59.6, 108.9)	125.5 (86.7, 164.3)
41	50.1 (32.1, 68.0)	69.1 46.6, 91.7)	94.8 65.7, 123.9)	132.4 (92.5, 172.3)	199.7 (136.5, 262.8)
42	72.5 (51.5, 93.5)	94.2 (69.4, 119.0)	121.6 (91.4, 151.7)	159.1 (120.5, 197.8)	221.6 (165.4, 277.7)
43	69.5 (50.6, 88.5)	88.1 (66.6, 109.7)	111 (85.9, 136.2)	141.9 (110.5, 173.4)	192.9 (146.5, 237.7)
44	83.7 (62.7, 104.7)	104 (80.6, 127.4)	128.5 (90.8, 131.4)	160.9 (110.6, 159.4)	212.4 (166.2, 257.9)
45	77.1 (61.5, 92.7)	92.7 (75.2, 110.2)	111.1 (90.8, 131.4)	135 (110.6, 159.4)	173.4 (141.7, 205.0)
46	52.8 (38.6, 67.0)	70.3 (51.5, 89.0)	92.8 (68.1, 117.4)	124.5 (91.5, 157.4)	178.4 (131.5, 225.8)
47	42.9 (31.4, 54.4)	57.1 (41.8, 72.3)	75.3 (55.3, 95.4)	101.1 (74.3, 127.8)	145.1 (106.8,183.4)
48	38 (27.8, 48.2)	50.5 (37.0, 64.0)	66.7 (49.0, 84.4)	89.5 (65.8, 113.2)	128.5 (94.6, 162.4)

The result showed in Table 5.9 indicated that there is no much heavy rainfall at the beginning of the SIM as SWM. Also, it is noted that low variability at the withdrawal of the monsoon. However, the Weeks 41-45 showed the high rainfall amount with the large variability result cause to form the extreme rainfall events. The weeks 42 and 44 record the great amount of rainfall for this region during SIM than others. Based on the analysis of the running weekly totals, it can be expected much heavy rainfall with high variability during the time span of 16^{th}-22^{nd} October. Figure 5.14 represented the 90^{th} percentile of running total during SIM and its 95% confidence bands.

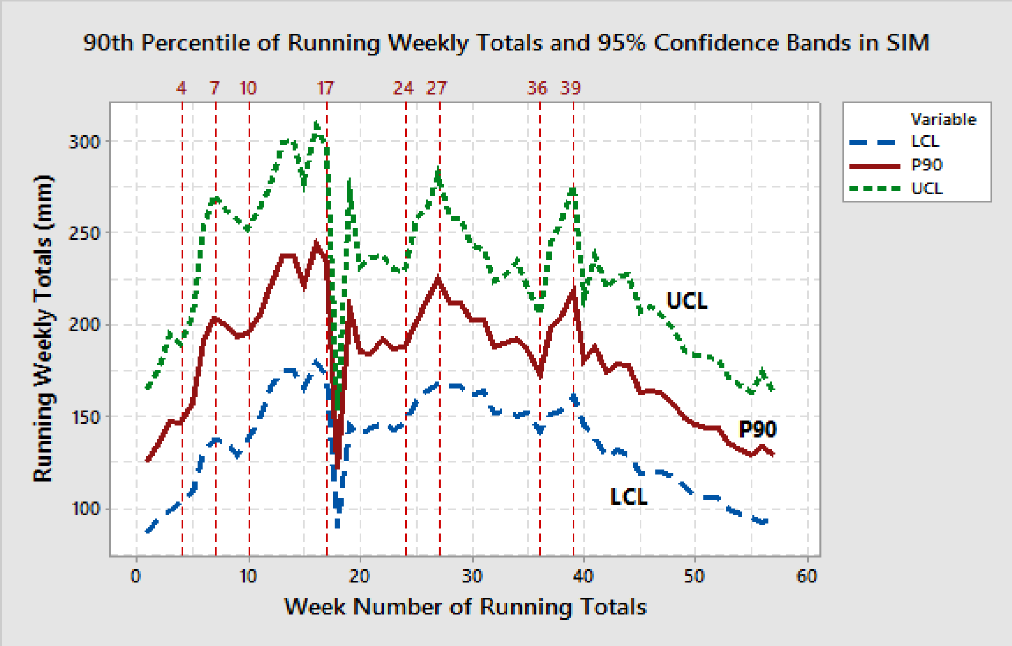

Figure 5.14: The 90th percentiles of running total of weekly rainfall and 95% confidence intervals during SIM in Colombo

Based on the analysis of both weekly totals as well as running weekly totals during the SIM period, there is a high possibility to have heavy rainfall events during the Weeks 41-45 (08th October to 11th November). There is a 90% chance to have 221.6 mm maximum rainfall in the week 42. This value can be varied between 165.4mm to 277.7 mm at 95% confidence level. Thus, week 42 (15th-21st October) has much chance to form extreme rainfall events during this monsoon period.

5.4. The 95% Confidence Intervals for the Weekly Rainfall Percentiles using Bootstrapping Approach

Confidence intervals for quantiles of a random variable mostly depend on the distribution function. However, according to the Burn (2003) there are several shortcomings of this approach. The number of assumptions with respect to the distribution and necessity of larger data series to make inferences are the main drawbacks of this approach. A bootstrapping approach has been proposed as an alternative approach for calculating confidence intervals through the resampling process. Dunn (2001) made an attempt to build bootstrap confidence intervals for predicting rainfall quantities. Simultaneous confidence intervals for a daily minimum rainfall total using a bootstrap resampling method considering of serial dependency have been produced by Ferro et al., (2005). Lucio (2007) adapted bootstrap method for the purpose of evaluating of small sample inferences for monthly rainfall extreme quantiles.

Three approaches Bayesian, Bootstrap and Profile Likelihood were employed to construct confidence intervals of extreme rainfall quantiles by Chen Si et al., (2016). There are many bootstrapping approaches to calculate the confidence intervals for the population parameters. The percentile bootstrap method, parametric bootstrapping method, the bootstrap-t intervals and the bias corrected accelerated percentile (BCa) method are the some of the methods often used for making confidence intervals as alternative to the parametric approach. In this study, the percentile bootstrap method is utilized to calculate the 95% confidence intervals for the weekly rainfall percentiles. The percentile method is more popular among applied statistician (Hall, 1992).

5.4.1. Percentile Bootstrap Method

The bootstrap method is used to make inferences by using the information based on a number of resample from the same sample. This is a nonparametric technique that assists to make conclusions about the characteristics of a population based on the existing sample unlike the parametric approach which makes assumptions about the estimators. The procedure creates simulated data set by drawing observations from the original sample with replacement. If a parameter can be expressed as a function of an unknown distribution, then its bootstrap estimator is also the function of the same distribution function.

Suppose a random sample of size n, $X = (X_1, X_2, X_3, ..., X_n)$ from an unknown population with probability distribution function f(x) and let θ be the parameter and $\hat{\theta}$ be the estimator for θ based on data set. B is the number of samples with size n generated from the f(x) then $X^* = (X^*_1, X^*_2, X^*_3 ..., X^*_n)$ denote the bootstrap random sample of size n. Let $\hat{\theta}^*$ be an estimator computed using the bootstrapping sample of X*.

Suppose we generate B number of bootstrap samples with size n from the original sample data and for each sample we computed the statistic of interest $\hat{\theta}^* = \left(\hat{\theta}_1^*, \hat{\theta}_2^*, ..., \hat{\theta}_B^*\right)$. In our study, the rainfall percentile is the interest in statistic. The ordered bootstrap values are used to compute the bootstrap confidence intervals from the Percentile method. Suppose 1000 bootstrap replications of $\hat{\theta}$ denoted by $\left(\theta_1^*, \theta_2^*, ..., \theta_{1000}^*\right)$ and after ranking ascending order it can be denoted as $\left(\theta_{[1]}^*, \theta_{[2]}^*, ..., \theta_{[1000]}^*\right)$. Then the bootstrap percentile confidence intervals at the 95% level of confidence would be $\left[\theta_{(25)}^*, \theta_{(975)}^*\right]$ (Singh and Xie, 2008).

5.4.1.1. CI for Weekly Percentiles in SWM

Table 5.10 depicts weekly rainfall percentiles and the corresponding 95% bootstrap confidence intervals. Those intervals also were made for the weekly rainfall percentiles at 50, 60, 70, 80 and 90 based on the 1000 bootstrap samples.

Table 5.10: The 95% confidence intervals of weekly rainfall percentiles (based on 1000 bootstrap samples) pertaining to SWM (week 18-39)

Week	**PERCENTILES**				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
18	42.0 (30.9, 66.3)	61.6 (34.8, 100.6)	100.0 (54.1, 130.3)	129.4 (96.4, 207.1)	223.8 (128.6, 343.4)
19	55.6 (32.6 ,86.4)	81.1 (48.1, 105.1)	101.7 (78., 144.0)	142.2 (99.2, 191.8)	198.1 (144.6, 351.1)
20	48.5 (26.2, 59.7)	57.9 (47.4, 84.6)	82.0 (56.7, 126.8)	121.6 (68.2, 178.6)	197.5 (116.5, 306.7)
21	50.7 (41.2, 64.5)	61.3 (50.0, 84.2)	82.6 (56.3, 101.8)	99.6 (72.0, 143.9)	154.6 (102.7, 242.6)
22	48.0 (29.2 , 78.1)	75.7 (43.6,97.2)	85.2 (67.0, 143.5)	142.8 (84.6, 163.2)	164.6 (144.8, 184.4)
23	33.1 (24.0, 44.4)	42.6 (31.7, 51.5)	49.1 (41.5, 78.1)	76.5 (49.0, 105.3)	114.9 (76.5, 135.6)

Table 5.10: (Continued...)

24	32.0 (27.2, 41.6)	38.8 (31.0,48.3)	48.0 (37.0, 63.6)	61.9 (47.8, 85.5)	86.6 (60.4, 116.5)
25	20.8 (14.0, 36.3)	35.6 (17.5, 49.9)	49.0 (32.7, 66.9)	63.5 (48.0, 84.9)	91.1 (68.0, 114.8)
26	28.2 (14.6, 49.4)	47.5 (22.8, 57.2)	54.0 (36.7, 65.2)	65.0 (53.4, 80.6)	83.2 (66.3, 143.4)
27	17.6 (11.9, 31.9)	30.4 (16.8, 49.9)	47.3 (27.6, 59.4)	59.0 (46.0, 78.7)	86.1 (61.8, 123.8)
28	16.2 (10.1, 25.9)	24.2 (13.4, 34.5)	33.7 (20.0, 59.)	57.6 (31.8, 83.1)	86.3 (57.9, 92.6)
29	17.9 (11.7, 35.4)	33.2 (15.4, 41.0)	40.4 (27.9, 55.7)	53.5 (39.1, 70.2)	76.7 (58.4, 164.2)
30	11.9 (7.7, 21.5)	18.1 (11.0, 24.6)	23.7 (15.3, 34.3)	33.9 (23.5, 45.5)	47.5 (35.4, 104.2)
31	8.6 (3.3, 17.2)	13.3 (6.4, 26.3)	25.1 (12.1, 31.5)	31.4 (24.2, 52.3)	62.0 (32.0, 89.0)
32	11.6 (4.3, 18.1)	17.1 (9.7, 24.3)	23.7 (16.2, 45.8)	42.7 (21.6, 51.9)	57.9 (46.9, 82.0)
33	17.9 (10.0, 29.3)	27.4 (15.0, 46.9)	41.7 (26.4, 55.7)	54.6 (41.2, 86.3)	89.3 (56.0, 106.9)
34	20.1 (9.6, 24.0)	22.9 (14.0, 30.3)	30.0 (20.9, 54.7)	54.2 (29.6, 72.3)	72.8 (54.8, 115.0)
35	19.4 (7.6, 25.9)	21.8 (18.6, 29.1)	28.3 (22.4, 47.7)	47.2 (28.0, 62.4)	64.9 (48.9, 145.9)
36	19.3 (7.9, 33.7)	27.4 (15.8, 58.2)	45.9 (25.3, 68.2)	68.0 (43.9, 105.0)	114.7 (69.0, 142.6)
37	28.8 (16.2, 37.5)	34.7 (25.1,45.5)	44.4 (34.3, 70.7)	69.9 (44.2, 116.3)	118.1 (77.8, 153.0)
38	32.1 (22.3, 42.9)	40.4 (29.1, 64.1)	58.9 (38.6, 130.9)	122.8 (54.8, 164.1)	167.2 (116.1, 215.4)
39	56.6 (32.2, 89.6)	85.8 (48.3,104.3)	102.8 (77.1, 158.5)	155.4 (101.0, 216.1)	225.6 (164.4, 316.8)

*The values in parenthesis represent the corresponding 95% confidence intervals.

The result indicated that the heavy rainfall at the beginning of the SWM. Furthermore, it can be expected much rainfall from week 18 to week 23. It is evident from Table 5.10 that 80% or more chances to have 207.1mm maximum weekly rainfall in the weeks 18 - 23. However, after the 23rd week it can be seen clear decline of weekly rainfall up to week 35. It can be expected 86.3 mm maximum week rainfall with 80% probability at week 33 which showed maximum rainfall variability out of weeks 24-35. The week 31 and 32 marked much lower rainfall during the SWM. Since week 36 it can be seen much rainfall till end of the season. The week 39 marked highest rainfall amount during the SWM. The table depicts a high variability at the weeks 18-23, 29 and 38-39. Also, there is a much higher possibility to have extreme rainfall during the weeks 18-23, 29 and 38-39.

However, almost similar conclusion can be made based on the parametric as well as percentile bootstrap approach since both emphasis the same time period to have much possibility to form extreme rainfall events to the region during the SWM.

5.4.1.2. CI for Weekly Percentiles in SIM

Table 5.11 depicts weekly rainfall percentiles and the corresponding 95% confidence intervals which made using percentiles bootstrap approach.

Table 5.11: The 95% confidence intervals of weekly rainfall percentiles (based on 1000 bootstrap samples) pertaining to SIM (week 40-48)

Week	PERCENTILES				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
40	35.1 (20.1, 50.3)	48.8 (32.3, 73.4)	67.1 (45.7, 99.6)	99.5 (64.4, 118.9)	122.7 (100.4, 165.9)
41	46.4 (29.8, 71.7)	64.8 (44.6, 107.4)	87.7 (61.7, 143.0)	135.6 (84.6, 222.6)	239.3 (137.9, 291.7)
42	71.1 (52.8, 91.5)	83.1 (70.0, 131.8)	107.4 (79.3, 157.7)	156.0 (105.1, 211.4)	220.1 (161.9, 366.0)
43	73.5 (51.9, 96.6)	90.8 (70.0, 119.3)	114.4 (86.7, 138.2)	135.4 (108.7, 166.0)	185.7 (139.7, 265.2)
44	83.0 (59.7, 119.8)	107.9 (78.7, 134.1)	132.5 (97.6, 169.1)	168.7 (128.9, 201.3)	212.0 (169.7, 280.7)
45	73.7 (61.0, 86.1)	79.2 (71.0, 112.0)	109.5 (78.7, 126.1)	125.8 (107.0, 178.9)	182.7 (125.7, 314.8)
46	53.4 (33.0, 80.9)	75.6 (45.9, 109.5)	100.8 (72.0, 142.7)	142.6 (97.0, 185.5)	192.7 (144.4, 226.2)
47	53.8 (29.1, 74.7)	73.3 (53.2, 83.9)	80.2 (66.2, 90.2)	90.1 (79.2, 114.8)	125.6 (90.5, 178.7)
48	32.0 (24.1, 65.6)	62.7 (30.5, 77.0)	76.2 (55.2, 107.2)	105.7 (71.8, 129.4)	132.3 (107.7, 152.1)

*The values in parenthesis represent the corresponding 95% confidence intervals.

The results indicated that the heavy rainfall over the SIM compared with the SWM. It can be expected 135.6 mm maximum rainfall at 80% probability at week 41 and the value can be varied from 84.6mm to 222.6mm which indicated that a high variability. It is noted that high rainfall variability at the weeks 41-45 in SIM. Thus, there is a much higher chance to have extreme rainfall events at above weeks. Based on the result of Table 5.10, it is clear that the beginning as well as the withdrawal of the monsoon season showed much low rainfall amount along with the low variation with compared to the middle of the weekly rainfall. Here also, a good agreement is seen with the result of parametric approach.

However, it is noted that the width of the 95% confidence intervals made based on the percentile bootstrapping approach get much high value than the parametric approach. Small sample size and positive skewed distribution of weekly rainfall are some reasons for the high width of confidence intervals of percentiles which made using those methods. To make accurate confidence intervals, the coverage probability should be taken into account.

5.5. Accurate Confidence Interval Bands

The coverage probability of the confidence interval is one of the imperative factors that should be considered when making inferences using confidence limits. Accurate confidence bands enhance the degree of the awareness level of rainfall variability at high uncertainty. To calculate the accurate confidence interval bands, the parametric bootstrapping approach is used by utilizing the coverage probability which can made bootstrapping calibration.

The main aim of this analysis is to find the accurate level of confidence intervals for weekly rainfall percentiles derived from Weibull distributions based on the real coverage probability which formed using Bootstrap Calibration. Accurate estimates, either point or intervals are essential since many decisions might be depend on those values. In such situation, sample size is the other main factor which influences to accurate inferences. In fact, it is more complicated to compose inferences and make decisions at the small size of the sample. Most of the time, estimates derived from the fitted theoretical probability distributions becomes inaccurate at the small sample size. To overcome this problem, the bootstrapping technique can be used. The Weibull distribution and its properties are explained by the following section.

5.5.1. Weibull Distribution

The Weibull distribution was invented by Waloddi Weibull and this is widely used continuous probability distribution. Moreover, life and climatic data are analyzed using this versatility distribution. The characteristic of the Weibull distribution is varied based on the values of the scale and shape parameters. The shape of the density function of the Weibull distribution changes drastically with the value of the shape parameter. The Weibull distribution can be approximated to the normal distribution when shape parameter is about 3.6 (Johnson and Kotz,1970). Figure 5.12 describes the shape of the Weibull distribution with different scale and shape parameters.

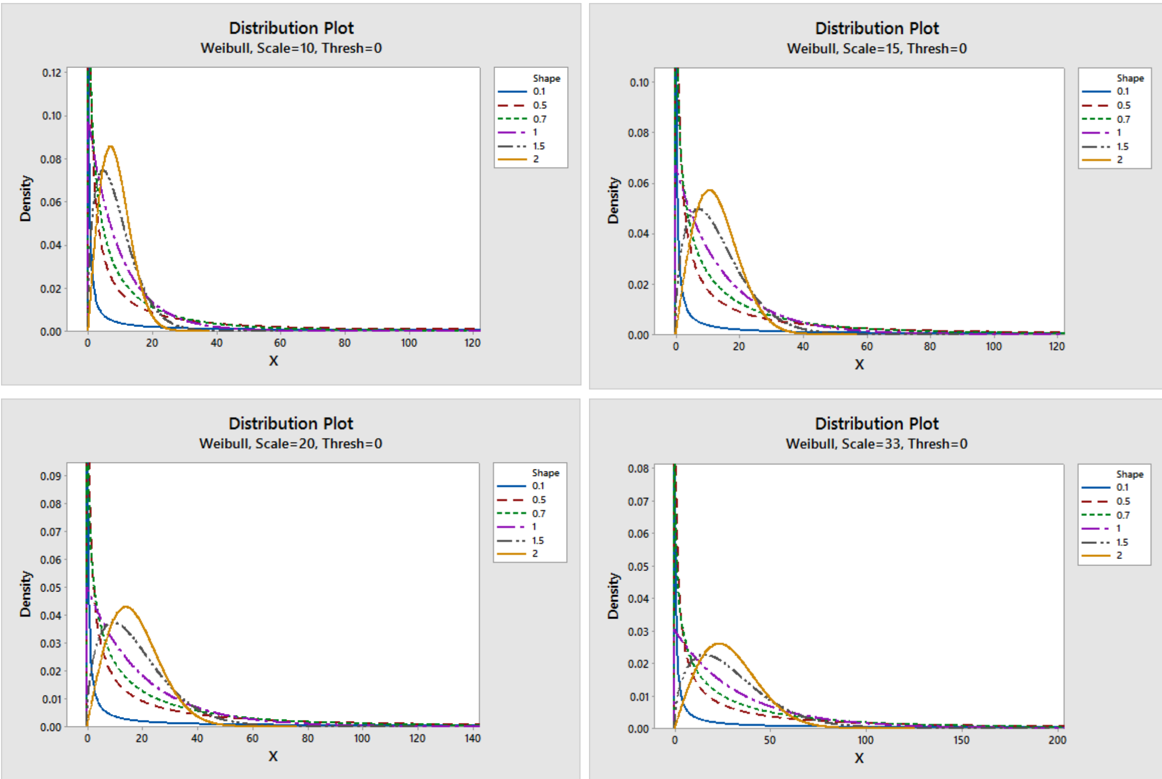

Figure 5.15: Density functions of Weibull distribution with different scale and shape parameters

The p^{th} percentile of the Weibull distribution (X_p) and its variance [Var (X_p)] is defined as follows (Heo et al., 2001). $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimators for the scale and shape parameters of the Weibull distribution.

The p^{th} Percentile of Weibull Distribution - X_p

$$\hat{X}_p = \hat{\alpha}\left[-\ln(1-p)\right]^{1/\hat{\beta}}$$

$$\operatorname{Var}(\hat{X}_p) = \frac{\hat{X}_p^{\,2}}{\hat{\alpha}^2}\operatorname{Var}(\hat{\alpha}) + \frac{\hat{X}_p^2}{\hat{\beta}^4} Z_p^2 \operatorname{Var}(\hat{\beta}) - 2\frac{Z_p \hat{X}_p^2}{\hat{\alpha}\hat{\beta}^2}\operatorname{Cov}(\hat{\alpha},\hat{\beta}) \qquad \text{Where} \quad Z_p = \ln\left[-\ln(1-p)\right]$$

The equations that used to calculate the confidence intervals for the Weibull percentiles are given as;

$$\left[\exp\left(\ln\left(\hat{X}_p\right) - Z_{\alpha/2} \; \frac{\sqrt{\operatorname{Var}\left(\hat{X}_p\right)}}{\hat{X}_p} \right), \; \exp\left(\ln\left(\hat{X}_p\right) + Z_{\alpha/2} \; \frac{\sqrt{\operatorname{Var}\left(\hat{X}_p\right)}}{\hat{X}_p} \right) \right]$$

5.5.2. The Coverage Probability

The coverage probability of a confidence interval can be briefly explained as the proportion of the time that interval that contains the true value of interest. The coverage probability of a confidence interval can be calculated using simulation method; firstly, many samples of size n should be simulated from the population and compute the confidence intervals for interest parameter for each sample. After that, the proportion of samples should be computed for the known population parameters is contained in the confidence interval. That proportion is an estimate for the coverage probability for the confidence interval. However, a discrepancy can be occurred between the computed coverage probability and the nominal coverage probability due to many reasons such as approximating a discrete distribution with a continuous distribution, when the population is not normal etc. In this study, our interest parameter is percentile and based on the confidence intervals of percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) the coverage probability will be calculated.

5.5.3. Data for the Simulation

To work out the coverage probability, here considered one weekly data series during the SWM as the population. In this study, the data which belong to the week 24 (11-17 June) in SWM (The weekly data of 46 years for the time span from 1970 to 2015) was considered as the population. Here, small sample size (n=46) was considered for this analysis to distinguish the real and nominal confidence bands clearly. The summary statistic of the total rainfall during week 24 is presented in Table 5.12 along with the histogram (Figure 5.16).

Table 5.12: Descriptive statistics of the weekly rainfall data (week 24)

Variable	No. of Data	Mean	Median	Min	Max	C.V (%)	Skewness
Week 24	46	36.1	18.4	0.1	146.3	104.9	1.37

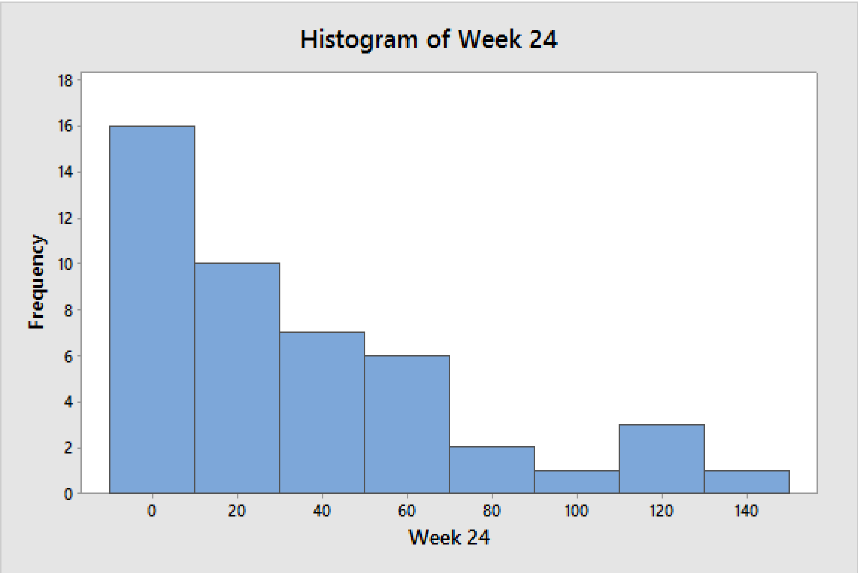

Figure 5.16: Histogram of weekly rainfall data (week 24)

From 1970 to 2015 total weekly rainfall in the 24th week varied from 0.1mm to 146.3mm with a mean 36.1mm. Figure 5.16 illustrates the weekly rainfall as positively skewed with a longer tail to the right and the result was further confirmed as the coefficient of skewness is 1.37. The large coefficient of variance (104.9%) gives evidence to high fluctuations in weekly rainfall (week 24).

The total rainfall during week 24 was fitted to different type of probability distributions and those data were well fitted with the two parameter Weibull distribution. Corresponding Anderson Darling and Kolmogorov-Smirnov test statistics were 0.258 (P-value =0.256) and 0.0673 (P-value = 0.941) respectively. The maximum likelihood estimates for the scale and shape parameters of the fitted Weibull distribution were 33.9286 and 0.8775 respectively.

5.5.4. The Simulation Procedure

Weekly rainfall series for the 24th week from 1970 to 2015 was taken as population having 46 points and well fitted with the two parameter Weibull distribution. Based on the population data (N=46), 2000 random samples (each sample size is also equal to 46) were generated using bootstrapping approach called as Sample1, Sample2, Sample3, ... Sample2000 from the Weibull distribution (α, β). Furthermore, it is estimated the maximum likelihood estimates (MLE) for the scale and shape parameters of data sets pertaining to the Sample1 $(\hat{\alpha}_1, \hat{\beta}_1)$, Sample2 $(\hat{\alpha}_2, \hat{\beta}_2)$, Sample3 $(\hat{\alpha}_3, \hat{\beta}_3)$ and so on. Five percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) were calculated for each sample (Sample1 to Sample2000).

Again 300 samples were generated (Same sample size (n=46)) based on the generated Sample1, Sample2, Sample3 etc. Let denote those 300 samples derived from the Sample1, as $Sam1_1$, $Sam1_2$...., $Sam1_{300}$. Here, it describes only the coverage probability of randomly selected four samples (Sample68, Sample423, Sample802 and Sample1551). Let us consider the 300 samples generated based on the Sample1. Firstly, it is calculated the 50th percentile and corresponding 95% confidence intervals of $Sam1_1$, $Sam1_2$, $Sam1_3$ and so on. The coverage probability was calculated based on the 300 confidence intervals (95%). The same procedure was carried out to calculate the coverage probability of confidence intervals at 95.2%, 95.4%, 95.6%, 95.8%, 96%, 96.2%, 96.4%, 96.6%, 96.8%, 97%, 97.2%, 97.4%, 97.6%, 97.8% and 98% confidence levels. Other samples which generated from the Sample1, Sample2...., Sample2000 were applied using the above procedure and calculated the corresponding coverage probabilities for each confidence level listed above.

5.5.5. Results Obtained from the Simulation

For the interpretation purpose the computed coverage probabilities of confidence intervals of randomly selected four samples (Samples 68, 423, 802 and 1551)) at different uncertainty levels are presented in Tables 5.13-5.16 respectively.

Table 5.13: The coverage probabilities of five percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample68

Confidence Level (%)	Coverage Probability				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
95.0	93.00	93.00	93.00	92.00	92.67
95.2	93.33	93.33	93.00	92.67	93.00
95.4	93.33	93.67	93.00	93.00	93.00
95.6	93.67	93.67	93.00	93.00	93.67
95.8	94.00	93.67	93.00	93.00	93.67
96.0	94.33	93.67	93.67	93.33	94.00
96.2	**95.33**	94.33	94.00	94.00	94.67
96.4	95.67	**95.00**	94.00	94.67	**95.00**
96.6	96.33	95.33	94.67	94.67	95.00
96.8	96.33	95.67	**95.00**	**94.67**	95.00
97.0	96.67	96.00	95.00	**95.33**	95.00
97.2	96.67	96.33	95.33	96.00	95.00
97.4	96.67	96.33	95.67	96.00	95.33
97.6	96.67	96.33	96.00	96.00	95.33
97.8	96.67	96.33	96.00	96.33	95.67
98.0	96.67	96.33	96.67	96.33	96.33

Thus, 95% confidence interval for 95% coverage probability for the sample size 46 was found as (0.89, 1.00). The Table 5.13 shows that the 95% coverage probability of P_{50} can be attained at 96.2 % confidence level. Similarly, the 95% coverage probability of P_{60}, P_{70}, P_{80} and P_{90} can be reached at the confidence levels 96.4, 96.6, 96.8 and 96.4 respectively.

Table 5.14: The coverage probabilities of five percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 423

Confidence Level (%)	**Coverage Probability**				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
95.0	94.00	93.67	90.67	87.67	90.67
95.2	94.00	94.00	91.00	88.67	91.67
95.4	94.00	94.00	91.00	89.00	91.67
95.6	94.00	94.33	91.67	89.00	91.67
95.8	94.33	94.33	91.67	89.00	92.00
96.0	94.33	94.33	92.00	89.33	92.00
96.2	94.33	94.67	92.00	89.33	92.33
96.4	94.67	94.67	92.33	90.00	93.00
96.6	**95.00**	94.67	92.33	91.33	93.67
96.8	95.67	94.67	92.67	91.67	94.00
97.0	95.67	**95.00**	93.00	92.00	94.67
97.2	96.00	96.00	93.33	92.67	94.67
97.4	96.67	96.67	93.33	93.00	94.67
97.6	97.00	97.00	93.67	93.67	**95.00**
97.8	97.33	97.67	94.67	94.33	95.33
98.0	97.67	97.67	**95.00**	**95.00**	95.33

As explained above (Table 5.13), the real 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} can be obtained at the 96.6, 97.0, 98.0,98.0, and 97.6 respectively.

Table 5.15: The coverage probabilities of five percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 802

Confidence Level (%)	**Coverage Probability**				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
95.0	94.33	94.33	90.67	89.33	89.33
95.2	94.33	94.33	90.67	89.33	89.67
95.4	94.33	**94.67**	90.67	89.67	90.00
95.6	**95.33**	**95.33**	91.33	89.67	90.00
95.8	95.67	95.33	91.67	89.67	91.33
96.0	95.67	95.33	91.67	90.00	91.33
96.2	96.00	95.67	92.00	90.00	91.67
96.4	96.00	95.67	92.67	90.00	92.00
96.6	96.33	96.00	92.67	90.67	93.33
96.8	96.33	96.00	92.67	91.00	93.33
97.0	96.33	96.33	93.33	91.33	94.00
97.2	96.67	96.33	94.00	91.67	94.67
97.4	97.33	96.67	94.33	92.00	94.67
97.6	97.33	96.67	94.67	93.00	94.67
97.8	97.67	97.00	**95.00**	94.00	**95.00**
98.0	97.67	97.00	95.00	**95.00**	95.33

Table 5.15 illustrates the 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} obtained at the 95.6, 95.4, 97.8, 98.0, and 97.8 confidence levels respectively.

Table 5.16: The coverage probabilities of five percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 1551

Confidence Level (%)	Coverage Probability				
	P_{50}	**P_{60}**	**P_{70}**	**P_{80}**	**P_{90}**
95.0	92.67	92.33	92.00	92.00	89.67
95.2	93.00	92.33	92.00	92.33	89.67
95.4	93.33	92.67	92.00	92.67	89.67
95.6	93.33	92.67	93.33	92.67	92.33
95.8	**94.67**	93.00	93.33	93.00	92.33
96.0	95.33	93.00	93.33	93.00	92.33
96.2	96.00	93.00	93.67	93.67	93.33
96.4	96.00	93.33	93.67	93.67	93.33
96.6	96.00	93.33	94.00	94.00	93.67
96.8	96.00	93.67	94.00	94.33	93.67
97.0	96.33	93.67	94.67	94.67	94.67
97.2	96.33	93.67	94.67	94.67	94.67
97.4	96.67	94.67	**95.00**	94.67	94.67
97.6	96.67	94.67	95.00	**95.00**	**95.00**
97.8	96.67	**95.00**	95.33	95.00	95.33
98.0	97.67	95.00	95.33	95.33	95.33

As explained above (Table 5.13), the Table 5.16 shows that the 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} obtained at the 95.8, 97.8, 97.4, 97.6, and 97.6 confidence levels respectively. Same procedure was carried out for the remaining samples and calculated the average accurate coverage probability based on the 300 samples derived from each 2000 samples presented in Table 5.17.

Table 5.17: Average accurate confidence level based on the 95% confidence level for Weibull percentiles

Percentiles	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
Coverage Probability	95.901	97.501	97.603	97.680	97.910

Results in Table 5.17 indicate that the real 95% CI does not attain for any of the percentile values under Weibull distributional though analysis found Weibull is the best fitted distribution. The accurate confidence level is closer to 95% at P_{50} and gap increases as percentile value increases. Based on the above real confidence levels, it is calculated the confidence bands of percentiles of week 24 as follows.

Table 5.18: The confidence bands of percentiles of week 24 (nominal and actual values)

Percentile	Value	Confidence limits at level of 95% (Nominal)		Accurate confidence levels attend the real coverage probability of 95%	Confidence limits at coverage probability (Actual)	
P_{50}	22.4	15.2	32.9	95.901	14.9	33.5
P_{60}	30.7	21.5	43.8	97.501	20.5	46.1
P_{70}	41.9	30.0	58.6	97.603	28.5	61.6
P_{80}	58.4	42.0	81.2	97.680	39.8	85.5
P_{90}	87.8	62.0	124.3	97.910	58.3	132.2

The asymptotic behavior of bootstrap confidence limits for weekly rainfall percentiles were found to be more useful than normal confidence intervals from practical and decision point of view. Due to development of computing power in statistics, it is not difficult task to compute coverage probabilities of percentiles if the distribution is known. However, based on the result formed from the simulation, there is a considerable difference between nominal and calculated coverage probabilities. Weibull distribution, drastically tends to be skewed to the right when the shape parameter less than one. Thus, the distribution of weekly rainfall deviates from the normal distribution with respect to the lower (less than one) value of shape parameter of the distribution. The deviation of the normality of the fitted distribution with the small size of sample could be the reason for the discrepancy of the nominal and calculated coverage probabilities.

5.6. Summary of the Chapter 5

Weekly rainfall data pertaining to SWM is skewed with a longer tail extending to the right to all the weeks in SWM. However, a common probability distribution was not found to represent all the weeks, but three parameter Weibull distribution was well fitted with the most of the weeks. Based on the results of percentiles and corresponding 95% confidence intervals analysis which derived using parametric and bootstrapping approach, it can be expected that much heavy rainfall with high variability during arrival of SWM in the weeks 18-23 (30th April to 10th June) and withdrawal of the SWM in the weeks of 38-39 (17-30 September).

Based on the parametric analysis of the both weekly totals as well as running weekly totals during the SIM period, there is a high possibility to have heavy rainfall events during the Weeks 41-45 (08th October to 11th November). A similar result was obtained from the bootstrapping approach also.

However, the lengths of the 95% confidence intervals were not in satisfactory level. Small sample size and strongly skewed distribution pattern might be a one of the reasons for wide confidence bands. In addition to the weekly rainfall percentile analysis, the 95% confidence bands of percentiles are utilized to compute the real coverage probability of the 95% confidence intervals. Rainfall total during the 24th week (11-17 June) in SWM was considered as the study population and those data series was well fitted with the Weibull distribution. Based on the simulation using bootstrapping approach, it is found that the most of the coverage probability of 95% confidence intervals of 50th percentile is less than 0.95 and the 95% accurate coverage probability is attained at the average level of 95.901%. The corresponding accurate coverage probabilities of 95% confidence intervals of 60th,70th, 80th and 90th percentiles are given at the average levels of 97.501%, 97.603%, 97.680% and 97.910% of respectively.

CHAPTER 6

MODELING OF WEEKLY RAINFALL: CLASSICAL TIME SERIES APPROACHES

The main goal of this chapter is to find the possibilities of forecasting weekly rainfall using conventional time series models. An attempt was made to model the weekly rainfall series with exogenous variables such as weekly temperature, relative humidity and vapor pressure in this chapter. The data series from 1990 to 2014 (1300 points) were used to train the models and an independent data set were used to validate the models. Furthermore, the draw backs of the time series modeling are discussed in detail in this chapter as those drawbacks would be useful for creating new types of models.

6.1. Variability of Weekly Rainfall during 1990-2014

The modeling weekly rainfall was done using 1352 data points during the time span from 1990 to 2015. Here, the series with length 1300 was considered model forming while rest was used for the model validation. In order to examine the temporal variability of the weekly rainfall, the time series plot was obtained and it is presented by Figure 6.1.

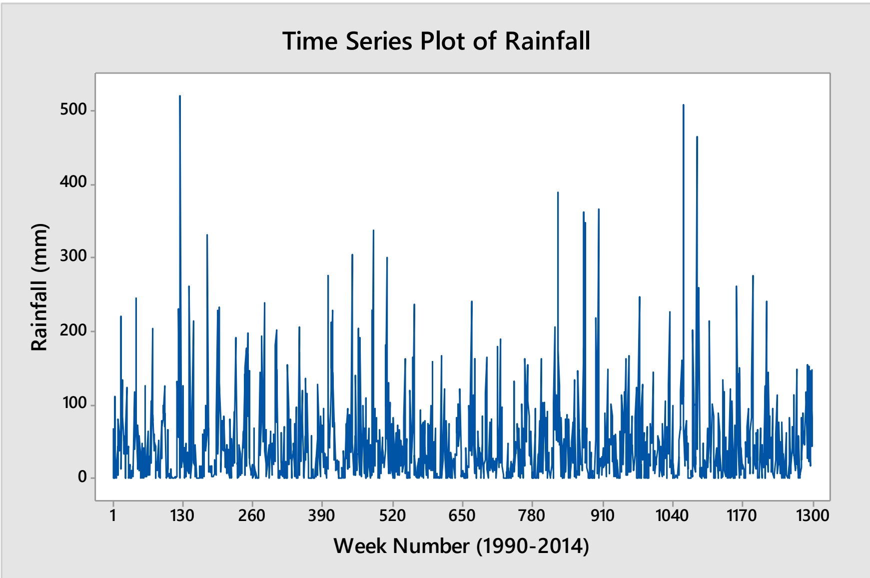

Figure 6.1: Time series plot of the weekly rainfall $\{Y_t\}$ from 1990 to 2014

Based on the above plot, random behavior of the rainfall pattern can be clearly observed. However, it cannot be identified decreasing or increasing trend in weekly rainfall during the considered time span.

6.2. Identification of ARIMA Model

In order to identify the correlation structure of the observed series, the autocorrelation was taken and those result is shown in Figure 6.2.

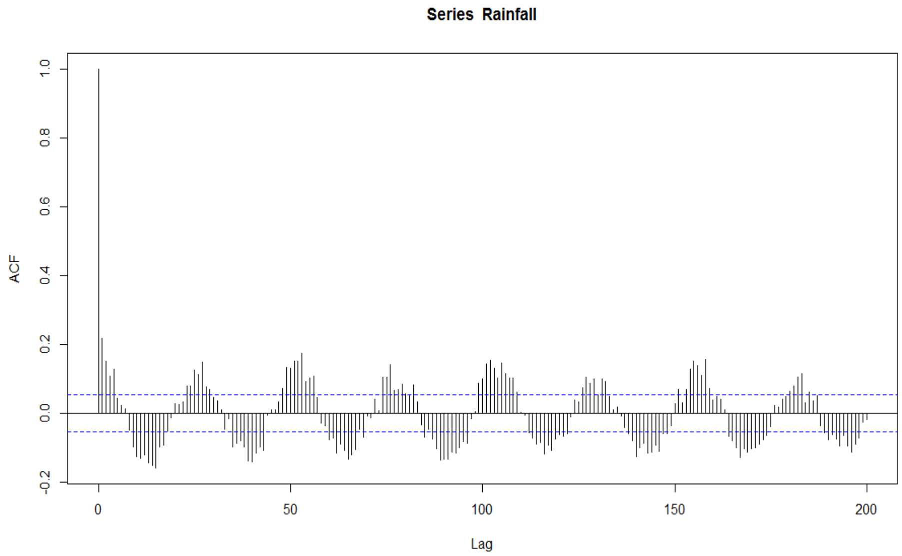

Figure 6.2: Autocorrelation plot of the series from 1990 to 2014

A seasonal behavior can be clearly recognized from this plot. Since the data were captured on a weekly basis and also seasonality of length 52 can be seen, new series was obtained by taking one long-term difference. That is Z_t, $\{Z_t\}=\{Y_t-Y_{t-52}\}$. Thus, to identifying the seasonal length, autocorrelation function (ACF) and Partial auto correlation function (PACF) were obtained with 52 lag difference. The plot of ACF of the new series $\{Z_t\}$ is shown in Figure 6.3.

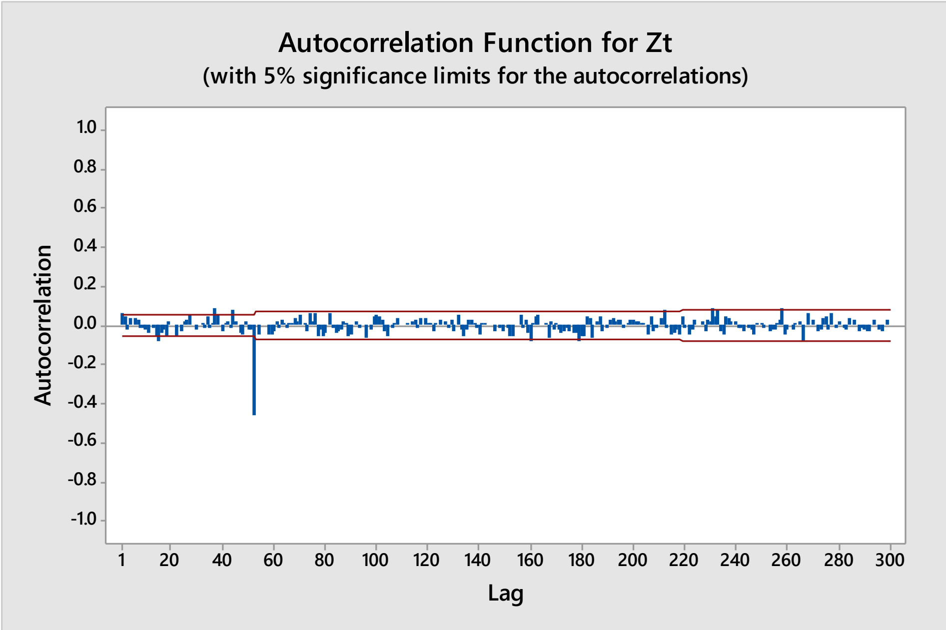

Figure 6.3: ACF of the Z_t series from 1990 to 2014 with 52 lag

The plot of ACF of the series with 52 lags showed one significant spike at 52 lag. This pattern identified that the new series can be considered as stationary series. In order to identify the suitable models for weekly rainfall, firstly, observed series was tested for the stationary using Argument Dickey Fuller Test (ADF). The result is illustrated in Table 6.1.

Table 6.1: Result of Dickey Fuller test

Null Hypothesis: ZT has a unit root
Exogenous: Constant

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-33.17186	0.0000
Test critical values:	1% level	-3.435381	
	5% level	-2.863649	
	10% level	-2.567943	

*MacKinnon (1996) one-sided p-values.

Based on the above result, the null hypothesis that the there is a unit root is rejected indicates that the series is stationary at 0.05 level of significance. The Partial Auto correlation plot of new series {Zt} is obtained and those result is presented in Figure 6.4.

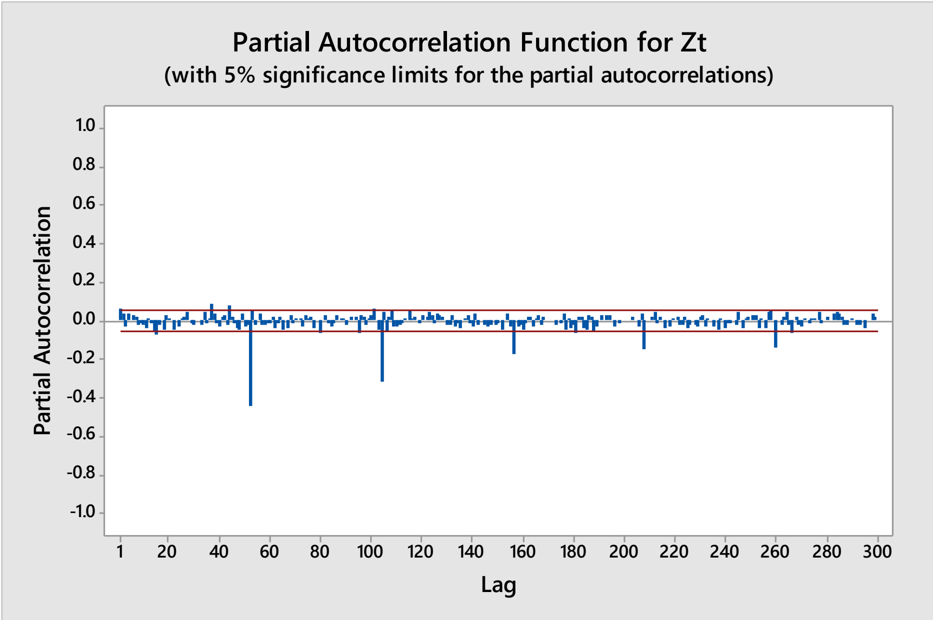

Figure 6.4: PACF of the series from 1990 to 2014 with 52 lag

According to the plot of PACF it can be clearly identified that the seasonal length is 52 since significant sample autocorrelation existed in the 52nd lag and lag multiplier of 52 (52,104,156, 208...). Thus, many Box and Jenkins models were developed for the new data series and models were selected based on the Akaike Information Criterion (AIC), Schwarz Criterian (SC), Durbin-Watson (DW) statistics, Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) criteria. Based on the criteria, selected models are listed in the Table 6.2.

Table 6.2: Selected models and values of the selection criteria

Model	AIC	SC	DW	RMSE	MAD	MAPE
SARIMA (1,0,0) $(0,1,0)_{52}$	11.55	11.56	2.00	77.54	49.46	885.57
SARIMA (1,0,2) $(0,1,0)_{52}$	11.55	11.57	2.00	77.42	49.48	897.37
SARIMA (2,0,1) $(0,1,0)_{52}$	11.55	11.57	2.00	77.45	49.50	896.94
SARIMA (1,0,0) $(1,1,0)_{52}$	11.55	11.56	2.00	77.40	49.01	895.24
SARIMA (1,0,1) $(1,1,0)_{52}$	11.55	11.57	2.00	77.45	49.90	896.94

Based on the model selected criteria, it is selected SARIMA (1,0,0) $(1,1,0)_{52}$ as a best fitted since this returns the smallest RMSE and MAD. Diagnostic tests were carried out for the best selected model. The residuals are random at 0.05 level of significant and the corresponding result is depicted by Figure 6.5. However, the assumption which is residuals are normally distributed are highly deviate at the level of 0.05 significance and the normality test result of the selected model SARIMA (1,0,0) $(1,1,0)_{52}$ is presented from the Figure 6.6.

Sample: 1 1300
Included observations: 1246
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.001	0.001	0.0014	
		2	0.000	0.000	0.0015	
		3	-0.02...	-0.02...	1.0314	0.310
		4	0.035	0.035	2.5826	0.275
		5	-0.00...	-0.00...	2.5848	0.460
		6	0.034	0.033	3.9932	0.407
		7	0.026	0.028	4.8562	0.434
		8	-0.01...	-0.01...	5.0236	0.541
		9	-0.00...	-0.00...	5.1182	0.646
		1...	-0.01...	-0.02...	5.5754	0.695
		1...	-0.03...	-0.03...	7.1194	0.625
		1...	0.011	0.011	7.2826	0.699
		1...	-0.00...	-0.00...	7.2846	0.776
		1...	-0.04...	-0.04...	10.011	0.615
		1...	-0.07...	-0.07...	17.314	0.185
		1...	-0.02...	-0.02...	18.223	0.197
		1...	-0.01...	-0.01...	18.374	0.243
		1...	-0.05...	-0.05...	21.596	0.157
		1...	0.026	0.028	22.440	0.168
		2...	0.007	0.011	22.506	0.210
		2...	0.009	0.013	22.600	0.255
		2...	-0.05...	-0.04...	26.374	0.154
		2...	-0.00...	-0.00...	26.387	0.192
		2...	-0.03...	-0.03...	27.555	0.191
		2...	0.017	0.007	27.905	0.219
		2...	0.023	0.018	28.596	0.236
		2...	0.050	0.048	31.794	0.164
		2...	0.001	0.003	31.797	0.200
		2...	-0.00...	-0.00...	31.797	0.240
		3...	-0.01...	-0.02...	32.267	0.264
		3...	0.007	-0.00...	32.328	0.306
		3...	0.008	-0.00...	32.417	0.348
		3...	-0.01...	-0.02...	32.640	0.386
		3...	0.044	0.047	35.128	0.322
		3...	-0.01...	-0.01...	35.511	0.351
		3...	0.008	0.006	35.583	0.394

Figure 6.5: The correlogram plot of the residual of the model SARIMA $(1,0,0)\times(1,1,0)_{52}$

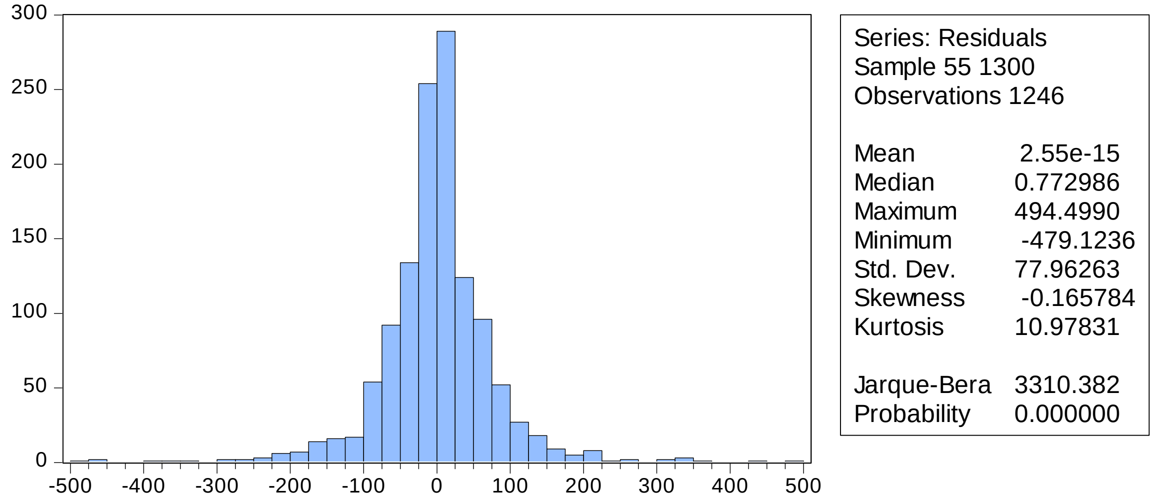

Figure 6.6: The normality test of the residuals of the model SARIMA $(1,0,0)(1,1,0)_{52}$

The best fitted model was tested for the serial correlation using the Breusch-Godfrey serial correlation LM test and the corresponding result is presented in Table 6.3.

Table 6.3: Test result of the Breusch-Godfrey serial correlation LM test

Model	F statistics	Prob. F(2,1241)	Prob. Chi-Square (2)
SARIMA $(1,0,0)\ (1,1,0)_{52}$	1.06	0.3467	0.3456

Based on the result shown in the Table 6.3, the residuals derived from the model do not show the any significant serial correlation. However, the assumption that the squared residuals are random is significantly deviated at 0.05 significant level and the corresponding test result is described by Figure 6.7 and this indicates the non constant variance. Thus, in order to examine the presence of the heteroskedasticity, ARCH effect was tested using the ARCH test and the test result is presented in Table 6.4.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.059	0.059	4.3284	0.037
		2	0.078	0.074	11.853	0.003
		3	0.070	0.062	17.955	0.000
		4	0.057	0.045	22.092	0.000
		5	-0.02...	-0.03...	22.580	0.000
		6	-0.01...	-0.02...	22.960	0.001
		7	0.000	-0.00...	22.960	0.002
		8	-0.02...	-0.01...	23.605	0.003
		9	-0.03...	-0.02...	25.277	0.003
		1...	-0.04...	-0.03...	27.526	0.002
		1...	-0.03...	-0.03...	29.467	0.002
		1...	-0.05...	-0.04...	33.744	0.001
		1...	-0.03...	-0.01...	35.027	0.001
		1...	-0.02...	-0.01...	35.860	0.001
		1...	-0.04...	-0.03...	38.223	0.001
		1...	-0.03...	-0.02...	39.745	0.001
		1...	-0.04...	-0.04...	42.588	0.001
		1...	-0.01...	-0.00...	42.886	0.001
		1...	-0.00...	0.000	42.984	0.001
		2...	0.001	0.003	42.985	0.002
		2...	-0.03...	-0.03...	44.460	0.002
		2...	0.012	0.006	44.637	0.003
		2...	0.069	0.066	50.683	0.001
		2...	0.063	0.054	55.768	0.000
		2...	0.194	0.181	103.94	0.000
		2...	0.055	0.016	107.79	0.000
		2...	0.102	0.056	121.04	0.000
		2...	0.049	0.009	124.06	0.000
		2...	0.057	0.024	128.16	0.000
		3...	-0.01...	-0.02...	128.29	0.000
		3...	0.004	-0.00...	128.31	0.000
		3...	-0.02...	-0.03...	129.16	0.000
		3...	-0.04...	-0.03...	131.60	0.000
		3...	-0.02...	-0.00...	132.54	0.000
		3...	-0.05...	-0.02...	136.19	0.000
		3...	-0.03...	-0.00...	137.47	0.000

Figure 6.7: The correlogram of squared residuals of the model SARIMA(1,0,0)$(1,1,0)_{52}$

Table 6.4: Test results of the heteroskedasticity ARCH effect

Model	F statistics	Prob. F(3,1289)	Prob. Chi-Square(3)
SARIMA (1,0,0) $(1,1,0)_{52}$	5.621	0.0037	0.0038

Based on the above result, there is no evidence to accept the null hypothesis that the there is no ARCH effect at the 0.05 level of significance. Thus, we can conclude that the ARCH effect is presented in the residuals which derived from the best fitted model.

However, despite the ARCH effect, the model was tested for the independent data set (The weeks in 2015) and the observed and predicted values is presented by the Figure 6.8.

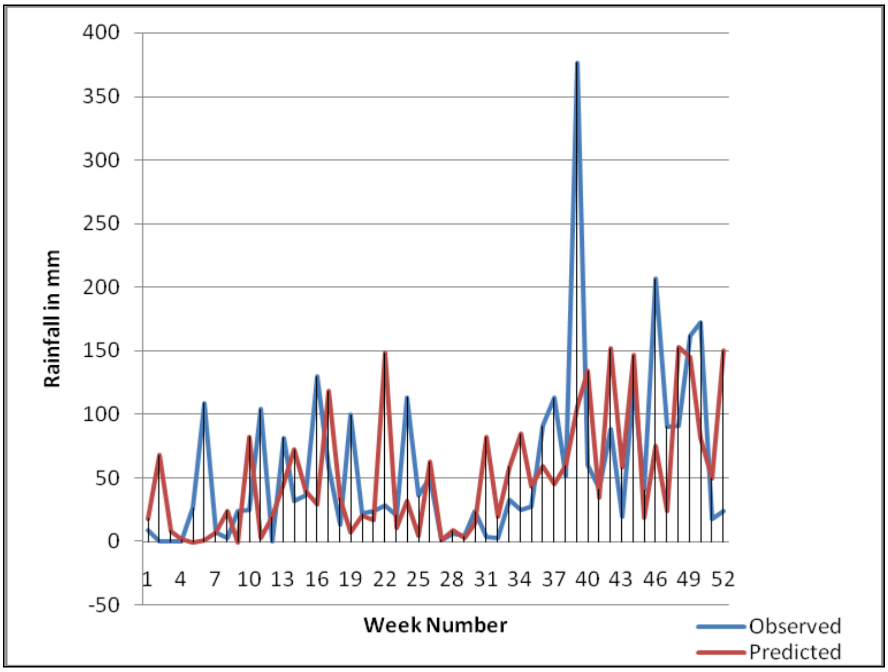

Figure 6.8: Observed and predicted weekly rainfall in 2015 using the model SARIMA (1,0,0) $(1,1,0)_{52}$

According to the above figure, there is not much good agreement with the forecasted and observed values in weekly rainfall. Since the heteroskedasticity existed in the mean models, thus it is required to fit a variance model in addition to the mean equation. The section 6.3 describes the model which can be used to capture the not only mean behavior but also variation in the weekly rainfall series.

6.3. Development of GARCH/ARCH Model

Many GARCH models were employed to capture the conditional variance existed from the best fitted mean model. The model SARIMA (1,0,0) $(1,1,0)_{52}$-GARCH (1,2) is selected as the best fitted hybrid model for the weekly rainfall series. The parameters estimation is presented in Table 6.5. Based on the result shown in the Table 6.5, all the model parameters are significant except constant term. Thus, model assumptions are tested and the residuals and squared residuals derived from the model are in random order at the 0.05 level of significance.

Table 6.5: Parameter estimation of the model SARIMA (1,0,0) $(1,1,0)_{52}$-GARCH (1,2)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.592860	1.775992	0.896885	0.3698
AR(1)	-0.189212	0.087144	-2.171262	0.0299
SAR(1)	0.245646	0.083604	2.938202	0.0033
Variance Equation				
C	813.5964	68.23715	11.92307	0.0000
RESID(-1)^2	0.191982	0.016442	11.67666	0.0000
GARCH(-1)	1.128340	0.032597	34.61500	0.0000
GARCH(-2)	-0.410790	0.019678	-20.87536	0.0000

The model was tested for the independent data set (The weeks in 2015) and the observed and the predicted values are presented by the Figure 6.9.

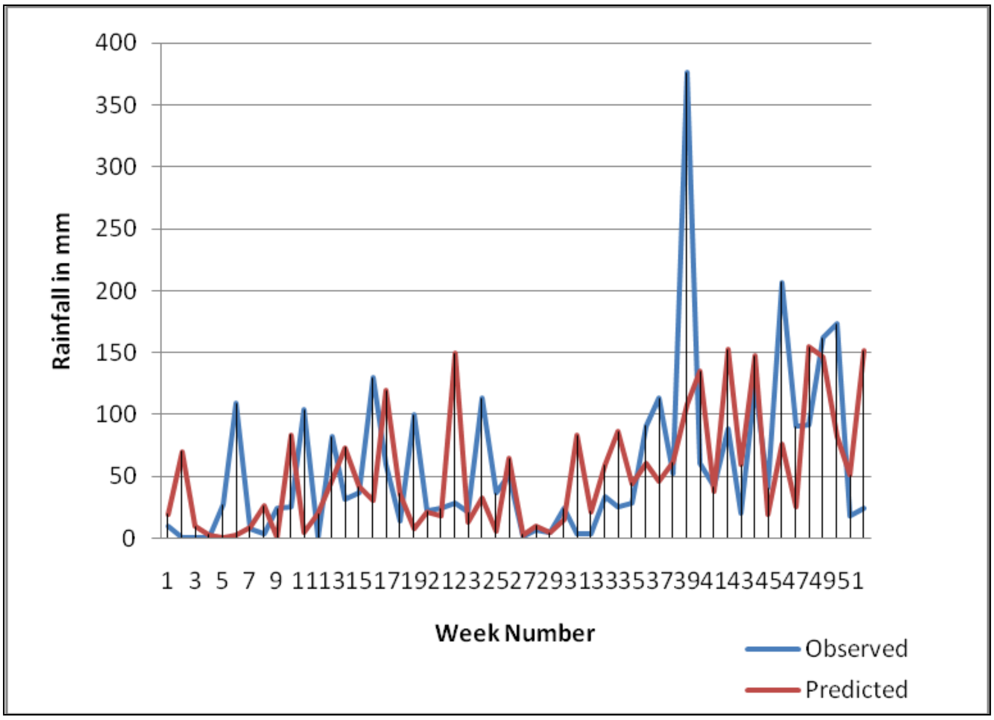

Figure 6.9: Observed and predicted weekly rainfall in 2015 using the model SARIMA (1,0,0) $(1,1,0)_{52}$- GARCH (1,2)

According to the Figure 6.9, we cannot see much agreement between the observed weekly rainfall with the predicted. Thus, to improve the power of the forecasting auto regressive integrated moving average models were fitted for the deseasonalized series. The corresponding details of the deseasonalized series is explained the next section.

6.4. Modeling for Deseasonalized Data

As it was found difficult to fit GARCH models with high forecasting performance, as an alternative method seasonality was removed from the original series, assuming that the observed series can be represented as

$$Y_t = S_{ti} + T_t + e_t \ (t=1,2\ldots\ 1300;\ i=1,2\ldots.52)$$

Initially, detrended series was calculated and by getting the averages for the 52 weeks over the period of 1990 to 2014 make the seasonal index by assuming the additional model. Thus, deseasonalized series was made based on the calculated seasonal index.

Different Box-Jenkins models were applied for the deseasonalized series to capture the mean behavior. Finally, AR (1) model is selected as the best fitted model to describe the deseasonalized data. Diagnostic tests were carried out and it was found that the selected model residual is random at 0.05 level of significant. But the assumption which is residual is normally distributed is still deviated at the 0.05 significant level. Also, the squared residuals are significantly deviated (Figure 6.10) and indicated that the time dependence variance. Thus, the heteroskedasticity test was applied to test the ARCH effect and the result is presented in Table 6.6. Furthermore, Table 6.7 describes the test result of the serial correlation of the selected AR (1) model for the deseaonalized data.

Table 6.6: The result of ARCH effect of AR (1) for deseasonalized data

Heteroskedasticity Test: ARCH (AR (1) for deseasonalized Data)			
F-statistic	4.056724	Prob. F(3,1292)	0.0070
Obs*R-squared	12.09393	Prob. Chi-Square(3)	0.0071

The result shown in Table 6.6 indicates that the presence of the ARCH effect at 0.05 level of significance.

Table 6.7: The result of serial correlation of AR (1) for deseasonalized data

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.242785	Prob. F(3,1294)	0.8665
Obs*R-squared	0.730758	Prob. Chi-Square(3)	0.8659

Based on the test result of the serial correlation LM test, there is no evidence to reject the null hypothesis that the no serial correlations at 0.05 level of significance. This implies that the there is no serial correlation.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.031	0.031	1.2274	0.268
		2	0.051	0.050	4.5642	0.102
		3	0.080	0.077	12.873	0.005
		4	0.060	0.054	17.643	0.001
		5	-0.01...	-0.02...	18.025	0.003
		6	-0.01...	-0.02...	18.266	0.006
		7	0.010	0.004	18.407	0.010
		8	-0.02...	-0.02...	19.201	0.014
		9	-0.02...	-0.02...	20.111	0.017
		1...	-0.03...	-0.02...	21.489	0.018
		1...	-0.03...	-0.02...	22.723	0.019
		1...	-0.03...	-0.02...	24.219	0.019
		1...	-0.02...	-0.01...	24.982	0.023
		1...	-0.03...	-0.02...	26.231	0.024
		1...	-0.03...	-0.02...	28.004	0.022
		1...	0.013	0.021	28.231	0.030
		1...	-0.03...	-0.02...	29.412	0.031
		1...	0.004	0.009	29.433	0.043
		1...	-0.00...	-0.00...	29.493	0.059
		2...	-0.00...	-0.00...	29.498	0.078
		2...	-0.01...	-0.01...	29.619	0.100
		2...	-0.00...	-0.00...	29.619	0.128
		2...	0.033	0.030	31.089	0.121
		2...	0.044	0.042	33.622	0.092
		2...	0.180	0.175	76.382	0.000
		2...	0.034	0.018	77.931	0.000
		2...	0.072	0.045	84.801	0.000
		2...	0.021	-0.01...	85.366	0.000
		2...	0.031	0.003	86.671	0.000
		3...	-0.00...	-0.01...	86.719	0.000
		3...	-0.00...	-0.00...	86.763	0.000
		3...	-0.02...	-0.02...	87.508	0.000
		3...	-0.02...	-0.01...	88.303	0.000
		3...	-0.01...	-0.00...	88.569	0.000
		3...	-0.03...	-0.01...	90.119	0.000
		3...	-0.04...	-0.02...	92.338	0.000

Figure 6.10: The correlogram of squared residuals of the model AR (1) for deseasonalized data

Since the squared residuals is not random and the ARCH effect is presented, the various GARCH models are utilized to capture the non constant variance by keeping the mean model as AR (1) for deseasonalized data series. Out of the various models, AR (1)-GARCH (1,1) model is selected as the best model for the forecasting weekly rainfall in Colombo city. The diagnostic test was carried out for the best fitted model as done for the pervious. The model parameter estimates are presented in Table 6.8.

Table 6.8: The result of the estimated AR (1)-GARCH (1,1) model for deseasonalized series

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	45.76302	1.285538	35.59835	0.0000
AR(1)	0.111689	0.028697	3.892002	0.0001
Variance Equation				
C	676.1237	56.75019	11.91403	0.0000
RESID(-1)^2	0.342132	0.029182	11.72395	0.0000
GARCH(-1)	0.515348	0.029095	17.71282	0.0000

Based on the above result all the parameters are significant at 0.05 level of significance. The residual analysis was carried out and the residuals as well as squared residuals are not significantly deviated from the randomness. The corresponding correlogram plots are presented by Figure 6.11 and Figure 6.12 respectively. Based on the test result of the ARCH test (Table 6.9), it can be concluded that the there is no ARCH effect moreover.

Table 6.9: The result of ARCH effect of AR(1)-GARCH(1,1) for deseasonalized data

Heteroskedasticity Test: ARCH (AR(1)-GARCH(1,1))			
F-statistic	0.509352	Prob. F(3,1292)	0.6759
Obs*R-squared	1.530976	Prob. Chi-Square(3)	0.6751

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob...
		1	0.000	0.000	0.0003	
		2	0.021	0.021	0.5715	0.450
		3	-0.01...	-0.01...	0.6971	0.706
		4	0.044	0.043	3.1691	0.366
		5	-0.00...	-0.00...	3.1693	0.530
		6	-0.00...	-0.00...	3.1693	0.674
		7	0.040	0.041	5.3105	0.505
		8	0.004	0.002	5.3337	0.619
		9	-0.04...	-0.04...	8.0276	0.431
		1...	-0.04...	-0.04...	10.320	0.325
		1...	-0.02...	-0.02...	10.902	0.365
		1...	0.010	0.011	11.044	0.440
		1...	-0.03...	-0.03...	12.635	0.396
		1...	-0.03...	-0.02...	13.822	0.387
		1...	-0.06...	-0.05...	18.738	0.175
		1...	-0.00...	0.003	18.738	0.226
		1...	-0.00...	-0.00...	18.835	0.277
		1...	0.002	0.003	18.838	0.338
		1...	0.028	0.029	19.839	0.342
		2...	0.016	0.015	20.193	0.383
		2...	-0.00...	-0.00...	20.202	0.445
		2...	-0.03...	-0.02...	21.516	0.428
		2...	0.004	-0.00...	21.542	0.487
		2...	-0.02...	-0.03...	22.383	0.497
		2...	0.006	-0.00...	22.432	0.553
		2...	0.003	0.001	22.444	0.610
		2...	0.006	0.002	22.488	0.662
		2...	-0.04...	-0.04...	25.200	0.563
		2...	-0.01...	-0.01...	25.673	0.591
		3...	-0.00...	-0.00...	25.767	0.638
		3...	0.019	0.018	26.235	0.663
		3...	0.010	0.013	26.356	0.704
		3...	-0.02...	-0.02...	27.153	0.711
		3...	0.043	0.047	29.642	0.635
		3...	-0.02...	-0.02...	30.474	0.641
		3...	0.012	0.011	30.673	0.677

Figure 6.11: The correlogram of residuals derived from the model AR (1)-GARCH (1,1) for the deseasonalized data

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob...
		1	-0.03...	-0.03...	1.1966	0.274
		2	-0.01...	-0.01...	1.4894	0.475
		3	-0.00...	-0.00...	1.4895	0.685
		4	0.022	0.022	2.1314	0.712
		5	-0.02...	-0.02...	2.8797	0.719
		6	-0.01...	-0.01...	3.2039	0.783
		7	0.018	0.016	3.6319	0.821
		8	-0.01...	-0.01...	4.0573	0.852
		9	-0.02...	-0.02...	4.8189	0.850
		1...	-0.03...	-0.04...	6.7078	0.753
		1...	-0.02...	-0.03...	7.6216	0.747
		1...	-0.02...	-0.02...	8.5398	0.742
		1...	-0.01...	-0.02...	9.0339	0.770
		1...	-0.02...	-0.02...	9.6627	0.786
		1...	-0.03...	-0.03...	11.277	0.733
		1...	0.015	0.011	11.593	0.772
		1...	-0.02...	-0.02...	12.318	0.781
		1...	0.013	0.009	12.530	0.819
		1...	-0.00...	-0.00...	12.531	0.862
		2...	-0.00...	-0.01...	12.553	0.896
		2...	-0.01...	-0.01...	12.752	0.917
		2...	-0.01...	-0.01...	12.905	0.936
		2...	0.025	0.017	13.744	0.934
		2...	0.020	0.016	14.248	0.941
		2...	0.123	0.120	34.174	0.104
		2...	0.042	0.049	36.523	0.082
		2...	0.026	0.030	37.408	0.088
		2...	-0.00...	-0.00...	37.460	0.109
		2...	-0.00...	-0.01...	37.495	0.134
		3...	0.004	0.005	37.516	0.163
		3...	0.008	0.012	37.594	0.193
		3...	-0.02...	-0.02...	38.450	0.200
		3...	-0.01...	-0.00...	38.619	0.231
		3...	-0.01...	-0.00...	38.787	0.263
		3...	-0.03...	-0.02...	40.544	0.239
		3...	-0.03...	-0.02...	42.411	0.214

Figure 6.12: The correlogram of squared residuals derived from the model AR (1)-GARCH (1,1) for the deseasonalized data

Thus, the best selected model was tested for the independent data set and the observed and the predicted weekly rainfall is depicted from the Figure 6.13. Furthermore, the forecasting result was evaluated by calculating the absolute error in mm as shown in the Table 6.10.

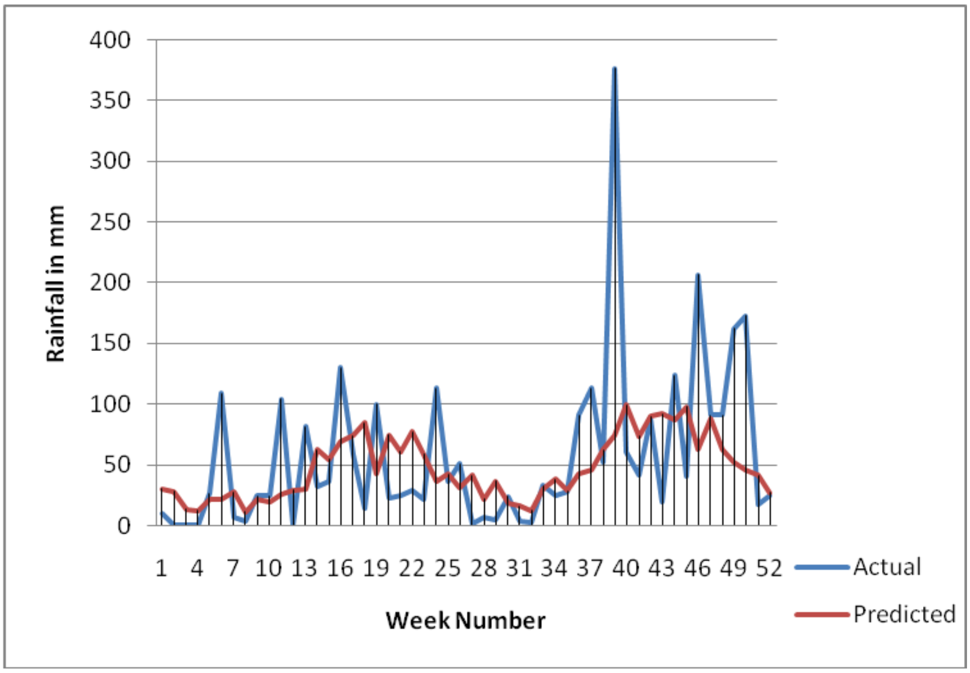

Figure 6.13: Actual and predicted weekly rainfall in 2015 using AR (1)-GARCH (1,1) for deseasonalized data

Table 6.10: The absolute error in mm for the weekly rainfall in 2015 [AR (1) - GARCH (1,1)] model for deseasonalized data

Absolute Error in mm	Number of weeks	Cumulative
00--10	11 (21.2)	11 (21.2)
11--15	6 (11.5)	17 (32.7)
16--20	4 (7.7)	21 (40.4)
21--25	3 (5.8)	24 (46.2)
26--30	3 (5.8)	27 (52.0)
31--35	3 (5.8)	30 (57.8)
36--40	5 (9.6)	35 (67.4)
41--45	0 (0.0)	35 (67.4)
46--50	2 (3.8)	37 (71.2)
More than 50	15 (28.8)	52 (100.0)

The Figure 6.13 depicts the much good agreement between the observed and the predicted except extreme values. According to the Table 6.10, 21.2% of weeks rainfall can be predicted very closely with less than 10mm error bound from the observed values. Also, it is clear that the weekly rainfall can be predicted with less than 15mm error bound of 32.7% weeks in 2015. However, in 28.8% weeks in 2015 showed more than 50mm error.

6.5. Modeling Weekly Rainfall with Exogenous Variables using VAR

Minimum, maximum and the average of the three variables temperature, relative humidity and vapor pressure were taken to account as exogenous predictor variables in modeling weekly rainfall. The dynamic relationships among the climatic variables including rainfall are assessed based on the Vector Autoregressive model (VAR). VAR is often used for multivariate time series modeling specially in the fields as finance and agriculture. Relatively few studies can be found in literature in modeling climatic variables using VAR (Farook and Kannan, 2015).

Initially all the variables are tested for the stationary using Augmented Dickey Fuller Test and those result are presented in the Table 6.11.

Table 6. 11: Result of Augmented Dickey Fuller (ADF) test for determining the stationary of the time series

Variable	ADF test Statistics	p-value	Integration of order
Rainfall	-20.11381	0.0000	I(o)
MinTemp	-8.064293	0.0000	I(o)
AvgTemp	-9.404993	0.0000	I(o)
MaxTemp	-7.361883	0.0000	I(o)
MinRH	-9.341557	0.0000	I(o)
AvgRH	-8.217019	0.0000	I(o)
MaxRH	-19.68546	0.0000	I(o)
MinVapp	-9.920362	0.0000	I(o)
AvgVapp	-11.39427	0.0000	I(o)
MaxVapp	-9.606267	0.0000	I(o)

Table 6.11 clearly indicated that those climatic variables including the rainfall are stationary in their level form. To examine the correlation structure among the variables at various lag lengths, Pearson correlation matrix along with the corresponding p values were taken and those result is presented in the Table 6.12.

Table 6.12: The correlation between rainfall and exogenous climatic variables at lag 1 and lag 2

Variable	**Correlation**	**pvalue**	**Variable**	**Correlation**	**pvalue**
Tmin(-1)	0.055	0.048	Tmin(-2)	0.065	0.020
AvgTemp(-1)	0.004	0.883	AvgTemp(-2)	0.012	0.677
Tmax(-1)	-0.064	0.020	Tmax(-2)	-0.066	0.017
MinRH(-1)	0.169	0.000	MinRH(-2)	0.148	0.000
AvgRH(-1)	0.241	0.000	AvgRH(-2)	0.153	0.000
MaxRH(-1)	0.134	0.000	MaxRH(-2)	0.124	0.000
MinVap(-1)	0.186	0.000	MinVap(-2)	0.181	0.000
AvgVap(-1)	0.213	0.000	AvgVap(-2)	0.184	0.000
MaxVap(-1)	0.184	0.000	MaxVap(-2)	0.131	0.000

Based on the above table, average temperature does not give significant correlation with rainfall at the lag 1 as well as lag 2. The average relative humidity at lag 1 showed maximum significant correlation to the rainfall out of the all other variables (0.241). The second highest correlation is presented in average vapor pressure at lag 1(0.213). It is noted that the correlation between the weekly rainfall series and exogenous climatic variables are fairly low but significant except the variable average temperature at 0.05 level of significance. Stepwise regression was carried out with the above variables with lag 1 and lag 2 except average temperature. The corresponding result is presented in Table 6.13 and Table 6.14 respectively.

Table 6.13: Analysis of Variance of stepwise regression at lag 1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	356212	89053	26.89	0.000
AvgRH(-1)	1	70652	70652	21.33	0.000
AvgVap(-1)	1	64496	64496	19.47	0.000
Tmax(-1)	1	7810	7810	2.36	0.125
Tmin(-1)	1	27914	27914	8.43	0.004
Error	1293	4282289	3312		
Total	1297	4638500			

Model Summary	S	R-sq	R-sq(adj)	R-sq(pred)
	57.5491	7.68%	7.39%	6.93%

According to the Table 6.13, only four variables at lag1 are selected as best fitted variables to the rainfall. Also, low R^2 value indicated that the 7.68% variation of weekly rainfall only explained by the other exogenous climatic variables which is very small.

Table 6.14: Analysis of Variance of stepwise regression at lag 2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	417558	52195	15.94	0.000
AvgRH(-1)	1	54809	54809	16.74	0.000
AvgVap(-1)	1	34235	34235	10.45	0.001
Tmax(-1)	1	14039	14039	4.29	0.039
Tmin(-1)	1	20203	20203	6.17	0.013
Tmin(-2)	1	7451	7451	2.28	0.132
MaxRH(-2)	1	16849	16849	5.15	0.023
MinVap(-2)	1	37323	37323	11.40	0.001
MaxVap(-2)	1	7776	7776	2.37	0.124
Error	1289	4220943	3275		
Total	1297	4638500			

Model Summary	S	R-sq	R-sq(adj)	R-sq(pred)
	57.2240	9.00%	8.44%	7.82%

Eight variables are selected as predictor variables to the weekly rainfall series. However, the low R^2 value indicated that the 9% of variation in weekly rainfall explained by the other exogenous climatic variables is also small.

To find the optimal lag value to select best order of the VAR model, the selection criteria are obtained and the result are presented in Table 6.15.

Table 6.15: Values of the selection criterion for selecting the optimal lag order

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-27140.15	NA	12685234	41.89684	41.93273	41.91031
1	-25787.94	2683.564	1783698.	39.93509	40.29391*	40.06974
2	-25582.75	404.3669	1472632.	39.74344	40.42520	39.99927*
3	-25465.23	229.9459	1392008.	39.68709	40.69179	40.06412
4	-25365.44	193.8826*	1352330.*	39.65809*	40.98574	40.15631

Table 6.15 provides the values of different criterion for the different lag length order and selected lag 2 as optimal lag length based on the HQ (Hannan-Quinn Information Criterian). Based on the optimal lag length, the VAR was applied and the corresponding result is presented by Table 6.16.

Table 6.16 : VAR model for weekly rainfall series

	RAINFALL	TMIN	TMAX	AVGRH	MAXRH	MINVAP	AVGVAP	MAXVAP
RAINFALL (-1)	0.114147	-0.001294	-0.002113	0.001444	-0.000764	0.001658	0.000727	0.001177
	(0.03255)	(0.00061)	(0.00043)	(0.00191)	(0.00157)	(0.00120)	(0.00068)	(0.00067)
	[3.50656]	[-2.1060]	[-4.8846]	[0.75592]	[-0.4864]	[1.37871]	[1.07647]	[1.74997]
RAINFALL (-2)	0.073526	-0.000359	-0.000164	1.38E-05	0.000372	-0.00059	6.81E-05	-0.000574
	(0.03285)	(0.00062)	(0.00044)	(0.00193)	(0.00159)	(0.00121)	(0.00068)	(0.00068)
	[2.23792]	[-0.5783]	[-0.3756]	[0.00717]	[0.23472]	[-0.4866]	[0.10001]	[-0.84505]
TMIN(-1)	-3.466618	0.331014	-0.001399	0.228930	0.135557	0.194435	0.028017	-0.029825
	(1.89205)	(0.03570)	(0.02515)	(0.11104)	(0.09132)	(0.06988)	(0.03924)	(0.03909)
	[-1.83220]	[9.27162]	[-0.0556]	[2.06167]	[1.48444]	[2.78237]	[0.71397]	[-0.76289]
TMIN(-2)	-3.664593	0.201666	0.015203	0.363688	0.247954	0.120818	0.035938	-0.002809
	(1.86433)	(0.03518)	(0.02478)	(0.10941)	(0.08998)	(0.06886)	(0.03867)	(0.03852)
	[-1.96564]	[5.73261]	[0.61355]	[3.32397]	[2.75564]	[1.75463]	[0.92945]	[-0.07291]
TMAX(-1)	2.717274	0.018661	0.459521	0.217488	0.732540	0.017819	0.079054	0.186847
	(2.08428)	(0.03933)	(0.02770)	(0.12232)	(0.10060)	(0.07698)	(0.04323)	(0.04307)
	[1.30370]	[0.47448]	[16.5878]	[1.77798]	[7.28195]	[0.23148]	[1.82877]	[4.33853]
TMAX(-2)	-3.85025	-0.008167	0.331459	-0.127484	0.294272	-0.17194	0.060786	0.188907
	(2.09235)	(0.03948)	(0.02781)	(0.12280)	(0.10099)	(0.07728)	(0.04340)	(0.04323)
	[-1.84016]	[-0.2068]	[11.9188]	[-1.0381]	[2.91399]	[-2.2249]	[1.40076]	[4.36946]
AVGRH(-1)	1.308890	-0.009177	-0.020952	0.379513	0.221560	-0.02097	-0.014687	-0.01269
	(0.63147)	(0.01192)	(0.00839)	(0.03706)	(0.03048)	(0.02332)	(0.01310)	(0.01305)
	[2.07277]	[-0.7701]	[-2.4963]	[10.2406]	[7.26963]	[-0.8990]	[-1.1214]	[-0.97255]
AVGRH(-2)	-1.498058	0.017092	0.009009	0.137909	0.090803	0.011202	-0.009271	0.016320
	(0.63678)	(0.01202)	(0.00846)	(0.03737)	(0.03073)	(0.02352)	(0.01321)	(0.01316)
	[-2.35256]	[1.42247]	[1.06442]	[3.69022]	[2.95450]	[0.47632]	[-0.701]	[1.24035]
MAXRH(-1)	-0.045762	0.015426	0.040348	0.109782	0.188727	-0.00033	0.002127	-0.001478
	(0.63110)	(0.01191)	(0.00839)	(0.03704)	(0.03046)	(0.02331)	(0.01309)	(0.01304)
	[-0.07251]	[1.29540]	[4.81018]	[2.96406]	[6.19601]	[-0.0140]	[0.16248]	[-0.11336]
MAXRH(-2)	0.875696	0.016573	0.049957	0.082326	0.187653	-0.03195	0.001197	-0.000951
	(0.62913)	(0.01187)	(0.00836)	(0.03692)	(0.03036)	(0.02324)	(0.01305)	(0.01300)
	[1.39191]	[1.39603]	[5.97442]	[2.22969]	[6.17997]	[-1.3751]	[0.09170]	[-0.07314]

Table 6.16 (Continued)

MINVAP(-1)	0.900472	-0.057213	0.017785	0.012091	-0.049677	0.003423	-0.0498	-0.028474
	(1.34247)	(0.02533)	(0.01784)	(0.07879)	(0.06479)	(0.04958)	(0.02784)	(0.02774)
	[0.67076]	[-2.2585]	[0.99674]	[0.15347]	[-0.7667]	[0.06904]	[-1.7886]	[- 1.02651]
MINVAP(-2)	2.316140	-0.040715	-0.035085	0.117999	0.055295	-0.01574	-0.00663	-0.052218
	(1.33987)	(0.02528)	(0.01781)	(0.07863)	(0.06467)	(0.04949)	(0.02779)	(0.02769)
	[1.72863]	[-1.6103]	[-1.9701]	[1.50060]	[0.85505]	[-0.3181]	[-0.2385]	[- 1.88612]
AVGVAP(-1)	3.569213	0.352814	-0.029773	-0.001409	-0.127685	0.817008	0.673464	0.430800
	(3.03359)	(0.05724)	(0.04032)	(0.17804)	(0.14641)	(0.11204)	(0.06292)	(0.06268)
	[1.17657]	[6.16355]	[-0.7384]	[-0.0079]	[-0.8720]	[7.29196]	[10.7041]	[6.87280]
AVGVAP(-2)	2.910134	-0.017603	-0.013006	-0.116066	-0.189934	0.239285	0.255720	0.260364
	(3.11163)	(0.05871)	(0.04136)	(0.18262)	(0.15018)	(0.11492)	(0.06454)	(0.06429)
	[0.93524]	[-0.2998]	[-0.3144]	[-0.6355]	[-1.2647]	[2.08210]	[3.96249]	[4.04955]
MAXVAP(-1)	-0.16089	-0.040276	0.000845	0.010000	-0.083543	-0.08684	-0.02291	0.075961
	(1.90928)	(0.03603)	(0.02538)	(0.11205)	(0.09215)	(0.07052)	(0.03960)	(0.03945)
	[-0.08427]	[-1.1179]	[0.03332]	[0.08924]	[-0.9066]	[-1.2315]	[-0.5785]	[1.92546]
MAXVAP(-2)	-3.052905	0.030094	0.015523	0.110554	0.140071	-0.0009	0.000219	0.000489
	(1.92530)	(0.03633)	(0.02559)	(0.11299)	(0.09292)	(0.07111)	(0.03993)	(0.03978)
	[-1.58568]	[0.82837]	[0.60660]	[0.97842]	[1.50739]	[-0.0126]	[0.00547]	[0.01229]
R-squared	0.098321	0.421221	0.532502	0.336899	0.031768	0.509402	0.624012	0.474170
Adj. R-squared	0.087771	0.414449	0.527032	0.329141	0.020440	0.503662	0.619613	0.468017
Sum sq. resids	4182438.	1489.177	738.8438	14405.55	9742.790	5705.335	1799.071	1785.674
S.E. equation	57.11770	1.077777	0.759158	3.352130	2.756752	2.109583	1.184623	1.180204
F-statistic	9.319479	62.20042	97.35051	43.42276	2.804234	88.74249	141.8454	77.06996
Log likelihood	-7084.291	-1930.954	-1476.075	-3403.788	-3149.973	-2802.68	-2053.646	-2048.795
Akaike AIC	10.94036	2.999929	2.299037	5.269319	4.878233	4.343107	3.188976	3.181502
Schwarz SC	11.00407	3.063640	2.362748	5.333030	4.941945	4.406819	3.252687	3.245213
Mean dependent	45.47195	23.49615	31.84831	80.97042	95.39599	27.03043	29.83451	32.19052
S.D. dependent	59.80243	1.408467	1.103866	4.092655	2.785365	2.994385	1.920733	1.618111

According to the Table 6.16, the effects of the exogenous climatic variables on the rainfall are considerably small. Though the adjusted R^2 8.8%, the fitted model explained only 9.8% (R^2=0.098) of the total variation in the rainfall. The Granger causality test was applied to examine the direction of causality among the variables.

This is the technique can be used to determine to one time series is useful in forecasting another. The test result is displayed from the Table 6.17.

Table 6.17: Result of Granger Causality test

VAR Granger Causality/Block Exogeneity Wald tests

Dependent variable: RAINFALL

Excluded	Chi-sq	df	Prob.
TMAX	11.12273	2	**0.0038**
TMIN	3.403311	2	0.1824
AVGRH	7.208089	2	**0.0272**
MAXRH	1.981949	2	0.3712
MINVAPPRES	3.675845	2	0.1591
AVGVAPPRES	3.063885	2	0.2161
MAXVAPPRE	2.565040	2	0.2773
All	260.4913	14	0.0000

According to the above test result, the most of the null hypothesis that the exogenous climatic variables do not Granger-cause the rainfall do not reject. Average Relative humidity and maximum temperature are the only two variables made significant impact on rainfall and useful in forecasting rainfall at the 0.05 level of significance. Impulse response function was taken and corresponding graphs are presented from the Figure 6.14.

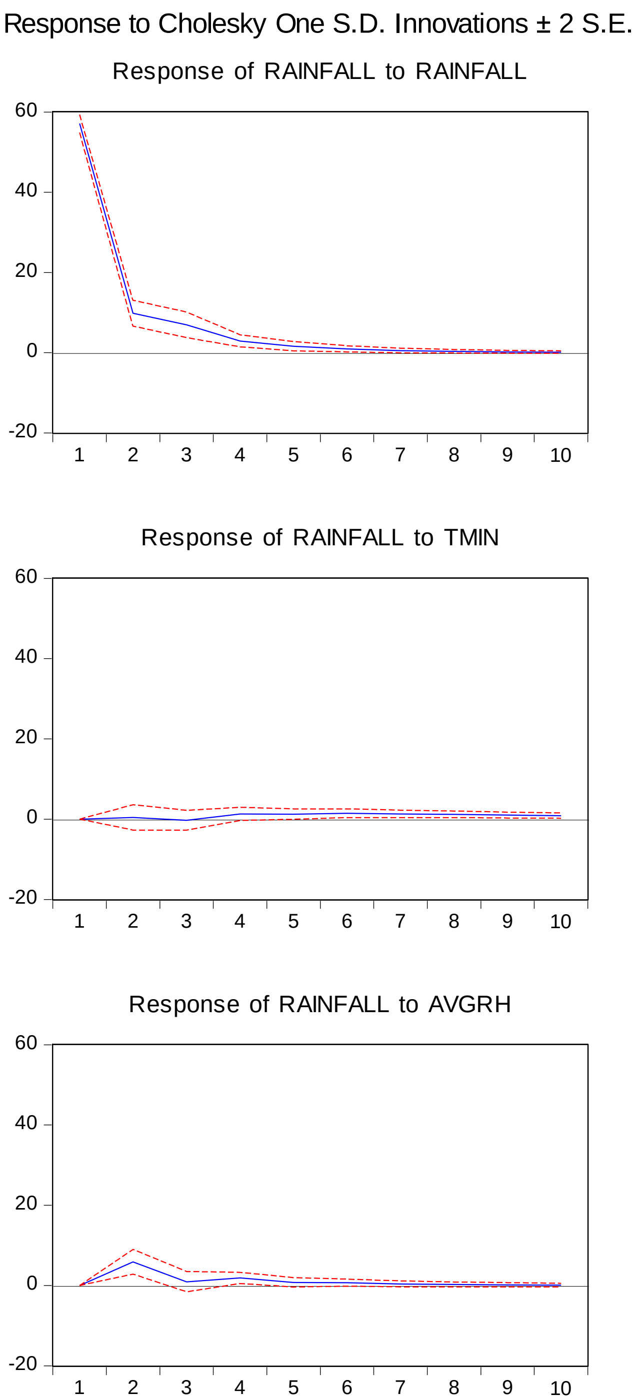

Figure 6.14: Impulse response function of average relative humidity to rainfall

According to the above figures, there is a positive effect on rainfall in the future with increasing of average relative humidity in the current period. Also, it can be seen a decreasing trend of positive effect until the 10th week. The forecasting in weekly rainfall is made based on the fitted VAR model. However, the forecasting performance is not in satisfactory level.

6.5.1. Modeling Deseasonalized Weekly Rainfall with Exogenous Variables

Initially deseasonalized rainfall is tested for the stationary using Augmented Dickey Fuller Test and those result is presented in the Table 6.18.

Table 6.18. Result of Augmented Dickey Fuller (ADF) test for the deseasonalized rainfall series

Null Hypothesis: DESEARAINFALL has a unit root
Exogenous: Constant

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-33.92381	0.0000
Test critical values:	1% level	-3.434980	
	5% level	-2.863472	
	10% level	-2.567848	

*MacKinnon (1996) one-sided p-values.

To find the most favorable lag value for the VAR model the selection criterion are used and those result are presented in Table 6.19.

Table 6.19: Values of the selection criterion for selecting the optimal lag order

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-23241.97	NA	527846.0	35.87958	35.91147	35.89155
1	-21955.63	2554.803	80036.91	33.99326	34.28032*	34.10098
2	-21760.72	384.7076	65396.96	33.79124	34.33346	33.99471*
3	-21652.87	211.5483	61119.13	33.72356	34.52094	34.02278
4	-21567.13	167.0985*	59106.29*	33.69002*	34.74257	34.08500

Table 6.19 shows of different selection criterion for the different lag length order and selected lag 2 as optimal lag length based on the HQ (Hannan-Quinn Information Criterion).

Table 6.20: VAR model for the deseasonalized data

	DESEARAINFALL	TMIN	TMAX	MAXRH	AVGRH	MINVAP	AVGVAP	MAXVAP
DESEARAINFALL(-1)	0.040687	0.000310	-0.001808	-0.001742	0.000250	0.001068	0.000216	0.000676
	(0.03160)	(0.00062)	(0.00043)	(0.00156)	(0.00193)	(0.00123)	(0.00069)	(0.00069)
	[1.28737]	[0.50187]	[-4.2404]	[-1.1180]	[0.12937]	[0.86629]	[0.31238]	[0.98111]
DESEARAINFALL(-2)	0.015567	0.000786	1.69E-05	-0.000418	-0.000926	-6.71E-05	3.77E-05	-0.000831
	(0.03161)	(0.00062)	(0.00043)	(0.00156)	(0.00193)	(0.00123)	(0.00069)	(0.00069)
	[0.49255]	[1.27230]	[0.03962]	[-0.2681]	[-0.4789]	[-0.0544]	[0.05449]	[-1.20602]
TMIN(-1)	-1.098819	0.282457	-0.065415	-0.139305	0.003032	0.166319	0.003570	-0.060858
	(1.82917)	(0.03575)	(0.02468)	(0.09015)	(0.11192)	(0.07133)	(0.04004)	(0.03987)
	[-0.60072]	[7.90041]	[-2.6503]	[-1.5452]	[0.02709]	[2.33167]	[0.08916]	[-1.52634]
TMIN(-2)	-2.733121	0.142952	-0.054216	-0.016182	0.152894	0.113821	0.021673	-0.026915
	(1.82685)	(0.03571)	(0.02465)	(0.09004)	(0.11177)	(0.07124)	(0.03999)	(0.03982)
	[-1.49608]	[4.00348]	[-2.1994]	[-0.1797]	[1.36788]	[1.59771]	[0.54194]	[-0.67590]
TMAX(-1)	2.434270	-0.082116	0.350049	0.354190	-0.076543	0.000707	0.059152	0.154855
	(2.13024)	(0.04164)	(0.02874)	(0.10499)	(0.13034)	(0.08307)	(0.04663)	(0.04643)
	[1.14272]	[-1.9721]	[12.1781]	[3.37358]	[-0.5872]	[0.00851]	[1.26847]	[3.33489]
TMAX(-2)	-1.213717	-0.098667	0.251906	0.027541	-0.333099	-0.189296	0.047133	0.168991
	(2.08046)	(0.04066)	(0.02807)	(0.10254)	(0.12729)	(0.08113)	(0.04554)	(0.04535)
	[-0.58339]	[-2.4264]	[8.97349]	[0.26860]	[-2.6168]	[-2.3332]	[1.03491]	[3.72639]
MAXRH(-1)	-0.240489	-0.010287	0.014838	0.106858	0.047498	-0.003666	-0.001391	-0.007652
	(0.62495)	(0.01222)	(0.00843)	(0.03080)	(0.03824)	(0.02437)	(0.01368)	(0.01362)
	[-0.38481]	[-0.8421]	[1.75958]	[3.46929]	[1.24217]	[-0.1504]	[-0.1016]	[-0.56174]
MAXRH(-2)	0.770743	-0.008909	0.024349	0.105139	0.019601	-0.036938	-0.003016	-0.007337
	(0.62278)	(0.01217)	(0.00840)	(0.03069)	(0.03810)	(0.02429)	(0.01363)	(0.01358)
	[1.23758]	[-0.7318]	[2.89756]	[3.42538]	[0.51441]	[-1.5209]	[-0.2212]	[-0.54049]
AVGRH(-1)	1.759521	-0.039324	-0.044925	0.154067	0.330562	-0.020755	-0.015322	-0.015443
	(0.61992)	(0.01212)	(0.00836)	(0.03055)	(0.03793)	(0.02417)	(0.01357)	(0.01351)
	[2.83830]	[-3.2454]	[-5.3706]	[5.04263]	[8.71519]	[-0.8585]	[-1.1290]	[-1.14280]
AVGRH(-2)	-1.32544	-0.012459	-0.017027	0.009466	0.076492	0.003344	-0.013847	0.010172
	(0.62910)	(0.01230)	(0.00849)	(0.03101)	(0.03849)	(0.02453)	(0.01377)	(0.01371)
	[-2.10688]	[-1.0132]	[-2.0058]	[0.30530]	[1.98727]	[0.13632]	[-1.0054]	[0.74174]

Table 6.20: (Continued...)

MINVAPPRES(-1)	-0.042881	-0.055376	0.019338	-0.041285	0.019411	0.004069	-0.04914	-0.027469
	(1.27248)	(0.02487)	(0.01717)	(0.06271)	(0.07786)	(0.04962)	(0.02786)	(0.02774)
	[-0.03370]	[-2.2264]	[1.12625]	[-0.6583]	[0.24932]	[0.08201]	[-1.7641]	[-0.99031]
MINVAPPRES(-2)	1.841315	-0.038222	-0.033937	0.063208	0.125414	-0.014315	-0.005843	-0.050938
	(1.27021)	(0.02483)	(0.01714)	(0.06260)	(0.07772)	(0.04953)	(0.02781)	(0.02769)
	[1.44961]	[-1.5395]	[-1.9800]	[1.00967]	[1.61373]	[-0.2889]	[-0.2101]	[-1.83971]
AVGVAPPRES (-1)	0.111385	0.397388	0.013577	0.043512	0.136947	0.832626	0.686051	0.447533
	(2.89384)	(0.05656)	(0.03905)	(0.14262)	(0.17706)	(0.11285)	(0.06335)	(0.06308)
	[0.03849]	[7.02573]	[0.34770]	[0.30508]	[0.77346]	[7.37826]	[10.8298]	[7.09471]
AVGVAPPRES (-2)	0.073927	0.013406	0.020954	-0.039661	0.008109	0.251368	0.267729	0.275898
	(2.95293)	(0.05772)	(0.03984)	(0.14554)	(0.18067)	(0.11515)	(0.06464)	(0.06437)
	[0.02504]	[0.23228]	[0.52590]	[-0.2725]	[0.04488]	[2.18291]	[4.14172]	[4.28628]
MAXVAPPRE(-1)	0.590670	-0.048671	-0.003547	-0.095244	0.001536	-0.089669	-0.02363	0.075440
	(1.81035)	(0.03538)	(0.02443)	(0.08922)	(0.11077)	(0.07060)	(0.03963)	(0.03946)
	[0.32627]	[-1.3754]	[-0.1452]	[-1.0674]	[0.01387]	[-1.2701]	[-0.5962]	[1.91171]
MAXVAPPRE(-2)	-1.799875	0.016848	0.001790	0.090411	0.071254	-0.005049	-0.003074	-0.004285
	(1.82766)	(0.03572)	(0.02466)	(0.09008)	(0.11182)	(0.07127)	(0.04001)	(0.03984)
	[-0.98480]	[0.47164]	[0.07257]	[1.00371]	[0.63720]	[-0.0708]	[-0.0768]	[-0.10757]
C	-6.908218	16.53217	16.26530	52.97403	40.44586	2.681531	2.519121	4.048911
	(116.606)	(2.27913)	(1.57340)	(5.74695)	(7.13446)	(4.54718)	(2.55260)	(2.54177)
	[-0.05924]	[7.25370]	[10.3377]	[9.21777]	[5.66909]	[0.58971]	[0.98689]	[1.59295]
R-squared	0.023866	0.442506	0.567446	0.093633	0.352998	0.509021	0.623969	0.474648
Adj. R-squared	0.011673	0.435543	0.562043	0.082312	0.344917	0.502889	0.619272	0.468086
Sum sq. resids	3754697.	1434.410	683.6181	9120.281	14055.80	5709.765	1799.277	1784.049
S.E. equation	54.13933	1.058186	0.730520	2.668269	3.312480	2.111225	1.185153	1.180127
F-statistic	1.957448	63.54890	105.0298	8.270919	43.68141	83.00456	132.8520	72.33536
Log likelihood	-7014.272	-1906.636	-1425.656	-3107.122	-3387.837	-2803.181	-2053.72	-2048.204
Akaike AIC	10.83401	2.964000	2.222891	4.813747	5.246282	4.345425	3.190631	3.182132
Schwarz SC	10.90170	3.031694	2.290584	4.881441	5.313975	4.413118	3.258325	3.249826
Mean dependent	45.43667	23.49615	31.84831	95.39599	80.97042	27.03043	29.83451	32.19052
S.D. dependent	54.45812	1.408467	1.103866	2.785365	4.092655	2.994385	1.920733	1.618111

Based on the Table 6.20, the effects of the exogenous climatic variables on the deseasonalized rainfall is considerably small. The Granger causality test was applied to examine the direction of causality among the variables. The test result is displayed from the Table 6.21.

Table 6.21: Result of Granger Causality test

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 01/30/19 Time: 16:17
Sample: 1 1300
Included observations: 1298

Dependent variable: DESEARAINFALL

Excluded	Chi-sq	df	Prob.
TMIN	3.352485	2	0.1871
TMAX	1.308047	2	0.5199
MAXRH	1.599714	2	0.4494
AVGRH	**10.07149**	**2**	**0.0065**
MINVAPPRES	2.110683	2	0.3481
AVGVAPPRES	0.002786	2	0.9986
MAXVAPPRE	1.028125	2	0.5981
All	21.88885	14	0.0809

According to the above test result, the average relative humidity is the only variable that gives significant impact on rainfall and useful in forecasting deseasonalized rainfall at the 0.05 level of significance. Impulse response function was taken and corresponding graphs are presented from the Figure 6.15.

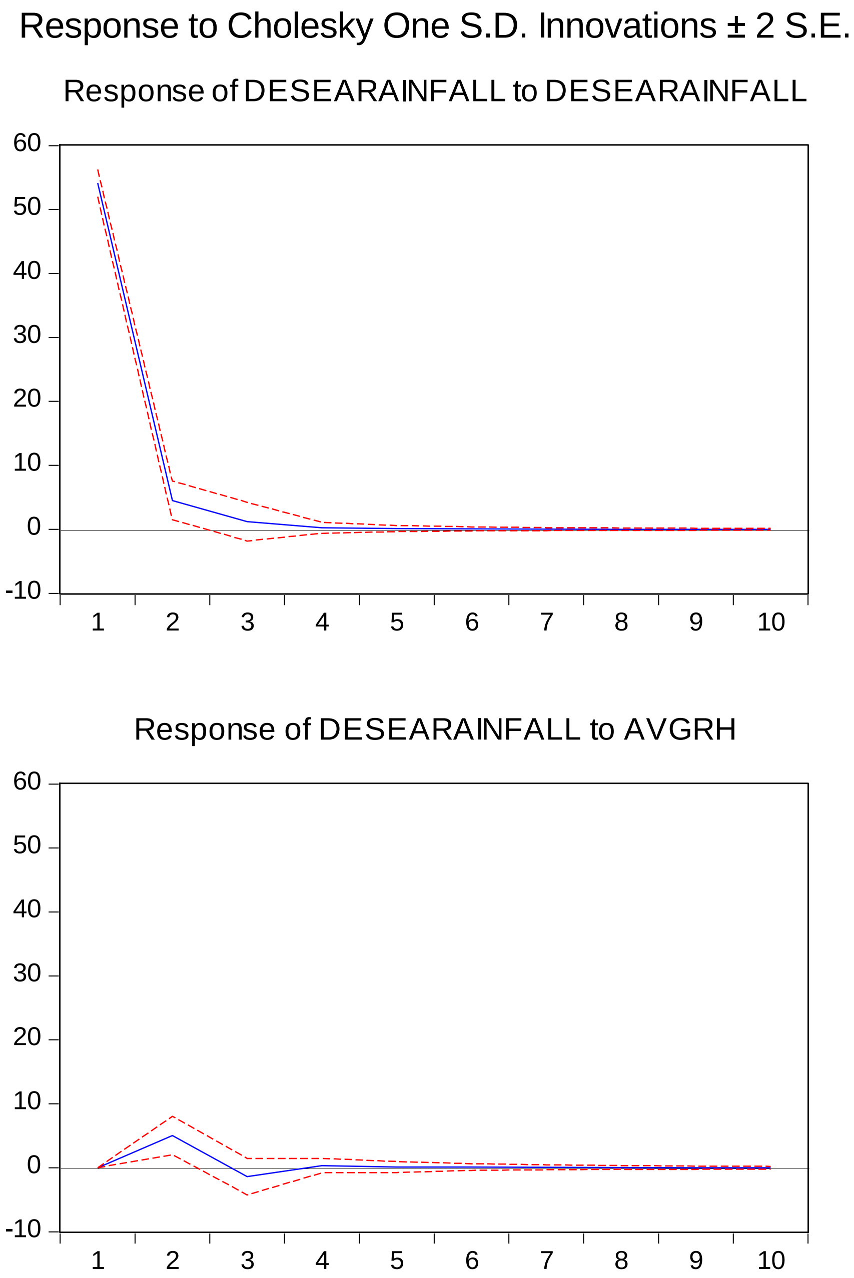

Figure 6.15: Impulse response function of average relative humidity to deseasonalized rainfall

According to the above figures, there is a positive effect on rainfall in the future with increasing of average relative humidity in the current period. Also, it can be seen that the decreasing trend of positive effect until the 10th week. The forecasting in weekly rainfall is made based on the fitted VAR model and actual and fitted weekly rainfall series is presented from the Figure 6.16.

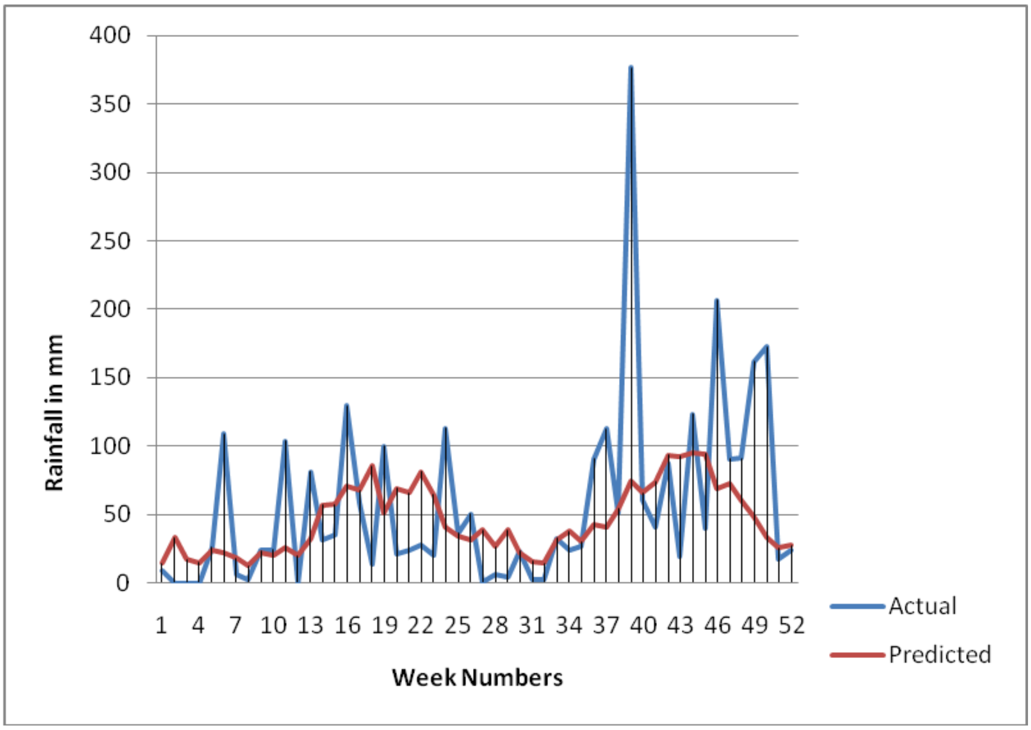

Figure 6.16: Actual and predicted rainfall in 2015 using VAR model for desesonalized data

According to the Figure 6.16, there is a much agreement of predicted values with the actual rainfall values. But still there is a noticeable gap in predicting weekly rainfall at extreme rainfall events.

6.6. Summary of the Chapter 6

The autocorrelation structure of the weekly rainfall provided the evidence to have a seasonal behavior with length of 52. Many autoregressive integrated moving average models were utilized to model weekly data series. Some models were identified as suitable models for weekly rainfall series based on the selection criteria and those models are successful in their linear domain. However, the weekly rainfall does not follow the simple linear regulations. Neither of the GARCH models were successful to model weekly data due to statistical complexity.

Thus, as an alternative, the seasonal index was calculated by assuming the additional model with length of 52 and AR (1)-GARCH (1,1) model was well fitted for the deseasonalized data series. However, forecasting accuracy was not satisfactory level. Though two exogenous variables namely, average relative humidity and maximum temperature significantly effect on weekly rainfall, give low contribution in forecasting weekly rainfall. Thus, inclusion of exogenous variables too did not improve forecasting accuracy much more and then the importance of new type of model was recommended.

CHAPTER 7

NOVEL APPROACH TO MODEL WEEKLY RAINFALL

7.1. Concept for New Modeling

As explained in the section 6.6 weekly rainfall exhibited mixed features of the non linearity phenomenon and thus, it is necessary to move to a new class of models which are beyond the conventional time series approaches. Then, long range dependency models which have been used to capture the blend features of the complex time series are considered as initial step in developing new models. The features of such models can be identified by two different approaches: (i) The spectral density function with an unbounded peak at the frequency is near to zero and (ii) The autocorrelation function decay the hyperbolically to zero. The periodigram of weekly rainfall series is shown in Figure 7.1.

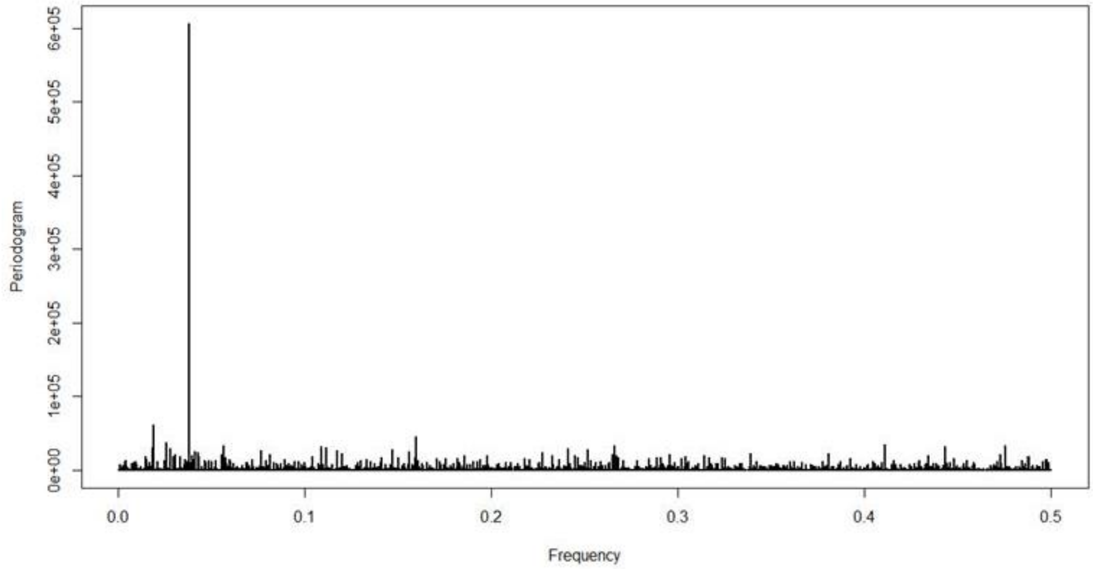

Figure 7.1: The periodgram of the rainfall series from 1990 to 2014

The Figure 7.1 exhibits unbounded spectral density at the near to zero. The maximum spectrum density of the weekly rainfall series can be seen at a frequency which is very close to zero (0.0385185). Consequently, different types of long-range dependency models are applied with some modifications to capture the real dynamic of weekly rainfall series. The exact maximum-likelihood method with Durbin-Levinson algorithm was utilized to estimate the long memory parameters and this approach has not been tested by the previous authors for rainfall studies. The exact maximum likelihood method is used to develop for all types of long-range dependency models. Five types of long-range dependency models: ARFIMA, ARFIMA for deseasonalized data, ARFIMA-GARCH, ARFIMA-GARCH for the deseasonalized data and adjusted SARFIMA-GARCH are developed to decide the best fitted model.

7.2. ARFIMA Long Range Dependency Model

The autoregressive fractionally integrated moving average (ARFIMA) long memory model is an extension of the conventional ARMA process. The model ARFIMA (p, d, q) allows the parameter "d" to take the fractional values for differencing and is known as long memory parameter.

The ARFIMA (p,d,q) model of a process $\{Y_t\}_{t\in Z}$ is given by (7.1)

$$\varphi(B)\ \nabla^d\ (Y_t - \mu) = \theta(B)\ \varepsilon_t \qquad (7.1)$$

Where μ is the mean of the process, $\{\varepsilon_t\}_{t\in Z}$ is a white noise process with zero mean and constant variance σ_ε^2, B is the backward shift operator, such that $y_{t-n} = B^n y_t$, $\varphi(B)$ and $\theta(B)$ are autoregressive and moving average polynomials of order p and q respectively such that

$$\varphi(B) = \sum_{i=1}^{p} \varphi_i\ B^i \qquad 1 \le i \le p \qquad (7.2)$$

$$\theta(B) = \sum_{j=1}^{q} \theta_j\ B^j \qquad 1 \le j \le q \qquad (7.3)$$

The differencing operator, such that $\nabla^d = (1-B)^d$ can be expressed by the binomial series as,

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!\,(d-1)!} B^k \qquad (7.4)$$

Where d is defined as the long memory parameter.

The ARFIMA (0, d,0) process is a discrete time series process $\{Y_t\}_{t\in Z}$ that satisfies the following equation.

$$\nabla^d \; Y_t = \; \varepsilon_t \qquad (7.5)$$

Where d < 1/2, $\{Y_t\}$ is a stationary process and has the infinite moving average representation

$$Y_t = \; \psi(B)\varepsilon_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \qquad (7.6)$$

Where $\psi_k = \dfrac{(k+d-1)!}{k!\,(d-1)!}$

It can be easily shown that $\psi_k \sim k^{d-1} \big/ (d-1)!$ as $k \to \infty$

Where d >-1/2, $\{Y_t\}$ is an invertible process and has the infinite autoregressive representation (7.7)

$$\pi(B)\,Y_t = \sum_{k=0}^{\infty} \pi_k Y_{t-k} = \varepsilon_t \qquad (7.7)$$

Where $\pi_k = \dfrac{(k-d-1)!}{k!\,(-d-1)!}$

As above it is clear $\pi_k \sim k^{-d-1} \big/ (-d-1)!$ as $k \to \infty$

Thus, $d \in \left(-\dfrac{1}{2}, \dfrac{1}{2}\right)$ exhibits the process is stationary and invertible. The spectral density function of the ARFIMA (0, d,0) process $\{Y_t\}_{t\in Z}$ is $f(\omega)$ that can be written as (7.8)

$$f(\omega) = \left(2\sin\frac{\omega}{2}\right)^{-2d} \quad 0 < \omega \le \pi \tag{7.8}$$

$$f(\omega) \approx \omega^{-2d} \quad \omega \to 0$$

The spectral density function $f(\omega)$ is unbounded when the frequency is near zero. Also, the auto covariance function at lag k (γ_k) and auto correlation function (ρ_k) of the process can be expressed as follows

$$\gamma_k = \text{Cov}(y_t, y_{t-k}) = E(y_t\, y_{t-k}) = \frac{(-1)^k(-2d)!}{(k-d)!(-k-d)!} \tag{7.9}$$

and

$$\rho_k = \text{Corr}(y_t, y_{t-k}) = \frac{d(1+d)\ldots(k-1+d)}{(1-d)(2-d)(3-d)\ldots(k-d)} \quad (k = 1, 2, 3, 4\ldots) \tag{7.10}$$

Hosking (1981) showed that the above auto correlation function of the process satisfies the expression $\rho(k) \approx k^{2d-1}$ when $0 < d < 1/2$ and it decays hyperbolically to zero as $k \to \infty$ for ARFIMA model. In contrast, ρ_k decays exponential for ARIMA model. The process with d = 0 reduces to a short memory ARMA model and under the assumption of stationary Gaussian with zero mean the log-likelihood function of this process is given by

$$L(\theta) = -\frac{1}{2}\log\ \det(\Gamma_\theta) - \frac{1}{2}y'\Gamma_\theta^{-1}y \tag{7.11}$$

Where $y = (y_1, y_2, y_3, \ldots\ldots y_n)'$, and Γ_θ is the variance co-variance matrix of $\{Y_t\}_{t\in Z}$ and θ is the parameter vector. The MLE for $\hat{\theta}$ was obtained by maximizing the $L(\theta)$

As explained in the section 7.1, the exact maximum likelihood estimation method was utilized for the parameter estimating and Monte Carlo simulation was carried out for different "d" values to evaluate the suitability of the estimation method before applying it for parameter estimation of the weekly rainfall series.

7.3. Results of Monte Carlo Simulation - ARFIMA (0, d,0)

A number of Monte Carlo experiments were carried out to evaluate the performance of the maximum likelihood method used for parameter estimation. The simulation was done based on various fractional differencing parameter values with 1000 replications. The four different series lengths (n=100, n=200, n=500 and n=1000) were considered for the simulation. The simulation results provided fractionally differenced parameter estimates and corresponding standard and mean square errors. Monte Carlo experiment was conducted on a simulated ARFIMA (0, d,0) series with parameter values: d=0.1, d=0.15, d=0.3 and d=0.45.

The simulation was carried out using the R programming language (Version 3.4.2) utilizing a HP11(8GB, 64bit) computer. The "arfima" package (Veenstra and Mcleod; 2012) in R optimized the log likelihood function and obtained the exact maximum likelihood estimators. Two algorithms namely Durbin-Levinson and Trench algorithms were utilized to maximize the log likelihood and obtain optimal simulation and forecasting results. The standard errors of the estimates SE ($\hat{d}$) and mean square error of the estimates MSE ($\hat{d}$) can be expressed as;

$$SE\left(\hat{d}\right) = \sqrt{\sum_{r=1}^{R}\left(\hat{d}_r - \hat{d}\right) \Big/ R} \qquad (7.12)$$

$$MSE\left(\hat{d}\right) = \sum_{r=1}^{R}\left(\hat{d}_r - d\right)^2 \Big/ R \qquad (7.13)$$

Where $\hat{d}_r$ is the MLE of d for the r^{th} replication and R is the number of replications. Tables 7.1 present the average of the estimated $(\hat{d})$, SE $(\hat{d})$ and MSE $(\hat{d})$ for d=0.1,0.15,0.3 and 0.45 for a generating process of ARFIMA(0,d,0) and for 1000 Monte Carlo replications.

Table 7.1: Result for exact maximum likelihood estimator of d for a generating process of ARFIMA (0, d,0)

d	**n**	$\hat{d}$	**SE($\hat{d}$)**	**MSE($\hat{d}$)**
0.1	100	0.05175	0.09127	0.01066
	200	0.07485	0.06268	0.00456
	500	0.08856	0.03679	0.00148
	1000	0.09499	0.02546	0.00067
0.15	100	0.10487	0.09158	0.01042
	200	0.12658	0.05936	0.00407
	500	0.14084	0.03680	0.00144
	1000	0.14560	0.02541	0.00067
0.3	100	0.24931	0.08773	0.01027
	200	0.27264	0.05750	0.00405
	500	0.28922	0.03624	0.00143
	1000	0.29474	0.02518	0.00067
0.45	100	0.37742	0.06959	0.01011
	200	0.40795	0.04772	0.00405
	500	0.43103	0.03142	0.00135
	1000	0.44359	0.02707	0.00077

The results in Tables 7.1 clearly indicates that the parameter bias has decreased with the increase in sample size irrespective of long memory parameter d. Furthermore, the results provide evidence that the parameters become consistent with the increase in series length. Also, as we expected the standard error and the MSE of the estimators have decreased with the increase in series length. Thus, it can be concluded that the performance of the maximum likelihood estimator is reasonably accurate.

7.4. Modeling Weekly Rainfall Using ARFIMA Model

The weekly rainfall series from 1990 to 2014 was used to train the model while the rest was used for validation. Various ARFIMA models were fitted for the data set and forecasting performance of the models were evaluated by using an independent sample size 52 (2015). The best fitted model was selected based on the minimum MAE (7.14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{7.14}$$

The ARFIMA (4,0.057924,4) was selected as the best fitted model and the corresponding estimates are presented in Table 7.2. The constant term was included irrespective of significant.

Table 7.2: The parameter estimates of the model ARFIMA (4, 0.05792421, 4)

Parameters	**Estimates**	**Standard Error**	**Pvalue**
φ_1	1.20698	0.024232	0.00000
φ_2	-0.24938	0.045421	0.00052
φ_3	0.57650	6.32e-07	0.00000
φ_4	-0.67522	6.32e-07	0.00000
θ_1	1.12444	0.023116	0.00000
θ_2	-0.11315	0.03651	0.00194
θ_3	0.52201	0.03542	0.00000
θ_4	-0.67435	0.02150	0.00000
Constant	-0.01633	0.03808	0.66787
d	0.05792	0.02765	0.03616

It can be concluded with 95% confidence that all model parameters except constant term are significantly different from zero. The best fitted model can be expressed by (7.15).

$$\left(1-1.206\,B+0.249\,B^2-0.576\,B^3+0.675\,B^4\right)(1-B)^{0.058}\left(Z_t+0.016\right)=\left(1+1.124\,B-0.113\,B^2+0.522\,B^3-0.674\,B^4\right)\varepsilon_t \tag{7.15}$$

Where Z_t is the standardized weekly rainfall series and B is the back-shift operator.

7.4.1. Residual Analysis for the Model ARFIMA (4, 0.05792421, 4)

The residual of the model was not significantly deviated from the random and the corresponding correlogram is depicted by Figure 7.2. The significance of p-values in Figure 7.3 confirms that there is a significant ARCH effect. The heteroskedasticity of the residual was also confirmed by ARCH LM test (Table 7.3)

Table 7.3: The result of ARCH LM test of ARFIMA (4,0.0579,4)

Heteroskedasticity Test: ARCH

F-statistic	3.721605	Prob. F(3,1293)	0.0111
Obs*R-squared	11.10348	Prob. Chi-Square(3)	0.0112

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.02...	-0.02...	0.8901	0.345
		2	0.060	0.060	5.6392	0.060
		3	-0.00...	-0.00...	5.6540	0.130
		4	0.016	0.012	5.9822	0.200
		5	0.009	0.010	6.0966	0.297
		6	0.013	0.012	6.3180	0.389
		7	0.010	0.010	6.4583	0.487
		8	0.008	0.007	6.5438	0.587
		9	-0.01...	-0.01...	6.8279	0.655
		1...	-0.03...	-0.04...	8.7697	0.554
		1...	-0.01...	-0.01...	8.9043	0.631
		1...	0.013	0.017	9.1252	0.692
		1...	-0.02...	-0.02...	9.9500	0.698
		1...	-0.02...	-0.02...	10.671	0.712
		1...	-0.05...	-0.04...	14.148	0.514
		1...	-0.00...	-0.00...	14.184	0.585
		1...	-0.03...	-0.02...	15.403	0.566
		1...	-0.00...	-0.00...	15.499	0.627
		1...	-0.00...	-0.00...	15.602	0.684
		2...	0.007	0.007	15.661	0.737
		2...	-0.02...	-0.02...	16.601	0.735
		2...	-0.05...	-0.05...	20.025	0.581
		2...	-0.01...	-0.01...	20.470	0.613
		2...	-0.04...	-0.04...	23.124	0.512
		2...	0.005	0.002	23.161	0.568
		2...	-0.02...	-0.01...	23.789	0.588
		2...	0.034	0.033	25.370	0.554
		2...	-0.04...	-0.03...	27.553	0.488
		2...	-0.01...	-0.02...	27.809	0.528
		3...	-0.02...	-0.02...	28.434	0.547
		3...	-0.00...	-0.00...	28.437	0.599
		3...	0.011	0.006	28.612	0.639
		3...	-0.03...	-0.03...	29.990	0.618
		3...	0.050	0.045	33.353	0.499
		3...	-0.02...	-0.02...	34.235	0.505
		3...	0.019	0.008	34.717	0.530

Figure 7.2: The correlogram of residuals of the model ARFIMA (4,0.0579,4)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.041	0.041	2.2125	0.137
		2	0.055	0.053	6.0893	0.048
		3	0.068	0.064	12.116	0.007
		4	0.046	0.039	14.884	0.005
		5	-0.01...	-0.02...	15.336	0.009
		6	-0.01...	-0.02...	15.696	0.015
		7	0.008	0.007	15.790	0.027
		8	-0.02...	-0.02...	16.615	0.034
		9	-0.02...	-0.02...	17.498	0.041
		1...	-0.03...	-0.02...	18.759	0.043
		1...	-0.03...	-0.02...	20.189	0.043
		1...	-0.03...	-0.02...	21.481	0.044
		1...	-0.02...	-0.01...	22.434	0.049
		1...	-0.03...	-0.02...	24.175	0.044
		1...	-0.04...	-0.03...	26.336	0.035
		1...	0.024	0.032	27.072	0.041
		1...	-0.02...	-0.02...	28.103	0.044
		1...	0.003	0.006	28.116	0.060
		1...	-0.01...	-0.01...	28.257	0.079
		2...	-0.00...	-0.00...	28.274	0.103
		2...	-0.01...	-0.01...	28.442	0.128
		2...	-0.00...	-0.00...	28.442	0.161
		2...	0.023	0.020	29.154	0.175
		2...	0.040	0.037	31.246	0.147
		2...	0.198	0.194	83.481	0.000
		2...	0.041	0.021	85.711	0.000
		2...	0.088	0.062	95.925	0.000
		2...	0.018	-0.01...	96.358	0.000
		2...	0.020	-0.00...	96.897	0.000
		3...	-0.00...	-0.00...	96.905	0.000
		3...	-0.01...	-0.01...	97.102	0.000
		3...	-0.02...	-0.02...	97.892	0.000
		3...	-0.02...	-0.00...	98.504	0.000
		3...	-0.01...	-0.00...	98.966	0.000
		3...	-0.03...	-0.01...	100.50	0.000
		3...	-0.03...	-0.01...	102.08	0.000

Figure 7.3: The correlogram of squared residuals of the model ARFIMA (4,0.0579,4)

In spite of the ARCH effect, the model was tested for an independent data set (weekly rainfall series in 2015). The observed values and the predicted values for the independent data set are shown in Figure 7.4. The absolute error was calculated to judge the forecasting power of the model (Table 7.4).

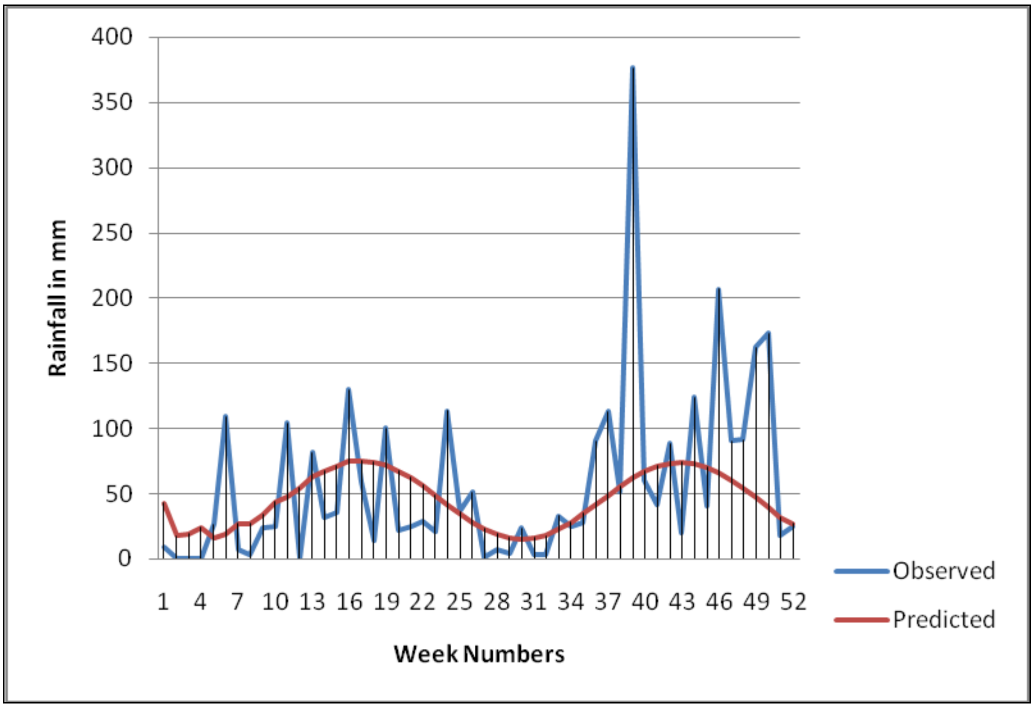

Figure 7.4: Observed and predicted weekly rainfall in 2015 using the ARFIMA (4,0.0579,4)

A comparison of result in Table 7.4 and Figure 7.4 claimed that the predicted values are in reasonably agreement with the observed rainfall values with exception for higher values of rainfall.

Table 7.4: The analysis of absolute error (in mm) for the weekly rainfall in 2015 - [ARFIMA (4,0.0579,4)]

Absolute Error in mm	**Number of weeks (%)**	**Cumulative (%)**
00--10	10 (19.2)	10 (19.2)
11--15	6 (11.5)	16 (30.7)
16--20	6 (11.5)	22 (42.2)
21--25	4 (7.7)	26 (49.9)
26--30	6 (11.5)	32 (61.4)
31--35	1 (1.9)	33 (63.3)
36--40	4 (7.7)	37 (71.0)
41--45	1 (1.9)	38 (72.9)
46--50	2 (3.9)	40 (76.8)
More than 50	12 (23.2)	52 (100.0)

The result in Table 7.4 indicated that the weeks with less than 10 mm error is 19.2% while 30.7% of weeks' absolue error was less than to 15mm. Moveover, the percentage points with absoulte error greater than 50 mm is 23.2% but this figure is lower than the corresponding percentage under the best fitted conventional model described in section 6.4 (28.8%).

7.5. ARFIMA Long Range Dependency Model for Deseasonalized Data

In the point of view of reducing variability of the original series, ARFIMA was developed to deseasonalized data series. The methodology of deseasonalization procedure was discussed in the section 6.4. The best fitted model identified for deseasonalized data is ARFIMA (5,0.05999,5) and the corresponding parameter estimates are presented in Table 7.5.

Table 7.5: The parameter estimates of the model ARFIMA (5,0.05999,5)

Parameters	**Estimates**	**Standard Error**	**Pvalue**
φ_1	-0.62733	0.06944	0.00000
φ_2	0.09028	0.04465	0.04272
φ_3	0.75049	0.02278	0.00000
φ_4	0.64047	0.05544	0.00000
φ_5	0.09181	0.04207	0.04352
θ_1	-0.61462	0.03889	0.00000
θ_2	0.12104	0.03908	0.00195
θ_3	0.79656	0.02380	0.00000
θ_4	0.61783	0.03909	0.00000
θ_5	0.07839	0.03909	0.04374
Constant	-0.01061	0.01038	0.30665
d	0.05999	0.04561	0.00000

All model parameters except constant term are significant at the 0.05 level of significance. The best fitted model is

$$\left(1+0.627\,B-0.090\,B^2-0.576\,B^3-0.751\,B^4-0.092\,B^5\right)(1-B)^{0.060}\left(Z_t+0.011\right)=$$
$$\left(1-0.615\,B+0.121\,B^2+0.797\,B^3+0.618\,B^4+0.078\,B^5\right)\varepsilon_t \quad (7.16)$$

Where Z_t is the standardized deseasonalized weekly rainfall series and B is the back-shift operator.

7.5.1. Residual Analysis

The correlogram plot of the residual (Figure 7.5) provided sufficient evidence to randomness. The correlogram of the squared residuals is depicted by Figure 7.6 indicates the assumption of the constant variance is significantly deviated. Thus, ARCH LM test is utilized to test the heteroskedasticity of the residuals and the corresponding test result is presented by the Table 7.6.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.018	0.018	0.4216	0.516
		2	0.009	0.009	0.5376	0.764
		3	0.009	0.009	0.6406	0.887
		4	0.008	0.007	0.7146	0.950
		5	-0.00...	-0.00...	0.7147	0.982
		6	0.008	0.008	0.7976	0.992
		7	0.012	0.012	0.9943	0.995
		8	0.002	0.001	1.0001	0.998
		9	-0.01...	-0.01...	1.2573	0.999
		1...	-0.04...	-0.04...	3.4117	0.970
		1...	-0.02...	-0.02...	4.1915	0.964
		1...	0.012	0.013	4.3674	0.976
		1...	-0.03...	-0.03...	5.5686	0.960
		1...	-0.03...	-0.03...	6.9676	0.936
		1...	-0.04...	-0.04...	9.3669	0.858
		1...	-0.00...	0.002	9.3696	0.897
		1...	-0.01...	-0.01...	9.5966	0.920
		1...	0.012	0.014	9.7810	0.939
		1...	0.022	0.022	10.427	0.942
		2...	0.031	0.029	11.708	0.926
		2...	-0.00...	-0.00...	11.708	0.947
		2...	-0.02...	-0.02...	12.471	0.947
		2...	0.002	-0.00...	12.476	0.962
		2...	-0.01...	-0.02...	12.923	0.967
		2...	0.019	0.015	13.392	0.971
		2...	0.004	0.001	13.412	0.980
		2...	0.042	0.040	15.759	0.957
		2...	-0.02...	-0.02...	16.485	0.958
		2...	-0.00...	-0.00...	16.581	0.968
		3...	-0.01...	-0.01...	16.757	0.975
		3...	0.008	0.009	16.843	0.982
		3...	0.017	0.015	17.210	0.985
		3...	-0.02...	-0.02...	17.861	0.985
		3...	0.051	0.055	21.338	0.955
		3...	-0.01...	-0.01...	21.596	0.963
		3...	0.011	0.013	21.749	0.971

Figure 7.5: The correlogram of residuals of the model ARFIMA (5,0.0599,5)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.029	0.029	1.0877	0.297
		2	0.052	0.051	4.6040	0.100
		3	0.084	0.081	13.723	0.003
		4	0.050	0.043	16.937	0.002
		5	-0.01...	-0.02...	17.348	0.004
		6	-0.01...	-0.02...	17.656	0.007
		7	0.009	0.005	17.768	0.013
		8	-0.02...	-0.02...	18.477	0.018
		9	-0.02...	-0.02...	19.322	0.023
		1...	-0.03...	-0.02...	20.517	0.025
		1...	-0.03...	-0.02...	21.780	0.026
		1...	-0.03...	-0.02...	23.211	0.026
		1...	-0.02...	-0.01...	23.944	0.032
		1...	-0.03...	-0.02...	25.206	0.033
		1...	-0.03...	-0.02...	26.983	0.029
		1...	0.011	0.018	27.139	0.040
		1...	-0.03...	-0.02...	28.322	0.041
		1...	0.005	0.010	28.358	0.057
		1...	-0.00...	-0.00...	28.444	0.075
		2...	0.002	-0.00...	28.448	0.099
		2...	-0.01...	-0.01...	28.624	0.123
		2...	0.001	-0.00...	28.625	0.156
		2...	0.033	0.030	30.069	0.147
		2...	0.043	0.041	32.497	0.115
		2...	0.181	0.177	76.137	0.000
		2...	0.033	0.017	77.586	0.000
		2...	0.074	0.047	84.811	0.000
		2...	0.025	-0.01...	85.634	0.000
		2...	0.025	0.000	86.494	0.000
		3...	-0.00...	-0.01...	86.567	0.000
		3...	-0.00...	-0.00...	86.605	0.000
		3...	-0.02...	-0.02...	87.391	0.000
		3...	-0.02...	-0.01...	88.152	0.000
		3...	-0.01...	-0.00...	88.410	0.000
		3...	-0.03...	-0.01...	90.000	0.000
		3...	-0.04...	-0.02...	92.158	0.000

Figure 7.6: The correlogram of squared residuals of the model ARFIMA (5,0.0599,5)

Table 7.6: The result of ARCH LM test of ARFIMA (5,0.0599,5)

Heteroskedasticity Test: ARCH

F-statistic	4.346847	Prob. F(3,1293)	0.0047
Obs*R-squared	12.95027	Prob. Chi-Square(3)	0.0047

Based on the results shown in Table 7.6, it can be concluded that the ARCH effect is significant (p-value < 0.05).

However, despite the ARCH effect, the model was tested for the same independent data set (weekly rainfall series in 2015). The observed and the predicted values for the independent data set is shown in Figure 7.7.

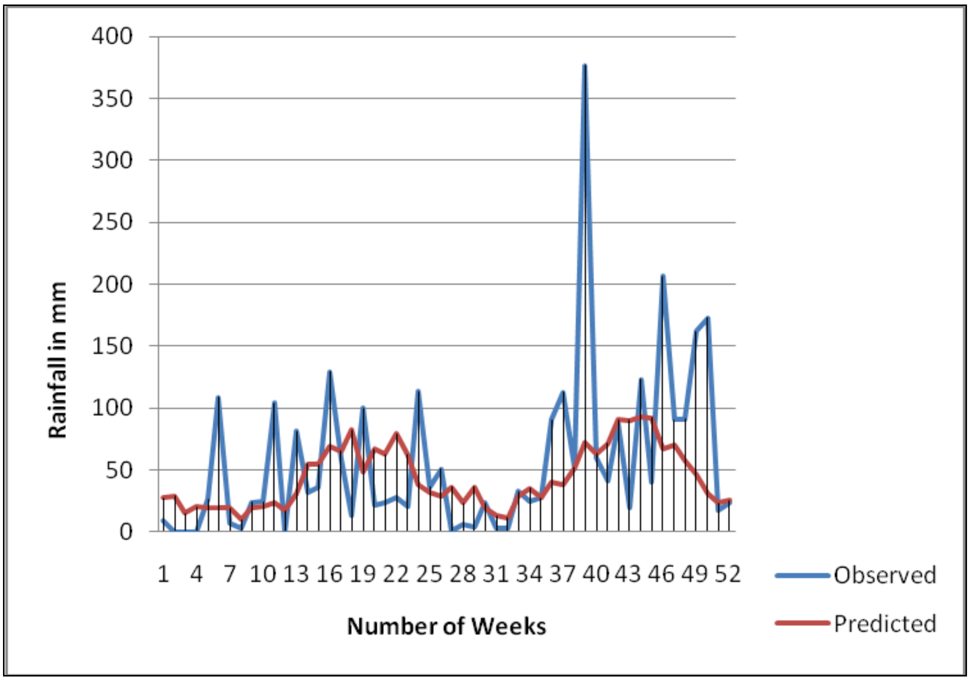

Figure 7.7: Observed and predicted weekly rainfall in 2015 using the ARFIMA (5,0.0599,5) for deseasonalized data

The Figure 7.7 depicts the predicted values are in good agreement with the observed rainfall values in 2015 than the previous model predictions. Consequently, the most of the predicted made based on the model is closer with the observed values. However, still there is a noticeable gap in capturing the extreme values with the model. To evaluate the degree of the forecasting performance of the model, the absolute error distribution was taken and the result is presented by the Table 7.7.

Table 7.7: The analysis of absolute error (in mm) for the weekly rainfall in 2015 [ARFIMA (5,0.0599,5)]

Absolute Error in mm	**Number of weeks (%)**	**Cumulative (%)**
00--10	17 (32.7)	17 (32.7)
11--15	2 (3.8)	19 (36.5)
16--20	6 (11.6)	25 (48.1)
21--25	2 (3.8)	27 (51.9)
26--30	2 (3.8)	29 (55.7)
31--35	4 (7.8)	33 (63.5)
36--40	1 (1.9)	34 (65.4)
41--45	2 (3.8)	36 (69.2)
46--50	3 (5.8)	39 (75.0)
More than 50	13 (25.0)	52 (100.0)

Based on the result of the Table 7.7, the weeks which with less than 10mm error is 32.7%. This is a considerable increment of the number of weeks (by 13.5%) compared with the previous model predicted weeks in the same category. However, the number of weeks which give more than 50 mm absolute error have been slightly increased (by 1.9%).

7.6. ARFIMA Long Range Dependency Model with Heteroskedasticity

The long-range dependency models so far discussed showed heteroskedasticity in the innovation. Thus, GARCH model is employed to capture the stochastic volatility of mean ARFIMA models. Some properties of the ARFIMA-GARCH model is discussed by the next section.

The ARFIMA (p, d, q)-GARCH (r, s) model of a discrete time series process $\{Y_t\}_{t\in Z}$ is defined by the following formula,

$$\varphi(B)\ \nabla^d\ (Y_t - \mu) = \theta(B)\ \varepsilon_t \tag{7.17}$$

$$\varepsilon_t / F_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i} \tag{7.18}$$

Where $\alpha_0 > 0,\ \alpha_1, \alpha_2, \alpha_3 .... \alpha_r \geq 0, \beta_1, \beta_2, \beta_3, ... \beta_s \geq 0$ r and s are positive integers, d is a real number and B is the backward-shift operator. The term F_{t-1} is the set of which derived by the σ field past information $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3} ...\}$. $\varphi(B)$ and $\theta(B)$ are autoregressive and moving average polynomials of order p and q respectively (Refer 7.2 and 7.3)

Where, d < 1/2 the $\{Y_t\}_{t\in Z}$ is the second order stationary and it can be written as the following (Ling and Li,1997).

$$Y_t = \varphi^{-1}(B)\, \theta(B) \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!\,(d-1)!}\ \varepsilon_{t-k} \tag{7.19}$$

If d > -1/2 the $\{Y_t\}_{t\in Z}$ is invertible and ε_t can be written as follows

$$\varepsilon_t = \varphi(B)\theta^{-1}(B) \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!\,(d-1)!}\ Y_{t-k} \tag{7.20}$$

The maximum likelihood estimates for the parameters λ of the model ARFIMA-GARCH is obtained by maximizing the conditional log likelihood l_t

$$L(\lambda) = \frac{1}{n} \sum_{t=1}^{n} l_t \qquad l_t = -\frac{1}{2} \ln\ h_t - \frac{\varepsilon_t^2}{2h_t} \tag{7.21}$$

Where, $\lambda = (\gamma^T, \delta^T)^T$ and $\gamma = (\varphi_1, \varphi_2, ...\varphi_p, \theta_1, \theta_2, ...\theta_p, d)^T$ and

$\delta = (\alpha_0, \alpha_1, \alpha_2, ...\alpha_r, \beta_1, \beta_2, ...\beta_s)^T$

The maximum likelihood estimation (MLE) method was employed to obtain estimates for model parameters. To evaluate the suitability of the method for parameter estimation, a Monte Carlo simulation was done with various fractional differencing values.

7.7. The Results of Monte Carlo Simulation - ARFIMA-GARCH

The simulation results provided fractionally differenced parameter estimations, variance model parameters estimations along with the corresponding standard error (SE) and mean square errors (MSE) of the parameters. It was carried out based on 1000 replications with different sizes of samples (n=100, n=200, n=500 and n=1000). The Monte Carlo experiment was conducted on a simulated ARFIMA (0, d,0)-GARCH (1,1) series with following parameter combinations.

$\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6 \qquad h_t = 0.15 + 0.2\varepsilon_{t-1}^2 + 0.6h_{t-1}$

d= 0.1,0.15,0.3 and 0.45

The simulation was carried out with R programming language (Version 3.4.2) using HP11 (8GB, 64bit) computer. The package "rugarch" in R optimized the log likelihood function and obtained the exact maximum likelihood estimators.

Table 7.8-7.11 present the average of the estimated d, α_0, α_1 and β_1 which were computed based on 1000 replications. Furthermore, the Tables report the standard error of the estimates SE ($\hat{d}$) (Refer 7.12), SE ($\hat{\alpha}_0$), SE ($\hat{\alpha}_1$)and SE($\hat{\beta}_1$) along with the mean square error of the estimates MSE($\hat{d}$) (Refer 7.13), MSE($\hat{\alpha}_0$), MSE($\hat{\alpha}_1$) and MSE($\hat{\beta}_1$) respectively such that:

$$SE(\hat{\alpha}_0) = \sqrt{\sum_{r=1}^{R} (\hat{\alpha}_{0r} - \hat{\alpha}_0) \Big/ R} \qquad (7.22)$$

$$SE(\hat{\alpha}_1) = \sqrt{\sum_{r=1}^{R}(\hat{\alpha}_{1r} - \hat{\alpha}_1) \Big/ R} \qquad (7.23)$$

$$SE(\hat{\beta}_1) = \sqrt{\sum_{r=1}^{R}(\hat{\beta}_{1r} - \hat{\beta}_1) \Big/ R} \qquad (7.24)$$

Where $\hat{\alpha}_{0r}, \hat{\alpha}_{1r}$ and $\hat{\beta}_{1r}$ are the MLE of α_0, α_1 and β_1 for the r^{th} replication. The value R denotes the number of replications. The relevant MSE can be expressed as follows,

$$MSE(\hat{\alpha}_0) = \sum_{r=1}^{R}(\hat{\alpha}_{r0} - \alpha_0)^2 \Big/ R \qquad (7.25)$$

$$MSE(\hat{\alpha}_1) = \sum_{r=1}^{R}(\hat{\alpha}_{r1} - \alpha_1)^2 \Big/ R \qquad (7.26)$$

$$MSE(\hat{\beta}_1) = \sum_{r=1}^{R}(\hat{\beta}_{r1} - \beta_1)^2 \Big/ R \qquad (7.27)$$

Table 7.8: The MLE of d, α_0, α_1 and β_1 of a generating process of ARFIMA (0,d,0) - GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.1.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.07505	0.08082	0.08952	0.09515
SE($\hat{d}$)	0.07783	0.06094	0.04054	0.02700
MSE($\hat{d}$)	0.00668	0.00408	0.00175	0.00075
$\hat{\alpha}_0$	0.07094	0.10044	0.13805	0.15129
SE($\hat{\alpha}_0$)	0.11041	0.11266	0.08602	0.05281
MSE($\hat{\alpha}_0$)	0.01844	0.01515	0.00754	0.00279
$\hat{\alpha}_1$	0.11152	0.14204	0.17934	0.19368
SE($\hat{\alpha}_1$)	0.15044	0.13060	0.08878	0.05363
MSE($\hat{\alpha}_1$)	0.03046	0.02041	0.00831	0.00292
$\hat{\beta}_1$	0.80155	0.72727	0.63664	0.60262
SE($\hat{\beta}_1$)	0.25925	0.25851	0.18672	0.11003
MSE($\hat{\beta}_1$)	0.10783	0.08303	0.03621	0.01211

Table 7.9: The MLE of d, α_0, α_1 and β_1 of a generating process of ARFIMA (0,d,0) - GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.15.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.11464	0.12679	0.13967	0.14532
SE($\hat{d}$)	0.08943	0.06719	0.04076	0.02699
MSE($\hat{d}$)	0.00925	0.00505	0.00177	0.00075
$\hat{\alpha}_0$	0.07373	0.10263	0.13947	0.15182
SE($\hat{\alpha}_0$)	0.11536	0.11276	0.08426	0.05165
MSE($\hat{\alpha}_0$)	0.01913	0.01496	0.00721	0.00267
$\hat{\alpha}_1$	0.11405	0.14460	0.18109	0.19436
SE($\hat{\alpha}_1$)	0.15295	0.13012	0.08719	0.05214
MSE($\hat{\alpha}_1$)	0.03078	0.02000	0.00796	0.00275
$\hat{\beta}_1$	0.79587	0.72149	0.63318	0.60122
SE($\hat{\beta}_1$)	0.26576	0.25888	0.18252	0.10656
MSE($\hat{\beta}_1$)	0.10899	0.08178	0.11619	0.12755

Table 7.10: The MLE of d, α_0, α_1 and β_1 of a generating process of ARFIMA (0,d,0) - GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.3.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.26080	0.27842	0.29125	0.29612
SE($\hat{d}$)	0.10273	0.06993	0.04059	0.02709
MSE($\hat{d}$)	0.01209	0.00536	0.00172	0.00075
$\hat{\alpha}_0$	0.07279	0.10395	0.14538	0.15580
SE($\hat{\alpha}_0$)	0.11401	0.11185	0.08353	0.04915
MSE($\hat{\alpha}_0$)	0.01896	0.01463	0.00700	0.00245
$\hat{\alpha}_1$	0.11525	0.14725	0.18548	0.19766
SE($\hat{\alpha}_1$)	0.15669	0.12942	0.08246	0.04560
MSE($\hat{\alpha}_1$)	0.03173	0.01953	0.00701	0.00209
$\hat{\beta}_1$	0.79651	0.71849	0.62025	0.59233
SE($\hat{\beta}_1$)	0.26847	0.25394	0.17527	0.09419
MSE($\hat{\beta}_1$)	0.11069	0.07853	0.03113	0.00893

Table 7.11: The MLE of d, α_0, α_1 and β_1 of a generating process of ARFIMA (0,d,0) - GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.45.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.40720	0.42797	0.44289	0.44770
SE($\hat{d}$**)**	0.08894	0.05994	0.03755	0.02646
MSE($\hat{d}$**)**	0.00974	0.00408	0.00146	0.00071
$\hat{\alpha}_0$	0.07601	0.12041	0.14941	0.15630
SE($\hat{\alpha}_0$**)**	0.11293	0.11916	0.08161	0.04862
MSE($\hat{\alpha}_0$**)**	0.01823	0.01507	0.00666	0.00240
$\hat{\alpha}_1$	0.12065	0.15949	0.18873	0.19784
SE($\hat{\alpha}_1$**)**	0.15684	0.12548	0.07863	0.04476
MSE($\hat{\alpha}_1$**)**	0.03090	0.01739	0.00631	0.00201
$\hat{\beta}_1$	0.78589	0.68169	0.61120	0.59154
SE($\hat{\beta}_1$**)**	0.26907	0.25938	0.16700	0.09219
MSE($\hat{\beta}_1$**)**	0.10696	0.07395	0.02835	0.00857

Tables 7.8-7.11 provide evidence to the parameter bias has decreased as with the increase of the series length irrespective of long memory parameter d. It is also noted that the parameters become consistent with the increase in series length. Standard error and the MSE of estimators decrease with the increase in series length as expected. Thus, we can conclude that a sensible estimation of the maximum likelihood estimator for the fractional differencing parameters and variance model parameters. It is noted that the parameters estimates for the d, α_0 ,and α_1 get much low value than real at small sample size while β_1 get much high values than real parameter value at small size of sample. This feature is highlighted in all above combinations.

7.8. Modeling Weekly Rainfall Using ARFIMA-GARCH Model

Many ARFIMA-GARCH models were fitted to the weekly rainfall series data with the size of the sample being 1300. Those fitted were employed to predict the weekly rainfall over the year 2015. The best fitted model is selected with minimum mean absolute error (MAE).

A model ARFIMA (4,0.116577,6)-GARCH (1,1) was found to be the best fitted model for the weekly rainfall series. The corresponding parameter estimates with standard errors are presented in Table 7.12.

Table 7.12: The parameter estimates of the model ARFIMA (4,0.116577,6)-GARCH (1,1)

Parameters	**Estimates**	**Standard Error**	**pvalue**
φ_1	2.986950	0.00062	0.00000
φ_2	-4.00000	0.000755	0.00000
φ_3	2.927562	0.000603	0.00000
φ_4	-0.968635	0.000313	0.00000
θ_1	-2.952466	0.00007	0.00000
θ_2	3.855760	0.00006	0.00000
θ_3	-2.678475	0.00006	0.00000
θ_4	0.729736	0.00003	0.00000
θ_5	0.138485	0.00013	0.00000
θ_6	-0.042142	0.000171	0.00000
Constant	44.6898	0.89140	0.00000
d	0.116577	0.027478	0.01482
α_0	829.99100	40.62404	0.00000
α_1	0.268774	0.046747	0.00000
β	0.525041	0.031188	0.00000

All the model parameters of the mean and variance model are significant at 0.05 level of significance. The mean and variance model equations can be expressed by (7.28) and (7.29) respectively.

$$\left(1-2.987\ B+4.000\ B^2-2.928\ B^3+0.969\ B^4\right)(1-B)^{0.1166}(Y_t-44.690)=$$
$$\left(1-2.953\ B+3.856\ B^2-2.678\ B^3+0.730\ B^4+0.138\ B^5-0.042\ B^6\right)\varepsilon_t \quad (7.28)$$

Where Y_t is the weekly rainfall series and

$$\varepsilon_t/F_{t-1} \sim N(0, h_t)$$
$$h_t = 829.991 + 0.269\ \varepsilon_{t-1}^2 + 0.525\ h_{t-1} \quad (7.29)$$

7.8.1. Residual Analysis for the Model ARFIMA (4,0.116577,6)-GARCH (1,1)

The residuals analysis was carried out and the corresponding test result of the residual and squared residual are presented in Table 7.13 and Table 7.14 respectively.

Table 7.13: The result of weighted Ljung-Box test on standardized residuals of the model ARFIMA (4,0.116577,6)-GARCH (1,1)

Lag order	Statistics	p-value
Lag [1]	0.01070	0.9176
Lag [29]	13.3572	0.9981
Lag [49]	22.5723	0.7354

Table 7.14: The result of weighted Ljung-Box test on standardized squared residuals of the model ARFIMA (4, 0.116577,6)-GARCH (1,1)

Lag order	Statistics	p-value
Lag [1]	0.8163	0.3663
Lag [29]	1.326	0.7824
Lag [49]	2.1558	0.8856

Based on the result of the Table 7.13 and Table 7.14, the residuals as well as squared residuals derived from the model are not significantly deviated from random.

However, the ARCH LM test is applied to test the heteroskedasticity of the residuals and the results is presented by the Table 7.15.

Table 7.15: The result of weighted ARCH LM test of the model ARFIMA (4,0.116577,6)-GARCH (1,1)

Lag order	**Statistics**	**p-value**
ARCH Lag [3]	0.003621	0.9520
ARCH Lag [5]	0.823101	0.7861
ARCH Lag [7]	1.198062	0.8794

Based on the results indicated by the Table 7.15, it can be concluded that the there is no ARCH effect moreover (pvalue > 0.05, the null hypothesis that the there is no ARCH effect is not rejected). Thus, the model was tested for weekly rainfall data in 2015 and the observed and predicted values are illustrated by the Figure 7.8.

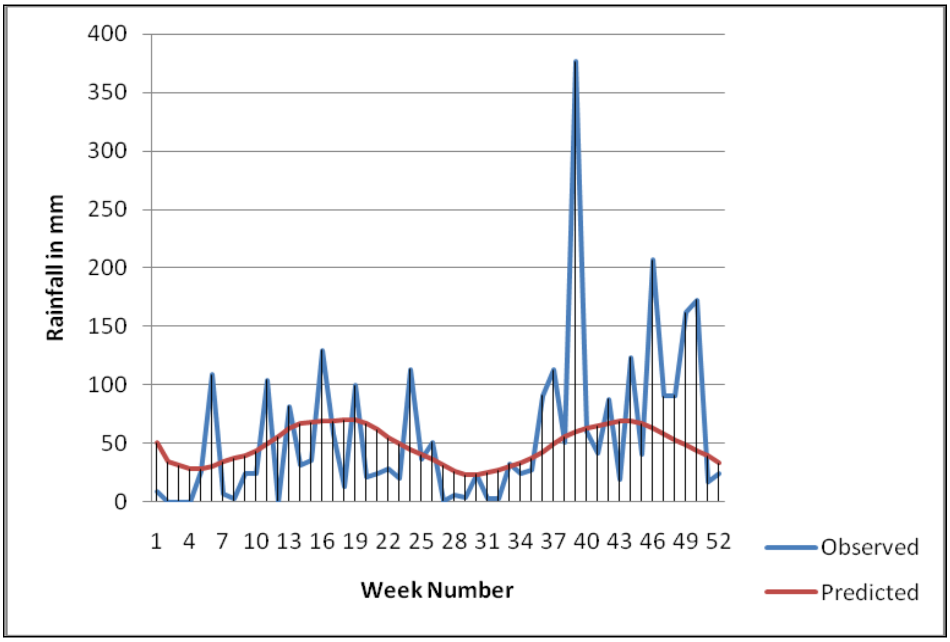

Figure 7.8: Observed and predicted weekly rainfall in 2015 using the model $\text{ARFIMA}\left(4,\, 0.116577, 6\right) - \text{GARCH}(1,1)$

The Figure 7.8 depicts that there is no much improvement in forecasting result compared with the ARFIMA model (Section 7.2). However, those predicted values made based on the model with good accuracy by accounting the heteroskedasticity. To assess the power of the forecasting the absolute error was calculated and result is presented by Table 7.16.

Table 7.16: The absolute error in mm for the weekly rainfall in 2015
$\text{ARFIMA}(4,\ 0.116577,\ 6) - \text{GARCH}(1,1)$

Absolute Forecasting Error in mm	**Number of weeks (%)**	**Cumulative (%)**
0-10	10 (19.2)	10 (19.2)
11-15	03 (5.8)	13 (25.0)
16-20	06 (11.5)	19 (36.5)
21-25	06 (11.5)	25 (48.0)
26-30	07 (13.6)	32 (61.6)
31-35	03 (5.8)	35 (67.4)
36-40	01 (1.9)	36 (69.3)
41-45	02 (3.8)	38 (73.1)
46-50	02 (3.8)	40 (76.9)
More than 50	12 (23.1)	52 (100.0)

The number of weeks which with less than 10 mm error and more than 50 mm error are 19.2% and 23.1% respectively. This seems to be much similar result which made based on the model of ARFIMA. Moreover, the percentage points with absoulte error less than 15 mm is 25.0% and this considerable lower than the percentage points with less than 15 mm error which predicted using only mean model ARFIMA (30.7%).

7.9. ARFIMA Long Memory Model for Deseasonalized data with Heteroskedasticity

To improve the power of forecast performance and model accuracy simultaneously, the model ARFIMA-GARCH was utilized for the deseasonalized weekly rainfall series. The best fitted model is identified as ARFIMA (6,0.243588,5)-GARCH (1,1) and the estimated parameter is presented by Table 7.17.

Table 7.17: The parameter estimates of the model ARFIMA (6,0.243588,5) - GARCH (1,1) for deseasonalized series

Parameters	**Estimates**	**Standard Error**	**pvalue**
φ_1	0.760948	0.040762	0.00000
φ_2	0.354948	0.000021	0.00000
φ_3	-0.019284	0.000003	0.00000
φ_4	0.367182	0.000022	0.00000
φ_5	-0.425125	0.000025	0.00000
φ_6	-0.038733	0.000004	0.00000
θ_1	-0.926677	0.000038	0.00000
θ_2	-0.324362	0.000022	0.00000
θ_3	0.074596	0.000013	0.00000
θ_4	-0.276101	0.000020	0.00000
θ_5	0.451520	0.000026	0.00000
Constant	3.625243	0.040762	0.00000
d	0.243588	0.027423	0.00000
α_0	52.99254	1.834374	0.00000
α_1	0.004739	0.000278	0.00000
β	0.977096	0.001344	0.00000

All the parameters, including both mean and variance model parameters are significant at 0.05 level of significance. The mean and variance model equations can be expressed by (7.30) and (7.31) respectively.

$$\left(1-0.761\,B+0.355\,B^2-0.019\,B^3+0.367\,B^4-0.425\,B^5-0.039\,B^6\right)(1-B)^{0.2436}\left(Y_t-3.625\right)=\left(1-0.927\,B-0.324\,B^2+0.075\,B^3-0.276\,B^4+0.452\,B^5\right)\varepsilon_t \tag{7.30}$$

Where Yt is the deseasonalized weekly rainfall series and

$\varepsilon_t/F_{t-1} \sim N(0, h_t)$

$$h_t = 52.9925 + 0.0047\,\varepsilon_{t-1}^2 + 0.9771\,h_{t-1} \tag{7.31}$$

7.9.1. Residual Analysis for the Model ARFIMA (6,0.243588,5) -GARCH (1,1) for Deseasonalized Series

The residual analysis was carried out and the residuals and squared residual are not deviated from the random at 0.05 level of significance. The corresponding results of the residual analysis are presented by Table 7.18 and 7.19.

Table 7.18: The result of weighted Ljung-Box test on standardized residuals of the model ARFIMA (6,0.243588,5)-GARCH (1,1)

Lag order	Statistics	p-value
Lag [1]	0.02113	0.8844
Lag [32]	8.20596	0.9999
Lag [54]	17.2109	0.9988

Table 7.19: The result of weighted Ljung-Box test on standardized squared residuals of the model ARFIMA (6,0.243588,5)-GARCH (1,1)

Lag order	Statistics	p-value
Lag [1]	0.5809	0.44595
Lag [19]	7.0749	0.57489
Lag [59]	9.4116	0.63265

According to the result of the Table 7.18 and Table 7.19, the residuals and squared residuals are not significantly deviated from random. However, the ARCH LM test is employed to test the heteroskedasticity of the residuals and the corresponding results is presented by the Table 7.20.

Table 7.20: The result of weighted ARCH LM test of the model ARFIMA (6,0.243588,5)-GARCH (1,1)

Lag order	Statistics	p-value
ARCH Lag [3]	0.006004	0.9382
ARCH Lag [5]	1.158850	0.6864
ARCH Lag [7]	1.670715	0.7865

Based on the results indicated by the Table 7.20, it can be concluded that the there is no ARCH effect moreover in this model also (pvalue > 0.05). Thus, the model was tested for weekly rainfall data in 2015 and the observed and predicted values are presented by the Figure 7.9.

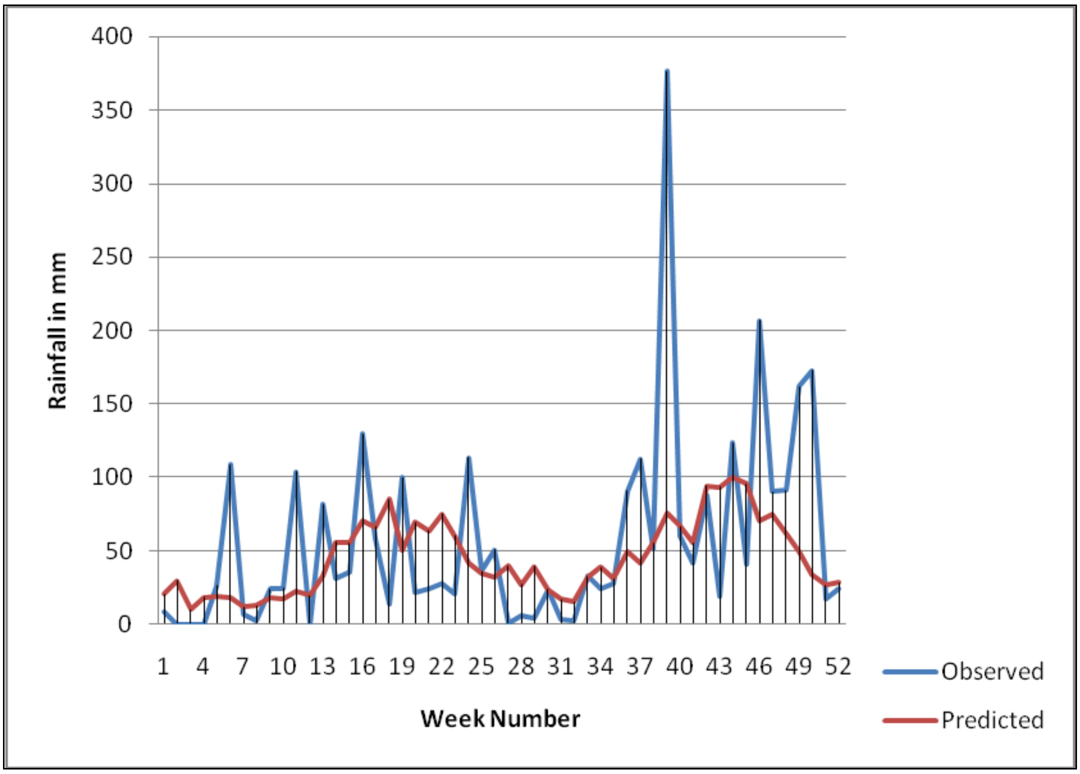

Figure 7.9: Observed and predicted weekly rainfall in 2015 using the model ARFIMA (6,0.243588,5)-GARCH (1,1) for deseasonalized data

According to the forecast result, which depicts from the Figure 7.9, there is a considerable good agreement with the observed and predicted. However, there is a still noticeable gap in predicting extreme rainfall events. Thus, to evaluate the degree of the forecasting performance the absolute errors were calculated and the corresponding result is presented in Table 7.21.

Table 7.21: The analysis of absolute error in (mm) for the weekly rainfall in 2015 ARFIMA (6,0.243588,5)-GARCH (1,1) for deseasonalized data

Absolute Forecasting Error in mm	Number of weeks (%)	Cumulative (%)
0-10	16 (30.8)	16 (30.8)
11-15	6 (11.6)	19 (42.4)
16-20	5 (9.6)	24 (52.0)
21-25	2 (3.8)	28 (55.8)
26-30	2 (3.8)	28 (59.6)
31-35	1 (1.9)	30 (61.5)
36-40	3 (5.8)	32 (67.3)
41-45	1 (1.9)	37 (69.2)
46-50	4 (7.7)	39 (76.9)
More than 50	12 (23.1)	52 (100.0)

The number of weeks which with less than 10 mm error is slightly decreased than the weeks in the same category predicted based on the ARFIMA for deseasonalized data. The percentage points less than 15 mm absolute error was 42.4%. This is a good confirmatory agreement with the observed. It is important to note that those predictions were made based on the model with good accuracy by accounting the heteroskedasticity and most of the weeks predictions are get closer values for the observed.

7.10. Adjusted SARFIMA -GARCH Long Range Dependency Model

A series that present long memory features and periodic behavior with the conditional variance, SARFIMA model with GARCH type innovation is much suitable for the modeling such a kind of a stochastic process. The adjusted SARFIMA-GARCH model was utilized to capture the real dynamic of weekly rainfall series are discussed in this section. Initially, Seasonal Autoregressive Fractional Moving Average Model (SARFIMA) was applied to capture the long memory features along with the seasonal behavior of the process and fitted GARCH model for the residual which derived from the SARFIMA.

The SARFIMA model is a natural extension of the ARFIMA process with an additional seasonal filter by Porter-Hudak (1990). The model consists of long memory dependency features with periodic behavior in terms of the data.

A SARFIMA (p,d,q)×(P,D,Q)$_s$ model of a process $\{Y_t\}_{t\in Z}$ is given by the formula (7.32) ,

$$\varphi(B)\ \psi\left(B^{S}\right) \nabla^{d}\ \nabla_{S}^{D}\left(Y_t-\mu\right)=\ \theta(B)\ \Theta\left(B^{S}\right)\varepsilon_t \qquad (7.32)$$

Where μ is the mean of the process, $\{\varepsilon_t\}_{t\in Z}$ is a white noise process with zero mean and constant variance σ_{ε}^{2}. B is the backward shift operator such that $y_{t-n}=B^{n}\ y_t$ and S is the seasonal length. $\varphi(B)$ and $\psi(B)$ are the non seasonal and seasonal autoregressive polynomials of order p and P respectively such that

$$\psi(B)=\sum_{k=1}^{P}\psi_k\ B^{k} \qquad 1\le k\le P \qquad (7.33)$$

$\theta(B)$ and $\Theta(B)$ are the non-seasonal and seasonal moving average polynomials of order q and Q respectively defined as

$$\Theta(B)=\sum_{m=1}^{Q}\Theta_m\ B^{m} \quad 1\le m\le Q \qquad (7.34)$$

The seasonal operator $\nabla_S^D=(1-B^S)^D$ can be expressed by the binomial series as,

$$(1-B^S)^D=\sum_{k=0}^{\infty}\frac{(k+D-1)\,!}{k!\ (D-1)\,!}B^{Sk} \qquad (7.35)$$

When d=0 and D=0, the model is reduced to a classical SARIMA model. If conditions 0 < d < 0.5 and 0 < D < 0.5 are satisfied, then the process becomes stationary. The spectral density function of the SARFIMA model can be written as follows:

$$f_s(\lambda) = \frac{\sigma_\varepsilon^2 \left|\theta_q\left(e^{-i\lambda}\right)\right|^2 \left|\Theta_Q\left(e^{-i\lambda\lambda}\right)\right|^2}{2\pi \left|\varphi_p\left(e^{-i\lambda}\right)\right|^2 \left|\psi_P\left(e^{-i\lambda\lambda}\right)\right|^2} \left|1-e^{i\lambda}\right|^{-2d} \left|1-e^{i\lambda\lambda}\right|^{-2D} \tag{7.36}$$

The maximum likelihood estimation (MLE) method was employed to obtain estimates for model parameters of adjusted SARFIMA-GARCH. To evaluate the suitability of the method for parameter estimation, a Monte Carlo simulation was done with various combination of seasonal and non-seasonal fractionally differenced parameter values.

7.11. The Result of Monte Carlo Simulation - Adjusted SARFIMA-GARCH

In order to evaluate the performance of the maximum likelihood method in estimating the parameters of the model, a number of Monte Carlo experiments were carried out. The simulation results provided non-seasonally and seasonally differenced parameter estimations and the corresponding standard error (SE) and mean square errors (MSE) of the parameters. As well as the variance model parameters estimations were done by fitting GARCH (1,1) model to the residuals of the SARFIMA model. It was carried out based on 1000 replications with different sizes of samples (n=100, n=200, n=500 and n=1000). Seasonal length was considered as 52 corresponding to weekly rainfall. Monte Carlo experiment was conducted on a simulated SARFIMA (0, d,0) x (0, D,0)$_{52}$-GARCH (1,1) series with following parameter combinations.

d=0.1 and D=0.45   d=0.15 and D=0.45

d=0.3 and D=0.3   d=0.45 and D=0.10.

$\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ $h_t = 0.15 + 0.2\varepsilon_{t-1}^2 + 0.6h_{t-1}$

Here also, the "arfima" package (Veenstra and Mcleod; 2012) in R optimized the log likelihood function and obtained the exact maximum likelihood estimators. The simulation was carried out with R programming language (Version 3.4.2) using

HP11 (8GB, 64bit) computer. Table 7.22-7.25 present the average of the estimated d, D, α_0, α_1 and β_1 which were computed based on 1000 replications. Furthermore, the tables report the standard error of the estimates SE ($\hat{d}$) (Refer the equation 7.12), SE ($\hat{D}$) and mean square error of the estimates MSE ($\hat{d}$) (Refer the equation 7.13) and MSE ($\hat{D}$) respectively such that:

$$SE(\hat{D}) = \sqrt{\sum_{r=1}^{R}(\hat{D}_r - \hat{D}) \Big/ R} \tag{7.37}$$

Where $\hat{D}_r$ is the MLE of D for the r^{th} replication. The value R denotes the number of replications. The relevant MSE can be expressed as

$$MSE(\hat{D}) = \sum_{r=1}^{R}(\hat{D}_r - D)^2 \Big/ R \tag{7.38}$$

Table 7.22: The MLE of D, d, α_0, α_1 and β_1 of a generating process of SARFIMA (0,d,0)×(0,D,0)-GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.1** and **D=0.45.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.01467	0.05484	0.0843	0.09292
SE($\hat{d}$)	0.08838	0.06595	0.04484	0.03621
MSE($\hat{d}$)	0.01509	0.00639	0.00226	0.00136
$\hat{D}$	0.45352	0.45605	0.45416	0.45195
SE($\hat{D}$)	0.01822	0.01225	0.00983	0.00946
MSE($\hat{D}$)	0.00034	0.00019	0.00011	0.00009
$\hat{\alpha}_0$	0.18195	0.22167	0.1819	0.16081
SE($\hat{\alpha}_0$)	0.18514	0.16380	0.09834	0.05415
MSE($\hat{\alpha}_0$)	0.03530	0.03197	0.01069	0.00305
$\hat{\alpha}_1$	0. 07455	0.12917	0.16441	0.17743
SE($\hat{\alpha}_1$)	0.10350	0.08670	0.06012	0.04205
MSE($\hat{\alpha}_1$)	0.02645	0.01253	0.00489	0.00228
$\hat{\beta}_1$	0.68152	0.56483	0.58804	0.60484
SE($\hat{\beta}_1$)	0.29227	0.26419	0.16632	0.09476
MSE($\hat{\beta}_1$)	0.09207	0.07103	0.0278	0.00900

Table 7.23: The MLE of D, d, α_0 , α_1 and β_1 of a generating process of SARFIMA (0,d,0)×(0,D,0)-GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.15** and **D=0.45.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.06048	0.10293	0.13361	0.14238
SE($\hat{d}$ **)**	0.08897	0.06597	0.04649	0.03320
MSE($\hat{d}$ **)**	0.01593	0.00657	0.00243	0.00116
$\hat{D}$	0.45308	0.45579	0.45375	0.45202
Se($\hat{D}$ **)**	0.01843	0.01234	0.00988	0.00914
MSE($\hat{D}$ **)**	0.00035	0.00019	0.00011	0.00008
$\hat{\alpha}_0$	0.18121	0.22116	0.18212	0.16279
SE($\hat{\alpha}_0$ **)**	0.18684	0.16349	0.09808	0.05617
MSE($\hat{\alpha}_0$ **)**	0.03588	0.03179	0.01065	0.00332
$\hat{\alpha}_1$	0.07384	0.12846	0.16441	0.17651
SE($\hat{\alpha}_1$ **)**	0.10309	0.08673	0.06031	0.04232
MSE($\hat{\alpha}_1$ **)**	0.02654	0.01264	0.0049	0.00234
$\hat{\beta}_1$	0.6848	0.56624	0.5878	0.60321
SE($\hat{\beta}_1$ **)**	0.29133	0.26447	0.16619	0.09872
MSE($\hat{\beta}_1$ **)**	0.09206	0.07109	0.02777	0.00976

Table 7.24: The MLE of D, d, α_0, α_1 and β_1 of a generating process of SARFIMA (0,d,0)×(0,D,0)-GARCH(1,1) with $\alpha_0 = 0.15$, $\alpha_1 = 0.2$, $\beta_1 = 0.6$ and **d=0.3** and **D=0.3.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.22345	0.25789	0.28404	0.29191
SE($\hat{d}$ **)**	0.09362	0.06618	0.04542	0.03169
MSE($\hat{d}$ **)**	0.01463	0.00615	0.00232	0.00107
$\hat{D}$	0.27587	0.29134	0.29554	0.29794
SE($\hat{D}$ **)**	0.07678	0.04156	0.02733	0.01963
MSE($\hat{D}$ **)**	0.00648	0.0018	0.00077	0.00039
$\hat{\alpha}_0$	0.1966	0.20161	0.17253	0.16032
SE($\hat{\alpha}_0$ **)**	0.18468	0.14769	0.08641	0.05143
MSE($\hat{\alpha}_0$ **)**	0.03628	0.02448	0.00798	0.00275
$\hat{\alpha}_1$	0.12848	0.15819	0.18011	0.18738
SE($\hat{\alpha}_1$ **)**	0.12329	0.09166	0.06233	0.04201
MSE($\hat{\alpha}_1$ **)**	0.02032	0.01015	0.00428	0.00192
$\hat{\beta}_1$	0.60489	0.56268	0.58478	0.59574
SE($\hat{\beta}_1$ **)**	0.31005	0.24688	0.15138	0.09276
MSE($\hat{\beta}_1$ **)**	0.09615	0.06234	0.02315	0.00862

Table 7.25: The MLE of D, d, α_0, α_1 and β_1 of a generating process of SARFIMA (0,d,0)×(0,D,0)-GARCH(1,1) with $\alpha_0 = 0.15, \alpha_1 = 0.2, \beta_1 = 0.6$ and **d=0.45** and **D=0.1.**

n	**100**	**200**	**500**	**1000**
$\hat{d}$	0.36677	0.39857	0.426	0.43861
SE($\hat{d}$**)**	0.08139	0.05613	0.03725	0.02884
MSE($\hat{d}$**)**	0.01356	0.0058	0.00196	0.00096
$\hat{D}$	0.02071	0.07194	0.09022	0.09427
SE($\hat{D}$**)**	0.14222	0.06721	0.03566	0.02765
MSE($\hat{D}$**)**	0.02652	0.0053	0.00137	0.0008
$\hat{\alpha}_0$	0.18613	0.18885	0.16614	0.15776
SE($\hat{\alpha}_0$**)**	0.15882	0.13244	0.07836	0.04814
MSE($\hat{\alpha}_0$**)**	0.02653	0.01905	0.0064	0.00238
$\hat{\alpha}_1$	0.17649	0.1903	0.19467	0.19621
SE($\hat{\alpha}_1$**)**	0.13421	0.09716	0.06178	0.042
MSE($\hat{\alpha}_1$**)**	0.01857	0.00953	0.00384	0.00178
$\hat{\beta}_1$	0.56087	0.54565	0.57909	0.59065
SE($\hat{\beta}_1$**)**	0.29032	0.23251	0.1397	0.0877
MSE($\hat{\beta}_1$**)**	0.08582	0.05701	0.01995	0.00778

The parameter bias has decreased as with the increase of size of the sample irrespective of the differencing parameters. Also, it is clearly seen, the parameters become consistent with the increase in series length. Standard error and the MSE of estimators decrease with the increase in series length as anticipated. Thus, Tables 7.22-7.25 give evident for a rational estimation of the maximum likelihood estimator for the non-seasonal and seasonal fractional differencing parameters along with the variance model parameters.

7.12. Modeling Weekly Rainfall Using Adjusted SARFIMA-GARCH Model

Several SARFIMA models were fitted to the weekly rainfall series data with the size of the sample being 1300. A model SARFIMA $(1, 0.116, 1)\times(1, 0.171, 0)_{52}$ was found to be the best fitted model for the weekly rainfall series. Since the heteroskedasticity existed of the residual derived from the best fitted SARFIMA mode1, GARCH model is employed to residual from the SARFIMA. The corresponding mean and variance model parameter estimates are presented by Table 7.26.

Table 7.26: The parameter estimates of the model SARFIMA (1,0.115677,1) $\times(1,0.170750,0)_{52}$ with GARCH(1,1)

Parameters	**Estimates**	**Standard Error**	**Pvalue**
φ_1	-0.911360	0.14272	0.00000
θ_1	-0.901880	0.14983	0.00000
ψ_1	-0.086060	0.03948	0.00000
Constant	0.004100	0.10213	0.00000
d	0.115677	0.02696	0.00000
D	0.170750	0.02912	0.00000
α_0	0.225230	0.03139	0.00000
α_1	0.234890	0.045240	0.00000
β	0.568650	0.042180	0.00000

All the model parameters of the mean as well as variance equation are significant at 0.05 level of significance. The model mean and variance equations can be expressed by (7.39) and (7.40) respectively.

$$(1+0.9114\ B)(1+0.0861\ B^{52})(1-B)^{0.11567}(1-B^{52})^{0.17075}(Z_t-0.004100\ )=(1-0.90188\ B)\ \varepsilon_t \tag{7.39}$$

Where Zt is the standard weekly rainfall series and

$$\varepsilon_t/F_{t-1} \sim N(0, h_t)$$

$$h_t = 0.22523 + 0.23489\ \varepsilon_{t-1}^2 + 0.56865h_{t-1} \tag{7.40}$$

The residual analysis was carried out for the mean model and the residuals are random at 0.05 level of significant. However, the squared residual is significantly deviated from the random. Thus, ARCH LM test was applied to test the ARCH effect and the based on the test result [Test statistic = 18.48707 (pvalue = 0.04228)], it can be concluded that the ARCH effect is presented at 0.05 level of significance. Thus, to capture the stochastic volatility a variance model [GARCH (1,1)] was utilized for the residual derived from the model SARFIMA. Since the SARFIMA $(1,\ 0.116,\ 1)\times(1,\ 0.171,\ 0)_{52}$ is selected as best fitted mean model to describe the weekly rainfall behavior, the model was utilized to predict the weekly rainfall over the year 2015. The Figure 7.10 illustrates the observed weekly rainfall over the year 2015 along with the predicted estimates.

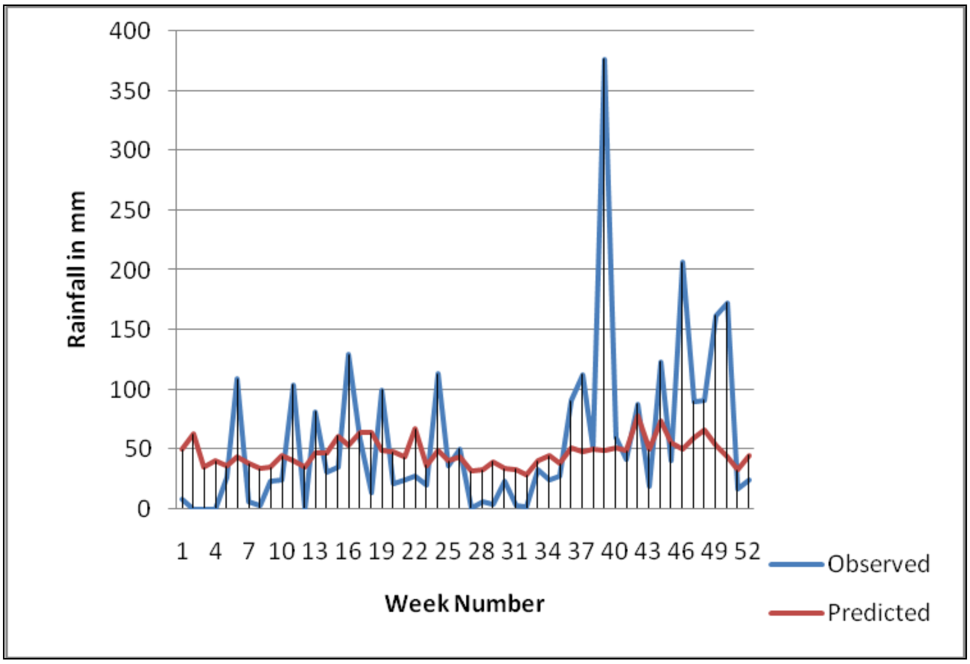

Figure 7.10: Observed and predicted weekly rainfall in 2015 using the SARFIMA $(1,\ 0.116,\ 1)\times(1,\ 0.171,\ 0)_{52}$

The Figure 7.10 depicts the much agreement with observed and predicted values except extreme rainfall events. Same as the previous sections, the absolute errors were calculated to test the forecasting power of the model and the corresponding result is presented in Table 7.27.

Table 7.27: The analysis of absolute error in mm for the weekly rainfall in 2015 SARFIMA $(1,\ 0.116,\ 1)\times(1,\ 0.171,\ 0)_{52}$

Absolute Forecasting Error in mm	**Number of weeks (%)**	**Cumulative (%)**
0-10	12 (23.1)	12 (23.1)
11-15	4 (7.7)	16 (30.8)
16-20	4 (7.7)	20 (38.5)
21-25	5 (9.6)	25 (48.1)
26-30	6 (11.5)	31 (59.6)
31-35	4 (7.7)	35 (67.3)
36-40	3 (5.8)	38 (73.1)
41-45	1 (1.9)	39 (75.0)
46-50	3 (5.8)	42 (80.8)
More than 50	10 (19.2)	52 (100.0)

Based on the above forecasted result, the weeks with less than 10 mm is 23.1%. Since seasonal pattern was accounted when building the model as an additional feature, we expect the considerable high percentage for this category. It is noted that the weeks which with more than 50 mm error is slightly lower than the others. However, the forecasting performance is not much differ from the ARFIMA.

7.13. Comparison of the Five Long Range Dependency Models

To select the best model, the forecasting performance of the models are evaluated based on the predicted result which made from 2015 to 2017. All the models were made using data from 1990 to 2014 and their forecasting performance were assessed by using an independent data set (from 2015 to 2017). The comparison was done by accounting model accuracy with the power of the forecasting of the models.

A simple index is introduced to measure the forecasting performance by assigning weights for the absolute error as follows (Table 7.28).

Table 7.28: The weights assigned for the absolute forecasting error category

Absolute Forecasting Error Category	**Weights**
00-10	9
11-15	7
16-20	5
21-25	3
26-30	1
31-35	-1
36-40	-3
41-45	-5
46-50	-7
More than 50	-9

An index (I) was developed by multiplying the frequency of the corresponding categories with their weights. The model which with highest values of the index was considered as the model having best forecasting performance. The analysis of the absolute errors using data from 2015 to 2017 were carried out and the corresponding results are shown in Appendix -2. The calculated indices are presented in Table 7.29. The observed and predicted values from the five models for an independent data set (from 2015 to 2017) are shown by Figures 7.11-Figure 7.15 respectively and the five long range dependency models are,

Model -1 - ARFIMA (4,0.05792,4)

Model -2 - ARFIMA (5,0.05999,5) for deseasonalized data

Model -3 - ARFIMA (4,0.116577,6)-GARCH (1,1)

Model -4 - ARFIMA (6,0.243588,5)-GARCH (1,1) for deseasonalized

Model -5 - Adjusted SARFIMA (1,0.115677,1) × (1,0.17075,0) -GARCH (1,1)

Table 7.29: The comparison of five long range dependency models

Indicators	Model -1	Model -2	Model -3	Model -4	Model -5
Index (using 2015 data)	40	50	28	68	48
Index (using 2016 data)	12	15	6	20	10
Index (using 2017 data)	26	44	12	52	24
Index (using 2015-2017 data)	78	109	46	140	82
Correlation (using 2015 to 2017) [pvalue]	0.4130 [0.0000]	0.4137 [0.0000]	0.4357 [0.0000]	0.4368 [0.0000]	0.4077 [0.0000]
Bias (using 2015 to 2017)	3.965	4.722	3.260	3.032	4.125
RMSE (using 2015 to 2017)	57.17	57.02	57.36	56.27	58.23
MAE (using 2015 to 2017)	37.71	37.43	38.275	36.998	38.456

Out of the five models, Model-4 gives the highest indices for the years 2015, 2016 and 2017. Consequently, the index is for the time span from 2015 to 2017 was 140 and this the highest out of five model. Moreover, all the other measurement confirm the Model-4 is the best model since the it shows the highest correlation value between the observed and predicted (0.4368). Also, Model-4 gives the least bias, MAE and RMSE. Furthermore, it is important to note that the model ARFIMA(6,0.243588,5)-GARCH(1,1)-[Model-4] is free from ARCH effect indicated that the no heteroskedasticity moreover, of the residuals which derived from the Model-4. Thus, by comparing the both aspect which are the high model accuracy and good forecasting performance, the Model-4 can be selected as the best fitted model in modeling weekly rainfall series.

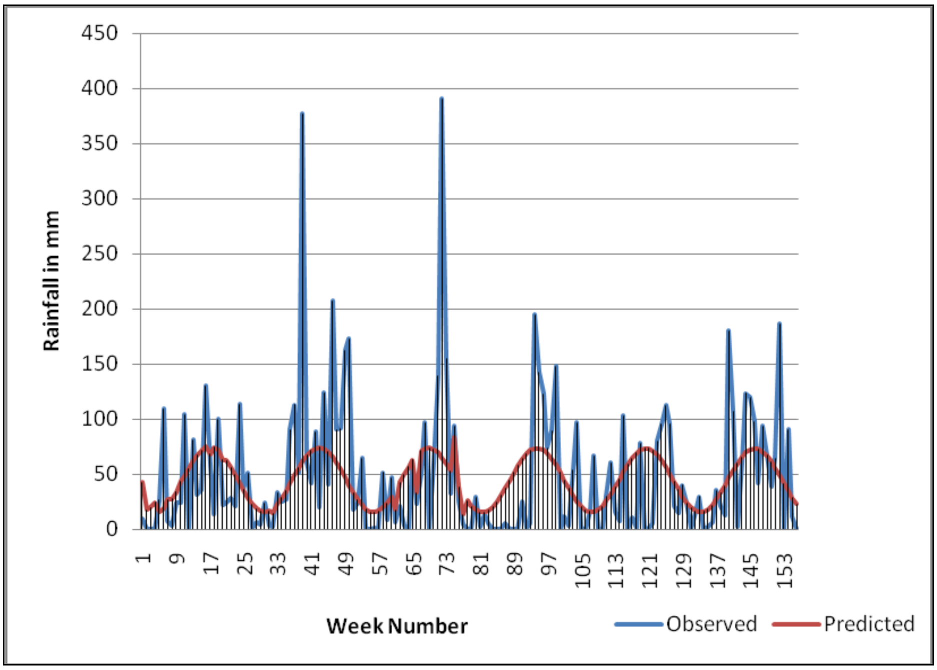

Figure 7.11: Observed and predicted weekly rainfall from 2015 to 2017 using the model ARFIMA (4,0.05792,4)

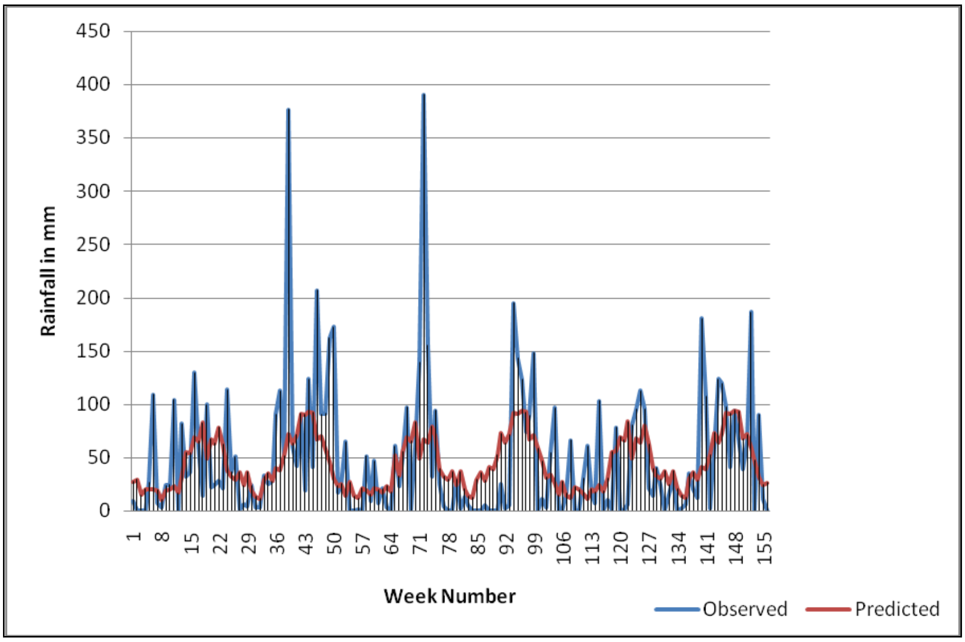

Figure 7.12: Observed and predicted weekly rainfall from 2015 to 2017 using the model ARFIMA (5,0.05999,5) for deseasonalized data

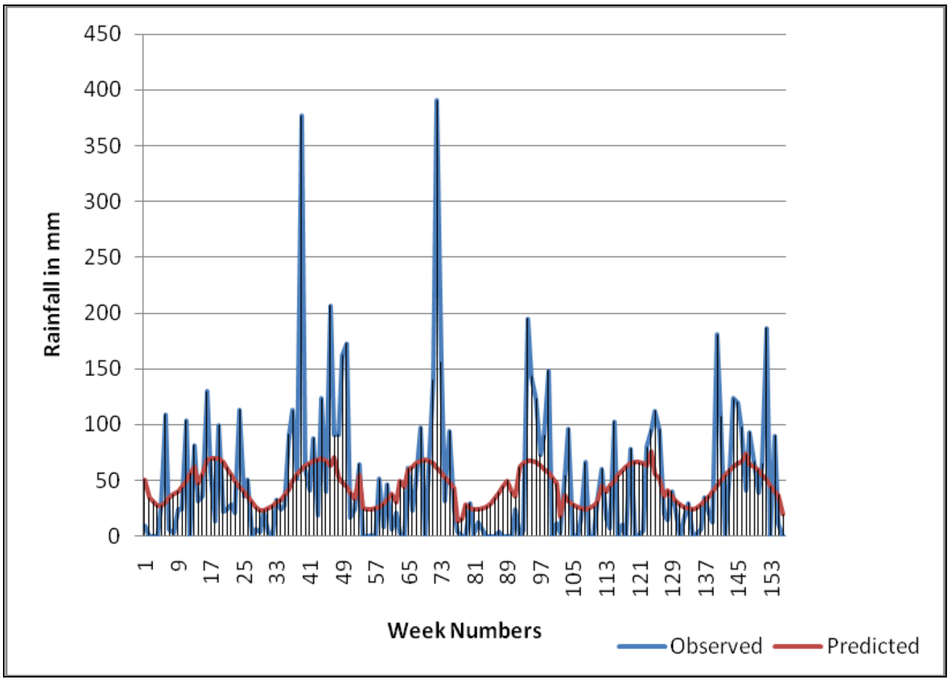

Figure 7.13: Observed and predicted weekly rainfall from 2015 to 2017 using the model ARFIMA (4,0.116577,6)-GARCH (1,1)

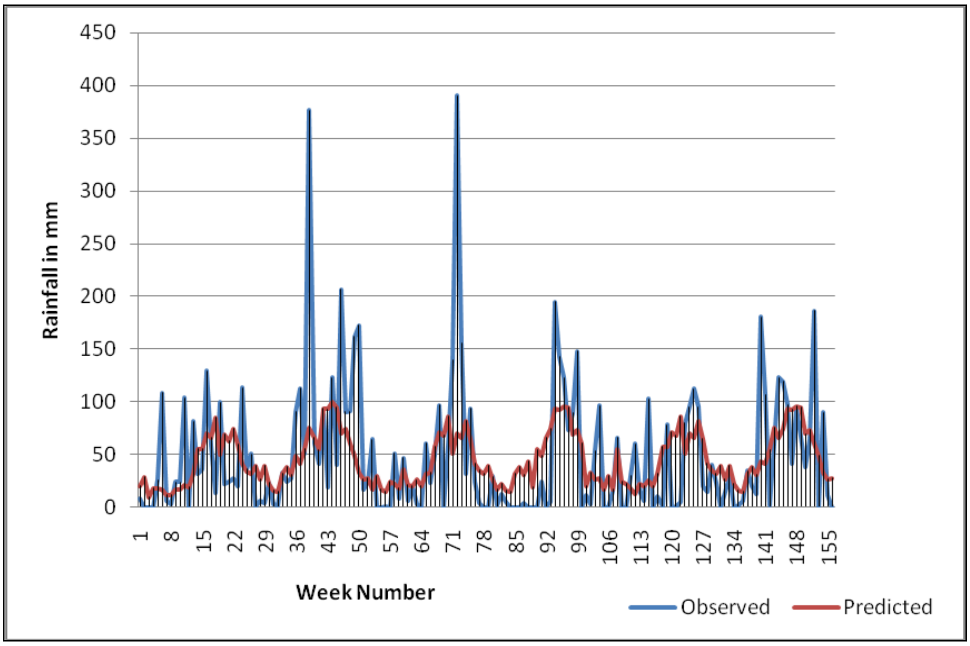

Figure 7.14: Observed and predicted weekly rainfall from 2015 to 2017 using the model ARFIMA (6,0.243588,5)-GARCH (1,1) for deseasonalized data

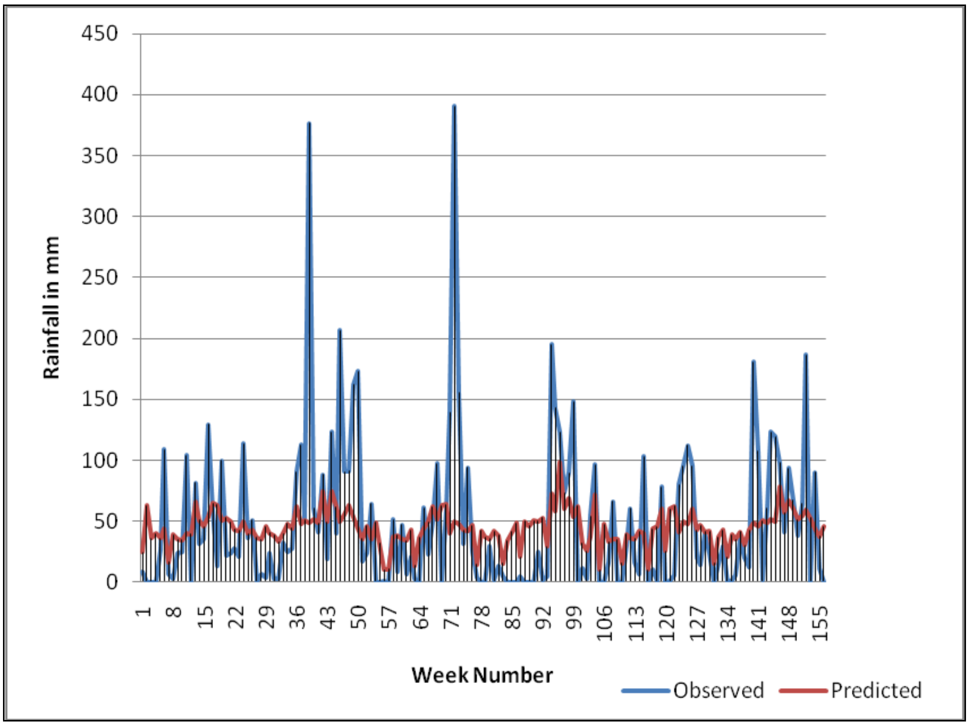

Figure 7.15: Observed and predicted weekly rainfall from 2015 to 2017 using the adjusted SARFIMA (1,0.115677,1) ×(1,0.17075,0) -GARCH(1,1) model

7.14. Summary of the Chapter 7

Long range dependency models are proposed to fit weekly rainfall series since it exhibits an unbounded spectral density at near to zero frequency. Non seasonal and seasonal long range dependency models were utilized to capture the persistence characteristics of the weekly rainfall series. All the models were tested using an independent data series while forecasting power was assessed through the absolute forecasting error.

The exact maximum likelihood estimation (MLE) method was utilized to estimate model parameters. It is important to note that this method was not tested previously for the model parameter estimation in context of long memory for the rainfall studies. However, to evaluate the suitability of the method for parameter estimation, Monte Carlo simulations were carried out with various non seasonally and seasonally fractionally differenced parameter values along with the variance model parameters. It is noted that the parameter bias has decreased as with the increase of the series length irrespective of the fractional differencing parameters in all the Monte Carlo Simulations. Also, it is evident from the estimated parameters in different models, that they become consistent with the increase in size of the sample. Thus, the result of the Monte Carlo simulations are exhibited the corresponding method is reasonably accurate.

It can be concluded that the ARFIMA- GARCH for deseasonalized data showed the highest forecasting performance in modeling weekly rainfall series in context of the long memory. However, the adjusted SARFIMA-GARCH also showed confirmatory agreement with the observed values. It is noted that the forecasting performance in 2017 is considerably high in all the models than in 2016. Thus, all the long memory models show encourage forecasting performance for the weekly rainfall series at such a high uncertainty level.

CHAPTER 8

CONCLUSIONS, RECOMMENDATIONS AND FUTURE STUDIES

The objectives of this study was to develop a statistical model to forecast weekly rainfall in Colombo city. After a comphensive study on past work carried out by various researchers, a novel model: ARFIMA(6,0.243588,5)-GARCH(1,1) was developed to model deseasonalized weekly rainfall to achive the objective. Based on the inferences derived in this study, the following conclusions, recommendations and suggestions for further investigation are given below.

8.1. Conclusions

- The temporal variability of the weekly rainfall was analyzed during the time span from 1960 to 2015 and found that the no significant linear or quadratic trend pattern in weekly rainfall series.

- Due to an unbounded spectrum peak existed near to zero frequency on periodgram of the weekly rainfall, it is not possible to expand the search space of the model.

- Thus, it was forced to restrict to five models within the class of ARFIMA and SARFIMA long range dependency models.

- The five selected long range dependency models are:
    a) ARFIMA (4,0.05792,4)
    b) ARFIMA (5,0.05999,5) for deseasonalized data
    c) ARFIMA(4,0.116577,6)-GARCH (1,1)
    d) ARFIMA(6,0.243588,5)-GARCH(1,1) for deseasonalized
    e) Adjusted SARFIMA(1,0.115677,1) ×$(1,0.17075,0)_{52}$-GARCH(1,1)

- By validating the forecasting performance and model accuracy of the five models for the training set as well as for an independent set, the best fitted model identified is ARFIMA(6,0.243588,5)-GARCH(1,1) for deseasonalized to forecast weekly rainfall in Colombo city. A separate table was developed for the weekly seasonal components.

- The above model ("the best fitted model") is more superior than other four models with respect to statistical aspects as well non statistical aspects.

- The forecasting performance of the best fitted model is not much diluted with the increase of the forecasting length.

- The long range dependency model parameters were estimated using exact maximum likelihood estimation method which has not been tested for the rainfall studies by the previous authors.

- The analysis of the results of the Monte Carlo simulations, found that the bias of the parameters decreases with the increase of the size of the sample irrespective of the long range dependency model parameters. Also, the parameters were consistent with the increase of the sample size. Moreover, the simulation result is shown that the maximum likelihood estimation method is more reasonably accurate to use to estimate the model parameters.

- Weekly rainfall data are positive skewed with longer tail to the right and those series behaviour were analyzed moreover in context of confidence interval by using two approaches parametric and bootstrapping. It is found that the three parameter Weibull distribution is most suitable for the many weekly rainfall series in SWM and SIM.

- Based on the result of the rainfall percentile analysis, it was identified that the there is a high possibility to form extreme rainfall events during the arrival and withdrawal of the SWM. The corresponding weeks were identified as weeks 18-23 (30th April to 10th June) and 38-39 (17th-30th September).

- There is a much more chance to form extreme rainfall events during the SIM too, the corresponding weeks identified was 41-45 (8th October to 11th November).

- According to the analysis of the result of simulation which used to compute the coverage probability of the confidence intervals, it can be concluded that the most of the coverage probability of 95% confidence intervals of percentiles is less than 0.95 and the 95% accurate coverage probability can be attained at the more than 95% average level.

- Of the various climate variables, average relative humidity and maximum temperature are only the two exogenous variable affect on weekly rainfall significantly, but accuracy of prediction did not significantly improved. Thus, it can be concluded that exogenous climate variables are not beneficial in forecasting weekly rainfall.

8.2. Recommendations

- When an unbounded spectrum density was formed near to zero frequency, the models should be selected among the class of ARFIMA and SARFIMA.

- The best fitted long range dependency model ARFIMA(6,0.243588,5)-GARCH(1,1) for deseasonalized is recommended to be used in short term forecasting weekly rainfall in Colombo city in Sri Lanka.

- When modeling weekly rainfall, the other exogenous climate variables are not required.

- Due to the less coverage probability of 95% confidence intervals for percentiles, on average more than 95% confidence level should be considered to get the accurate coverage probability with real confidence bands.

- Result obtained from the analysis of rainfall percentiles can be used to predict the time periods which can have high possibility to form extreme rainfall events specially two seasons such as SWM and SIM.

- The weekly rainfall variation information derived from the percentile analysis would be useful for policy planners in various fields such as constructions, climate monitoring, rain water harvesting etc.

- The developed novel model to forcast weekly rainfall can be used to make more inferences to highlight the features of the weekly rainfall.

8.3. Future Studies

- The novel model developed has to be improved to capture the high peaks in rainfall which is the challenging task to the statisticians.

- The possibility of Gegenbauer models (Gray et al., 1989) can be investigated if the periodagram of the rainfall series illustrates multiple unbounded spectral peaks away from the zero frequency.

- It would be better a more accurate model can be developed for non seasonal weekly rainfall, which I feel that is a another challenging task for an applied statistician.

- A study can be further extended to make more accurate confidence intervals for rainfall percentiles by accounting the real coverage probabilities which can derived by bootstrapping calibiration for all weeks distributions.

- ♦ A comparison study can be further carried out by changing the time and frequency domain methods of estimation for fractional differencing parameters to get best fitted model at different size of sample.

- ♦ Based on the methodology developed in this study, a user friendly computer software can be developed to estimate the parameters of the SARFIMA-GARCH models by maximizing the likelihood of the mean and variance equations jointly in a single step.

CHAPTER 9

PUBLICATIONS BASED ON THIS STUDY

The list of publications generated by this study are given below.

9.1. Publications

1. Silva, H.P.T.N., & Peiris, T.S.G. (2017). Analysis of Weekly Rainfall using Percentile Bootstrap Approach, *International Journal of Ecology & Development, 32* (3): 97-106.

2. Silva, H.P.T.N., & Peiris, T.S.G. (2017). Statistical Modeling of Weekly Rainfall: A case Study in Colombo City in Sri Lanka, *In Proceedings of the 3rd Moratuwa Engineering Research Conference (MERCon*), Sri Lanka, 10-12 May, 241-246, IEEE.

3. Silva, H.P.T.N., & Peiris, T.S.G. (2017). Modeling of Weekly Rainfall using Confidence Interval Approach: A Case Study, *In Proceedings of the 4th International Conference on Multidisciplinary Approaches*, Sri Lanka, 20-22 September, 28.

4. Silva, H.P.T.N., & Peiris, T.S.G. (2017). Modeling Weekly Rainfall: Problems Encountered, *In Proceedings of the International Statistics Conference.* Sri Lanka, 28-29 December, 91.

5. Silva, H.P.T.N., & Peiris, T.S.G. (2018). Accurate confidence intervals for Weibull percentiles using bootstrap calibration: A case study of weekly rainfall in Sri Lanka, *International Journal of Ecological Economics and Statistics, 39* (3): 67-76.

6. Silva, H.P.T.N., Dissanayake, G.S., & Peiris, T.S.G. (2019). Modeling persistent and periodic weekly rainfall in an environment of an emerging Sri Lankan economy, *In Proceedings of the 12th International Conference of the Thailand Econometric Society (TES 2019)*, 9-11 January. Thailand. Springer Book Series: Structural Changes and Their Econometric Modeling.

7. Silva, H.P.T.N., Dissanayake, G.S., & Peiris, T.S.G. (2019). The use of fractionally autoregressive integrated moving average for the rainfall forecasting, *In Proceedings of the 2ndh International Econometric Conference in Vietnam (ECONVN 2019)*, 14-16 January. Vietnam. Springer Book Series: Beyond Traditional Probabilistic Methods in Economics.

8. Silva, H.P.T.N and Peiris T.S.G. (2020). Development of long memory model to forecast weekly rainfall, *In Proceedings of the international Conference on Environmental and Medical Statistics at the University of Peradeniya, Sri Lanka, (ICEMS 2020)*, 9-10 January.

International Journal of Ecology & Development
Year 2017; Volume 32, Issue No. 3; Int. J. Ecol. Dev.
ISSN 0972-9984 (Print); ISSN 0973-7308 (Online)

Analysis of Weekly Rainfall using Percentile Bootstrap Approach

H. P. T. N. Silva[1] and T. S. G. Peiris[2]

[1]Department of Social Statistics, Faculty of Humanities and Social Science,
University of Sri Jayewardenepura, Sri Lanka
Email: thanuja@sjp.ac.lk

[2]Department of Mathematics, Faculty of Engineering,
University of Moratuwa, Sri Lanka
Email: sarathp@uom.lk

ABSTRACT

Analysis of pattern of weekly rainfall would enhance the management of water resource and capability to deal with water related issues, which enable us to mitigate the impact of climate change. This study evaluates the weekly rainfall variability during the South West Monsoon (SWM) and the Second Inter Monsoon (SIM) in the context of confidence intervals. The confidence intervals of rainfall quantiles were made using percentile bootstrap approach. Daily rainfall data from 1960 to 2015 of Colombo in Sri Lanka were used for this study. The Wald Wolfowiz test was used for the test of independence of weekly data series. It is noted that the 82% of the weeks pertaining to SWM marked more than 100% coefficient of variation result in the high fluctuations in weekly rainfall. Conversely, there is much lower variation in weekly rainfall in SIM than SWM. Based on the 95% confidence intervals for percentiles, the weeks 18-23 and 38-39 which belong to SWM and the weeks 41- 45 which pertaining to SIM showed not only high rainfall, but also high rainfall variation result cause to the high possibility to form extreme rainfall events.

Key Words: Weekly Rainfall, Rainfall Quantiles, Confidence Intervals, Colombo, Bootstrap

Mathematics Subject Classification: 62G15

Journal of Economic Literature (JEL) Classification : Q54

1. INTRODUCTION

The awareness of quantity of rainfall with the variability improves the ability of decision makers to deal with the consequences of rainfall. Moreover, information on changes in temporal variability of weekly rainfall of urban areas is useful for the number of fields such as constructions, tourism, health, electricity, plan urban traffic and sewer system, rainwater harvesting, management of water resources. Also, consciousness of variability of weekly rainfall constructive, particularly, for reducing of flood damages. According to Lo and Koralegedera (2015), the more cities including Colombo in Sri Lanka are in a risk of water related issues due to changes in rainfall patterns, urbanization and installation of complex infrastructure. Furthermore, they reported that the city Colombo more and more vulnerable to several extreme weather events, mainly to heavy rainfall events in the future. Some of the researchers made attempts to model the extreme rainfall events at the high uncertainty of atmospheric behavior using different techniques (Hao et al., 2011; Mayooran and Laheetharan., 2014; Wirnancy et al., 2017). However, It is more vital to consider rainfall modeling at short range scale as weekly to mitigate the circumstance which exists due to changes of rainfall pattern.

1.1. Rainfall in Sri Lanka

Sri Lanka receives rainfall throughout the year, with the mean annual rainfall varying from 900 mm in the dries parts to over 5000 mm in the wettest part. The annual rainfall pattern in many parts of Sri Lanka is bimodal and mainly governed by the seasonally varying monsoon system. Rainy periods of the country have been classified into four seasons (Domroes, 1974); First Inter Monsoon (FIM) from March to April, South West Monsoon (SWM) from May to September, Second Inter Monsoon (SIM) from October to November and North East Monsoon (NEM) from December to February.

1.2. Studies in Weekly Rainfall

Extremely very few attempts were made to analyze the weekly rainfall pattern in Sri Lanka. Weerasighe (1989) used Markov Chain probability analysis for weekly rainfall in Mapalana area in Sri Lanka with respect to the agricultural operational planning. Waidyarathne et al. (2006) analyzed weekly rainfall data to investigate the change of the onset of FIM rain in coconut growing agro ecological regions in Sri Lanka. However, a number of studies have been conducted by researchers in other countries to understand the weekly rainfall variability at given region. Most of the studies have been employed theoretical probability distributions to identify the pattern of rainfall at weekly scale (Sharma and Singh, 2010, Sharda and Das, 2005,Ghosh et al., 2016).

1.3. Use of Confidence Interval in Rainfall Studies

Confidence intervals for quantiles are depend on the distribution function. However, according to the Burn (2003) there are several shortcoming of this approach. It is necessary to make the number of assumptions with respect to the distribution and necessity to larger data series to make inferences are the main drawbacks of the traditional method. A bootstrapping approach can be proposed as an alternative approach for calculating confidence intervals through the resampling process. Dunn (2001) made an attempt to build bootstrap confidence intervals for predicting rainfall quantities. Simultaneous confidence intervals for a daily minimum rainfall total using a bootstrap resampling method considering of serial dependency have been produced by Ferro et al., (2005). Lucio (2007) adapted bootstrap method for the purpose of evaluating of small sample inferences for monthly rainfall extreme quantiles. Three approaches; Bayesian, Bootstrap and Profile Likelihood were employed to construct confidence intervals of extreme rainfall quantiles by Chen Si et al., (2016).

1.4. Importance of Weekly Rainfall Variability

In the last decade, more people in urban areas were affected by natural disaster resulting in large economic losses in the country. Sri Lanka is experiencing changes in climate particularly, rainfall events which make erratic variation. Weekly rainfall analysis is not only important for agricultural activities but also for other administrative purposes particularly, for urban areas. By looking only analysis of extreme rainfall events is not enable to confront the all human hardships during adverse rainy seasons. Forecasting weekly rainfall quantiles help to understand the pattern of variation of weekly rainfall which prominent for plan many activities, particularly, in urban areas. However, no studies were reported of weekly rainfall quantiles in Sri Lanka, in the context of confidence intervals. The main goal of this study is to construct reliable rainfall percentiles with the 95% confidence intervals during SWM and SIM using percentile bootstrap method.

2. MATERIAL AND METHODS

The Colombo is the commercial capital of Sri Lanka with latitudes 6^0 93□ N and Longitude 79^0 86□ E. The City Colombo is located in the Western part of the country which directly receives rainfall from SWM and SIM. Due to the Colombo is the commercial capital of Sri Lanka, large population density, huge construction projects, different industrial activities and various events can be enclosed. Daily rainfall data of Colombo were collected for fifty six years (1960-2015) from the Department of Meteorology, Sri Lanka for this study.

Explanatory analysis was carried out for the weekly rainfall pertaining to the SWM and SIM seasons while the Wald Wolfowiz test was used for the test of independence of weekly data series (Sharda and Das, 2005). It is formed the 95 % confidence intervals for weekly rainfall percentiles in SWM and SIM of the Colombo City using percentile bootstrap approach.

2.1. Weekly Rainfall Data

The daily rainfall (mm) data have been converted into weekly rainfall scale such that Week 1 corresponding to 1-7 January, Week 2 related to 8-14 January and so on. A year was divided into 52 weeks by ignoring leap years day (29thof February). The week 18 (30th of April to 06th of May) to week 48 (26th of November to 02nd of December) has been considered for the analysis. Out of those weeks, week 18 to week 39 pertaining to SWM while week 40 to week 48 belongs to SIM season (Table 1 and Table 2).

2.2. Confidence Interval for Parameters Using the Percentile Bootstrap Method

The bootstrap method is used to make inferences by using the information based on a number of resample from the same sample. This is a nonparametric technique that assists to make conclusions about the characteristics of a population based on the existing sample unlike the parametric approach which make assumptions about the estimators. The procedure creates simulated data set by drawing observations from the original sample with replacement. If a parameter can be expressed as a function of an unknown distribution, then its bootstrap estimator is also the function of the same distribution function. Suppose a random sample of size n, X= (X1, X2, X3,..., Xn) from an unknown population with probability distribution function F (X) and let θ be the parameter and $\hat{\theta}$ be the sample statistic (estimator) computed from the data set. B is the number of samples with size n generated from the F (X) then X*= (X*1, X*2, X*3,..., X*n) denote the bootstrap random sample of size n. Let $\hat{\theta}^*$ be a statistic (an estimator) computed using the bootstrapping sample of X*.

There are many methods that can be applied to obtain the bootstrapping confidence intervals. Some of those are the normal approximation, the percentile, the bias corrected accelerated percentile and percentile t method. The percentile method more popular among applied statistician (Hall, 1992). Suppose we generate B number of bootstrap samples with size n from the original sample data and for each sample we computed the statistic of interest $\hat{\theta}^* = (\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \ldots, \hat{\theta}_B^*)$. In our study, the rainfall percentile is the interest in statistic. The ordered bootstrap values are used to compute the bootstrap confidence intervals from the Percentile method. Suppose 1000 bootstrap replications of $\hat{\theta}$ denoted by $\left(\theta_1^*, \theta_2^*, \ldots, \theta_{1000}^*\right)$ and after ranking ascending order it can be denoted as $\left(\theta_{[1]}^*, \theta_{[2]}^*, \ldots, \theta_{[1000]}^*\right)$. Then the bootstrap percentile confidence intervals at the 95% level of confidence would be $\left[\theta_{(25)}^*, \theta_{(975)}^*\right]$ (Singh and Xie, 2008).

3. RESULTS

The methodology presented above was applied to the 56 years (1960-2015) weekly rainfall data. Based on the Wald-Wolfowitz test, it can be concluded that the weekly rainfall data series (Week 18-48) pertaining to the SWM and SIM were independent at the 5% level of significance.

3.1. Explanatory Data Analysis Result of Weekly Rainfall During SWM

The summary statistics of weekly rainfall data pertaining to SWM (Week 18-39) is presented in Table1. Also Figure.1 depicts the box plot of weekly rainfall in SWM.

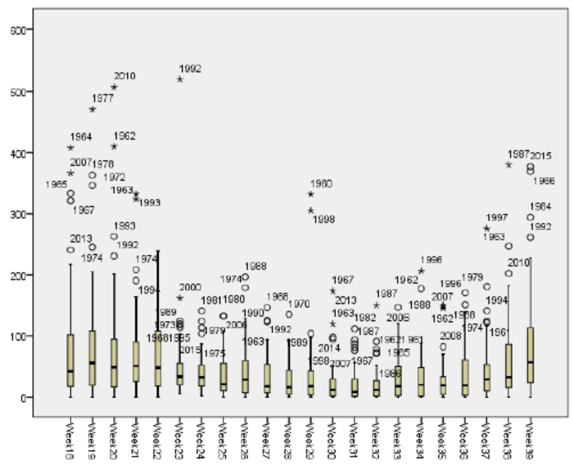

Figure 1. Weekly rainfall (Week 18-39)
(Here "O" denote the outliers and "*" denote the extreme values)

Table 1: Summary statistics of weekly rainfall total pertaining to SWM (Week 18-39) for the period of (1960-2015)

Week No	Date	Mean	Median	Minimum	Maximum	Coefficient of Variance
18	April 30 - May 06	81.6	42.0	0.0	407.7 (1964)	117.8
19	May 07-13	86.2	55.6	0.0	470.3 (1977)	111.9
20	May 14-20	75.5	48.5	0.0	506.8 (2010)	128.2
21	May 21-27	69.5	50.7	0.0	331.8 (1993)	99.5
22	May 28-June 03	69.0	48.0	0.5	239.1 (1995)	89.6
23	June 04-10	52.5	33.0	5.1	519.8 (1992)	138.4
24	June 11-17	40.6	32.0	2.0	141.0 (1981)	79.0
25	June 18-24	35.8	20.8	0.0	132.8 (2006)	99.3
26	June 25-July 01	39.5	28.2	0.0	196.4 (1988)	106.2
27	July 02-08	34.2	17.6	0.1	146.3 (1968)	106.0
28	July 09-15	29.0	16.2	0.0	135.2 (1970)	111.2
29	July 16-22	37.2	17.9	0.0	331.6 (1960)	163.6
30	July 23-29	22.7	11.9	0.0	173.0 (1967)	141.9
31	July 30-August 05	19.5	8.6	0.0	111.6 (1982)	135.9
32	August 06-12	21.5	11.6	0.0	150.4 (1987)	133.7
33	August 13-19	30.8	17.9	0.0	146.6 (2006)	113.7
34	August 20-26	29.4	20.0	0.0	205.9 (1996)	136.6
35	August 27-September 02	27.7	19.4	0.0	150.4 (1988)	128.1
36	September 03- 09	36.1	19.3	0.0	170.1 (1996)	120.7
37	September 10 - 16	44.3	28.8	0.0	275.6 (1997)	119.0
38	September 17-23	62.1	32.1	0.3	379.9 (1987)	120.1
39	September 24-30	86.6	56.6	0.0	376.4 (2015)	105.2

The values in parenthesis represent the corresponding years

Figure 1 illustrates that the distributions of the weekly rainfall data during SWM and all those distributions are positive skewed. The lowest (19.5mm) and highest (86.6mm) mean weekly rainfall were reported in the 31st week and 39th week respectively. It is noted that mean weekly rainfall in SWM gradually decreases after 19th week up to 31st week and then the pattern has changed to increase. Maximum weekly rainfall was recorded in the 23rd week (519.8mm) in 1992. It is noted that the 82% of the weeks in SWM marked more than 100% coefficient of variation which indicates the high variation in weekly rainfall. The figure1 depicts that there is a significant high variation in weekly rainfall in the week 18-23 and 29. Much lower variation can be seen in weeks 24 and 25. An almost similar pattern was observed in median weekly rainfall also.

3.2. Explanatory Data Analysis Result of Weekly Rainfall During SIM

Table 2 shows the summary statistics of weekly rainfall in SIM (Week 40-48) as well as Figure 2 indicates the box plot of weekly rainfall in SIM.

Table 2: Summary statistics of weekly rainfall total pertaining to SIM (Week 40-48) for the period of (1960-2015)

Week No	Date	Mean	Median	Minimum	Maximum	Coefficient of Variance
40	October 01-07	51.9	35.1	0.2	327.2 (2000)	100.3
41	October 08-14	81.7	46.4	0.0	370.1 (1976)	109.3
42	October 15-21	98.6	71.1	0.0	413.7 (1963)	94.7
43	October 22-28	89.7	73.5	0.0	362.4 (2006)	82.8
44	October 29-November 04	102.9	82.9	0.0	337.0 (1973)	76.0
45	November 05-11	91.1	73.7	0.0	464.0 (2010)	93.1
46	November 12-18	76.8	53.4	0.0	347.3 (2006)	99.7
47	November 19-25	62.4	53.8	0.0	388.5 (2005)	104.1
48	November 26- December 02	55.2	32.0	0.0	232.1 (2010)	97.2

The values in parenthesis represent the corresponding years

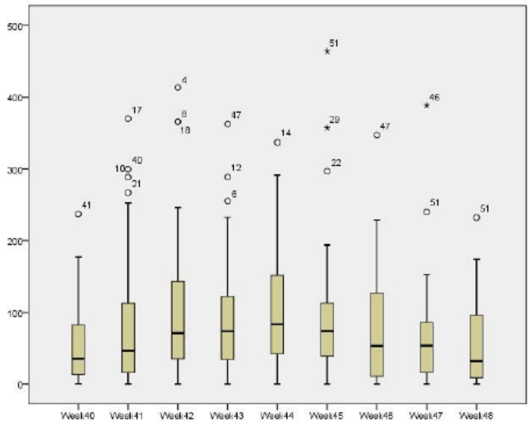

Figure 2. Weekly rainfall (Week 40-48)

Mean weekly rainfall of SIM varies from 51.9 mm to 102.90 mm. The lowest and highest mean week rainfall was reported the 40th week and the 44th week respectively. The highest weekly rainfall in SIM was reported in 2010. It can be seen in a similar pattern of weekly rainfall in mean and median in SIM. Also noted that in all the weeks in SIM higher mean rainfall than median rainfall.

Almost mean and median weekly rainfall during SIM is much higher than the weekly rainfall in SWM. It can be seen that the high variability in weekly rainfall at the beginning of the seasons. However, coefficient of variance values indicates that the considerable much low fluctuation in the weekly rainfall pertaining to SIM than SWM. Thus, this indicated that the much consistent with variation of weekly rainfall during the SIM season.

3.3. Confidence Intervals for Weekly Rainfall Percentiles in SWM

Table 3 depicts weekly rainfall percentiles and the corresponding 95% bootstrap confidence intervals. Those intervals were made for the weekly rainfall percentiles at 50, 60, 70, 80 and 90 based on the 1000 bootstrap samples.

Table 3: The 95% confidence intervals of weekly rainfall percentiles (Based on 1000 bootstrap samples) pertaining to SWM (Week 18-39)

Week Number	PERCENTILES				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
18	42.0 *(30.9, 66.3)*	61.6 *(34.8, 100.6)*	100.0 *(54.1, 130.3)*	129.4 *(96.4, 207.1)*	223.8 *(128.6, 343.4)*
19	55.6 *(32.6 ,86.4)*	81.1 *(48.1, 105.1)*	101.7 *(78., 144.0)*	142.2 *(99.2, 191.8)*	198.1 *(144.6, 351.1)*
20	48.5 *(26.2, 59.7)*	57.9 *(47.4, 84.6)*	82.0 *(56.7, 126.8)*	121.6 *(68.2, 178.6)*	197.5 *(116.5, 306.7)*
21	50.7 *(41.2, 64.5)*	61.3 *(50.0, 84.2)*	82.6 *(56.3, 101.8)*	99.6 *(72.0, 143.9)*	154.6 *(102.7, 242.6)*
22	48.0 *(29.2 , 78.1)*	75.7 *(43.6,97.2)*	85.2 *(67.0, 143.5)*	142.8 *(84.6, 163.2)*	164.6 *(144.8, 184.4)*
23	33.1 *(24.0, 44.4)*	42.6 *(31.7, 51.5)*	49.1 *(41.5, 78.1)*	76.5 *(49.0, 105.3)*	114.9 *(76.5, 135.6)*
24	32.0 *(27.2, 41.6)*	38.8 *(31.0,48.3)*	48.0 *(37.0, 63.6)*	61.9 *(47.8, 85.5)*	86.6 *(60.4, 116.5)*
25	20.8 *(14.0, 36.3)*	35.6 *(17.5, 49.9)*	49.0 *(32.7, 66.9)*	63.5 *(48.0, 84.9)*	91.1 *(68.0, 114.8)*
26	28.2 *(14.6, 49.4)*	47.5 *(22.8, 57.2)*	54.0 *(36.7, 65.2)*	65.0 *(53.4, 80.6)*	83.2 *(66.3, 143.4)*
27	17.6 *(11.9, 31.9)*	30.4 *(16.8, 49.9)*	47.3 *(27.6, 59.4)*	59.0 *(46.0, 78.7)*	86.1 *(61.8, 123.8)*
28	16.2 *(10.1, 25.9)*	24.2 *(13.4, 34.5)*	33.7 *(20.0, 59.)*	57.6 *(31.8, 83.1)*	86.3 *(57.9, 92.6)*
29	17.9 *(11.7, 35.4)*	33.2 *(15.4, 41.0)*	40.4 *(27.9, 55.7)*	53.5 *(39.1, 70.2)*	76.7 *(58.4, 164.2)*
30	11.9 *(7.7, 21.5)*	18.1 *(11.0, 24.6)*	23.7 *(15.3, 34.3)*	33.9 *(23.5, 45.5)*	47.5 *(35.4, 104.2)*
31	8.6 *(3.3, 17.2)*	13.3 *(6.4, 26.3)*	25.1 *(12.1, 31.5)*	31.4 *(24.2, 52.3)*	62.0 *(32.0, 89.0)*
32	11.6 *(4.3, 18.1)*	17.1 *(9.7, 24.3)*	23.7 *(16.2, 45.8)*	42.7 *(21.6, 51.9)*	57.9 *(46.9, 82.0)*
33	17.9 *(10.0, 29.3)*	27.4 *(15.0, 46.9)*	41.7 *(26.4, 55.7)*	54.6 *(41.2, 86.3)*	89.3 *(56.0, 106.9)*
34	20.1 *(9.6, 24.0)*	22.9 *(14.0, 30.3)*	30.0 *(20.9, 54.7)*	54.2 *(29.6, 72.3)*	72.8 *(54.8, 115.0)*
35	19.4 *(7.6, 25.9)*	21.8 *(18.6, 29.1)*	28.3 *(22.4, 47.7)*	47.2 *(28.0, 62.4)*	64.9 *(48.9, 145.9)*
36	19.3 *(7.9, 33.7)*	27.4 *(15.8, 58.2)*	45.9 *(25.3, 68.2)*	68.0 *(43.9, 105.0)*	114.7 *(69.0, 142.6)*
37	28.8 *(16.2, 37.5)*	34.7 *(25.1,45.5)*	44.4 *(34.3, 70.7)*	69.9 *(44.2, 116.3)*	118.1 *(77.8, 153.0)*
38	32.1 *(22.3, 42.9)*	40.4 *(29.1, 64.1)*	58.9 *(38.6, 130.9)*	122.8 *(54.8, 164.1)*	167.2 *(116.1, 215.4)*
39	56.6 *(32.2, 89.6)*	85.8 *(48.3,104.3)*	102.8 *(77.1, 158.5)*	155.4 *(101.0, 216.1)*	225.6 *(164.4, 316.8)*

The values in parenthesis represent the corresponding 95% confidence intervals.

The result indicated that the heavy rainfall at the begins of the SWM. Furthermore, it can be expected much rainfall from week 18 to week 23. It is evident from Table 3 that 80% or more chances to have 207.1mm maximum weekly rainfall in the weeks 18 - 23. However, after 23rd week it can be seen clear decline of weekly rainfall up to week 35. It can be expected 86.3 mm maximum week rainfall with 80% probability at week 33 which showed maximum rainfall variability out of weeks 24-35. The week 31 and 32 marked much lower rainfall during the SWM. Since week 36 it can be seen much rainfall till end of the season. The week 39 marked highest rainfall amount during the SWM. The table depicts a high variability at the weeks 18-23, 29 and 38-39. Also, there is a much higher possibility to have extreme rainfall during the weeks 18-23, 29 and 38-39.

3.4 Confidence Intervals for Weekly Rainfall Percentiles in SIM

Table 4 depicts weekly rainfall percentiles and the corresponding 95% confidence intervals. The results indicated that the heavy rainfall over the SIM compare with the SWM. It can be expected 135.6 mm maximum rainfall at 80% probability at week 41 and the value can be varied from 84.6mm to 222.6mm which indicated that a high variability. It is noted that high rainfall variability at the weeks 41-45 in SIM. Thus, there is a much higher chance to have extreme rainfall events at above weeks.

Table 4: The 95% confidence intervals of weekly rainfall percentiles (Based on 1000 bootstrap samples) pertaining to SIM (Week 40-48)

Week Number	PERCENTILES				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
40	35.1 *(20.1, 50.3)*	48.8 *(32.3, 73.4)*	67.1 *(45.7, 99.6)*	99.5 *(64.4, 118.9)*	122.7 *(100.4, 165.9)*
41	46.4 *(29.8, 71.7)*	64.8 *(44.6, 107.4)*	87.7 *(61.7, 143.0)*	135.6 *(84.6, 222.6)*	239.3 *(137.9, 291.7)*
42	71.1 *(52.8, 91.5)*	83.1 *(70.0, 131.8)*	107.4 *(79.3, 157.7)*	156.0 *(105.1, 211.4)*	220.1 *(161.9, 366.0)*
43	73.5 *(51.9, 96.6)*	90.8 *(70.0, 119.3)*	114.4 *(86.7, 138.2)*	135.4 *(108.7, 166.0)*	185.7 *(139.7, 265.2)*
44	83.0 *(59.7, 119.8)*	107.9 *(78.7, 134.1)*	132.5 *(97.6, 169.1)*	168.7 *(128.9, 201.3)*	212.0 *(169.7, 280.7)*
45	73.7 *(61.0, 86.1)*	79.2 *(71.0, 112.0)*	109.5 *(78.7, 126.1)*	125.8 *(107.0, 178.9)*	182.7 *(125.7, 314.8)*
46	53.4 *(33.0, 80.9)*	75.6 *(45.9, 109.5)*	100.8 *(72.0, 142.7)*	142.6 *(97.0, 185.5)*	192.7 *(144.4, 226.2)*
47	53.8 *(29.1, 74.7)*	73.3 *(53.2, 83.9)*	80.2 *(66.2, 90.2)*	90.1 *(79.2, 114.8)*	125.6 *(90.5, 178.7)*
48	32.0 *(24.1, 65.6)*	62.7 *(30.5, 77.0)*	76.2 *(55.2, 107.2)*	105.7 *(71.8, 129.4)*	132.3 *(107.7, 152.1)*

The values in parenthesis represent the corresponding 95% confidence intervals.

4. DISCUSSION AND CONCLUSION

Weekly rainfall data pertaining to SWM and SIM are positive skewed with a longer tail widen to the right. It is observed that a similar pattern of mean and median weekly rainfall in SWM and SIM. Also, it is noted that the 82% of the weeks pertaining to SWM marked more than 100% coefficient of variation result in the high fluctuations in weekly rainfall during this time span. Conversely, there is much lower variation in weekly rainfall in SIM than SWM. Based on the analysis of rainfall percentiles and confidence intervals which constructed using bootstrapping approach, it can be expected much heavy rainfall at the beginning of the SWM. Also, a similar pattern can be identified at the withdrawal of the monsoon. Thus, there is a high possibility to form extreme rainfall events in the weeks 18-23 (30th April to 10th June) and 38-39 (17-30 September) during SWM. It is noted that the high intensity shower over the SIM. The Weeks 41-45 (8th October to 11th November) which pertaining to SIM showed not only high rainfall, but also high rainfall variation result cause to the high likelihood to form severe rainfall events.

Confidence intervals of weekly rainfall quantiles which can be made using probability distributions is not much accuracy due to those forms by assuming normal approximation. However, this can be employed to describe the mean weekly rainfall variability. Small sample size and positive skewed distribution of weekly rainfall are some reasons for the poor accuracy rate of confidence intervals of percentiles which made using traditional methods. As an alternative, distribution free method such as percentile bootstrap approach can be utilized to describe the uncertainty of rainfall percentiles. However, we cannot satisfy about the length of the 95% confidence intervals of rainfall percentiles which made based on the percentiles bootstrap approach also. This can be developed using parametric bootstrapping approach at an optimal confidence level which can be made using bootstrapping calibration.

ACKNOWLEDGEMENT

This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant (ASP/01/RE/HSS/2016/75).

REFERENCES

Burn, D. H., 2003, The use of resampling for estimating confidence intervals for single site and pooled frequency analysis, *Hydrological sciences journa*l, 48(1), 25-38.

Chen, S., Li, Y., Shin, J., and Kim, T., 2016, Constructing confidence intervals of extreme rainfall quantiles using Bayesian, bootstrap, and profile likelihood approaches. *Science China Technological Sciences*, 59(4), 573-585.

Domroes, M., 1974, *The Agroclimatic of Sri Lanka*. Wiesbaden.

Dunn, P. K., 2001, Bootstrap confidence intervals for predicted rainfall quantiles. *International Journal of Climatology*, 21(1), 89-94.

Ferro, C. A., Hannachi, A., and Stephenson, D. B., 2005, Simple nonparametric techniques for exploring changing probability distributions of weather. *Journal of Climate*, 18(21), 4344-4354.

Ghosh, S., Roy, M. K., and Biswas, S. C., 2016, Determination of the Best Fit Probability Distribution for Monthly Rainfall Data in Bangladesh. *American Journal of Mathematics and Statistics*, 6 (4), 170-174.

Hall, P., 1992, *The Bootstrap and Edgeworth Expansion*. SpringerVerlag, New York.

Hao, J., Bathke, A., Skees, J.R.,Dai, H., 2011, Weather Risks, Ratemaking and Modeling the Tail Distribution: An application of Extreame Value Theory, *International Journal of Economics and Statistics*, Vol 20, W11.

Lo, K. F. A. and Koralegedara, S. B., 2015, Effects of climate change on urban rainwater harvesting in Colombo city, Sri Lanka. *J. Environments*, 2(1), 105-124.

Lucio, P., 2007, Bootstrap for statistical evaluation of small sample inference for precipitation extreme quantiles. Paper presented at the Geophysical Research Abstracts.

Mayooran,T. and Laheetharan, A., 2014, The Statistical Distribution of Annual Maximum Rainfall in Colombo District, Sri Lankan *J.Applied Statistics*, vol. 15, no. 2.

Sharda, V., & Das, P., 2005, Modelling weekly rainfall data for crop planning in a sub-humid climate of India. *Agricultural water management*, 76(2), 120-138.

Sharma, M. A., & Singh, J. B., 2010, Use of probability distribution in rainfall analysis. *New York Science Journal*, 3 (9), 40-49.

Singh, K., & Xie, M., 2008, Bootstrap: a statistical method. Unpublished manuscript, Rutgers University, USA. Retrieved from: http://www.stat.rutgers.edu/home/mxie/RCPapers/bootstrap. pdf.

Waidyarathne, K. P., Peiris, T.S.G. & Samita, S., 2006, Shift in onset of first inter monsoon rain in coconut growing areas in Sri Lanka. *Journal of Tropical Agricultural Research*, 18, 1-12.

Weerasinghe, K.,1989, The rainfall probability analysis of Mapalana and its application to agricultural production of the area, *J Natl Sci Counc Sri Lanka*, 17(2), 173-186.

Wirnancy,J.S., Wigenac, A., Djuraidah, A., 2017, Quantile Regression with Functional Principal Component in Statistical Downscaling to predict Extreame Rainfall, *International Journal of Ecological Economics and Statistics*, Vol 38 (1).

Statistical Modeling of Weekly Rainfall: A Case Study in Colombo City in Sri Lanka

H. P. T. N. Silva
Department of Social Statistics
University of Sri Jayewardenepura
Nugegoda, Sri Lanka
Email: thanuja@sjp.ac.lk

T. S. G. Peiris
Department of Mathematics
University of Moratuwa
Moratuwa, Sri Lanka
Email: sarathp@uom.lk

Abstract— **The modeling of the weekly rainfall percentile is imperative for better understanding of rainfall patterns in any region. This study focuses on selecting the most appropriate probability distributions for weekly rainfall and use those to make reliable rainfall percentile with the 95% confidence intervals. Daily rainfall data of 56 years (1960-2015) during the South West Monsoon in Colombo City were used for this analysis. The three parameter Weibull distribution has been found most probable distribution for most of weekly rainfall totals. Weibull, two parameter Exponential, Exponential and Lognormal distributions were well fitted distributions for remaining totals. Based on the 95% confidence intervals of percentiles, the weeks 18-23, and 38-39 during SWM showed not only high rainfall, but also high rainfall variation results which caused high possibility to form extreme rainfall events. Heavy rainfall with great variation during the period of 30th April to 10th June and 17-30th of September was further confirmed by the result of running total of weekly rainfall.**

Keywords— Weekly Rainfall; Distribution; Colombo; Percentile; Confidence Intervals

I. Introduction

Rainfall percentiles are employed in designing of water related structures in many fields. Sound awareness about the rainfall pattern is vital to mitigate the various issues derived from heavy rainfall and long dry spell existence due to climate change. The probability distribution of the rainfall is essential to examine the pattern of rainfall specially in short range scale to get the maximum benefit from the rainfall by minimizing the damages which would be caused by changes of atmospheric behavior. Numerous people who live in urban areas are faced with many difficulties due to extreme rainfall events, especially from floods which occur from time to time [1]. Thus, prior knowledge of weekly rainfall behavior will be helpful to minimize such damages. By analyzing the rainfall characteristics on a weekly scale would be helpful to plan many activities which enclose with the urban areas, such as industrial, constructions, rain water harvesting, health and climate monitoring.

Sri Lanka is a tropical country which is vulnerable to climate change specially, from erratic rainfall events. The rainfall of Sri Lanka is strongly governed by the four seasonal varying monsoon system. Four major monsoon periods; First Inter Monsoon (FIM) from March to April, South West Monsoon (SWM) from May to September, Second Inter Monsoon (SIM) from October to November and North East Monsoon (NEM) from December to February can be identified in Sri Lanka [2].

Most of the researchers use point estimates derived from different theoretical probability distributions for rainfall percentiles and attempt to make inferences of rainfall amount. [3] used the Generalized Extreme Value distribution, Gamma and Log Pearson distributions for the maximum weekly rainfall in the monsoon period at the Pantnagar region in India to study the temporal variability of maximum weekly rainfall. According to the review of [4] the Weibull distribution is more likely fitted for describing weekly rainfall at Dehradun in India. Also, they used the probability distribution models for computing minimum assured amount of rainfall at different probability levels. Beta and Weibull distributions were fitted for the weekly rainfall during the monsoon and non monsoon periods, respectively, and those best fit distributions are employed for computing minimum assured amount of rainfall at different probability levels for the Command area by [5].

Moreover, many researchers have fitted theoretically probability distribution for the rainfall data at different time scales mainly monthly, seasonally and annually ([6], [7], [8] ,[9], [10]). However, extremely few studies were reported in Sri Lanka with respect to the rainfall variation at weekly scale. As noted in [11], weekly rainfall data were analyzed to investigate the change of the onset of FIM rain in coconut growing agro ecological regions in Sri Lanka.

However, it might be more risky depending on a single value formed from probability distributions to mitigate the circumstances which would be existed due to climate change. Confidence interval is one of the most popular technique that can be used to measure the uncertainty. Some researchers had made attempts to construct confidence intervals for rainfall amounts using different approaches such as Bootstrap and Bayesian. Bootstrap confidence intervals were made for the predicted rainfall quantities to show the effects of the Southern Oscillation Index Phase on rainfall quantiles by [12]. The three approaches; Bayesian, Bootstrap and Profile Likelihood were applied to construct confidence intervals of extreme rainfall quantiles by [13]. A study [14] was carried out to obtain reliable rainfall quantiles estimates for several return

This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant (ASP/01/RE/HSS/2016/75).

978-1-5090-6491-5/17/$31.00 ©2017 IEEE

periods by using Wakeby Distribution with the method of L-moments estimates. Also, the 90% confidence intervals for the quantiles determined by Wakeby Distribution were constructed by using bootstrap resampling technique. To the best of the authors' knowledge, no study has been conducted for weekly rainfall quantiles in context of the parametric confidence interval approach in Sri Lanka.

The main goal of this study is to select the most appropriate probability distribution of weekly rainfall and use those selected distributions to make reliable rainfall percentile with 95% confidence intervals.

II. Materials and Methods

The City of Colombo is the commercial capital of Sri Lanka. It is situated with latitudes 6^0 93′ N and Longitude 79^0 86′ E and is selected as the study site. Daily rainfall data of Colombo were collected from 1960 to 2015 from the Department of Meteorology, Sri Lanka for this study. Weekly rainfall pertaining to SWM is considered for this analysis due to this monsoon brings rainfall directly to the Colombo area during May to September. The Wald Wolfowiz test was used for the test of independence of weekly data series (Sharda and Das, 2005). Two goodness of fit test; Anderson Darling and Kolmogorov-Smirnov were used to identify the best fitted distributions for weekly rainfall data separately. Rainfall percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) were derived using best fitted distribution and constructed the 95% confidence bands for corresponding rainfall percentiles.

Furthermore, running totals of weekly rainfall were obtained to identify the pattern of weekly rainfall which start on any day during SWM. Moreover, 95% confidence intervals for percentiles based on the best fitted distributions of running totals were constructed.

A. Weekly Rainfall Data

The daily rainfall (mm) data has been converted into weekly rainfall by dividing a year into 52 weeks as Week 1, Week 2, Weeks 3 and others corresponding to 1-7 January, 8-14 January, 15-21 January and so on respectively. It is noted that the February 29^{th} wasn't taken into account when marking 52 weeks. The weeks pertaining to SWM is presented in Table I. Also, running totals of weekly rainfall were obtained during SWM period with Week 1 of the running total corresponding to 30^{th} of April to 06^{th} of May, Week 2 represent the period, 1-7 May, Week 3 of the running total corresponding to 2-8 May, Week 4 related to 3-9 May and so on. It is calculated total of 148 running totals of weekly rainfall during the SWM.

B. Fittings the Probability Distributions

Weekly rainfall data as well as running weekly rainfall totals were fitted to various theoretical probability distributions such as Normal, Lognormal, Gamma, Weibull, Exponential, Smallest Extreme Value, Largest Extreme Value, Logistic, Log logistics and also tried different forms of some distributions such as 3- parameter Gamma, 2- Parameter Exponential, 3-Parameter Log logistic and 3-Parameter Weibull distributions.

TABLE I. WEEKS PERTAINING TO THE SWM

Weeks	Date	Weeks	Date
18	April 30-May 06	29	July 16-22
19	May 07-13	30	July 23-29
20	May 14-20	31	July 30-August 05
21	May 21-27	32	August 06-12
22	May 28-June 03	33	August 13-19
23	June 04-10	34	August 20-26
24	June 11-17	35	August 27-September 02
25	June 18-24	36	September 03-09
26	June 25- July 01	37	September 10-16
27	July 02-08	38	September 17-23
28	July 09-15	39	September 24-30

Anderson Darling test and Kolmogorov-Smirnov test were used as goodness of fit tests for parametric distributions. The computations were done using statistical software, namely Minitab 17 and Stata 12.1. Selected probability distribution functions are described by considering X as a random variable representing weekly rainfall as presented in Table II. The formula used for the percentile and its variance calculation is also shown in Table III. Furthermore, Table IV depicts the formulas that were employed for the confidence bands of percentiles.

III. Results and Discussion

A. Modeling Weekly Rainfall

Histogram of dataset provides clear evidence that the distributions of the weekly rainfall are skewed to the right. Four randomly selected weeks 18, 24, 30 and 37 are depicted in Fig1. An almost similar pattern was observed in remaining data series also. Before fitting various probability distributions to data set, data were tested for normality using Anderson Darling test and it was revealed that, no data series followed a normal distribution. Furthermore, according to the result of the Wald-Wolfowitz test, there is no evidence to reject the null hypothesis which data are independent at the 5% level of significance for all week. Table V illustrates the best fitted distribution for weekly rainfall total with estimated maximum Likelihood estimators (MLE). Also the corresponding Anderson Darling test statistics (AD) and Kolmogorov-Smirnov test statistic (KS) were presented in the Table V. Same procedure was carried out for the running totals and obtained a similar result.

It is noted that the most of the week belongs to the SWM were well fitted with the 3 parameter Weibull distribution. However, weeks 22-24, Exponential, Lognormal and Weibull distributions were found to be most appropriate distributions. Two parameter Exponential distributions were most probable distribution for the Weeks 26, 29, 31 and 34. Moreover, 68% of the running weekly totals are well fitted with the 3 parameter Weibull distribution while 22% are fitted with the two parameter Exponential distribution and the remaining are well fitted with the Exponential, Largest Extreme Value, Weibull and Lognormal distributions.

TABLE II. PROBABILITY DENSITY FUNCTIONS

Distribution	Probability Density Function	Parameters
Lognormal	$f(x)=\frac{1}{\sigma x\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)}{2\sigma^2}\right)$	μ - Location Parameter, σ -Scale Parameter $\mu \geq 0, \sigma > 0, X \geq 0$
Exponential	$f(x)=\frac{1}{\alpha}\exp\left(-\frac{x}{\alpha}\right)$	α -Scale Parameter $\alpha > 0$
2 Parameter Exponential	$f(x)=\frac{1}{\alpha}\exp\left(-\frac{(x-\lambda)}{\alpha}\right)$	α -Scale Parameter, λ -Threshold parameter $\alpha > 0, \lambda < X$
Largest Extreme Value	$f(x)=\frac{1}{\sigma}\exp\left[\left(\frac{x-\mu}{\sigma}\right)\right]\exp\left\{-\exp\left(\frac{(x-\mu)}{\sigma}\right)\right\}$	μ - Location Parameter, σ -Scale Parameter $\mu \geq 0, \sigma > 0, X \geq 0$
Weibull	$f(x)=\frac{\beta}{\alpha^\beta}x^{\beta-1}\exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right)$	α -Scale Parameter, β -Shape Parameter $\alpha > 0, \beta > 0, X \geq 0$
3-Pataramete Weibull	$f(x)=\frac{\beta}{\alpha^\beta}(x-\lambda)^{\beta-1}\exp\left(-\left(\frac{x-\lambda}{\alpha}\right)^\beta\right)$	α -Scale Parameter, β -Shape Parameter, λ -Threshold parameter $\alpha > 0, \beta > 0, \lambda < X$

TABLE III. THE FORMULAS USED FOR PERCENTILES AND VARIANCE ESTIMATES

Distribution	Percentiles $\left(\hat{X}_p\right)$	Variance of Percentile Var $\left(\hat{X}_p\right)$
Lognormal	$\hat{\mu}+z_p\hat{\sigma}$	$Var(\hat{\mu})+z_p{}^2\,Var(\hat{\sigma})+2Z_pCov(\hat{\mu},\hat{\sigma})$
Exponential	$-\ln(1-p)\hat{\alpha}$	$[-\ln(1-p)]^2\,Var(\hat{\alpha})$
2 Parameter Exponential	$\hat{\lambda}+[-\ln(1-p)\hat{\alpha}]$	$Var(\hat{\lambda})+[-\ln(1-p)]^2\,Var(\hat{\alpha})+2[-\ln(1-p)]Cov(\hat{\lambda},\hat{\alpha})$
Largest Extremes Value	$\hat{\mu}+z_p\hat{\sigma}$	$Var(\hat{\mu})+z_p{}^2\,Var(\hat{\sigma})+2Z_pCov(\hat{\mu},\hat{\sigma})$
Weibull	$\hat{\alpha}\,[-\ln(1-p)]^{1/\beta}$	$\frac{\hat{X}_p^2}{\hat{\alpha}^2}Var(\hat{\alpha})+\frac{\hat{X}_p^2}{\hat{\beta}^4}z_p{}^2\,Var(\hat{\beta})-2Z_p\frac{\hat{X}_p^2}{\hat{\alpha}\hat{\beta}^2}Cov(\hat{\alpha},\hat{\beta})$
3-Parameter Weibull	$\hat{\lambda}+\hat{\alpha}\,[-\ln(1-p)]^{1/\beta}$	$Var(\hat{\lambda})+\hat{\omega}^2\,Var(\hat{\alpha})+\frac{\hat{\alpha}^2}{\hat{\beta}^4}\hat{\omega}Z_p^2\,Var(\hat{\beta})-2\frac{\hat{\alpha}}{\hat{\beta}^2}\hat{\omega}^2z_p{}^2\,Cov(\hat{\alpha},\hat{\beta})+2\hat{\omega}Cov(\hat{\alpha},\hat{\lambda})-2\frac{\hat{\alpha}}{\hat{\beta}^2}Z_p\hat{\omega}\,Cov(\hat{\beta},\hat{\lambda})$

TABLE IV. THE FORMULAS USED FOR CONFIDENCE LIMITS FOR PERCENTILES

Distribution	Confidence Bands
Lognormal Exponential Weibull	$\left\{\exp\left[\ln(\hat{X}_p)-Z_{\alpha/2}\frac{\sqrt{Var(\hat{X}_p)}}{\hat{X}_p}\right],\exp\left[\ln(\hat{X}_p)+Z_{\alpha/2}\frac{\sqrt{Var(\hat{X}_p)}}{\hat{X}_p}\right]\right\}$
2-Parameter Exponential 3- Parameter Weibull	If $\lambda < 0$ $\left\{\hat{X}_p-Z_{\alpha/2}\sqrt{Var(\hat{X}_p},\quad \hat{X}_p+Z_{\alpha/2}\sqrt{Var(\hat{X}_p}\right\}$ $\lambda > 0$ $\left\{\hat{X}_p-Z_{\alpha/2}\sqrt{\frac{Var(\hat{X}_p)}{\hat{X}_p}},\quad \hat{X}_p+Z_{\alpha/2}\sqrt{\frac{Var(\hat{X}_p)}{\hat{X}_p}}\right\}$
Largest Extreme Value	$\left\{\hat{X}_p-Z_{\alpha/2}\sqrt{Var(\hat{X}_p},\quad \hat{X}_p+Z_{\alpha/2}\sqrt{Var(\hat{X}_p}\right\}$

TABLE V. BEST FITTED STATISTICAL MODELS AND MAXIMUM LIKELIHOOD ESTIMATES FOR WEEKLY RAINFALL DURING SWM

Week No.	Best Fitted Distribution	AD	KS	Estimated Parameters (MLE)
18	3 - Parameter Weibull	0.317 (0.501)	0.0782 (0.884)	$\alpha = 77.061$, $\beta = 0.878$, $\lambda = -0.838$
19	3 - Parameter Weibull	0.131 (0.520)	0.0526 (0.996)	$\alpha = 82.249$, $\beta = 0.888$, $\lambda = -0.935$
20	3 - Parameter Weibull	0.247 (0.510)	0.0684 (0.956)	$\alpha = 67.331$, $\beta = 0.804$, $\lambda = -0.508$
21	3 - Parameter Weibull	0.362 (0.461)	0.1027 (0.596)	$\alpha = 73.570$, $\beta = 1.086$, $\lambda = -1.752$
22	Exponential	0.457 (0.540)	0.0857 (0.773)	$\alpha = 68.989$
23	Lognormal	0.319 (0.526)	0.0700 (0.928)	$\mu = 3.518$, $\sigma = 0.912$
24	Weibull	0.291 (0.257)	0.0691 (0.934)	$\alpha = 43.645$, $\beta = 1.267$
25	3 - Parameter Weibull	0.498 (0.222)	0.0752 (0.910)	$\alpha = 34.182$, $\beta = 0.884$, $\lambda = -0.383$
26	2- Parameter Exponential	0.912 (0.103)	0.1099 (0.110)	$\alpha = 40.204$, $\lambda = -0.718$
27	3 - Parameter Weibull	0.275 (0.521)	0.073 (0.926)	$\alpha = 32.535$, $\beta = 0.887$, $\lambda = -0.269$
28	3 - Parameter Weibull	0.531 (0.186)	0.069 (0.952)	$\alpha = 24.822$, $\beta = 0.741$, $\lambda = -0.131$
29	2- Parameter Exponential	0.873 (0.107)	0.1813 (0.182)	$\alpha = 37.875$, $\lambda = -0.676$
30	3 - Parameter Weibull	0.596 (0.210)	0.1066 (0.548)	$\alpha = 16.711$, $\beta = 0.626$, $\lambda = -0.038$
31	2- Parameter Exponential	0.841 (0.126)	0.1823 (0.232)	$\alpha = 19.853$, $\lambda = -0.355$
32	3 - Parameter Weibull	0.617 (0.113)	0.1002 (0.627)	$\alpha = 15.263$, $\beta = 0.602$, $\lambda = -0.029$
33	3 - Parameter Weibull	0.445 (0.531)	0.094 (0.706)	$\alpha = 23.975$, $\beta = 0.651$, $\lambda = -0.067$
34	2- Parameter Exponential	0.694 (0.101)	0.1193 (0.194)	$\alpha = 29.964$, $\lambda = -0.535$
35	3 - Parameter Weibull	0.607 (0.120)	0.1186 (0.410)	$\alpha = 22.408$, $\beta = 0.698$, $\lambda = -0.089$
36	3 - Parameter Weibull	0.544 (0.328)	0.0888 (0.770)	$\alpha = 26.012$, $\beta = 0.602$, $\lambda = -0.049$
37	3 - Parameter Weibull	0.246 (0.531)	0.0662 (0.967)	$\alpha = 40.709$, $\beta = 0.838$, $\lambda = -0.366$
38	3 - Parameter Weibull	0.438 (0.315)	0.0979 (0.656)	$\alpha = 57.303$, $\beta = 0.855$, $\lambda = -0.261$
39	3 - Parameter Weibull	0.397 (0.394)	0.0679 (0.958)	$\alpha = 81.654$, $\beta = 0.863$, $\lambda = -0.831$

* The value in parenthesis represent the corresponding P value

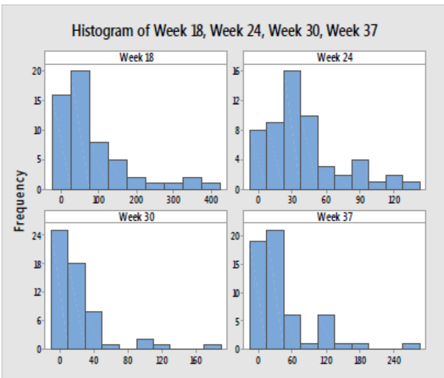

Fig. 1. Histograms of Weeks 18, 24, 30 and 37

B. *Percentile and confidence intervals*

Weekly rainfall percentiles and the corresponding 95% confidence intervals are presented in Table VI. Those intervals were made for the weekly rainfall percentiles at 50, 60, 70, 80 and 90 based on the probability distributions which were selected as best fitted for corresponding weeks.

The result indicated that there was much heavy rainfall at the begins of the SWM. Also, Weeks 18-23 marked considerable rainfall with high variability. It is noted that 90th percentiles of Weeks 18-23 varies between 108.4mm to 209.6 mm which bring a greater amount of rainfall to this region. According to the Table VI, there is a 90% chance to have 209.6 mm maximum rainfall, during the 19th week and this value can be varied between 144.2 mm and 274.9 mm at 95% confidence level. However, a clear decreasing pattern of weekly rainfall can be identified after the 23rd week.

TABLE VI. PERCENTILES OF WEEKLY RAINFALL AND THE CORRESPONDING 95% CONFIDENCE INTERVAL DURING SWM IN COLOMBO

Week Number	**PERCENTILES**				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
18	49.9 *(32.1, 67.7)*	68.9 *(46.5, 91.4)*	94.4 *(65.4, 123.3)*	131.7 *(92.1, 171.3)*	198.5 *(136.3, 260.7)*
19	53.5 *(34.6, 72.4)*	73.6 *(49.9, 97.3)*	100.45 *(70.0, 130.9)*	139.7 *(98.1, 181.2)*	209.6 *(144.2, 274.9)*
20	42.2 *(25.9, 58.5)*	59.9 *(38.7, 81.0)*	84.3 *(56.2, 112.4)*	121.2 *(81.5, 160.9)*	189.5 *(124.5, 254.6)*
21	50.8 *(35.9, 65.6)*	66.1 *(48.5, 83.8)*	85.5 *(64.1, 107.0)*	112.3 *(84.8, 139.8)*	156.8 *(116.8, 196.8)*
22	47.8 *(36.8, 62.1)*	63.2 *(48.6, 82.1)*	83.1 *(63.9, 107.9)*	111.0 *(85.4, 144.3)*	158.9 *(122.3, 206.4)*
23	33.72 *(26.6, 42.8)*	42.5 *(33.3, 54.1)*	54.4 *(42.1, 70.2)*	72.6 *(55.0, 95.9)*	108.4 *(78.5, 149.8)*
24	32.7 *(25.6, 41.7)*	40.7 *(32.6, 50.9)*	50.5 *(41.0, 62.3)*	63.5 *(51.7, 78.2)*	84.3 *(67.8, 104.8)*
25	22.2 *(14.3, 30.1)*	30.6 *(20.7, 40.4)*	41.8 *(29.1, 54.5)*	58.2 *(40.8, 75.6)*	87.4 *(59.8, 115.1)*
26	27.2 *(19.9, 34.4)*	36.1 *(26.5, 45.8)*	47.7 *(35.0, 60.4)*	64.0 *(47.0, 80.9)*	91.9 *(67.6, 116.1)*
27	21.3 *(13.8, 28.7)*	29.2 *(19.8, 38.6)*	39.8 *(27.8, 51.9)*	55.4 *(38.9, 71.8)*	83.0 *(57.1, 108.9)*
28	15.0 *(8.7, 21.3)*	21.9 *(13.6, 30.3)*	31.8 *(20.3, 43.2)*	47.0 *(30.3, 63.8)*	76.4 *(47.6, 105.2)*
29	25.5 *(18.7, 32.5)*	34.0 *(24.9, 43.1)*	44.9 *(33.0, 56.9)*	60.3 *(44.3, 76.2)*	86.5 *(63.7, 109.4)*
30	9.3 *(4.7, 13.8)*	14.5 *(8.0, 21.0)*	22.5 *(12.9, 32.0)*	35.7 *(20.7, 50.7)*	63.4 *(35.1, 91.6)*
31	13.4 *(9.8, 17.0)*	17.8 *(13.1, 22.6)*	23.6 *(17.3, 29.8)*	31.6 *(23.2, 40.0)*	45.4 *(33.4, 57.3)*
32	8.3 *(4.0, 12.5)*	13.2 *(7.0, 19.3)*	20.8 *(11.6, 29.9)*	33.6 *(18.9, 48.3)*	61.0 *(32.7, 89.2)*
33	13.6 *(7.1, 20.0)*	20.9 *(11.9, 29.9)*	31.8 *(18.8, 44.8)*	49.8 *(29.6, 69.9)*	86.4 *(49.1, 123.6)*
34	20.2 *(14.8, 25.7)*	26.9 *(19.7, 34.1)*	35.5 *(26.1, 45.0)*	47.7 *(35.1, 60.3)*	68.5 *(50.4, 86.5)*
35	13.2 *(7.3, 19.0)*	19.7 *(11.7, 27.6)*	29.2 *(18.0, 40.3)*	44.2 *(27.6, 60.9)*	74.0 *(44.4, 103.5)*
36	14.1 *(6.9, 21.3)*	22.5 *(12.0, 32.9)*	35.4 *(19.8, 51.0)*	57.3 *(32.3, 82.3)*	103.9 *(155.6, 152.2)*
37	25.9 *(16.3, 35.6)*	36.3 *(24.0, 48.7)*	50.4 *(34.3, 66.6)*	71.5 *(49.0, 93.9)*	109.8 *(73.6, 146.0)*
38	37.1 *(23.6, 50.5)*	51.5 *(34.4, 68.6)*	70.9 *(48.7, 93.1)*	99.7 *(69.1, 130.3)*	151.7 *(103.1, 200.3)*
39	52.6 *(33.6, 71.6)*	73.0 *(48.9, 97.0)*	100.4 *(69.3, 131.6)*	140.9 *(97.8, 184.0)*	213.7 *(144.6, 282.8)*

The Weeks 31 and 32 marked lower rainfall amount than others during SWM. After 35th week, again it can be seen an increasing trend of weekly rainfall till the end of the season. The Week 39 records the highest rainfall amount in the SWM. The median rainfall of the 39th weeks was 52.6 mm while the 70th percentile of this week marked more than 100 mm rainfall amount which is a large quantity for the area. Week 38 also brings much heavy rainfall with noticeable variation in this season for this region.

The rainfall percentiles and corresponding 95% confidence intervals for running totals of weekly rainfall were constructed during the SWM in Colombo. Fig.2 represented only 90th percentile of running total and its 95% confidence bands. It also depicts the high rainfall variation with the arrival of SWM. Also, Fig.2 illustrates the much heavy rainfall due to the withdrawal of the SWM. Based on the result of the running total of the weekly rainfall, it can be further confirmed that there was heavy rainfall with great variation during the period of weeks 18-23 (30th April to 10th June) and weeks 38-39 (17th-30th of September).

IV. Conclusion

Weekly rainfall data pertaining to SWM is skewed with a longer tail extending to the right. One probability distribution has not been found to represent all the week. However, three parameter Weibull distribution was well fitted with the most of the week. Two parameter Exponential distribution, Exponential, Weibull and Lognormal are the other best fitted probability distributions for weekly rainfall data. Based on the percentiles and corresponding 95% confidence intervals which were derived using selected probability distributions, much heavy rainfall during the weeks 18-23 and 38-39 can be expected. Founded on the analysis of running weekly totals of rainfall, it can be further confirmed that there is a high possibility of extreme rainfall events forming within this period. Based on the analysis of past extreme rainfall events in Colombo area during SWM, it can be identified that the many floods occurred in the months May and June. Most recently (on 15 May 2016) Sri Lanka was hit by a severe tropical storm that caused heavy flooding in Colombo. Furthermore, floods occurred in Colombo in the past years; 1975, 1989, 1992, 2008 from May to June period [15].

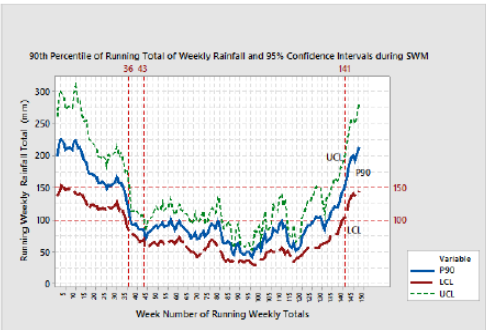

Fig. 2. 90th Percentiles of running total of weekly rainfall and 95% confidence intervals during SWM in Colombo

As shown Fig 2, much heavy variation in weekly rainfall can be identified with the arrival of the monsoon. Thus, the time onset of the monsoon is also important to mark extreme rainfall events.

However, we cannot be satisfied about the length of the 95% confidence intervals of rainfall percentiles as the intervals are somewhat wider. Small sample size and strongly skewed distribution pattern might be a one of the reasons for wide confidence bands. Furthermore, heavy skewed distributions have deviated from the normal distribution which can affect intervals bands as those are calculated based on the normality assumption. As an alternative, parametric bootstrapping approach with an optimal confidence level which can be made by bootstrapping calibration can be employed.

References

[1] M.Hettiarachchi et al.,"Urban Wetlands and Disaster resilience of Colombo, Sri Lanka," *Int. J. Disaster Resilience in the Built Environment*, vol.5, no.1, pp.79-89, 2014.

[2] M. Domroes, *The Agroclimatic of Sri Lanka*, Wiesbaden, 1974.

[3] M. A. Sharma and J. B. Singh, "Use of probability distribution in rainfall analysis," *J.New York Science*, vol. 3, no. 9, pp. 40-49, 2010.

[4] V. Sharda, and P. Das, "Modelling weekly rainfall data for crop planning in a sub-humid climate of India," *J. Agricultural Water Manage.*, vol. 76, no. 2, pp. 120-138, 2005.

[5] P. Mishra et al.,"Satistical and probability analysis of rainfall for crop planning in a canal command,"*J.Agriculture for sustainable Develop.*, vol.1, no.1, pp.95-102, 2013.

[6] N. O. S. Alghazali and D. A. H. Alawadi, "Fitting statistical distributions of monthly rainfall for some Iraqi stations," *Civil and Environmental Research*, vol. 6, no. 6, pp. 40-46, 2014.

[7] T. Mayooran and A. Laheetharan, "The Statistical Distribution of Annual Maximum Rainfall in Colombo District," *Sri Lankan J.Applied Statistics*, vol. 15, no. 2, 2014.

[8] S. Ghosh et al., "Determination of the Best Fit Probability Distribution for Monthly Rainfall Data in Bangladesh," *American J. Mathematics and Statistics*, vol. 6, no. 4, pp. 170-174, 2016.

[9] B.Singh et al., "Probability Analysis for Estimation of Annual one day maximum Rainfall of Jhalarapatan area of Rajasthan, India, *Plant Archives*, vol.12, no.2, pp. 1093-1100, 2012.

[10] N.Varathan, K.Perera, and N.Wikramanayake, "Statistical modeling of daily extreame Rainfall in Colombo", *Int. Conf. Sustainable Built Environment*, 2010.

[11] K. P. Waidyarathne et al., "Shift in onset of first inter monsoon rain in coconut growing areas in Sri Lanka," *J. Tropical Agricultural Research*, vol. 18, pp. 1-12, 2006.

[12] P. K. Dunn, "Bootstrap confidence intervals for predicted rainfall quantiles," *Int. J. Climatology*, vol. 21, no. 1, pp. 89-94, 2001.

[13] S. Chen et al., "Constructing confidence intervals of extreme rainfall quantiles using Bayesian, bootstrap, and profile likelihood approaches," *Sci. China Technological Sci.*, vol. 59, no. 4, pp. 573-585, 2016.

[14] J. S. Park, H. S. Jung, R. S. Kim, and J. H. Oh, "Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution," *Int. J. Climatology*, vol. 21, no. 11, pp. 1371-1384, 2001.

[15] V.Jegarascsingam,"*Sri Lanka Country Report*," 1998.

MODELING OF WEEKLY RAINFALL USING CONFIDENCE INTERVAL APPROACH: A CASE STUDY

Silva H.P.T.N.[1]* and Peiris T.S.G.[2]

[1]University of Sri Jayewardenepura, Sri Lanka,
[2]University of Moratuwa, Sri Lanka
thanuja@sjp.ac.lk

The degree of uncertainty of atmospheric behavior has been increased from time to time. Rainfall is one of the key climatic variable for surviving the diverse set of human and natural systems in the world. Awareness about the pattern of rainfall is essential to mitigate effects derived from climate change which cause to sustainable development of the country. Modeling rainfall percentile is one of the successful technique that can be used to describe the rainfall characteristics and its behavior. The main goal of this study is to model weekly rainfall percentile in the context of confidence intervals by developing probability distribution functions. Daily rainfall data from 1960 to 2015 during the period of Second Inter Monsoon (October to November) in Colombo City were used. Preliminary analysis found that there was no trend in weekly series. Based on the best fitted probability distributions reliable rainfall percentiles and corresponding 95% confidence bands were computed. Three parameter Weibull distribution has been found most probable for many weeks in considered time span while the rest were well fitted with the two parameter Exponential and Largest Extreme Value distributions. Based on the analysis, the beginning of the Second Inter Monsoon showed low shower with a consistent pattern. Also, a similar pattern was identified with the withdrawal of the monsoon. However, it is noted that the Weeks 41-45 (08th October to 11th November) marked heavy rainfall with high variability result which caused high possibility to form extreme rainfall events. Out of the above weeks, the Week 42 (15th -21st October) has a much higher chance to occur extreme rainfall events during this monsoon period. A similar approach was carried out for weekly running totals during Second Inter Monsoon and found consistent result. This information would be very useful for various stakeholders to plan many activities which influence the intensity of rainfall.

***Keywords**: Weekly Rainfall, Percentile, Confidence Intervals, Colombo, Distribution*

Acknowledgement: Financial assistance by the University of Sri Jayewardenepura, Sri Lanka, Research Grant (ASP/01/RE/HSS/2016/75).

Modeling Weekly Rainfall: Problems Encountered

H P T N Silva[1] and T S G Peiris[2]

1. Department of Social Statistics,
University of Sri Jayewardenepura, Sri Lanka

2. Department of Mathematics, University of Moratuwa, Sri Lanka

(**Email:** thanuja@sjp.ac.lk)

Abstract

Rainfall is the main source of the hydrological cycle and able to get more practical benefits by forecasting. However, modeling rainfall is a challenging task due to increase in the degree of uncertainty of atmospheric behavior from time to time. Relatively, few efforts have been done in modeling weekly rainfall in hydrologic time series. This study mainly focuses on the difficulties that arise in modeling weekly rainfall. The rainfall data of the commercial capital of Sri Lanka during the time span from 1990 to 2015 were employed for this analysis. By studying the various properties of weekly rainfall three types of models: (1) Seasonal autoregressive integrated moving average (SARIMA) (2) Generalized autoregressive conditional heteroscedasticity (GARCH) for deseasonalized data (3) Hybrid SARIMA-GARCH were identified as the most suitable models to forecast weekly rainfall. However, each model has statistical drawbacks which needs to pay attention of the applied statistician.

Keywords: GARCH, Heteroscedasticity, SARIMA, Weekly Rainfall

International Journal of Ecological Economics and Statistics;
Volume. 39, Issue No. 3; Year 2018; Int. J. Ecol. Econ. Stat.;
ISSN 0973-1385 (Print), ISSN 0973-7537 (Online)

Accurate Confidence Intervals for Weibull Percentiles Using Bootstrap Calibration: A Case Study of Weekly Rainfall in Sri Lanka

H. P. T. N. Silva[1] and **T. S. G. Peiris**[2]

[1]Department of Social Statistics, Faculty of Humanities and Social Science,
University of Sri Jayewardenepura, Sri Lanka
Email: thanuja@sjp.ac.lk

[2]Department of Mathematics, Faculty of Engineering,
University of Moratuwa, Sri Lanka
Email: sarathp@uom.lk

ABSTRACT

Modeling rainfall percentiles in the context of the confidence interval is an appropriate technique that can be employed to make inferences about the rainfall characteristic. The coverage probability of the confidence interval is one of the imperative factor that should be considered when making inferences. Accurate confidence bands enhance the degree of the awareness level of rainfall variability at high uncertainty. The main aim of this study is to find the accurate level of confidence intervals for weekly rainfall percentiles derived from Weibull distributions based on the real coverage probabilities which are formed using bootstrap calibration. Weekly rainfall data from 1970 to 2015 in the Colombo city were used for this analysis. A simulation was carried out based on the one weekly series (week 24; 11-17 June) using the bootstrapping approach. It was found that the data series pertaining to the week 24 is well fitted with the two parameter Weibull distribution. Furthermore, the result reveals that the real coverage probabilities of 95% confidence intervals of 50^{th}, 60^{th}, 70^{th}, 80^{th} and 90^{th} weekly rainfall percentiles which were derived using maximum likelihood estimators of Weibull distribution can be attained on average at the levels 95.901%, 97.501%, 97.603%, 97.680% and 97.910% respectively.

Key Words: Percentiles, Coverage Probability, Bootstrapping, Confidence Intervals, Weekly Rainfall

Mathematics Subject Classification: 62F25

Journal of Economic Literature (JEL) Classification : C15

1. INTRODUCTION

An accurate analysis of the pattern of rainfall is essential to make effective decisions by utilizing water resources in many fields. Modelling rainfall percentile is one of the successful technique that can be applied to describe the rainfall characteristics in any region. Information on rainfall percentiles would enhance the level of awareness of rainfall behaviour, which can be used to reduce the difficulties that exist due to changes of atmospheric behaviour. The complexity of the temporal pattern of the rainfall is high in the short range scales of weekly and hourly than monthly, seasonal and annual. However, it is more important to have prior information on weekly rainfall in the urban areas which those were engaged with the large number of activities related in many fields such as industrial, constructions, health, rain water harvesting etc.. Moreover, the occurrences of extreme rainfall events in high population density areas have had a significant negative impact on the lives of the people and the infrastructure of the city. Most of the urban cities including the city of Colombo in Sri Lanka were vulnerable to many water related issues derived from the erratic rainfall events caused by changes in rainfall patterns, urbanization and installation of complex infrastructure by Lo and Koralegedera (2015).

The rainfall of the country is strongly governed by the seasonal varying monsoon system. According to Domroes (1974), a seasonal monsoon system of the country can be mainly divided into four major periods; First Inter Monsoon (FIM) from March to April, South West Monsoon (SWM) from May to September, Second Inter Monsoon (SIM) from October to November and North East Monsoon (NEM) from December to February.

Many researchers have made attempts to describe the temporal behaviour of the rainfall based on the point estimates for rainfall percentiles which derived fitted probability distributions (Sharda and Das, 2005, Sharma and Singh, 2010, Mishra et al., 2013), However, it is more imperative and practical to form a range for percentile rather than express it by a single value at the high uncertainty of climatic behaviour. Confidence intervals can be used to make inferences not only for the rainfall quantity, but also to have an idea about the its variability at the particular level of uncertainty. Some of the researchers employed confidence intervals to describe the characteristics of the rainfall quantile (Dunn, 2002, Park et al., 2001, Silva and Peiris, 2017).

Accurate estimates, either point or intervals is essential for better planning. The coverage probability of confidence intervals is one of the essential aspects to be considered to make more accurate interval estimates. In such a situation, the sample size is the other main factor which influences accurate inferences. In fact, it is more complicated to compose inferences and make decisions based on the small sample size. Most of the time, estimates derived from the fitted theoretical probability distributions becomes inaccurate due to the small sample size. To overcome this problem, the bootstrapping technique can be used. Past studies have shown three parameter and two parameter Weibull distributions were well fitted to the rainfall data, especially on the weekly scales (Sharda and Das, 2005; Silva and Peiris, 2017). In view of the above, the main objective of this study is to find the accurate confidence interval levels for weekly rainfall percentiles formed from two parameter Weibull distributions based on the real coverage probability derived from bootstrap calibration.

2. MATERIAL AND METHODS

2.1 Study area and data description

The city Colombo is situated with latitudes 6^0 93' N and Longitude 79^0 86' E in Sri Lanka and is selected as the study site . The city Colombo is the commercial capital of Sri Lanka. Daily rainfall data from 1970 to 2015 during the period of SWM in Colombo city were used for this analysis. These data were obtained from the Colombo meteorology station which is the only station with available meteorology data of Colombo from the Department of Meteorology in Sri Lanka. The daily rainfall (mm) data has been converted into weekly rainfall as Week 1 corresponding to 1-7 January, Week 2, Week 3 and so on corresponding to 8-14 January, 15-21 January and so on. Based on the above classification, the weeks 18-39 (30th of April to 30th of September) belongs to the SWM.

To work out the coverage probability, one weekly data series should be considered during the SWM as the main data series. In this study, the data which belongs to the week 24 (11-17 June) in SWM (The weekly data of 46 years for the time span from 1970 to 2015) was considered as the main data set. Those data were fitted to many theoretical probability distributions and two tests; Anderson Darling and Kolmogorov -Smirnov were used to test the goodness of fit of the parametric distributions.

2.2. Weibull distribution

The Weibull distribution is widely used for climatic data analysis due to its properties. The probability density function (Pdf) of a weibully variable, X with scale parameter 'α' and shape parameter 'β' is given by equations (1).

$$f(x)=\frac{\beta}{\alpha^{\beta}}x^{\beta-1}\exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) \qquad X \geq 0,\ \alpha > 0,\ \beta > 0 \qquad (1)$$

where, 'α' is the scale parameter and 'β' is the shape parameter of the distribution.

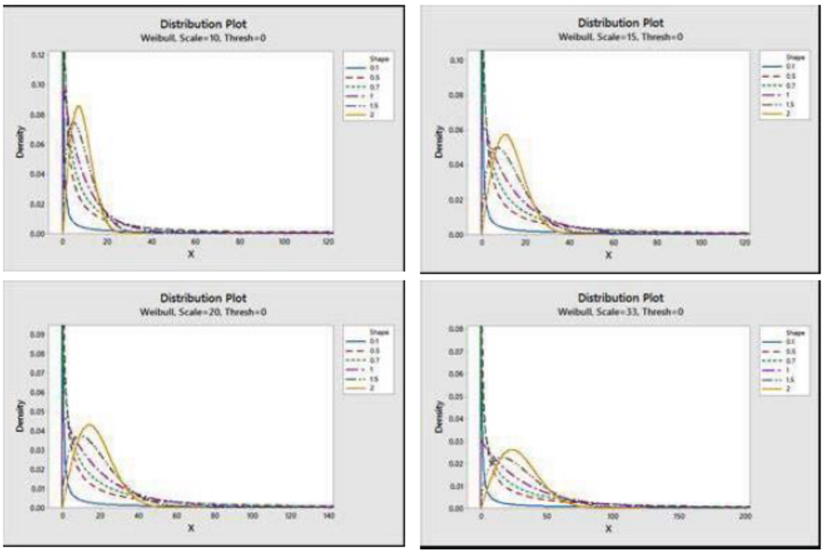

Figure 1: Probability density functions of Weibull distribution with different scale and shape parameters

The shape of the density function of the Weibull distribution changes drastically with the value of the shape parameter as shown in Figure 1.

2.3. (1-α) % Confidence intervals for the p^{th} percentile of Weibull distribution

The Weibull distribution can be approximated to the normal distribution when shape parameter is about 3.6 (Johnson and Kotz,1970). The p^{th} percentile of the Weibull distribution (Xp) and its variance are defined by equations (2) and (3) (Heo, et al., 2001).

$$\hat{X}_p = \hat{\alpha}\left[-\ln(1-p)\right]^{1/\hat{\beta}} \tag{2}$$

$$Var(\hat{X}_P) = \frac{\hat{X}_P^{\,2}}{\hat{\alpha}^2} Var(\hat{\alpha}) + \frac{\hat{X}_P^2}{\hat{\beta}^4} Z_P^2 Var(\hat{\beta}) - 2\frac{Z_P \hat{X}_P^2}{\hat{\alpha}\hat{\beta}^2} Cov(\hat{\alpha},\hat{\beta}) \tag{3}$$

where $Z_P = \ln[-\ln(1-p)]$ and $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimators for the α and β respectively. The equations that used to calculate the (1- α)% confidence intervals for the Weibull percentiles under the normal approximation is given by equation (4);

$$\left[\exp\left(\ln(\hat{X}_P) - Z_{\alpha/2}\,\frac{\sqrt{Var(\hat{X}_P)}}{\hat{X}_P}\right),\ \exp\left(\ln(\hat{X}_P) + Z_{\alpha/2}\,\frac{\sqrt{Var(\hat{X}_P)}}{\hat{X}_P}\right)\right] \tag{4}$$

2.4. Coverage probability

The coverage probability of a confidence interval can be briefly explained as the proportion of the time that interval contains the true value of interest. The coverage probability of a confidence interval can be calculated using simulation method. Firstly, number of samples of size n are simulated based on the main data series to compute the confidence intervals for interest parameter for each sample. After that, it should be computed for the proportion of samples for the known population parameters contained in the confidence intervals. That proportion is an estimate for the coverage probability for the confidence interval. However, a discrepancy can occur between the computed coverage probability and the nominal coverage probability. In this study, our interest parameter is percentile. The coverage probability will be calculated based on the confidence intervals of percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) .

2.5. Simulation

Assume that the major data series (The week 24) was well fitted with the two parameter Weibull distribution with scale parameter α and shape parameter β. Based on the size of the main data series (N=46), 2000 random samples (each sample size is also equal to 46) were generated using a bootstrapping approach called as Sample1, Sample2, Sample3, ... Sample2000 from the Weibull distribution (α, β). Furthermore, the scale and shape parameters of data sets pertaining to the Sample1 $(\hat{\alpha}_1, \hat{\beta})$, Sample2$(\hat{\alpha}_2, \hat{\beta}_2)$, Sample3$(\hat{\alpha}_3, \hat{\beta}_3)$and so on were estimated using maximum likelihood (MLE) method. Five percentiles (P_{50}, P_{60}, P_{70}, P_{80} and P_{90}) were calculated for each sample (Sample1 to Sample2000). It again generated 300 samples (Same sample size (n=46)) based on the generated Sample1, Sample2, Sample3 etc.. Those 300 samples which derived from the Sample1 could be indicated as $Sam1_1$, $Sam1_2$,...., $Sam1_{300}$. Here it describes only the coverage probability of randomly selected four samples (Sample68, Sample423, Sample802 and Sample1551). When considering the 300 samples generated based on the Sample1, firstly, the 50th percentiles and corresponding 95% confidence intervals were calculated of $Sam1_1$, $Sam1_2$, $Sam1_3$ and so on. The coverage probability was calculated based on the 300 confidence intervals (95%). The same procedure was carried out to calculate the coverage probability of confidence intervals at 95.2%, 95.4%, 95.6%, 95.8%, 96%, 96.2%, 96.4%, 96.6%, 96.8%, 97%, 97.2%, 97.4%, 97.6%, 97.8% and 98% confidence levels. Other samples which were generated from the Sample2,....., Sample2000 were applied the above procedure and the corresponding coverage probabilities for each confidence level listed above were calculated.

3. RESULTS

The summary statistic of the total rainfall during week 24 is presented in Table 1 along with the histogram (Figure 2).

Table 1: Descriptive Statistics of the weekly rainfall data (week 24)

Variable	No. of Data	Mean	StDev	Median	Mini	Max	Coefficient of Variance(%)	Skewness
Week 24	46	36.1	37.9	18.4	0.1	146.3	104.9	1.37

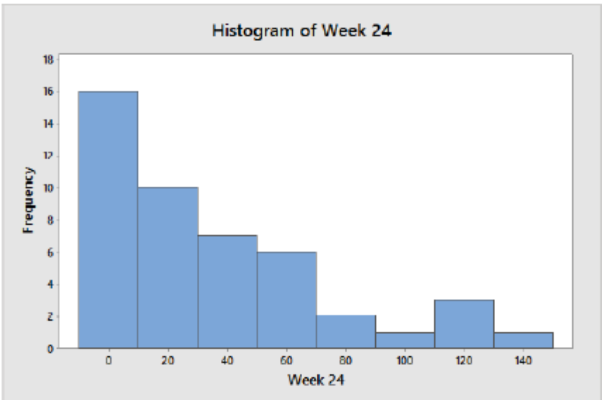

Figure 2: Histogram of weekly rainfall data (week 24)

From 1970 to 2015 total weekly rainfall in the 24^{th} week varied from 0.1mm to 146.3mm with a mean 36.1mm. Figure 2 illustrates the weekly rainfall as are positively skewed with a longer tail to the right and the result was further confirmed as the coefficient of skewness is 1.37. The large coefficient of variance (104.9%) gives evidence to high fluctuations in weekly rainfall (week24).

The total rainfall during week 24 was fitted to different type of probability distributions such as Normal, Exponential, Gamma, Weibull and so on and those data were well fitted with the two parameter Weibull distribution. Corresponding Anderson Darling and Kolmogorov -Smirnov test statistics were 0.258(P-value =0.256) and 0.0673 (P-value = 0.941) respectively. The maximum likelihood estimates for the scale and shape parameters of the fitted Weibull distribution were 33.9286 and 0.8775 respectively. The coverage probabilities of confidence intervals of Sample68 at different uncertainty levels are presented in Table 2. The corresponding coverage probabilities of the confidence limits of Sample423, Sample802 and Sample1551 are presented in Table 3, Table 4 and Table 5 respectively.

Table 2 : Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample68

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	93.00	93.00	93.00	92.00	92.67
95.2	93.33	93.33	93.00	92.67	93.00
95.4	93.33	93.67	93.00	93.00	93.00
95.6	93.67	93.67	93.00	93.00	93.67
95.8	94.00	93.67	93.00	93.00	93.67
96.0	94.33	93.67	93.67	93.33	94.00
96.2	**95.33**	94.33	94.00	94.00	94.67
96.4	95.67	**95.00**	94.00	94.67	**95.00**
96.6	96.33	95.33	**94.67**	94.67	95.00
96.8	96.33	95.67	**95.00**	**94.67**	95.00
97.0	96.67	96.00	95.00	**95.33**	95.00
97.2	96.67	96.33	95.33	96.00	95.00
97.4	96.67	96.33	95.67	96.00	95.33
97.6	96.67	96.33	96.00	96.00	95.33
97.8	96.67	96.33	96.00	96.33	95.67
98.0	96.67	96.33	96.67	96.33	96.33

According to the above table, it is clear that 95% coverage probability of P_{50} can be attained at 96.2 % confidence level. The 95% coverage probability of P_{60}, P_{70}, P_{80} and P_{90} can be reached at the confidence levels 96.4%, 96.6%, 96.8% and 96.4% respectively.

Table 3: Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 423

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	94.00	93.67	90.67	87.67	90.67
95.2	94.00	94.00	91.00	88.67	91.67
95.4	94.00	94.00	91.00	89.00	91.67
95.6	94.00	94.33	91.67	89.00	91.67
95.8	94.33	94.33	91.67	89.00	92.00
96.0	94.33	94.33	92.00	89.33	92.00
96.2	94.33	94.67	92.00	89.33	92.33
96.4	94.67	94.67	92.33	90.00	93.00
96.6	**95.00**	94.67	92.33	91.33	93.67
96.8	95.67	94.67	92.67	91.67	94.00
97.0	95.67	**95.00**	93.00	92.00	94.67
97.2	96.00	96.00	93.33	92.67	94.67
97.4	96.67	96.67	93.33	93.00	94.67
97.6	97.00	97.00	93.67	93.67	**95.00**
97.8	97.33	97.67	94.67	94.33	95.33
98.0	97.67	97.67	**95.00**	**95.00**	95.33

Based on the Table 3, the real 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} can be obtained at the 96.6%, 97.0%, 98.0%, 98.0%, and 97.6% respectively.

Table 4: Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 802

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	94.33	94.33	90.67	89.33	89.33
95.2	94.33	94.33	90.67	89.33	89.67
95.4	94.33	**94.67**	90.67	89.67	90.00
95.6	**95.33**	**95.33**	91.33	89.67	90.00
95.8	95.67	95.33	91.67	89.67	91.33
96.0	95.67	95.33	91.67	90.00	91.33
96.2	96.00	95.67	92.00	90.00	91.67
96.4	96.00	95.67	92.67	90.00	92.00
96.6	96.33	96.00	92.67	90.67	93.33
96.8	96.33	96.00	92.67	91.00	93.33
97.0	96.33	96.33	93.33	91.33	94.00
97.2	96.67	96.33	94.00	91.67	94.67
97.4	97.33	96.67	94.33	92.00	94.67
97.6	97.33	96.67	94.67	93.00	94.67
97.8	97.67	97.00	**95.00**	94.00	**95.00**
98.0	97.67	97.00	95.00	**95.00**	95.33

Table 4 illustrates the 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} obtained at the 95.6%, 95.4%, 97.8%, 98.0%, and 97.8% confidence levels respectively.

Table 5: Coverage Probabilities of five Percentiles (P_{50}-P_{90}) based on the 300 samples derived from the Sample 1551

Confidence Level (%)	Coverage Probability				
	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
95.0	92.67	92.33	92.00	92.00	89.67
95.2	93.00	92.33	92.00	92.33	89.67
95.4	93.33	92.67	92.00	92.67	89.67
95.6	93.33	92.67	93.33	92.67	92.33
95.8	**94.67**	93.00	93.33	93.00	92.33
96.0	95.33	93.00	93.33	93.00	92.33
96.2	96.00	93.00	93.67	93.67	93.33
96.4	96.00	93.33	93.67	93.67	93.33
96.6	96.00	93.33	94.00	94.00	93.67
96.8	96.00	93.67	94.00	94.33	93.67
97.0	96.33	93.67	94.67	94.67	94.67
97.2	96.33	93.67	94.67	94.67	94.67
97.4	96.67	94.67	**95.00**	94.67	94.67
97.6	96.67	94.67	95.00	**95.00**	**95.00**
97.8	96.67	**95.00**	95.33	95.00	95.33
98.0	97.67	95.00	95.33	95.33	95.33

Table 5 shows that the 95% coverage probability of P_{50}, P_{60}, P_{70}, P_{80} and P_{90} obtained at the 95.8%, 97.8%, 97.4%, 97.6%, and 97.6% confidence levels respectively.

The same procedure was applied for the remaining samples and the average accurate coverage probability were calculated of 300 samples derived from the each 2000 sample and result is presented in Table 6.

Table 6 : Average accurate confidence level based on the 95% confidence level for Weibull percentiles

Percentiles	P_{50}	P_{60}	P_{70}	P_{80}	P_{90}
Coverage probability	95.901	97.501	97.603	97.680	97.910

Based on the above real confidence levels, the calculated confidence bands of percentiles of week 24 are as follows.

Table 7: Confidence bands of percentiles of week 24 (nominal and actual)

Percentile	Value	Confidence limits at level of 95% (nominal)		Accurate confidence levels	Confidence limits (Actual)	
P_{50}	22.4	15.2	32.9	95.901	14.9	33.5
P_{60}	30.7	21.5	43.8	97.501	20.5	46.1
P_{70}	41.9	30.0	58.6	97.603	28.5	61.6
P_{80}	58.4	42.0	81.2	97.680	39.8	85.5
P_{90}	87.8	62.0	124.3	97.910	58.3	132.2

Based on the result formed from the simulation, there is a considerable difference between nominal and calculated coverage probabilities. Weibull distribution, drastically tends to be skewed to the right when the shape parameter less than one. Thus, the distribution of weekly rainfall deviates from the normal distribution with respect to the lower (less than one) value of shape parameter of the distribution. The deviation of the normality of the fitted distribution with the small size of sample is might be one of the reason for the discrepancy of the nominal and calculated coverage probabilities.

4. DISCUSSION AND CONCLUSION

Accurate estimates are important to draw reliable decisions which would be helped to minimize the issues that are caused due to heavy rainfall events and to utilize the water resources of the country. Weekly rainfall percentiles with 95% confidence bands are utilized in this study to compute the real coverage probability of the 95% confidence intervals. Rainfall total during the 24th week (11-17 June) in SWM was considered as the main data series with size as 46. The data set were well fitted with the two parameter Weibull distribution and the maximum likelihood estimates for the scale and shape parameters of the fitted Weibull distribution were 33.9286 and 0.8775 respectively.

Based on the simulation carried out by bootstrapping approach, it is found that the most of the real coverage probabilities of 95% confidence intervals of percentiles (50th, 60th, 70th, 80th and 90th) is less than 0.95. The accurate coverage probability of 95% confidence interval for the 50th percentile is attained at the average level of 95.901%. The corresponding accurate coverage probabilities of 95% confidence intervals of 60th,70th, 80th and 90th percentiles are given at the average levels of 97.501%, 97.603%, 97.680% and 97.910% respectively. Based on the above result, it can be concluded that the most of the coverage probabilities of the 95% confidence intervals for the rainfall percentiles get less value than the 0.95. This implies that the confidence interval of the percentile which derived from the skewed distribution as two parameter Weibull distribution at small sample size is not always give actual coverage probability. As a result of this, inferences make based on this facts do not provide much accurate estimates which require to get decisions at the high uncertainty. Thus, it is more suitable to consider much greater value for confidence level to get 0.95 coverage probability for percentiles of skewed distribution as two parameter Weibull distribution along with the small sample size.

ACKNOWLEDGEMENT

This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant (ASP/01/RE/HSS/2016/75).

REFERENCES

Dunn, P. K., 2001, Bootstrap confidence intervals for predicted rainfall quantiles. *Int.J. Climatol*, **21(1)**, 89-94.

Domroes, M., 1974, The Agroclimatic of Sri Lanka. Wiesbaden.

Heo, J.H., Salas, J.D. , Kim, K.D., 2001, Estimation of Confidence Intervals of Quantiles for the Weibull Distribution. *Stochastic Envirnmental Reserach and Risk Assesment*, **15**, 284-309.

Johnson, N.L., Kotz,S.,1970, Continous Univariate Distribution-1. houghton Mifflin Company, Boston.

Lo, K. F. A., Koralegedara, S. B., 2015, Effects of climate change on urban rainwater harvesting in Colombo city. *Sri Lanka. Journal of Environments,* **2(1),** 105-124.

Mayooran, T., Laheetharan, A., 2014, The Statistical Distribution of Annual Maximum Rainfall in Colombo District. *Sri Lankan Journal of Applied Statistics,* **15(2),** 107-130.

Mishra, P., Khare, D., Mondal, A., Kundu, S., Shukla, R., 2013, Satistical and probability analysis of rainfall for crop planning in a canal command. *Agriculture and Sustainable Development*, **1(1),** 95-102.

Sharda, V., Das, P., 2005, Modelling weekly rainfall data for crop planning in a sub-humid climate of India. *Agricultural water management*, **76 (2)**, 120-138.

Sharma, M. A., Singh, J. B., 2010, Use of probability distribution in rainfall analysis. *New York Science Journal*, **3 (9)**, 40-49.

Silva, H.P.T.N, Peiris, T.S.G., 2017, Analysis of Weekly Rainfall using Percentile Bootstap Approach. *International Journal of Ecology & Development*, **32(3)**, 97-100.

Silva, H.P.T.N, Peiris, T.S.G., 2017, Statistical Modeling of Weekly Rainfall: A case study in Colombo city in Sri Lanka, Proceedings of the 3rd International Multidisciplinary Engineering Research Conference; 2017 May 29-30; Moratuwa, Sri Lanka: IEEE.

Park, J. S., Jung, H. S., Kim, R. S., O.H, J. H., 2001, Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution. *International Journal of Climatology*, **21(11)**, 1371-1384.

Modeling Persistent and Periodic Weekly Rainfall in an Environment of an Emerging Sri Lankan Economy

H. P. T. N. Silva[1(✉)], G. S. Dissanayake[2], and T. S. G. Peiris[3]

[1] Department of Social Statistics, Faculty of Humanities and Social Science, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
thanuja@sjp.ac.lk
[2] University of Sydney, Sydney, Australia
[3] Department of Mathematics, Faculty of Engineering, University of Moratuwa, Moratuwa, Sri Lanka

Abstract. The quantity of rainfall and its related events have become more and more uncertain due to climatic variability. The complexity of the rainfall pattern increases due to the changes of the atmospheric behavior from time to time. Relatively, few measures have been taken to perform the modeling of rainfall in the context of long memory. This paper provides an assessment of such a phenomenon by fitting an appropriate time series model. A long range dependency model is proposed to fit weekly rainfall data to explore characteristics of persistence through an unbounded spectral density. Careful examination of the data exhibits periodic fluctuations as an additional feature. Since, the rainfall series exhibits periodic variations and persistence, a seasonal autoregressive fractionally integrated moving average (SARFIMA) model is fitted. Parameters of it are estimated using maximum likelihood estimation (MLE) method. A Monte Carlo simulation was carried out with different seasonal and non seasonal fractionally differing parameters to measure the suitability of the method for parameter estimation. Best fitted model is chosen based on the minimum of the mean absolute error and the forecasting performance are compared with the result of Seasonal autoregressive integrated moving average (SARIMA) using an independent sample as a creative contribution.

Keywords: Seasonality · Rainfall · Fractional differencing
Long-memory · Maximum likelihood estimators · Forecasting

1 Introduction

Sri Lanka is highly vulnerable to the impacts of climate change which includes severe droughts, heavy flash floods and landslides over the past years. These whether related disasters affects many sectors in the economy and threaten to sustainable development of the country. Temporal variability of rainfall in the

V. Kreinovich and S. Sriboonchitta (Eds.): TES 2019, SCI 808, pp. 314–328, 2019.
https://doi.org/10.1007/978-3-030-04263-9_24

country is changing frequently which affect on staple food production such as rice, coconut and tea (Baba [1], Fernando et al. [17], Wejerathne et al. [41]). Sri Lanka is on a path of rapid urbanization. Thus, many damages can occur due to changes in rainfall behavior in urban areas with a high population density and modern infrastructure. Lo and Koralegedera [25] claimed that the more cities including Colombo in Sri Lanka are at a risk of water related issues due to changes in rainfall patterns. Also, each year the government of Sri Lanka spends huge amount of money to reconstruct and renovate the infrastructures which damage caused by floods in the wet zone (Sri Lanka Rapid Post Disaster Needs Assessment [37]). Accurate information on rainfall predictions enhances the ability to utilize the water resource in a productive manner. Those details of temporal variability of rainfall is not only important for agricultural activities, but also for important subject domains in urban areas such as construction, industrial planning, urban traffic and sewer systems, health, tourism, rainwater harvesting and climate monitoring. Furthermore, the importance of analysis of weekly rainfall pattern is highlighted by Silva and Peiris [34] and they analyzed weekly rainfall in Sri Lanka using percentiles bootstrap approach. Moreover, the same authors (Silva and Peiris [35]) carried out another study to explain the behavior of the south west monsoon rainfall by utilizing the weekly rainfall percentiles along with the 95% confidence interval bands using best fitted distribution for weekly rainfall in Colombo city. However, they emphasized precise rainfall prediction is very difficult in the tropical country like Sri Lanka with the low technology. According to Luk [26] also, quantitative forecasting of rainfall is extremely difficult. Silva and Peiris [36] discussed the problems faced when analyzing rainfall amount which has heavy skewed distribution using Weibull confidence interval for rainfall percentiles. The main goal of this study is to suggest a best fit long-memory model to capture weekly rainfall behavior. SARFIMA model is utilized in such a context. The outline of this paper is as follows. Section 2 describes past works related to long memory models. The functional form of the SARFIMA model with the maximum likelihood estimation procedure is described by Sect. 3. It is followed by Sect. 4 that will present the Monte Carlo simulation results to assess the accuracy and reliability of the estimation procedure. Section 5 will provide results of weekly rainfall modeling. Finally, Sect. 6 provides some concluding remarks.

2 Literature Related to Long Memory Models

Time series models have been developed with an increasing degree of accuracy over the last few decades. The short memory autoregressive moving average (ARMA) model introduced by Box and Jenkins [5] has been extensively used for a variety of applications. Recently, time series models with long memory features became very popular among researchers in many fields such as statistics and econometrics. Features of a fractionally integrated autoregressive moving average (ARFIMA) long memory model was initially introduced by Granger and Joyeux [20] and Hosking [24]. It was an extension of the traditional ARMA process with a fractional differencing parameter. The hyperbolic decay of the autocorrelation function and an unbounded spectral density are two key features of the ARFIMA process. A number of estimation methods of the fractional differencing parameter were proposed by Porter-Hudak and Geweke [19], Fox and Taqqu [18], Dahlhaus [10], Sowell [38], Chen et al. [7] and Robinson [33]. Comparison study assessments were done by Cheung and Diebold [8] on maximum likelihood estimators for fractionally differenced parameters using two types of maximum likelihood (ML) estimators in the form of frequency-domain ML and exact domain ML of time series processes with an unknown mean. Small sample properties of four ML estimators of the ARFIMA model was investigated through a Monte Carlo simulation by Hauser [23]. A study done by Wang et al. [40] evaluated the ability of detecting existence of long-memory in time series using four methods: Lo's modified rescale adjusted range test, Geweke and Porter Hudak test and two other approximate maximum likelihood estimation methods. Some of the research done by Chan and Palma [6], Palma [28] and Beran et al. [3] carried out an assessment of ARFIMA model parameters and their properties. Dissanayake [11] introduced a rapid lag order detection mechanism of the standard long memory ARFIMA process. Due to the practical success of the ARFIMA model, a more generalized fractionally differenced long memory time series model called the Gegenbauer ARMA (GARMA) was probed in detail by Gray et al. [21]. Chung [9] extended the work in introducing a grid based parameter estimation procedure of an elementary GARMA process. Fresh interest in the econometric community infused into the process the introduction of a new class of models with heteroskedasticity in Dissanayake and Peiris [12]. It was followed by the casting of the process driven by Gaussian white noise in state space by Dissanayake et al. [13] to establish a parameter estimation based optimal lag order validated by predictive accuracy. A similar experiment in which the process was driven by Generalized Autoregressive Conditionally heteroskedastic (GARCH) errors (instead of Gaussian white noise) was presented in Dissanayake et al. [14] with the validation of parameter estimation based optimal lag order done through log likelihood measures. A concise summary of fractionally differenced Gegenbauer processes with long memory was provided in Dissanayake [15]. An extensive review of fractionally differenced Gegenbauer processes with long memory is found in Dissanayake et al. [16]. It refers to certain conceptual paradigms presented in the survey on long memory by Guegan [22] in which an extended k-factor Gegenbauer process becomes a highlight of rigour. Though the ARFIMA model was able to capture the long range dependency, it does not take into account the seasonal variation patterns present in some real data set. The seasonal autoregressive fractionally integrated moving average (SARFIMA) of Porter-Hudak [30] is a natural extension of the ARFIMA process with an additional seasonal filter. The model consists of long memory dependency features with periodic behavior in terms of the data. SARFIMA model was utilized for forecasting of the monthly IBM product revenue in Ray [32]. Peiris and Singh [29] suggested a convenient method to calculate predictors for seasonal and non seasonal fractional parameters of long memory models under certain conditions. The work done by Bisognin and Lopes [4] described number of properties of seasonally fractional ARMA process in detail. SARFIMA model was applied to forecast Iraqi oil production and model parameters were estimated using conditional sum of squares by Mostafaei and Sakhabakhsh [27]. Additionally, Reisen et al. [31] proposed a semi parametric approach to estimate two seasonal fractional parameters in a SARFIMA model and the performance was evaluated through a Monte Carlo experiment. Very few attempts have been made to study the rainfall behavior in context of long memory. A study done by Yaya and Fashae [42] made an attempt to fit SARFIMA models for rainfall data in six rainfall zones of Nigeria. However, they could not find significant SARFIMA models which can capture the seasonal behavior with the long range dependency of the real data. Utilizing a SARFIMA model to assess seasonal and persistent properties of weekly rainfall in an emerging Asian economy such as Sri Lanka is missing in the current literature. Theoretical concepts linked with the long memory SARFIMA model are provided in the next section.

3 SARFIMA Long Memory Model

Long range dependency features can be identified by two different approaches but equivalent forms given below defined in two distinct domains called time and frequency (Bary [2]). In time domain, the auto correlation function $\rho_X(.)$ of the time series decays hyperbolically to zero. The correlation function, $\rho_X(k) \approx k^{2d-1}$ when $k \to \infty$ and $0.0 < d < 0.5$. The frequency domain, spectral density function $f_X(.)$ is unbounded when the frequency is near zero, that is, $f_X(\omega) \approx \omega^{-2d}$ when $\omega \to 0$. A process $\{Y_t\}_{t\in Z}$ is a stationary stochastic process given by the formula,

$$\phi(B)\psi(B^S)\nabla^d\nabla_S^D(Y_t-\mu) = \theta(B)\Theta(B^S)\epsilon_t \tag{1}$$

where μ is the mean of the process, $\{\epsilon_t\}_{t\in Z}$ is a white noise process with zero mean and variance σ_ϵ^2. B is the backward shift operator such that $y_{t-n} = B^n y_t$ and s is the seasonal length. $\phi(B)$ and $\psi(B)$ are the non seasonal and seasonal autoregressive polynomials of order p and P respectively such that

$$\phi(B) = \sum_{i=1}^{p} \phi_i B^i \quad 1 \le i \le p \quad \psi(B) = \sum_{k=1}^{P} \psi_k B^k \quad 1 \le k \le P \tag{2}$$

$\theta(B)$ and $\Theta(B)$ are the non-seasonal and seasonal moving average polynomials of order q and Q respectively defined as

$$\theta(B) = \sum_{j=1}^{q} \theta_j B^j \quad 1 \le j \le q \quad \Theta(B) = \sum_{m=1}^{Q} \Theta_m B^m \quad 1 \le m \le Q \tag{3}$$

The differencing operator ∇^d can be expressed as,

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k}(-B)^k \tag{4}$$

where $\binom{d}{k} = \frac{\Gamma(1+d)}{\Gamma(1+k)\Gamma(1+d-k)}$.

The seasonal operator ∇_S^D can be expressed as,

$$\nabla_S^D = (1 - B^S)^D = \sum_{k=0}^{\infty} \binom{D}{k} (-B^S)^k \tag{5}$$

where $\binom{D}{k} = \frac{\Gamma(1+D)}{\Gamma(1+k)\Gamma(1+D-k)}$.

The model (1) is specified by $SARFIMA(p, d, q)x(P, D, Q)_S$. When $d = 0$ and $D = 0$, the model is reduced to a classical seasonal SARFIMA model. If conditions 0 < d < 0.5 and 0 < D < 0.5 are satisfied the process becomes stationary. The spectral density function of the SARFIMA model can be written as follows:

$$f_S(\lambda) = \frac{\sigma_\epsilon^2 |\theta_q e^{(-i\lambda)}|^2 |\Theta_Q e^{(-i\lambda)S}|^2}{2\pi |\phi_p e^{(-i\lambda)}|^2 |\psi_P e^{(-i\lambda)S}|^2} |1 - e^{-i\lambda}|^{-2d} |1 - e^{-i\lambda S}|^{-2D} \tag{6}$$

An unbounded spectral density around the origin assessed through the utilization of (6) coupled with the hyperbolic decay of autocorrelation and partial autocorrelation blended with seasonality features (as shown in Figs. 2, 3 and 4) prompted towards parameter assessment being done using the exact maximum likelihood estimation method. It was done by maximizing the log likelihood function numerically. The package arfima in R was used to calculate the maximum likelihood estimators. Durbin-Levinson and Trench algorithms were utilized to maximize the likelihood and obtain optimal simulation and forecasting results (Veenstra and McLeod [39]).

4 Results of Monte Carlo Simulation

In order to evaluate the performance of the maximum likelihood method in estimating the parameters of the model, a number of Monte Carlo experiments were carried out. The simulation results provided non-seasonally and seasonally differenced parameter estimations and the corresponding standard and mean square errors (MSE) of the parameters. It was carried out based on 1000 replications with different sizes of samples ($n = 100$, $n = 200$, $n = 500$ and $n = 1000$). Seasonal length was considered as 52 corresponding to weekly rainfall. Monte Carlo experiment was conducted on a simulated $SARFIMA(0, d, 0)x(0, D, 0)_{52}$ series with following parameter combinations.

$d = 0.1$ and $D = 0.45$, $d = 0.15$ and $D = 0.45$, $d = 0.3$ and $D = 0.3$, $d = 0.45$ and $D = 0.10$.

The simulation was carried out using the R programming Language (Version 3.4.2) utilizing a HP11 (8 GB, 64 bit) computer. The standard errors of the estimates $SD(\hat{d})$, $SD(\hat{D})$ and mean square error of the estimates $MSE(\hat{d})$, $MSE(\hat{D})$ respectively such that:

$$SD(\hat{d}) = \sqrt{\sum_{r=1}^{R} (\hat{d}_r - \hat{d})/R}, \qquad SD(\hat{D}) = \sqrt{\sum_{r=1}^{R} (\hat{D}_r - \hat{D})/R},$$

$$MSE(\hat{d}) = \sum_{r=1}^{R} (\hat{d}_r - \hat{d})^2/R, \qquad MSE(\hat{D}) = \sum_{r=1}^{R} (\hat{D}_r - \hat{D})^2/R$$

Where $\hat{d}_r$ and $\hat{D}_r$ are the MLE of d and D for the rth replication. The value R denotes the number of replications ($R = 1000$ for the simulation).

Tables 1, 2 3 and 4 present the average of the estimated d, corresponding standard error and MSE of the estimator.

Table 1. MLE of d and D of a generating process of $SARFIMA(0, d, 0)x(0, D, 0)_{52}$ with $d = 0.1$ and $D = 0.45$. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	$SD(\hat{d})$	$MSE(\hat{d})$	$\hat{D}$	$SD(\hat{D})$	$MSE(\hat{D})$
100	0.0333	0.0826	0.0112	0.4423	0.0160	0.0003
200	0.0674	0.0590	0.0045	0.4462	0.0112	0.0001
500	0.0860	0.0359	0.0014	0.4475	0.0091	0.00008
1000	0.0927	0.0297	0.0009	0.4503	0.0118	0.0001

Table 2. MLE of d and D of a generating process of $SARFIMA(0, d, 0)x(0, D, 0)_{52}$ with $d = 0.15$ and $D = 0.45$. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	$SD(\hat{d})$	$MSE(\hat{d})$	$\hat{D}$	$SD(\hat{D})$	$MSE(\hat{D})$
100	0.0671	0.0835	0.0138	0.4502	0.0136	0.0001
200	0.1033	0.0570	0.0054	0.4539	0.0087	0.00009
500	0.1358	0.0357	0.0014	0.4530	0.0076	0.00006
1000	0.1429	0.0260	0.0007	0.4516	0.0070	0.00005

It can be seen from Tables 1, 2, 3 and 4 a reasonable assessment of the maximum likelihood estimator for the seasonal as well as non-seasonal fractional differencing parameters. It is noticeable that the parameter bias has decreased as with the increase of the series length. Also, it is evident from the parameters in Tables 1, 2, 3 and 4 that they become consistent with the increase in series length. Standard deviation and the MSE of estimators decrease with the increase in series length as expected.

Table 3. MLE of d and D of a generating process of $SARFIMA(0,d,0)x(0,D,0)_{52}$ with $d = 0.3$ and $D = 0.3$. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	$SD(\hat{d})$	$MSE(\hat{d})$	$\hat{D}$	$SD(\hat{D})$	$MSE(\hat{D})$
100	0.2236	0.0831	0.0127	0.2738	0.0690	0.0127
200	0.2607	0.0586	0.0049	0.2890	0.0386	0.0016
500	0.2846	0.0360	0.0015	0.2947	0.0260	0.0007
1000	0.2907	0.0262	0.0007	0.2978	0.0190	0.0003

Table 4. MLE of d and D of a generating process of $SARFIMA(0,d,0)x(0,D,0)_{52}$ with $d = 0.45$ and $D = 0.1$. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	$SD(\hat{d})$	$MSE(\hat{d})$	$\hat{D}$	$SD(\hat{D})$	$MSE(\hat{D})$
100	0.3697	0.0733	0.0118	0.0205	0.1414	0.0263
200	0.4015	0.0493	0.0047	0.0723	0.0667	0.0052
500	0.4283	0.0312	0.0014	0.0902	0.0365	0.0014
1000	0.4391	0.0244	0.0007	0.0936	0.0268	0.0007

5 Application for Real Data

5.1 Description of Dataset

Colombo city is the commercial capital of Sri Lanka, situated with latitudes 6 55 N and Longitude 79 51 E and is chosen as the study site. Daily rainfall data of Colombo were collected from 1990 to 2015 from the Department of Meteorology, Sri Lanka for this study. The daily rainfall (mm) data has been converted into weekly rainfall by dividing a year into 52 weeks such that week 1 corresponds to 1–7 January, Week 2 corresponds to 8–14 January and so on. The data during the time span from 1990 to 2014 was used as to build the models while the rest was used for the model validation.

5.2 Model Development

To examine the temporal variability of the rainfall series, time series plots was taken and it is presented in Fig. 1.

Random behavior of the rainfall pattern can be clearly observed in Fig. 1. However, it cannot be identified as a decreasing or increasing trend in weekly rainfall within the considered time span. In order to find out the seasonal behavior of the data, autocorrelation analysis was carried out and the result is presented in Figs. 2 and 3. A seasonal behavior can be clearly recognized from this plot. Since the data was captured on a weekly basis, identifying the seasonal length, ACF and PACF were done with 52 lag difference.

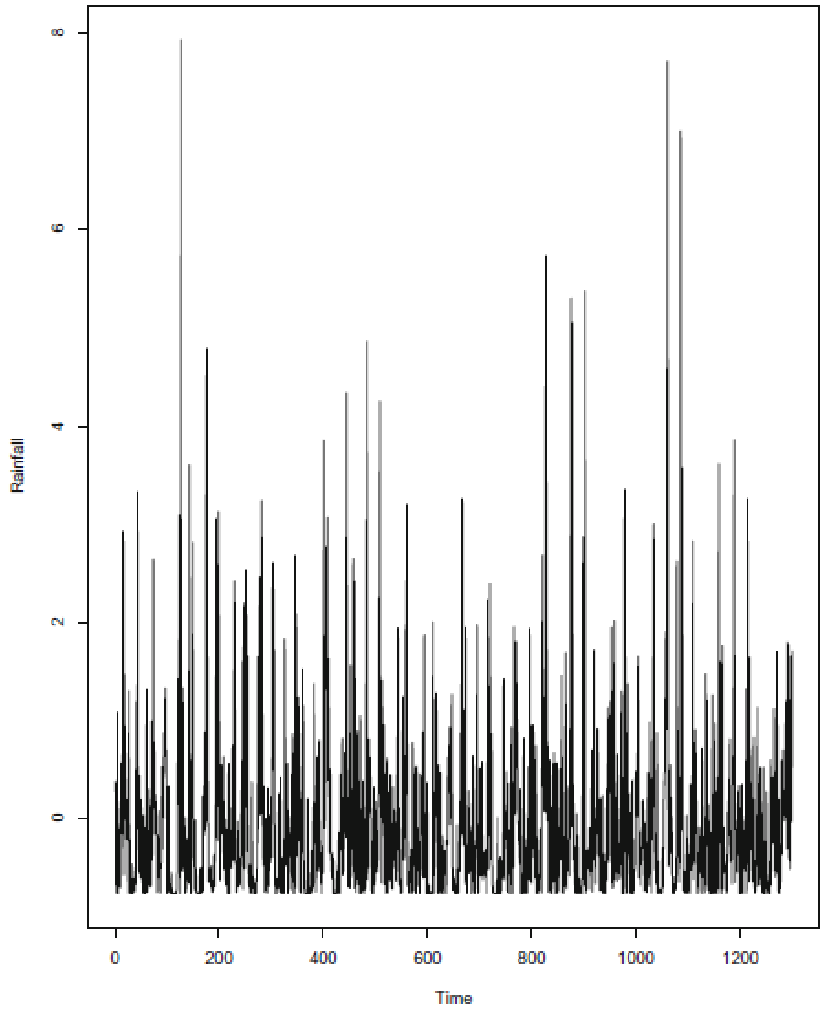

Fig. 1. Time series plot of weekly rainfall series from 1990 to 2014

Based on the ACF and PACF it can be clearly identified that the seasonal length is 52 since significant sample autocorrelation existed in the 52[nd] Lag. Figure 5 illustrate the sample spectrum which has the peak at frequency very closer to zero. The corresponding frequency gives a maximum spectrum density of 0.0385185. This value is not far from zero. (0.0385185 * 100/0.5 = 0.0770 = 7.7%). Thus, we can conclude that the SARFIMA series is suitable for this data set. The long term serial correlation in the data are accounted for in long memory modeling. It is very imperative to consider the long memory features to capture the real dynamics of rainfall. Thus, several SARFIMA models were fitted to the data with the size of the sample being 1300. Those fitted were utilized to predict the weekly rainfall over the year 2015. The best fitted model is

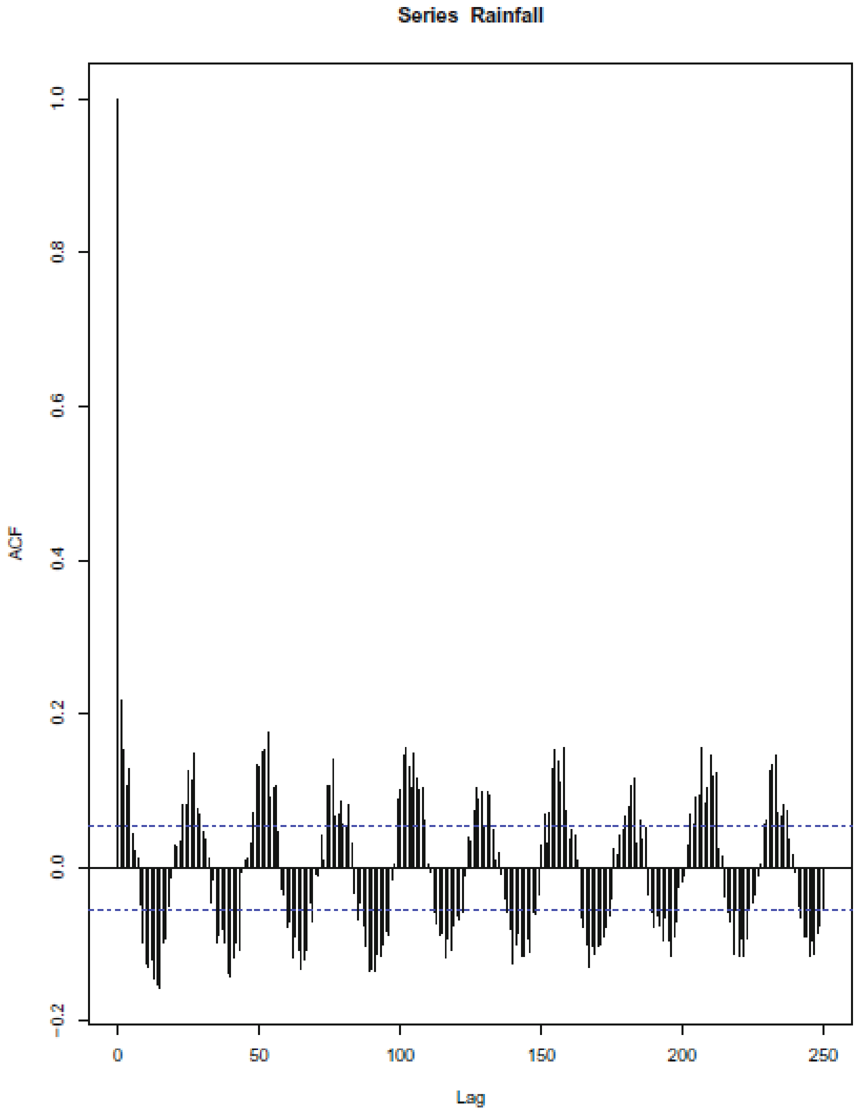

Fig. 2. Autocorrelation plot of the series from 1990 to 2014

selected with minimum mean absolute error (MAE). The MAE can be written as;

$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i|$.

Where e_i is the forecasting error and n is the length of the forecasting series. The corresponding result of the fitted long memory model is as follows.
A model $SARFIMA\ (1, 0.116, 1) \times (1, 0.171, 0)_{52}$ was found to be the best fitted model for the weekly rainfall series. The corresponding parameter estimates with standard errors are presented in Table 5.

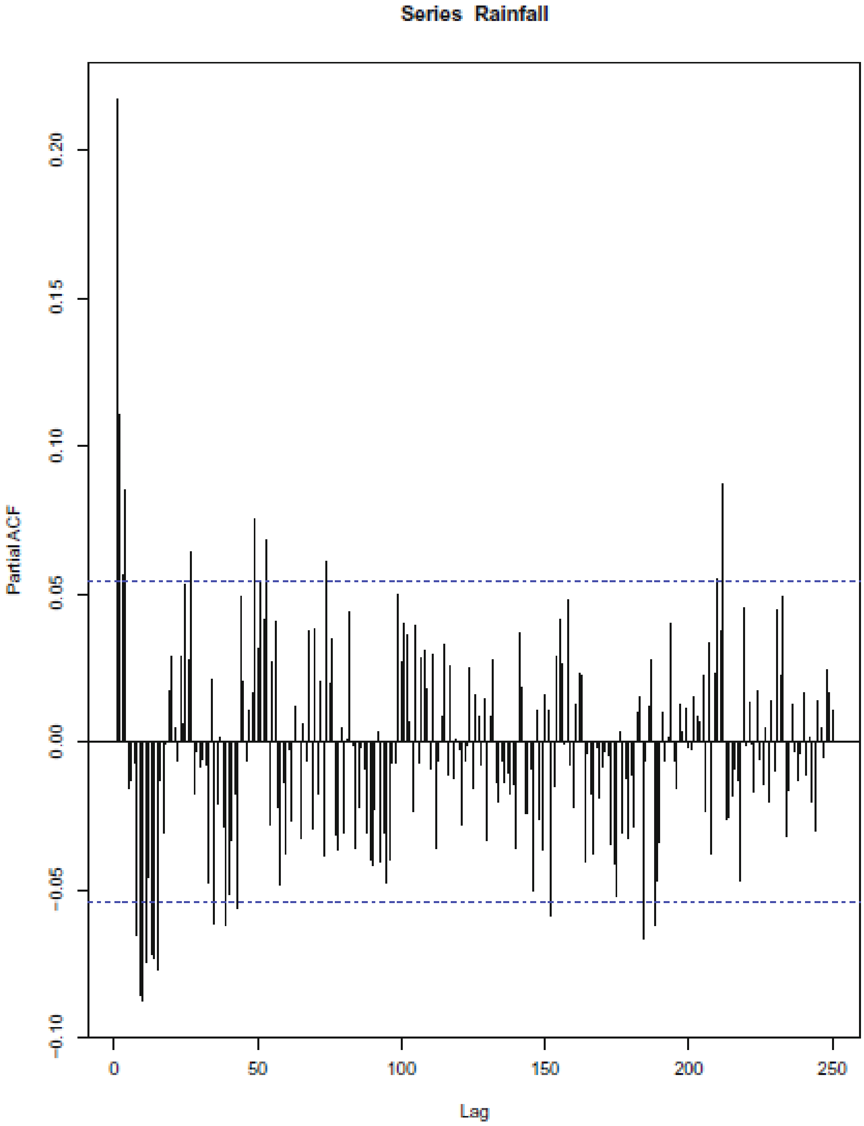

Fig. 3. Partial autocorrelation plot of the series from 1990 to 2014

Table 5. Fitted model for the weekly rainfall series $SARFIMA(p, d, q)x(P, D, Q)_S$ with $p = 1$, $q = 1$, $d = 0.1156735$, $P = 1$, $D = 0.1707546$, $Q = 0$ and $S = 52$

Coefficients	ϕ_1	θ_1	ψ_1	Constant	d	D
Estimate	−0.9113	−0.9018	−0.0860	0.0041	0.1156	0.1707
Standard Error	0.1427	0.1498	0.0394	0.1021	0.0269	0.0291
Z-value	−6.3854	−6.0192	−2.1796	0.0401	4.2900	5.8644
Pr(>|Z|)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

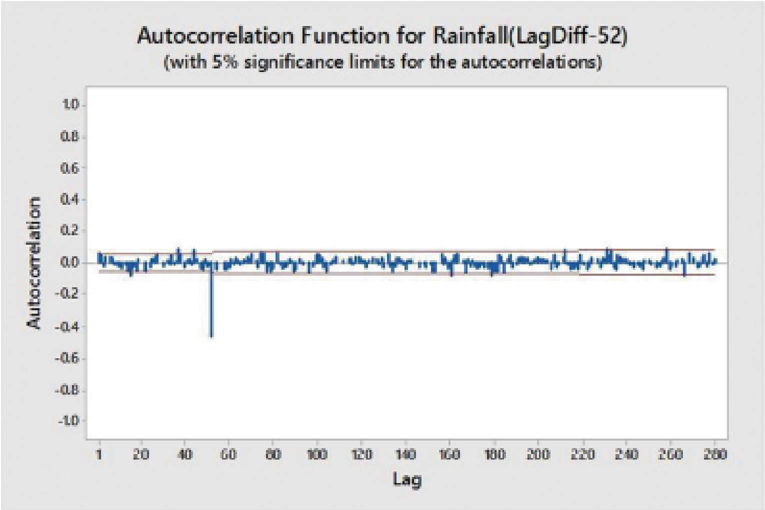

Fig. 4. ACF of the series from 1990 to 2014 with 52 Lag

All model parameters are significant at the 0.05 level of significance. The residual analysis of the fitted model was performed and found the uncorrelated at a 5% level of significance. Furthermore, the model was tested for weekly rainfall data in 2015 and the result is presented in Table 6. Various SARIMA models also fitted to the same dataset for the purpose of the comparison with the long memory model SARFIMA. A model $SARIMA(1,0,0)x(1,0,1)_{52}$ was found to be the best fitted model for the weekly rainfall series.

Table 6. Absolute Forecasting Error in mm for independent sample 2015

Absolute Forecasting Error in mm	SARIMA weekly percentage	SARFIMA weekly percentage
0–10	7(13.5)	12(23.1)
11–15	5(9.6)	4(7.7)
16–20	3(5.8)	4(7.7)
21–25	4(7.7)	5(9.6)
26–30	2(3.8)	6(11.5)
31–35	7(13.5)	4(7.7)
36–40	8(15.4)	3(5.8)
41–45	4(7.7)	1(1.9)
46–50	1(1.9)	3(5.8)
More than 50	11(21.1)	10(19.2)

Based on the above forecasted result, the model SARFIMA is outperform SARIMA. It can be seen that the more than 30% of the weeks' forecasting error

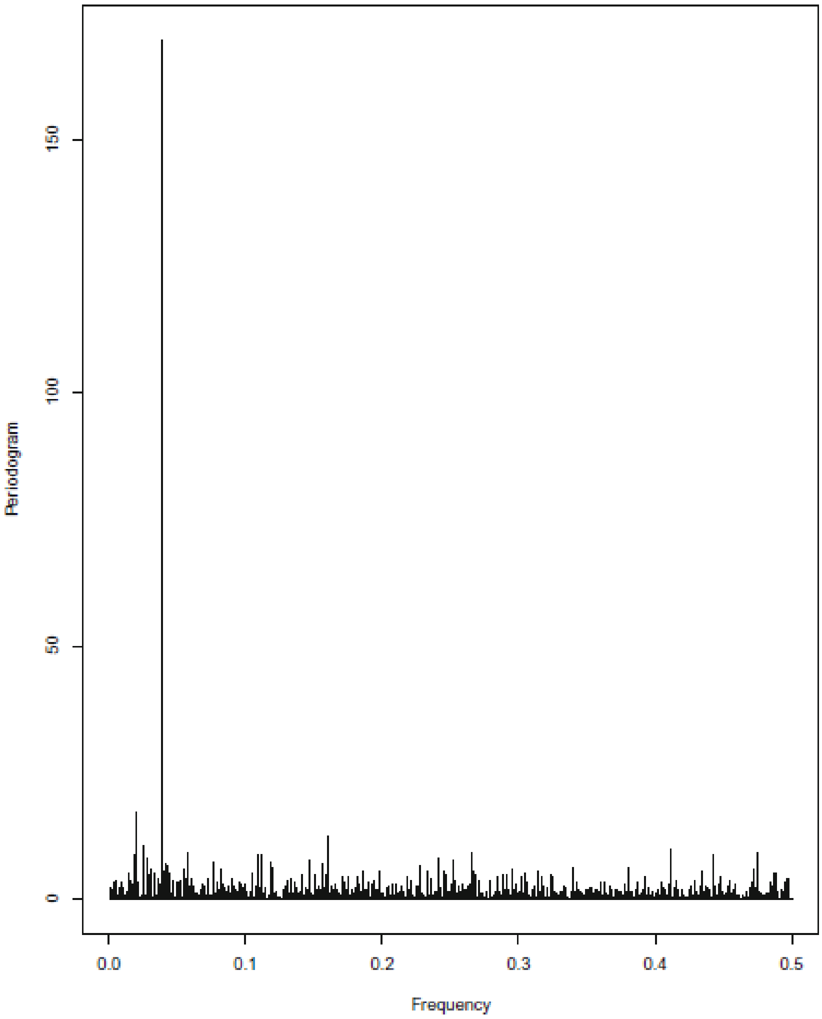

Fig. 5. The periodogram of the rainfall series from 1990 to 2014

less than to 15 mm indicated that a good agreement of the real and forecasted which derived from the SARFIMA. Then the novel model can be considered as best fitted for the weekly rainfall series. However, still there is a considerable number of weeks' forecasting error more than to 50 mm which need further improvement to this model.

6 Conclusion

It is evident from the results of this paper that long range dependency characteristics could couple with periodic variation in a weekly rainfall series. SARFIMA model can be considered to capture both long memory and seasonality. The

Monte Carlo simulation provides evidence towards optimal accuracy of parameter estimation. Furthermore, accuracy of the estimators improved with increasing of the series length. The effectiveness of the SARFIMA model was represented by using real datasets in the form of weekly rainfall from 1990 to 2014 (1300 size of the sample). $SARFIMA\ (1, 0.116, 1) \times (1, 0.171, 0)_{52}$ model was found to be the best model to forecast weekly rainfall in Colombo city. Thereafter, model was used to make independent sample long-range seasonal predictions of the weekly rainfall for the year 2015. Those result was compared with the $SARIMA(1, 0, 1)x(1, 0, 1)_{52}$ and found that the SARFIMA is superior than the SARIMA based on the predicted performance which has done for the independent data set.

Acknowledgement. This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant ASP/01/RE/HSS/2016/75).

References

1. Baba, N.: Sinking the pearl of the Indian ocean: climate change in Sri Lanka. Glob. Majority E-J. **1**(1), 4–16 (2010)
2. Bary, N.K.: A Treatise on Trigonometric Series. Pergamon Press (1964)
3. Beran, J., Feng, Y., Ghosh, S., Kulik, R.: Long memory Processes Probabilistic Properties and Statistic properties and Statistical Methods. Springer, Berlin/Heidelberg (2013)
4. Bisognin, C., Lopes, S.R.C.: Properties of Seasonal long memory processes. Math. Comput. Model. **49**, 1837–1851 (2009)
5. Box, G.E.P., Jenkins, G.M.: Time Series Analysis Forecasting and Control, Rev edn. Holden-Day, Merrifield (1976)
6. Chan, N.H., Palma, W.: Estimation of long-memory time series models. Adv. Econom. **20**, 89–121 (2006)
7. Chen, G., Abraham, B., Peiris, S.: Lag window estimation of the degree of differencing in fractionally integrated time series models. J. Time Series Anal. **15**(5), 473–487 (1994)
8. Cheung, Y., Diebold, F.X.: On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. J. Econometrics **62**, 301–316 (1994)
9. Chung, C.F.: A generalized fractionally integrated autoregressive moving-average process. J. Time Series Anal. **17**(2), 111–140 (1996)
10. Dahlhaus, R.: Efficient parameter estimation for self-similar processes. Ann. Stat. **17**(4), 1749–1766 (1989)
11. Dissanayake, G.S.: Rapid Optimal Lag Order Detection and Parameter Estimation of Standard Long Memory Time Series. Causal Inference in Econometrics, pp. 17–28. Springer, Cham (2016a)
12. Dissanayake, G.S., Peiris, S.: Generalized fractional processes with conditional heteroskedasticity. Sri Lankan J. Appl. Stat. **12**(1), 1–12 (2012)
13. Dissanayake, G.S., Peiris, M.S., Proietti, T.: State space modeling of Gegenbauer processes with long memory. Comput. Stat. Data Anal. **100**, 115–130 (2016)

14. Dissanayake, G.S., Peiris, M.S., Proietti, T., Ruiz, I.R. Garcia, G.R.: Estimation of generalized fractionally differenced processes with conditionally heteroskedastic errors. In: International Work Conference on Time Series. Proceedings ITISE, pp. 871–890 (2014)
15. Dissanayake, G.S.: Advancement of fractionally differenced Gegenbauer processes with long memory. Bull. Aust. Math. Soc. **94**(1), 173–174 (2016b)
16. Dissanayake, G.S., Peiris, M.S., Proietti, T.: Fractionally Differenced Gegenbauer Processes with Long Memory: A Review. Statistical Science (2018, to appear)
17. Fernando, M.T.N., Zubair, L., Peiris, T.S.G., Ranasinghe, C.S. Ratnasiri, J.: Economic Value of climate variability impacts on coconut production in Sri Lanka. AIACC working paper No. 45 (2007)
18. Fox, R., Taqqu, M.S.: Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Statist. **14**(2), 517–532 (1986)
19. Geweke, J., Hudak, S.P.: The estimation and application of long memory time series models. J. Time Series Anal. **4**, 221–238 (1983)
20. Granger, C.W.J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. J. Time Series Anal. **1**, 15–29 (1980)
21. Gray, H.L., Zhang, N., Woodward, W.A.: On generalized fractional process. J. Time Series Anal. **10**(3), 233–257 (1989)
22. Guégan, D.: How can we define the concept of long memory? An econometric survey. Econometric reviews **24**(2), 113–149 (2005)
23. Hauser, M.A.: Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo Study. J. Stat. Planning Infer. **80**, 229–255 (1999)
24. Hosking, J.R.M.: Fractional differencing. Biometrika **68**, 165–176 (1981)
25. Lo, K.F.A., Koralegedara, S.B.: Effects of climate change on urban rainwater harvesting in Colombo city, Sri Lanka. J. Environ. **2**(1), 105–124 (2015)
26. Luk, K.C., Ball, J.E., Sharma, A.: An application of artificial neural networks for rainfall forecasting. Math. Comput. Model. **33**(6), 683–693 (2001)
27. Mostafaei, H., Sakhabakhsh, L.: Using SARFIMA and SARIMA models to study and predict the Iraqi Oil production. J. Basic Appl. Sci. Res. **1**(10), 1715–1723 (2011)
28. Palma, W.: Long Memory Time Series Theory and Methods. Wiley, New Jersey (2007)
29. Peiris, M.S., Singh, N.: Predictors for seasonal and nonseasonal fractionally integrated ARIMA models. J. Biom. **38**(6), 741–752 (1996)
30. Porter-Hudak, S.: An application of the seasonal fractionally differenced model to the monetary aggregates. J. Am. Stat. Assoc. **85**(410), 338–344 (1990)
31. Reisen, V.A., Zamprogno, B., Palma, W., Arteche, J.: A semi parametric approach to estimate two seasonal fractional parameters in the SARFIMA model. Math. Comput. Simul. **98**, 1–17 (2014)
32. Ray, B.K.: Long range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. Int. J. Forecast. **9**, 255–269 (1993)
33. Robinson, P.M.: Gaussian semi parametric estimation of long range dependence. Ann. Stat. **23**, 1630–1661 (1995)
34. Silva, H.P.T.N., Peiris, T.S.G.: Analysis of weekly rainfall using percentile bootstrap approach. Int. J. Ecol. Dev. **32**(3), 97–106 (2017)
35. Silva, H.P.T.N., Peiris, T.S.G.: Statistical Modeling of weekly rainfall: a case study in Colombo city in Sri Lanka. Proceedings of the 3rd Moratuwa Engineering Research Conference (MER Con), 29–31 May, Sri Lanka, pp. 241–246. IEEE (2017)

36. Silva, H.P.T.N., Peiris, T.S.G.: Accurate confidence intervals for Weibull percentiles using bootstrap calibration: a case study of weekly rainfall in Sri Lanka. Int. J. Ecol. Econ. Stat. **39**(3), 67–76 (2018)
37. Sri Lanka Rapid post disaster needs Assessment: Floods and Landslides (2017). http://www.lk.undp.org/content/srilanka/en/home/library/environment_energy/SriLanka-Rapid-Post-Disaster-Needs-Assessment-2017.html
38. Sowell, F.: Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J. Econometrics **53**, 165–188 (1992)
39. Veenstra, J.Q., McLeod, A.I.: Persistence and Anti-persistence: Theory and Software (Ph.D. Thesis) (2013)
40. Wang, W., Gelder, P.H.A.J.M.V., Vriling, J.K., Chen, X.: Detecting long-memory: Monte Carlo Simulations and application to daily stream flow processes. Hydrol. Earth Syst. Sci. **11**, 851–862 (2007)
41. Wijeratne, M.A., Anandacoomaraswamy, A., Amarathunga, M.K.S.L.D., Ratnasiri, J., Basnayake, B.R.S.B., Kalra, N.: Assessment of impact of climate change on productivity of tea (Camellia Sinensis L) plantations in Sri Lanka. J. Natn. Sci. Foundation Sri Lanka **35**(2), 119–126 (2007)
42. Yaya, O.S., Fashae, O.A.: Seasonal fractional integrated time series models for rainfall data in Nigeria. J. Theoret. Appl. Climatol. **120**(2), 99–108 (2014)

The Use of Fractionally Autoregressive Integrated Moving Average for the Rainfall Forecasting

H. P. T. N. Silva[1(✉)], G. S. Dissanayake[2], and T. S. G. Peiris[3]

[1] Department of Social Statistics, Faculty of Humanities and Social Science, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
thanuja@sjp.ac.lk
[2] University of Sydney, Sydney, Australia
[3] Department of Mathematics, Faculty of Engineering, University of Moratuwa, Moratuwa, Sri Lanka

Abstract. A study of rainfall pattern and its variability in South Asian countries is vital as those regions are frequently vulnerable to climate change. Models for rainfall have been developed with different degrees of accuracy, since this key climatic variable is of importance at local and global level. This study investigates the rainfall behaviour using the long memory approach. Since the observed series consists of an unbounded spectral density at zero frequency, a fractionally integrated auto regressive model (ARFIMA) is fitted to explore the pattern and characteristics of the weekly rainfall in the city of Colombo. The maximum likelihood estimation (MLE) method was utilized to obtain estimates for model parameters. To evaluate the suitability of the method for parameter estimation, a Monte Carlo simulation was done with various fractionally differenced parameter values. Model selection was done based on the minimum of the mean absolute error and validated by the forecasting performance that was evaluated using an independent sample. The experimental result yielded a good prediction accuracy with a best fitted long range dependency model and a coverage probability of 95% in terms of prediction intervals that resulted in closer nominal coverage.

Keywords: Rainfall · Fractional differencing · Long-memory
Maximum likelihood estimators · Forecasting

1 Introduction

Modelling rainfall is a challenging task for researchers due to the high degree of uncertainty in atmospheric behaviour. Observational evidence indicates that the climate change has significantly affected global community at a different level. Climate vulnerabilities are expected to be critical in Sri Lanka in the various sectors as agriculture, fisheries, water, health, urban development, human settlement, economic infrastructure, biodiversity and ecosystem in the country [22].

V. Kreinovich et al. (Eds.): ECONVN 2019, SCI 809, pp. 567–580, 2019.
https://doi.org/10.1007/978-3-030-04200-4_40

Information on key climatic variable predictions allow to various stakeholders to prompt themselves for action in order to reduce adverse impacts and enhance positive effects of climatic variation. Rainfall is the one of the most important climatic variable to tropical country like Sri Lanka and this is the variable which give erratic variation at any time in the country. Sri Lanka receives rainfall during the year, with a mean annual rainfall varying from 900 mm in the dry zone to over 5000 mm in the wet zone. Annual rainfall pattern in many parts of Sri Lanka are bimodal and predominantly governed by a seasonally varying monsoon system. Sri Lanka needs to address climate change adaptation to ensure the economic development by the careful investigating of the information on rainfall pattern and its variability which resulting from the predictions of the best fitted rainfall models in various regions. Rainfall analysis is not only important for agricultural areas but also for the urban areas since those areas engage with many activities such as construction, industrial planning, urban traffic, sewer systems, health, rainwater harvesting and climate monitoring. Rainfall is the main source of the hydrological cycle and provides practical benefits through its analysis. Thus, modelling rainfall is one of the key requirements in the country, some of the researchers made attempt to analyse weekly rainfall in Sri Lanka using percentile bootstrap approach to identify the extreme rainfall events [24]. Another study was carried out by the Silva and Peiris [25] to identify the most likelihood time period to form the extreme rainfall events during the South west moons time span by fitting best probability distribution for the weekly rainfall percentiles. Since the Sri Lanka is a developing country which hasn't high technology to sensitive to some important climatic information with related to rainfall is one of the reason cause to low prediction accuracy. However, researchers made effort to model rainfall of the country with increasing degree of accuracy using different techniques. Silva and Peiris [26] discussed problems faced in modelling rainfall which showed positive skewed distribution with longer tail to the right. Rainfall is one of the most difficult variables of the hydrological cycle to understand and model due to its high variability in both space and time [13]. However, several modelling strategies have been applied for the forecasting of rainfall in different areas all over the world. Box-Jenkins autoregressive integrated moving average (ARIMA) model has been widely used for rainfall modelling ([11,20,29,30]). Some of the researchers have made attempts to model rainfall using artificial neural networks ([10,18]). However, very few studies on rainfall in context of long memory can be identified in literature. Granger and Joyeux [15] and Hosking [17] initially proposed a long memory class of models, known as the fractionally integrated autoregressive moving average (ARFIMA) process for stochastic processes. The model defined as ARFIMA (p, d, q) allows the parameter "d" to take fractional values for differencing. There is a fundamental change in the correlation structure of the ARFIMA model, when compared with the correlation structure of the conventional ARIMA model ([6]). According to Granger and Joyeux [15], the slowly decaying autocorrelation exhibited in long range dependency or long memory models differ from stationary ARIMA models that decay exponentially. Many researchers proposed different methods to estimate the fractional differencing parameter. Porter-Hudak and Geweke [14] proposed a method for estimating the long memory differencing parameters based on a simple linear regression of the log periodogram. An approximate maximum likelihood method for parameter “d” was proposed by Fox and Taqqu [12]. An exact maximum likelihood estimation method for differencing parameter was introduced by Sowell [27]. Chen et al. [6] developed a regression type estimator of “d” using lag window spectral density estimators. Number of studies were carried out by comparing various properties of the ARFIMA model based on the estimation method used for the fractionally differencing parameter. (See [2,3,7,16,23]). Dissanayake [9] established a methodology to find an optimal lag order of a standard long memory ARFIMA series within a short process time duration and applied the theory to Nile river data.

Though short memory models have been developed for rainfall still there is a noticeable gap modeling persistent rainfall in view of long memory. The main goal of this study is to fit an ARFIMA model for a weekly rainfall data series in the city of Colombo by capturing the long range dependency features. The paper outline is shaped as follows. In Sect. 2, the long memory ARFIMA model is introduced and some properties of the model are discussed. The model parameter estimation procedure is also described within the section. The results of the Monte Carlo simulation which was used to evaluate the suitability and reliability of the parameter estimation procedure is presented in Sect. 3. Section 4 provides brief details on prediction intervals for forecasting values relevant to the utilized series. The results of weekly rainfall modelling are presented in Sect. 5. Final section, comprises of the conclusion and proposed suggestions.

2 ARFIMA Long Range Dependency Model

ARFIMA is a natural extension of the Box and Jenkins model with non-integer values assigned for d. The ARFIMA (p, d, q) model of a process $\{Y_t\}_{t\in Z}$ is given by the formula

$$\phi(B)\nabla^d(Y_t-\mu)=\psi(B)\varepsilon_t \tag{1}$$

Where μ is the mean of the process, $\{\epsilon_t\}_{t\in Z}$ is a white noise process with zero mean and variance σ_ϵ^2. B is the backward shift operator such that $y_{t-n}=B^n y_t$, $\phi(B)$ and $\theta(B)$ are autoregressive and moving average polynomials of order p and q respectively.

$$\phi(B)=\sum_{i=1}^{p}\phi_i B^i \quad 1\le i\le p \tag{2}$$

$$\psi(B)=\sum_{j=1}^{P}\psi_j B^j \quad 1\le j\le q \tag{3}$$

where d is called as the long memory parameter and differencing operator ∇^d is defined as,

$$\nabla^d=(1-B)^d=\sum_{k=0}^{\infty}\binom{d}{k}(-B)^k \tag{4}$$

Where $\binom{d}{k} = \frac{\Gamma(1+d)}{\Gamma(1+k)\Gamma(1+d-k)}$.

If $d > -0.5$ then the process is invertible and if $d < 0.5$ then the process is stationary. Therefore $d \in (-\frac{1}{2}, \frac{1}{2})$ shows that the process is stationary and invertible. The spectral density function of the $\{Y_t\}_{t \in Z}$ is $f(\omega)$ that can be written as

$$f(\omega) = (2sin\frac{\omega}{2})^{-2d} \quad 0 < \omega \leq \pi \tag{5}$$

$f(\omega) \approx \omega^{-2d} \quad \omega \to 0$

The spectral density function $f(\omega)$ is unbounded when the frequency is near zero. Also, the autocovariance function and correlation function of the process can be expressed as follows

$$\gamma_k = \frac{(-1)^k(-2d)!}{(k-d)!(-k-d)!} \tag{6}$$

$$\rho_k = \frac{d(1+d)...(k-1+d)}{(1-d)(2-d)(3-d)...(k-d)}(k = 1, 2, 3, 4...) \tag{7}$$

Hosking (1981) showed that the auto correlation function of the process satisfies the expression $\rho_k \approx k^{2d-1}$ when $0 < d < 1/2$. Thus, the autocorrelation of the ARFIMA process decays hyperbolically to zero as $k \to \infty$ and in contrast, the auto correlation function of the ARIMA process has a exponential decay. The process with $d = 0$ reduces to a short memory ARMA model.

Let Z denote a series of "n" observations with mean μ and variance σ_Z^2. If the decay parameter is considered as α, then the natural fractional differencing parameter "d" can be written as $d = (1-\alpha)/2$.

The log likelihood function of the Exact Gaussian can be written as

$$l(\alpha, \sigma_y^2) = -\frac{1}{2}(log \quad det(\Gamma_n) + Z'\Gamma_n^{-1}Z') \tag{8}$$

The arfima package (See [28]) in R optimized the log likelihood function and obtained the exact maximum likelihood estimators. Two algorithms namely Durbin-Levinson and Trench algorithms were utilized to maximize the likelihood and obtain optimal simulation and forecasting results.

3 Result of the Monte Carlo Simulation

A number of Monte Carlo experiments were carried out to evaluate the performance of the maximum likelihood method used for parameter estimation. The simulation was done based on various fractional differencing parameter values with 1000 replications. The four different series lengths ($n = 100$, $n = 200$, $n = 500$ and $n = 1000$) were considered for the simulation. The simulation results provided fractionally differenced parameter estimates and corresponding standard and mean square errors. Monte Carlo experiment was conducted on a simulated ARFIMA(0,d,0) series with parameter values: $d = 0.1$, $d = 0.15$, $d = 0.3$ and $d = 0.45$.

The simulation was carried out using the R programming Language (Version 3.4.2) utilizing a HP11(8 GB, 64 bit) computer. The standard errors of the estimates $SD(\hat{d})$ and mean square error of the estimates $MSE(\hat{d})$ can be expressed as;

$$SD(\hat{d}) = \sqrt{\sum_{r=1}^{R}(\hat{d}_r - \hat{d})/R} \qquad MSE(\hat{d}) = \sum_{r=1}^{R}(\hat{d}_r - d)^2/R$$

Where $\hat{d}_r$ is the MLE of d for the r^{th} replication. The value R denotes the number of replications ($R = 1000$ for all tabulated simulation results of this paper). Tables 1, 2, 3 and 4 present the average of the estimated d, corresponding standard error and MSE of the estimator.

According to the results in Tables 1, 2, 3 and 4, the performance of the maximum likelihood estimator is reasonably accurate. It can be clearly seen that the parameter bias has decreased with the increase in sample size. Furthermore,

Table 1. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.1. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.0517	0.0912	0.0106
200	0.0748	0.0626	0.0045
500	0.0885	0.0367	0.0014
1000	0.0949	0.0254	0.0006

Table 2. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.15. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.1048	0.0915	0.0104
200	0.1265	0.0593	0.0040
500	0.1408	0.0367	0.0014
1000	0.1456	0.0254	0.0006

Table 3. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.3. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.2493	0.0877	0.0102
200	0.2726	0.0575	0.0040
500	0.2892	0.0362	0.0014
1000	0.2947	0.0251	0.0006

Table 4. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.45. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.3774	0.0695	0.0101
200	0.4079	0.0477	0.0040
500	0.4310	0.0314	0.0013
1000	0.4435	0.0270	0.0007

the results provide evidence that the parameters become consistent with the increase in series length. As we expected the standard deviation and the MSE of the estimators have decreased with the increase in series length.

4 Forecast and Prediction Intervals

Forecasts are obtained based on the best fitted long memory model. However, predicting of future values along with their prediction intervals become more beneficial in long memory time series analysis. The lower (L) and upper (U) boundaries covering the forecast values with known probability are simply called prediction intervals of the form [L, U]. A detailed review of approaches in calculating interval forecast using time series was described in Chatfield [5]. Charles et al. [4] made an effort to make prediction intervals to forecast US core inflation values that provided a unique fractional model. Prediction intervals were utilized to forecast tourism demand by Chu [8]. Zhou et al. [31] suggested a prediction interval method to predict aggregates of future values derived from a long memory model. A new bootstrap method for autoregressive models was proposed by Hwang and Shin [19]. Ali et al. [1] suggested a Sieve bootstrap approach to construct intervals for a long memory model. Prediction interval approach was utilized to measure the uncertainty about long-run predictions by Muller and Watson [21].

5 Application

Sri Lanka is a tropical country in South Asian region located at the latitudes of 5° 55 N and 9° 51 N and the longitudes of 79° 41 E and 81° 53 E with an area of 65610 km^2 and the Colombo city is the commercial capital of Sri Lanka. Daily rainfall data of Colombo were collected from 1990 to 2015 from the Department of Meteorology, Sri Lanka for this analysis. The daily rainfall (mm) data has been converted into weekly rainfall by dividing a year into 52 weeks such that week 1 corresponds to 1–7 January, Week 2 corresponds to 8–14 January and so on. The data during the time span from 1990 to 2014 was used to build the model while the rest was used for model validation. To examine the temporal

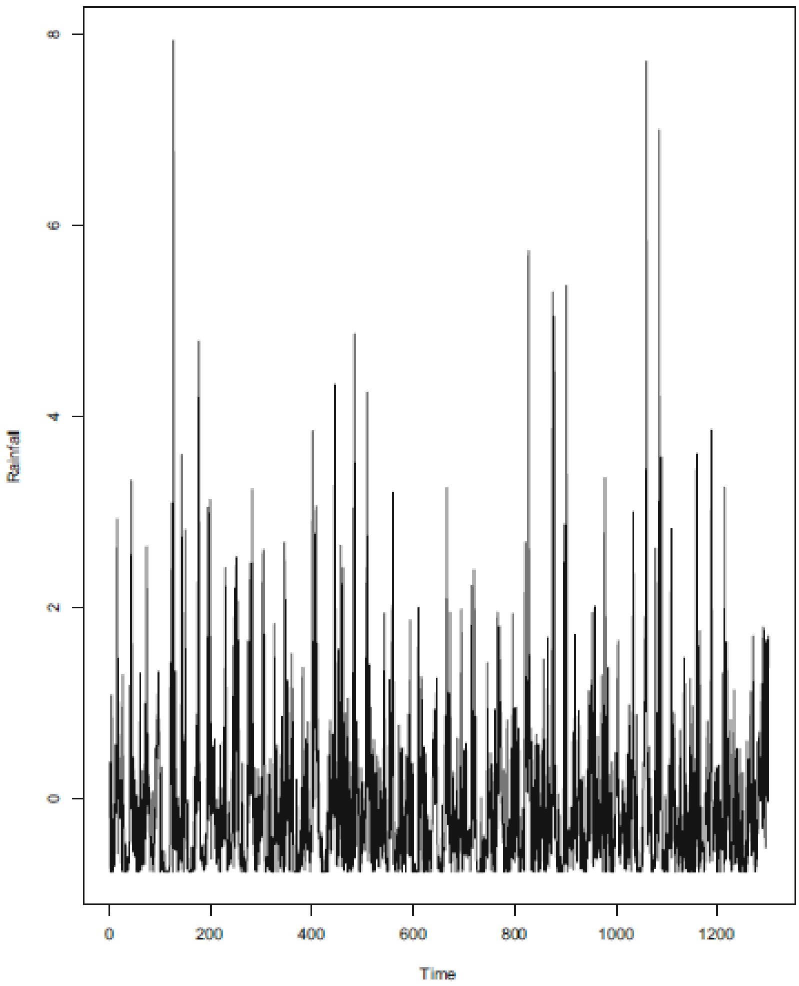

Fig. 1. Time series plot of weekly rainfall series from 1990 to 2014

variability of the rainfall series, time series plots were taken and presented in Fig. 1.

The time series plot explores the random behaviour of weekly rainfall during the considered time span from 1990 to 2014. In order to identify the correlation structure of the observed series, the autocorrelation and partial auto correlation plots were taken and those results are shown in Figs. 2 and 3 respectively.

In order to study the long memory features of the weekly rainfall series, the periodgram was obtained and presented in Fig. 4. The maximum spectrum density is 0.0385185 given at a frequency which is very close to zero. Based on those characteristics the series displays long memory. Thus, we conclude that the ARFIMA standard long memory model may be suitable for the observed weekly rainfall series. Long range correlation of observed data were considered

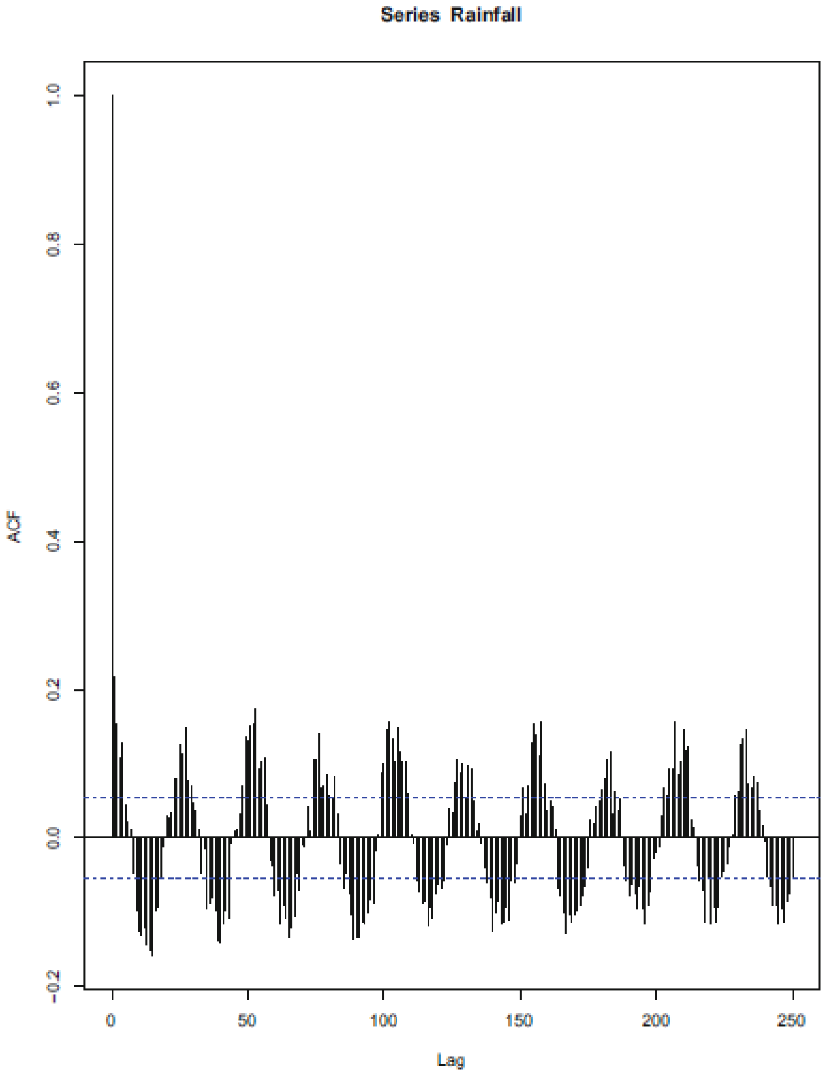

Fig. 2. Autocorrelation plot of the series from 1990 to 2014

in long memory modelling. Various ARFIMA models were fitted for the data set that vary from 1990 to 2014 (series length = 1300).

Those fitted models were employed to predict the weekly rainfall during the time span from 2014 to 2015 and best fitted model is selected with the minimum mean absolute error (MAE). The MAE can be written as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

Where e_i is the forecasting error and n is the length of the forecasting series.

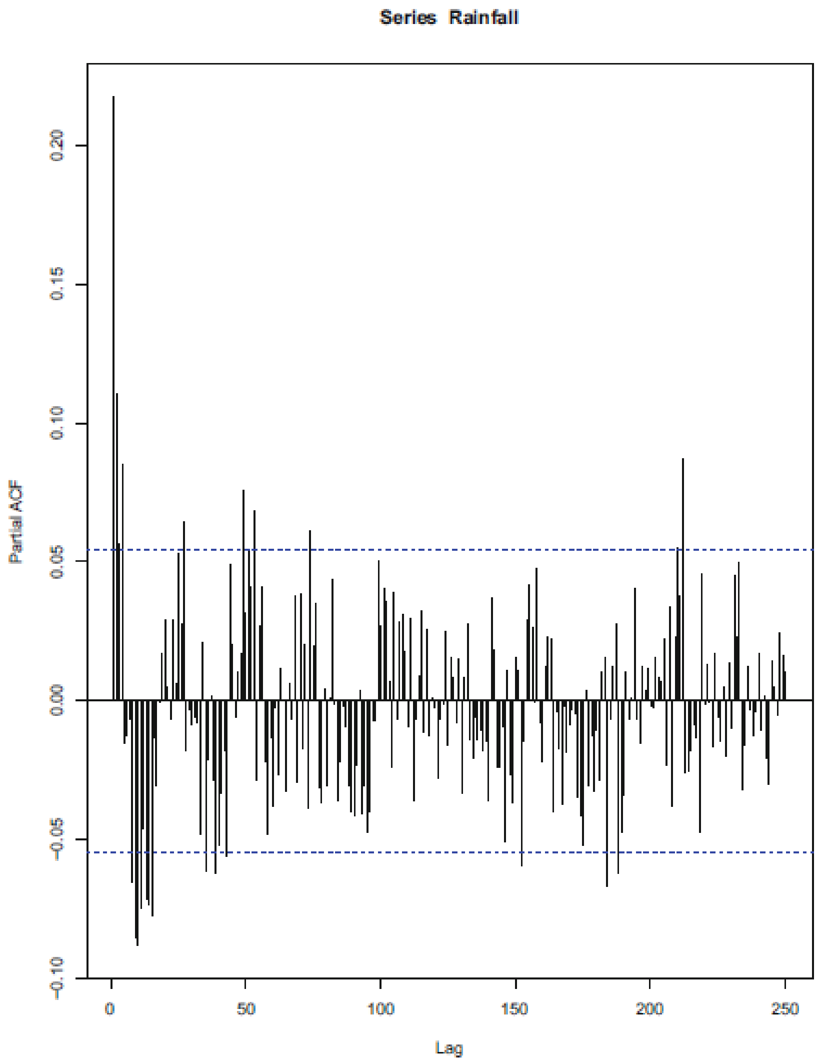

Fig. 3. Partial autocorrelation plot of the series from 1990 to 2014

The best fitted model and the corresponding parameter estimates are presented in Table 5. The ARFIMA (4,0,4) model was found to be the best fit for the weekly rainfall series returning the smallest MAE.

All model parameters except the constant are significant at the 0.05 level of significance. The residual analysis of the fitted model was performed and found the uncorrelated at a 5% level of significance. Furthermore, the model was tested for weekly rainfall data in 2015 and the result is presented in Table 6. Figure 5 illustrates the weekly rainfall over the year 2015 along with the predicted estimates.

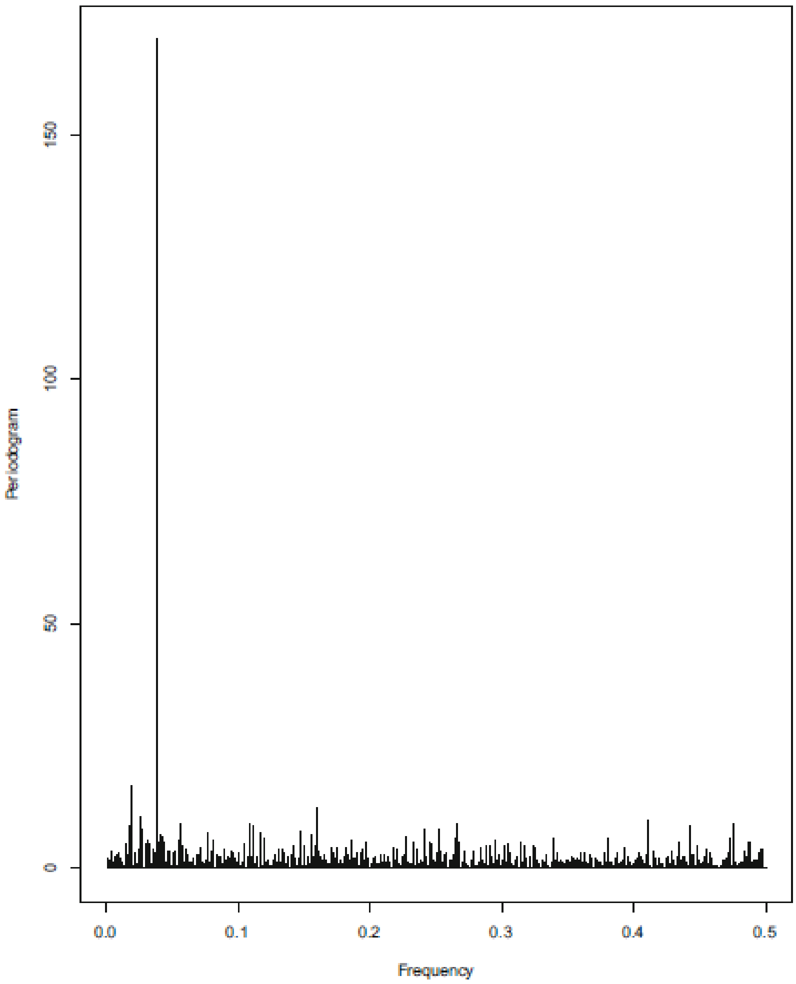

Fig. 4. The periodgram of the rainfall series from 1990 to 2014

Table 5. Fitted model for the weekly rainfall series ARFIMA (4,0,4) with p = 4, q = 4, d = 0.05792421

Coefficients	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ_1
Estimate	1.2059	−0.2493	0.5765	−0.6752	1.1243
Standard error	0.0242	0.0454	6.324e-07	6.324e-07	0.0231-Correct value (CV)
Z-value	4.9768e01	5.4903	9.1153e05	−1.0676e06	4.8638e01
Pr(>|Z|)	0.0000	0.0005	0.0000	0.0000	0.0000

Table 5. *(continued)*

Coefficients	θ_2	θ_3	θ_4	Constant	d
Estimate	−0.1131	0.5220	−0.6743	−0.0163	0.0579
Standard error	0.0365 (CV)	0.0354 (CV)	0.0215	0.0380	0.0276
Z-value	−3.0992	1.4735e01	−3.1363e01	−4.2907e-01	2.0950
Pr(>|Z|)	0.0019	0.0000	0.0000	0.6678	0.0361

Table 6. Absolute Forecast Error for independent sample (2015)

Absolute forecasting error in mm	ARFIMA number of weeks percentage
0–10	10(19.2)
11–15	6(11.5)
16–20	6(11.5)
21–25	4(7.7)
26–30	6(11.5)
31–35	1(1.9)
36–40	4(7.7)
41–45	1(1.9)
46–50	2(4.0)
More than 50	12(23.1)

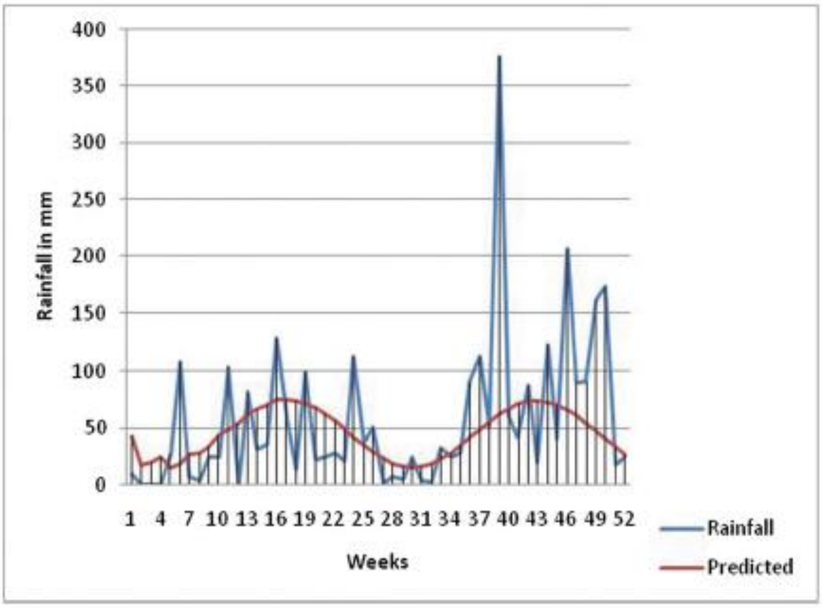

Fig. 5. Forecasted and actual weekly rainfall in 2015

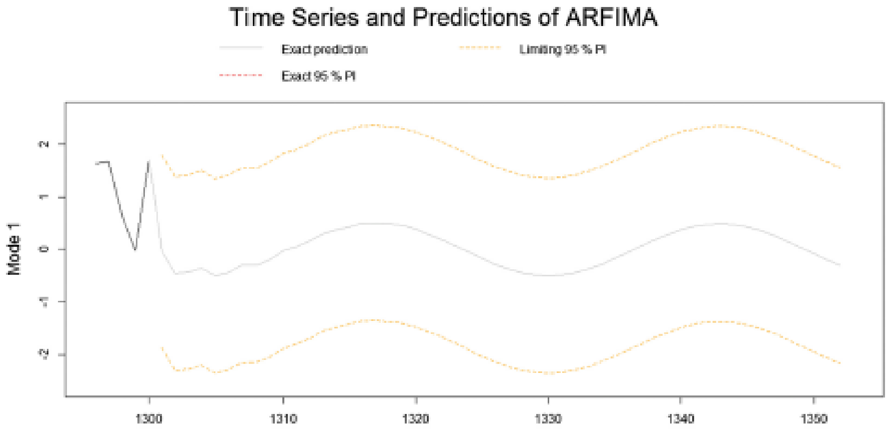

Fig. 6. Prediction intervals for forecasted rainfall values in 2015

According to Fig. 5, it can be seen that the predicted values are in considerable good agreement with the actual rainfall values. The result of the 95% prediction interval also provides encouraging prediction accuracy with a 93.23% coverage probability (Fig. 6).

6 Conclusion

Observed rainfall series illustrates long memory features with an unbounded spectral density. Therefore a standard long memory ARFIMA model was fitted to capture the rainfall pattern and its variability. The Monte Carlo simulation results prove the accuracy of the maximum likelihood method used to estimate the parameters of the model. Furthermore, it is noticed that the parameter bias has decreased and the parameters become consistent with the increase of the simulated series length. ARFIMA(4,0.0579,4) model was found to be the best fitted model that provided a minimum MAE. The out of sample prediction values give good conformity with the actual weekly rainfall in 2015. The 95% prediction intervals also give a promising result to capture the real dynamics of the persistent rainfall. For future work it is suggested that prediction intervals using the bootstrap re-sampling approach may forecast estimates with a higher degree of accuracy.

Acknowledgement. This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant (ASP/01/RE/HSS/2016/75).

References

1. Alamgir, A.A., Kalil, U., Khan, S.A., Khan, D.M.: A Sieve bootstrap approach to constructing prediction intervals for long memory time series. Res. J. Recent Sci. **4**(7), 93–99 (2015)

2. Beran, J., Feng, Y., Ghosh, S., Kulik, R.: Long memory Processes Probabilistic Properties and Statistic properties and Statistical Methods. Springer, Heidelberg (2013)
3. Chan, N.H., Palma, W.: Estimation of long-memory time series models. Adv. Econ. **20**, 89–121 (2006)
4. Charles, S.B., Franses, P.H., Ooms, M.: Inflation, forecast intervals and long memory regression models. Int. J. Forecast. **18**, 243–264 (2002)
5. Chatfield, C.: Calculating interval forecasts. J. Bus. Econ. Stat. **11**(2), 121–135 (1993)
6. Chen, G., Abraham, B., Peiris, S.: Lag window estimation of the degree of differencing in fractionally integrated time series models. J. Time Ser. Anal. **15**(5), 473–487 (1994)
7. Cheung, Y., Diebold, F.X.: On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. J. Econ. **62**, 301–316 (1994)
8. Chu, F.L.: A fractionally integrated autoregressive moving average approach to forecast tourism demand. Tour. Manag. **29**, 79–88 (2008)
9. Dissanayake, G.S.: Rapid optimal lag order detection and parameter estimation of standard long memory time series. In: Causal Inference in Econometrics, pp. 17–28. Springer, Cham (2016)
10. Dubey, A.D.: Artificial neural network models for rainfall prediction in Pondicherry. Int. J. Comput. Appl. **120**(3), 30–35 (2015)
11. Eni, D., Adeyeye, F.J.: Seasonal ARIMA modelling and forecasting of rainfall in Warri Town, Nigeria. J. Geosci. Environ. Protection **3**, 91–98 (2015)
12. Fox, R., Taqqu, M.S.: Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Stat. **14**(2), 517–532 (1986)
13. French, M.N., Krajewski, W.F., Cuykendall, R.R.: Rainfall forecasting in space and time using a neural network. J. Hydrol. **137**, 1–31 (1992)
14. Geweke, J., Hudak, S.P.: The estimation and application of long memory time series models. J. Time Ser. Anal. **4**, 221–238 (1983)
15. Granger, C.W.J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. J. Time Ser. Anal. **1**, 15–29 (1980)
16. Hauser, M.A.: Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study. J. Stat. Plan. Inference **80**, 229–255 (1999)
17. Hosking, J.R.M.: Fractional differencing. Biometrika **68**, 165–176 (1981)
18. Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K.: An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrol. Earth Syst. Sci. **13**, 1413–1425 (2009)
19. Hwang, E., Shin, D.W.: New bootstrap method for autoregressive models. Commun. Stat. Appl. Methods **20**(1), 85–96 (2013)
20. Momani, P.E.N.M.: Time series analysis model for rainfall data in Jordan: case study for using time series analysis. Am. J. Environ. Sci. **5**(5), 559–604 (2009)
21. Muller, U.K., Watson, M.W.: Measuring uncertainty about long-run predictions. Rev. Econ. Stud. **83**, 1711–1740 (2016)
22. National Climate Change Adaptation Strategy for Sri Lanka 2011 to 2016. Ministry of Environment (2010)
23. Palma, W.: Long Memory Time Series Theory and Methods. John Wiley and Sons, New Jersey (2007)
24. Silva, H.P.T.N., Peiris, T.S.G.: Analysis of weekly rainfall using percentile bootstrap approach. Int. J. Ecol. Dev. **32**(3), 97–106 (2017)

25. Silva, H.P.T.N., Peiris, T.S.G.: Statistical modeling of weekly rainfall: a case study in Colombo city in Sri Lanka. In: Proceedings of the 3rd Moratuwa Engineering Research Conference (MER Con), Sri Lanka, 29–31 May, pp. 241–246. IEEE (2017)
26. Silva, H.P.T.N., Peiris, T.S.G.: Accurate confidence intervals for Weibull percentiles using bootstrap calibration: a case study of weekly rainfall in Sri Lanka. Int. J. Ecol. Econ. Stat. **39**(3), 67–76 (2018)
27. Sowell, F.: Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J. Econ. **53**, 165–188 (1992)
28. Veenstra, J.Q., McLeod, A.I.: Persistence and Anti-persistence: Theory and Software (Ph.D. thesis) (2013)
29. Wang, S., Feng, J., Liu, G.: Application of seasonal time series model in the precipitation forecast. Math. Comput. Model. **58**, 677–683 (2013)
30. Zakaria, S., Al-Ansari, N., Knutsson, S., Al-Badrany, T.: ARIMA models for weekly rainfall in the semi-arid Sinjar District at Iraq. J. Earth Sci. Geotech. Eng. **2**(3), 25–55 (2012)
31. Zhou, Z., Xu, Z., Wu, W.B.: Long-term prediction intervals of time series. IEEE Trans. Inf. Theory **56**(3), 1436–1446 (2010)

Abstract No: 13

DEVELOPMENT OF LONG MEMORY MODEL TO FORECAST WEEKLY RAINFALL

H.P.T.N. Silva* **and T.S.G. Peiris**

[1]*Department of Social Statistics, Faculty of Humanities and Social Sciences, University of Sri Jayewardenepura, Nugegoda, 10250, Sri Lanka*
[2]*UDepartment of Mathematics, Faculty of Engineering, University of Moratuwa, Moratuwa, 10400, Sri Lanka*
thanuja@sjp.ac.lk

Awareness of pattern of weekly rainfall and its variability facilitate to make effective decisions with respect to the climate monitoring. Though various statistical and non statistical techniques have been developed for rainfall modeling with increasing degree of accuracy is still a noticeable gap for prediction of rainfall. The aim of this study was to model weekly rainfall in context of long memory along with the conditional heteroskedasticity. Weekly rainfall data (1990-2017) in Colombo city was obtained from the Department of Meteorology, Sri Lanka. Of the various types of long memory models developed for weekly series, the best fitted model is ARFIMA-GARCH for deseasonalized data. The model was trained using weekly rainfall data from 1990 to 2014 and validated using weekly data from 2015 to 2017. The forecasting performance of the new model is not much diluted with the increase of the forecasting length. The exact maximum likelihood estimation method was utilized to estimate the model parameters and Monte Carlo simulation was carried out with various fractional differencing parameters to evaluate the suitability of the estimation method. The simulation study is provided the empirical evidence to optimal accuracy of parameter estimation. The best fitted model developed is ARFIMA-GARCH for deseasonalized data. The forecasting performance of the model was evaluated based on the novel index developed using absolute error for an independent data set in addition to the classical indicators. The novel long range dependency model is recommended to be used in forecasting weekly rainfall in Colombo city in Sri Lanka.

Financial assistance from the University of Sri Jayewardenepura, Sri Lanka, *(Grant No ASP/01/RE/HSS/2016/75) is acknowledged.*

Keywords: ARFIMA-GARCH, Forecasting, Fractional differencing, Long-memory, Weekly rainfall

REFERENCE

Abbot, J., & Marohasy, J. (2014). Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmospheric Research, 138*, 166-178.

Abdul-Aziz, A., Anokye, M., Kwame, A., Munyakazi, L., & Nsowah-Nuamah, N. (2013). Modelling and Forecasting Rainfall Pattern in Ghana as a Seasonal Arima Process: The Case of Ashanti Region. *International Journal of Humanities and Social Science, 3*(3), 224-233.

Adenomon, M., Ojehomon, V., & Oyejola, B. (2013). Modelling the dynamic relationship between rainfall and temperature time series data in Niger State, Nigeria. *Mathematical Theory and Modelling, 3*(4), 53-70.

Ahmed, M., & Suphachalasai, S. (2014). Assessing the costs of climate change and adaptation in South Asia.

Alghazali, N. O. S., & Alawadi, D. A. H. (2014). Fitting statistical distributions of monthly rainfall for some Iraqi stations. *Civil and Environmental Research, 6*(6), 40-46.

Alhashimi, S. A. M. (2014). Prediction Of Monthly Rainfall In Kirkuk Using Artificial Neural Network And Time Series Models. *Journal of Engineering and Development, 18(1),* 129-143.

Ali, S. M. (2013). Time series analysis of Baghdad rainfall using ARIMA method. *Iraqi Journal of Science, 54*(4), 1136-1142.

Alimirzaie, H., Saremi, A., & Soleymani, M. M. (2015). Precipitation and Runoff Forecasting and Analysis Using Stochastic Models (Case Study: Urmia Lake Catchment area). *Bulletin of Envirnonment Pharmacology and Life Science,4(*6*),* 150-154.

Amarasingha, R., Suriyagoda, L., Marambe, B., Galagedara, L., Silva, G., Punyawardena, R, Howden, M. (2015). Modelling the impact of changes in rainfall distribution on the irrigation water requirement and yield of short and medium duration rice varieties using APSIM during Maha season in the dry zone of Sri Lanka. *Tropical Agricultural Research, 26*(2),274-284.

Ampaw, E. M., Akuffo, B., Larbi, S. O., & Lartey, S. (2013). Time Series Modelling of Rainfall in New Juaben Municipality of the Eastern Region of Ghana. *International Journal of Business and Social Science,4*(8), 116-129.

Anděl, J. (1986). Long memory time series models. *Kybernetika, 22*(2), 105-123.

Armesh, M., & Negaresh, H. (2013). Maximum Daily Rainfall Simulation by using Artificial Neural Network (Case Study: Saravan-Iran). *Research Journal of Environmental and Earth Sciences, 5*(11), 651-659.

Ashfaq, M., Shi, Y., Tung, W. w., Trapp, R. J., Gao, X., Pal, J. S., & Diffenbaugh, N. S. (2009). Suppression of South Asian summer monsoon precipitation in the 21st century. *Geophysical Research Letters, 36*(1).

Baba, N. (2010). Sinking the Pearl of the Indian Ocean: Climate Change in Sri Lanka. *Global Majority E-Journal, 1*(1), 4-16.

Babazadeh, H., & Shamsnia, S. (2014). Modeling climate variables using time series analysis in arid and semi arid regions. *African Journal of Agricultural Research, 9*(26), 2018-2027.

Bari, S., Rahman, M., Hussain, M., & Ray, S. (2015). Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research, 7*(1), 69-77.

Basnayake, B. R. S. B., Rathnasiri, J., and Vithanage, J.C.,. (2004). *Rainfall and temperature scenarios for Sri Lanka under the anticipated climate change*. Paper presented at the 2nd AIACC, Reaginal workshop for Asia and the Pacific, Manila, Philippines.

Bisognin, C., & Lopes, S. R. C. (2009). Properties of seasonal long memory processes. *Mathematical and Computer Modelling, 49*(9-10), 1837-1851.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics, 31*(3), 307-327.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis, control, and forecasting. *San Francisco, CA: Holden Day.*

Burn, D. H. (2003). The use of resampling for estimating confidence intervals for single site and pooled frequency analysis/Utilisation d'un rééchantillonnage pour l'estimation des intervalles de confiance lors d'analyses fréquentielles mono et multi-site. *Hydrological sciences journal, 48*(1), 25-38.

Burt, T., & Weerasinghe, K. (2014). Rainfall distributions in Sri Lanka in time and space: An analysis based on daily rainfall data. *Climate, 2*(4), 242-263.

Chakraborty, S., Denis, D., & Sherring, A. (2010). Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh‖. *International Journal of Advances in Engineering Science and Technology, 2*, 153-163.

Charaniya, N., & Dudul, S. (2013). Design of Neural Network Models for Daily Rainfall Prediction. *International Journal of Computer Applications*, *61*(14), 23-27.

Chen, G., Abraham, B., & Peiris, S. (1994). Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of time series analysis*, *15*(5), 473-487.

Chen, S., Li, Y., Shin, J., & Kim, T. (2016). Constructing confidence intervals of extreme rainfall quantiles using Bayesian, bootstrap, and profile likelihood approaches. *Science China Technological Sciences*, *59*(4), 573-585.

Cheung, Y.-W., & Diebold, F. X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of econometrics*, *62*(2), 301-316.

Chonge, M., Nyongesa,K., Mulati, O., Makokha, L., & &Tireito, F. (2015). A Time Series Model of Rainfall Pattern of Uasin Gishu County. *LOSR Journals of Mathematics*, *11*(5), 77-84.

Chung, C.F. (1996). A generalized fractionally integrated autoregressive moving-average process. *Journal of Time Series Analysis*, 17(2),111-140.

Dai, A. (2006). Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, *19*(18), 4605-4630.

Dastorani, M., Mirzavand, M., Dastorani, M. T., & Sadatinejad, S. J. (2014). Comparative study among different time series models applied to monthly rainfall forecasting in semi-arid climate condition. *Natural Hazards*, *81*(3), 1811-1827.

De Silva, C., Weatherhead, E., Knox, J. W., & Rodriguez-Diaz, J. (2007). Predicting the impacts of climate change—A case study of paddy irrigation water requirements in Sri Lanka. *Agricultural water management*, *93*(1), 19-29.

De Silva, C. S. (2006). *Impacts of climate change on water resources in Sri Lanka.* Paper presented at the 32nd WEDC International Conference.

De Zoysa, M., & Inoue, M. (2014). Climate change impacts, agroforestry adaptation and policy environment in Sri Lanka. *Open Journal of Forestry, 4*(5), 439.

Delleur, J. W., & Kavvas, M. L. (1978). Stochastic models for monthly rainfall forecasting and synthetic generation. *Journal of Applied Meteorology*, *17*(10), 1528-1536.

Department of Meteorology, Sri Lanka (2016). www.meteo.gov.lk.

Deshpande, R. R. (2012). On the rainfall time series prediction using multilayer perceptron artificial neural network. *International Journal of emerging technology and advanced engineering, 2*(1), 2250-2459.

Dhiman, R. C., Pahwa, S., Dhillon, G., & Dash, A. P. (2010). Climate change and threat of vector-borne diseases in India: are we prepared? *Parasitology Research, 106*(4), 763-773.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366a), 427-431.

Dissanayake, G., Peiris, M., & Proietti, T. (2016). State space modeling of Gegenbauer processes with long memory. *Computational Statistics & Data Analysis, 100,* 115-130.

Dissanayake, G., Peiris, M., & Proietti, T. (2018). Fractionally differenced Gegenbauer processes with long memory: A review. *Statistical Science, 33*(3), 413-426.

Dissanayake, G., Peiris, M., Proietti, T., Ruiz, I., & Garcia, G. (2014). *Estimation of generalized fractionally differenced processes with conditionally heteroskedastic errors.* Paper presented at the International Work Conference on Time Series. Proceedings ITISE.

Dissanayake, G., & Peiris, S. (2012). Generalized fractional processes with conditional heteroskedasticity. *Sri Lankan Journal of Applied Statistics, 12*(1), 1-12.

Dissanayake, G. S. (2016a). Advancement of fractionally differenced Gegenbauer processes with long memory. *Bulletin of the Australian Mathematical Society, 94*(1), 173-174.

Domroes, M. (1974). *The Agroclimatic of Sri Lanka*. Wiesbaden.

Dubey, A. D. (2015). Artificial neural network models for rainfall prediction in Pondicherry. *International Journal of Computer Applications, 120*(3), 30-35.

Dunn, P. K. (2001). Bootstrap confidence intervals for predicted rainfall quantiles. *International Journal of Climatology, 21*(1), 89-94. doi: 10.1002/joc.596.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society,* 987-1007.

Eni, D., & Adeyeye, F. J. (2015). Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection, 3*(6), 91-98.

Eriyagama, N., Smakhtin, V., Chandrapala, L., & Fernando, K. (2010). *Impacts of climate change on water resources and agriculture in Sri Lanka: a review and preliminary vulnerability mapping* (Vol. 135): IWMI.

Etuk, E. H., & Mohamed, T. M. (2014). Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by SARIMA Methods. *International Journal of Scientific Research in Knowledge, 2*(7), 320-330.

Farook, A. J., & Kannan, K. S. (2016). Climate change impact on rice yield in india–vector autoregression approach. *Sri Lankan Journal of Applied Statistics, 16*(3), 161-178.

Faulina, R. (2013). Suhartono," Hybrid ARIMA-ANFIS for Rainfall Prediction in Indonesia,". *nternational Journal of Science and Research, 2,* 159-162

Ferro, C. A., Hannachi, A., & Stephenson, D. B. (2005). Simple nonparametric techniques for exploring changing probability distributions of weather. *Journal of Climate, 18*(21), 4344-4354.

Fernando, M., Zubair, L., Peiris, T., Ranasinghe, C., & Ratnasiri, J. (2007). Economic value of climate variability impacts on coconut production in Sri Lanka: AIACC Working Paper No.45.

Fox, R., & Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics, 14*(2), 517-532.

French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. *Journal of Hydrology, 137*(1-4), 1-31.

Gerretsadikan, A., & Sharma, M. (2011). Modeling and forecasting of rainfall data of mekele for Tigray region (Ethiopia). *Statistics and Applications, 9*(1-2), 31-53.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis, 4*(4), 221-238.

Ghalhari, G. A. F., Bayatani, F., Fahiminezhad, E., (2015). Comparing the Forecasting Accuracy of the Box–Jenkins Models in Modeling Seasonal Precipitation (Case Study: The South of Kerman Province, Iran). *Journal of Applied Environmental and Biological Sciences, 5*(12), 64-78.

Ghosh, S., Roy, M. K., & Biswas, S. C. (2016). Determination of the Best Fit Probability Distribution for Monthly Rainfall Data in Bangladesh. *American Journal of Mathematics and Statistics, 6*(4), 170-174.

Golabi, M., Radmanesh, F., Akhondali, A., & Kashefipoor, M. (2013). Prediction of Seasonal Precipitation using Artificial Neural Networks (Case Study: Selected

Stations of (Iran) Khozestan Province). *Journal of Basic and Applied Scientific Research, 3*(1), 589-595.

Goswami, P., & Gouda, K. (2007). Objective determination of the date of onset of monsoon rainfall over India based on duration of persistence. *CSIR Centre for Mathematical Modelling and Computer Simulation, Research Report RR CM, 711*.

Granger, C. W., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis, 1*(1), 15-29.

Gray, H.L., Zhang, N. &Woodward, W.A. (1989). On generalized fractional Process. *Journal of Time Series Analysis*, 10(3), 233-257.

Gupta, A., Gautam, A., Jain, C., Prasad, H., & Verma, N. (2013). Time Series Analysis of Forecasting Indian Rainfall. *International Journal of Inventive Engineering and Sciences (IJIES), 1*(6), 42-45.

Gupta, P., Mishra, S., & Pandey, S. (2014). Time series data mining in rainfall forecasting using artificial neural network. *IJSET (ISSN: 2277-1581), 3*(8), 1060-1065.

Hall, P. (1992). T he Bootstrap and Edgeworth Expansion. *SpringerVerlag, New York*.

Henry, M. (2001). Averaged periodogram spectral estimation with long-memory conditional heteroscedasticity. *Journal of time series analysis, 22*(4), 431-459.

Heo, J.-H., Salas, J., & Kim, K.-D. (2001). Estimation of confidence intervals of quantiles for the Weibull distribution. *Stochastic environmental research and risk assessment, 15*(4), 284-309.

Hidayah, E., Iriawan, N.,, Anwar, N., & Edijatno, E. (2011). Generating Hourly Rainfall Model using Bayesian Time Series Model (A Case Study at Sentral Station, Bondowoso). *IPTEK The Journal for Technology and Science*, *22*(1), 50-56.

Hosking, J. (1981). Fractional differencing. Biometrika 68 165–176. *Mathematical Reviews (MathSciNet): MR614953 Zentralblatt MATH, 464*.

Hung, N. Q., Babel, M. S., Weesakul, S., & Tripathi, N. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences, 13*(8), 1413-1425.

IPCC. (2014).Climate Change: Impacts, Adaptation, and Vulnerability: Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change: Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jayawardene, H., Sonnadara, D., & Jayewardene, D. (2005). Trends of rainfall in Sri Lanka over the last century. *Sri Lankan Journal of Physics, 6*, 7-17.

Jegarascsingam, V. (1998). Sri Lanka Country Report.

Jensen, M. J. (2005). Semiparametric Bayesian inference of long-memory stochastic volatility models. *Journal of time series analysis, 25*(6), 895-922.

Johnson, N.L., & Kotz, S. (1970). Continous Univariate Distribution-1, Houghton Miffin Company, Boston.

Kane, I. L., & Yusof, F. (2013). Assessment of Risk of Rainfall Events with a Hybrid of ARFIMA-GARCH. *Modern Applied Science*, *7*(12), 78.

Kannan, M., Prabhakaran, S., Ramachandran, P.,. (2010). Rainfall forecasting using Data Mininig Technique. *International Journal of Engineering and Technology, 2*(6), 397-401.

Kaushik, I., & Singh, S. M. (2008). Seasonal ARIMA model for forecasting of monthly rainfall and temperature. *Journal of Environmental Research And Development, 3*(2), 506-514.

Khodashenas, S. R., Khalili, N., & Davari, K. (2010). Monthly Precipitation Prediction by Artificial Neural Networks (Case study: Mashhad synoptic station). *NOVATECH 2010*.

Krishnankutty, N. (2006). Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network. *Current Science, 90*(6),773-779.

Kumar, N. D., Janga, Reddy M, & Rajib, M. (2007). Regional Rainfall Forecasting using Large Scale Climate Teleconnections and Artificial Intelligence Techniques. *Journal of Intelligent Systems, 16*(4), 307-322.

Kumar, R., & Yadav, G. (2013). Forecasting of Rain Fall in Mirzapur District, Uttar Pradesh, India Using Feed-Forward Artificial Neural Network. *International Journal of Engineering Science Invention,2*(8), 87-93.

Kumarasiri, A., & Sonnadara, D. (2006). *Rainfall forecasting: an artificial neural network approach.* Paper presented at the Proceedings of the Technical Sessions.

Lin, G.-F., & Wu, M.-C. (2009). A hybrid neural network model for typhoon-rainfall forecasting. *Journal of Hydrology, 375*(3), 450-458.

Ling, S., & Li, W. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association, 92*(439), 1184-1194.

Lo, K. F. A., & Koralegedara, S. B. (2015). Effects of climate change on urban rainwater harvesting in Colombo city, Sri Lanka. *Environments, 2*(1), 105-124.

Lucio, P. (2007). *Bootstrap for statistical evaluation of small sample inference for precipitation extreme quantiles.* Paper presented at the Geophysical Research Abstracts.

Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modelling, 33*(6), 683-693.

Mahalakshmi, N., Umarani, P.R. & Selvaraj, S.R. (2014). Forecasting The Tamil Nadu Rainfall Using Hybrid ARIMA- ANN Model. *Recent Scientific Research, 5*(3), 566-569.

Mahsin, M. (2012). Modeling rainfall in Dhaka division of Bangladesh using time series analysis. *Journal of Mathematical Modelling and Application, 1*(5), 67-73.

Malmgren, B. A., Hulugalla, R., Hayashi, Y., & Mikami, T. (2003). Precipitation trends in Sri Lanka since the 1870s and relationships to El Niño–Southern Oscillation. *International Journal of Climatology, 23*(10), 1235-1252.

Manawadu, L., & Fernando, N. (2008). Climate Change in Sri Lanka.

Mar, K. W., & Naing, T. T. (2008). Optimum neural network architecture for precipitation prediction of Myanmar. *World Academy of Science, Engineering and Technology, 48*, 130-134.

Mathugama, S. C., & Peiris, T.S.G. (2011). Critical Evaluation of Dry Spell Research. *International Journal of Basic and Applied Sciences, 11*(6), 153-160.

Mathugama, S. C., & Peiris, T.S.G. (2012). Spatial and Temporal Analysis of Critical Dry Spells in Sri Lanka. *International Journal of Climate Change: Impacts and Responses, 3*(3), 71-88.

Mayooran, T., & Laheetharan, A. (2014). The statistical distribution of annual maximum rainfall in Colombo district. *Sri Lankan Journal of Applied Statistics, 15*(2), 1765-1784.

Mawilmada, N., Atapattu, S., Gunawardene, N., Weerasinghe, B., Nandana, M., Bellanawithana, A., Wimalasiri, R., & Kumari, N. (2010). National Climate Change Adaptation Strategy for Sri Lanka 2011 to 2016. *Ministry of Environment, Sri Lanka.*

Mekanik, F., & Imteaz, M. (2012). *A multivariate artificial neural network approach for rainfall forecasting: case study of Victoria, Australia.* Paper presented at the Proceedings of the world congress on engineering and computer Science.

Mekanik, F., Lee, T., & Imteaz, M. (2011). *Rainfall modeling using Artificial Neural Network for a mountainous region in West Iran.* Paper presented at the 19th

International Congress on Modelling and Simulation–Sustaining Our Future: Understanding and Living with Uncertainty, MODSIM2011.

Mesgari, E., Asheri, E.,Hooshyar,M.,Hemmesy,M.S.,. (2015). Rainfall Modeling and Forecasting using Neural Networks: A case study of Zab Watershed. *International Journal of Bulletin of Water Resources and Development, 3*(2).

Mishra, P., Khare, D., Mondal, A., Kundu, S., & Shukla, R. (2013). Satistical and probability analysis of rainfall for crop planning in a canal command. *Journal of Agriculture for Sustainable Development, 1*(1), 95-102.

Mohamed, T. M., & Ibrahim, A. A. (2016). Time Series Analysis of Nyala Rainfall Using ARIMA Method. *Journal of Engineering and Computer Science, 17*(1), 1-11.

Momani, P., & Naill, M. (2009). Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. *American Journal of Environmental Sciences, 5*(5), 599.

Montanari, A., Rosso, R., & Taqqu, M. S. (1997). Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water resources research*, *33*(5), 1035-1044.

Mostafaei, H., & Sakhabakhsh, L. (2011). Using SARFIMA and SARIMA models to study and predict the Iraqi Oil production. *Journal of Basic and Applied Scientific Research*, *1*(10), 1715-1723.

Najim, M. (2012). Impact of climate change on culture-based fisheries of seasonal reservoirs in Sri Lanka. *A case study report.*

Nam, W. H., Hong, E. M., & Baigorria, G. A. (2016). How climate change has affected the spatio-temporal patterns of precipitation and temperature at various time scales in North Korea. *International Journal of Climatology, 36*(2), 722-734.

Nanda, S. K., Tripathy, D. P., Nayak, S. K., & Mohapatra, S. (2013). Prediction of rainfall in India using Artificial Neural Network (ANN) models. *International Journal of Intelligent Systems and Applications, 5*(12), 1-22.

Narayanan, P., Basistha, A., Sarkar, S., & Kamna, S. (2013).Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. *Comptes rendus geoscience*, *345*(1), 22-27.

Nirmala, M. (2015). Computational Models for Forecasting Annual Rainfall in Tamilnadu. *Applied Mathematical Sciences*, *9*(13), 617-621.

Nirmala, M., & Sundaram, S. (2010). A Seasonal Arima Model for forecasting monthly rainfall in Tamilnadu. *National Journal on Advances in Building Sciences and Mechanics*, *1*(2), 43-47.

Nirmala, M., & Sundaram, S. (2010).Modeling and predicting the monthly rainfall in Tamilnadu as a seasonal multivarate ARIMA Process. *International Journal of Computer Engineering and Technology,1*(1), 103-111.

Nugroho, A., & Simanjuntak, B. H. (2014). ARMA (Autoregressive Moving Average) Model for Prediction of Rainfall in Re gency of Semarang-CentralJava-Republic of Indonesia. *IJCSI International Journal of Computer Science, 11*(3), 27-32.

OCHA. (2016). Sri Lanka; Floods and Landslides Situation. *Report No:02*. Retrieved June, 2016.

Ogallo, L. A. (1986). Stochastic modelling of regional annual rainfall anomalies in East Africa. *Journal of Applied Statistics, 13*(1), 49-56.

Omidvar, K. (2015). Daily Precipitation Forecasting using Artificial Neural Networks in the Province of Kerman: A case study of Station of Kerman,Baft, and Miandeh, Jiroft. *Journal of Geography and Reggional Development, 12*(23), 47-50.

Omotosho, J. B., Balogun, A., & Ogunjobi, K. (2000). Predicting monthly and seasonal rainfall, onset and cessation of the rainy season in West Africa using only surface data. *International Journal of Climatology, 20*(8), 865-880.

Pallavi, Singh, G. (2016). Review on rainfall forecasting using different Techniques and Algorithms. *International Journal of Innovative Research in computer and Communication Engineering, 4*(3), 2901-2903.

Partheepan, K., Jeyakumar, P., & Manobavan, M. (2005). Development of a Time-Series Model to Forecast Climatic Data in the Batticaloa District, Sri Lanka. *In proceeding of Water Professionals' Day symposium.*

Patel, J., & Parekh, F. (2014). Forecasting Rainfall Using Adaptive Neuro-Fuzzy Inference System (ANFIS). *International Journal of Application or Innovation in Engineering & Management (IJAIEM) ISSN*, 2319-4847.

Pathirana, S., Kawabata, M., & Goonatilake, R. (2009). Study of potential risk of dengue disease outbreak in Sri Lanka using GIS and statistical modelling. *Journal of Rural and Tropical Public Health, 8,* 8-17.

Patrick, D. K., Edmond, P.P., Jean-Marie, T.M., Louis,E.E., Ngbolua, K.N.,. (2014). Prediction of rainfall using autoregressive integrated moving average model;Case of Kinshasa city (Democratic Republic of the Congo) from the period of 1970 to 2009. *Journal of Computation in Biosciences and Engineering, 2*(1),1-6.

Peiris, M. S., & Singh, N. (1996). Predictors for seasonal and nonseasonal fractionally integrated ARIMA models. *Biometrical journal, 38*(6), 741-752.

Peiris, T.S.G, Peiris, T.U.S., & Rajapaksha, S. (2000). Prediction of annual national coconut production-a stochastic approach. *Sri Lanka Journal of Applied Statistics, 1*(1), 25-32.

Peiris, T.U.S, Peiris, T.S.G., & Samita, S. (2000). Arrival dates of southwest monsoon rains-a modelling approach.*Tropical Agricultural Research, 12*, 265-275.

Peiris, T.S.G., Wijeratne, M., Ranasinghe, C., Aanadacumaraswamy, A., Fernando, M., Jayakody, A., & Ratnasiri, J. (2004). *Impact of climate change on coconut and tea industry in Sri Lanka.* Paper presented at the 2nd AIACC regional workshop for Asia and the Pacific, Manila, the Philippines.

Piyasiri, C., Peiris, T., & Samita, S. (2004). Change in Rainfall and Temperature of Low Country Wet Intermediate (IL1) Region of Sri Lanka. *Tropical Agricultural Research, 16*, 271-281.

Popale, P., & Gorantiwar, S. (2014). Stochastic Generation and Forecasting of Weekly Rainfall for Rahuri Region. *International Journal of Innovative Research in science, Engineering and Technology*, *3*(4),185-196.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association, 85*(410), 338-344.

Portmann, R. W., Solomon, S., & Hegerl, G. C. (2009). Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proceedings of the National Academy of Sciences, 106*(18), 7324-7329.

Premalal, K. H. M. S. (2013). Change and Behavior of Rainfall Pattern in Sri Lanka with Southern Oscillation (SO) -El Nino and La Nina *SMRC Report -No.52*: SAARC Meteorological Resrach Centre(SMRC).

Ramesh, K., Basu, S., & Begum, Z. (1996). Objective determination of onset, advancement and withdrawal of the summer monsoon using large-scale forecast fields of a global spectral model over India. *Meteorology and Atmospheric Physics, 61*(3-4), 137-151.

Rasel, H. M., Imteaz, M. A., Hossain, I., & Mekanik, F. (2015). Comparative study between linear and non-linear modelling techniques in rainfall forecasting for South Australia.

Rathnayake, V., Premaratne, H., & Sonnadara, D. (2011). Performance of neural networks in forecasting short range occurrence of rainfall. *Journal of the National Science Foundation of Sri Lanka, 39*(3), 251-260.

Ray, B. K. (1993). Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. *International Journal of Forecasting, 9*(2), 255-269.

Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA (p, d, q) model using the smoothed periodogram. *Journal of time series analysis, 15*(3), 335-350.

Reisen, V. A., Sarnaglia, A. J. Q., Reis Jr, N. C., Lévy-Leduc, C., & Santos, J. M. (2014). Modeling and forecasting daily average PM10 concentrations by a seasonal long-memory model with volatility. *Environmental Modelling & Software, 51*, 286-295.

Reisen, V. A., Zamprogno, B., Palma, W., & Arteche, J. (2014a). A semiparametric approach to estimate two seasonal fractional parameters in the SARFIMA model. *Mathematics and Computers in Simulation, 98*, 1-17.

Reiter, A., Weidinger, R., & Mauser, W. (2012). Recent climate change at the upper Danube—a temporal and spatial analysis of temperature and precipitation time series. *Climatic change, 111*(3-4), 665-696.

Selvaraj, R. S., Sachithananthem,C.P.,&Thamizharasan,K.,. (2013). Modeling and Predicting Total Ozone Column and Rainfall in Kodaikanal,Tamilnadu by ARIMA process. *International Journal of Engineering and Computer Science 2*(8), 2521-2526.

Sena Jr, M., Reisen, V., & Lopes, S. (2006). Correlated errors in the parameters estimation of the ARFIMA model: A simulated study. *Communications in Statistics-Simulation and Computation, 35*(3), 789-802.

Shamsnia, S. A., Shahidi, N., Liaghat, A., Sarraf, A., & Vahdat, S. F. (2011). *Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran).* Paper presented at the International Conference on Environment and Industrial Innovation. IPCBEE.

Sharda, V., & Das, P. (2005). Modelling weekly rainfall data for crop planning in a sub-humid climate of India. *Agricultural water management, 76*(2), 120-138.

Sharma, M. A., & Singh, J. B. (2010). Use of probability distribution in rainfall analysis. *New York Science Journal, 3*(9), 40-49.

Shi, W., Tao, F., Liu, J., Xu, X., Kuang, W., Dong, J., & Shi, X. (2014). Has climate change driven spatio-temporal changes of cropland in northern China since the 1970s? *Climatic change, 124*(1-2), 163-177.

Singh, B., Rajpurohit, D., Vasishth, A., & Singh, J. (2012). Probability analysis for estimation of annual one day maximum rainfall of Jhalarapatan area of Rajasthan, India. *Plant Archives, 12*(2), 1093-1100.

Singh, K., & Xie, M. (2008). Bootstrap: a statistical method. *Unpublished manuscript, Rutgers University, USA. Retrieved from http://www. stat. rutgers. edu/home/mxie/RCPapers/bootstrap. pdf*.

Solgi, A., Nourani, V., & Pourhaghi, A. (2014). Forecasting Daily Precipitation Using Hybrid Model of Wavelet-Artificial Neural Network and Comparison with Adaptive Neurofuzzy Inference System (Case Study: Verayneh Station, Nahavand). *Advances in Civil Engineering, 2014*.

Solomon, S., Qin, D., Manning, M., Averyt, K., & Marquis, M. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC* (Vol. 4): Cambridge university press.

Soltani, S., Modarres, R., & Eslamian, S. (2007). The use of time series modeling for the determination of rainfall climates of Iran. *International Journal of Climatology, 27*(6), 819-829.

Sovoe, S. (2015). Modelling West African Total Precipitation Depth: A Statistical Approach*. Journal of Science and Technology, 3.*

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53,165–188.

Teimoorzadeh, K., Shiri, J., & Shiri, A. A. (2015). Forecasting Daily Precipitation Values, Using Wavelet Conjunction Models (Case Study: Tabriz and Marageh Stations, Iran). *Science, 34*(4), 265-269.

Terzi, Ö., & Çevik, E. (2012). Rainfall estimation using Artificial Neural Network Method, *SDU International Journal of Technological Science*, *4*(1).

Trenberth, K. E. (2005). The impact of climate change and variability on heavy precipitation, floods, and droughts. *Encyclopedia of hydrological sciences*.

Tularam, G. A. (2010). Relationship between EL NIno Southern Oscillation Index and Rainfall (Queensland, Australia). *International Journal of Sus.Dev.Plann, 5*(4), 378-391.

Vamsidhar, E., Varma, K., Rao, P. S., & Satapati, R. (2010). Prediction of rainfall using backpropagation neural network model. *International Journal on Computer Science and Engineering*, *2*(4), 1119-1121.

Varathan, N., Perera, K., & Wikramanayake, N. (2010). Statistical modeling of daily extreme rainfall in Colombo. *In Proceeding of International Conference of Sustainable Built Environment.*

Veenstra, J.Q. & McLeod, A.I. (2013). Persistence and Anti-persistence : Theory and Software (PhD Thesis).

Waidyarathne, K. P., Peiris, T.S.G. & Samita, S.,. (2006). Shift in onset of first inter monsoon rain in coconut growing areas in Sri Lanka. *Journal of Tropical Agricultural Research, 18,* 1-12.

Wang, H., Wang, C., Lin, X., & Kang, J. (2014). An improved ARIMA model for precipitation simulations. *Nonlinear Processes in Geophysics, 21*(6), 1159-1168.

Wang, S., Feng, J., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer Modelling, 58*(3), 677-683.

Weerasinghe, H., Premaratne, H., & Sonnadara, D. (2010). Performance of neural networks in forecasting daily precipitation using multiple sources. *Journal of the National Science Foundation of Sri Lanka, 38*(3), 163-170.

Wickramagamage, P. (2010). Seasonality and spatial pattern of rainfall of Sri Lanka: Exploratory factor analysis. *International Journal of Climatology, 30*(8), 1235-1245.

Wijeratne, M., Anandacoomaraswamy, A., Amarathunga, M., Ratnasiri, J., Basnayake, B., & Kalra, N. (2007). Assessment of impact of climate change on productivity of tea (Camellia sinensis L.) plantations in Sri Lanka. *Journal of the National Science Foundation of Sri Lanka, 35*(2), 119-126.

Yaya, O. S., & Fashae, O. A. (2014). Seasonal fractional integrated time series models for rainfall data in Nigeria. *Theoretical and applied climatology, 120*(1-2), 99-108.

Yu, J. W., & Yu, J. (2012). Rainfall time series forecasting based on Modular RBF Neural Network model coupled with SSA and PLS. *Journal of Theoretical and Applied Computer Science, 6*(2), 3-12.

Yusof, F., & Kane, I. L. (2012). Modeling monthly rainfall time series using ets state space and sarima models. *International Journal of Current Reseach,4*(9), 195-200.

Yusof, F., & Kane, I. L. (2013). Volatility modeling of rainfall time series. *Theoretical and applied climatology, 113*(1-2), 247-258.

Zafor, M. A., Chakraborty, A., Muniruzzaman, S. M., & Mojumdar, S. R. (2016). Rainfall Forecasting in Northeastern part of Bangladesh Using Time Series ARIMA Model. *Research Journal of Engineering Sciences, 5(*3), 17-31.

Zakaria, S., Al-Ansari, N., Knutsson, S., & Thafer, A.-B. (2012). ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. *Journal of Earth Sciences and Geotechnical Engineering, 2*(3), 25-55.

Zubair, L. (2002). El Nino–southern oscillation influences on rice production in Sri Lanka. *International Journal of Climatology, 22*(2), 249-260.

Zubair, L., Rao, S. A., & Yamagata, T. (2003). Modulation of Sri Lankan Maha rainfall by the Indian Ocean Dipole. *Geophysical Research Letters, 30*(2).

Zubair, L., & Ropelewski, C. F. (2006). The strengthening relationship between ENSO and northeast monsoon rainfall over Sri Lanka and southern India. *Journal of Climate, 19*(8), 1567-1575.

Zubair, L., Siriwardhana, M., Chandimala, J., & Yahiya, Z. (2008). Predictability of Sri Lankan rainfall based on ENSO. *International Journal of Climatology, 28*(1), 91-101.

APPENDIX - 1

WEEKLY RAINFALL PATTERNS AND THEIR DISTRIBUTIONS

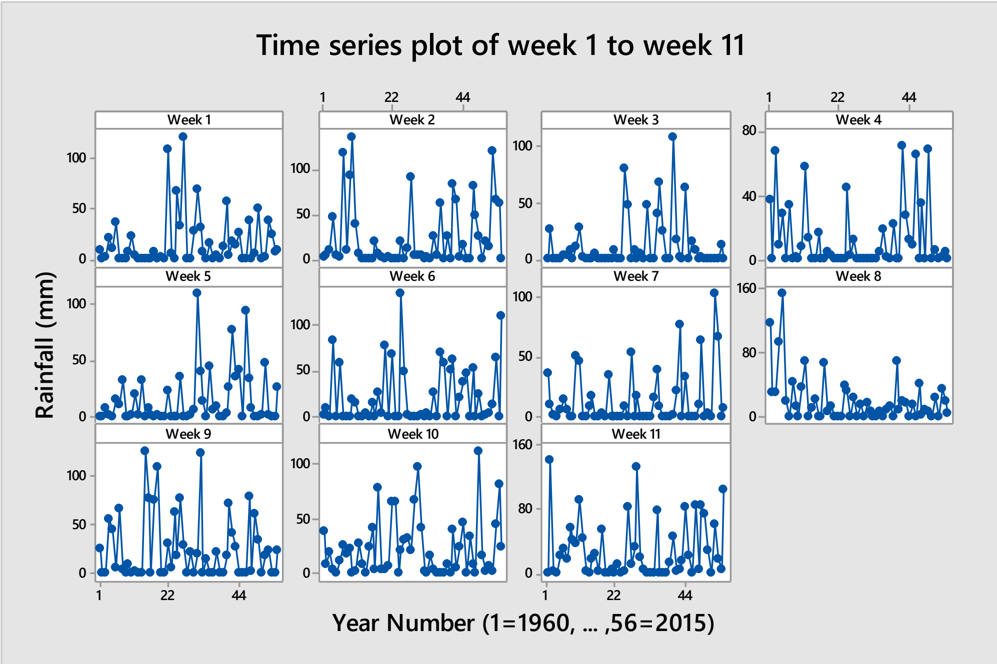

Figure A1.1: The time series plots of the weekly rainfall of week 1 -11

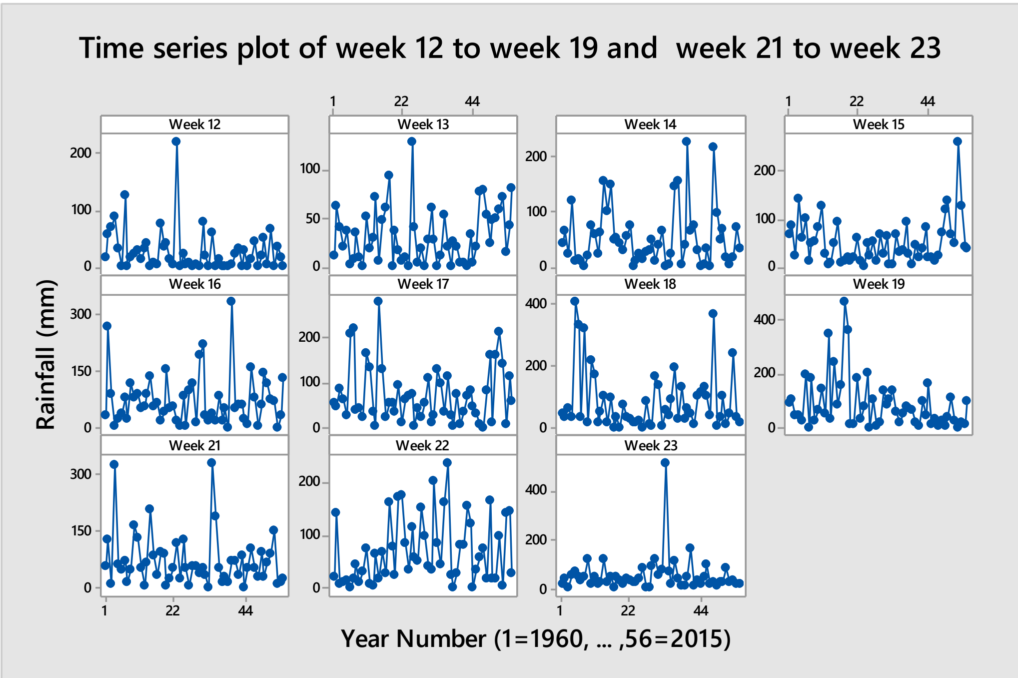

Figure A1.2: The time series plots of the weekly rainfall of weeks 12-19 and week 21 - 23.

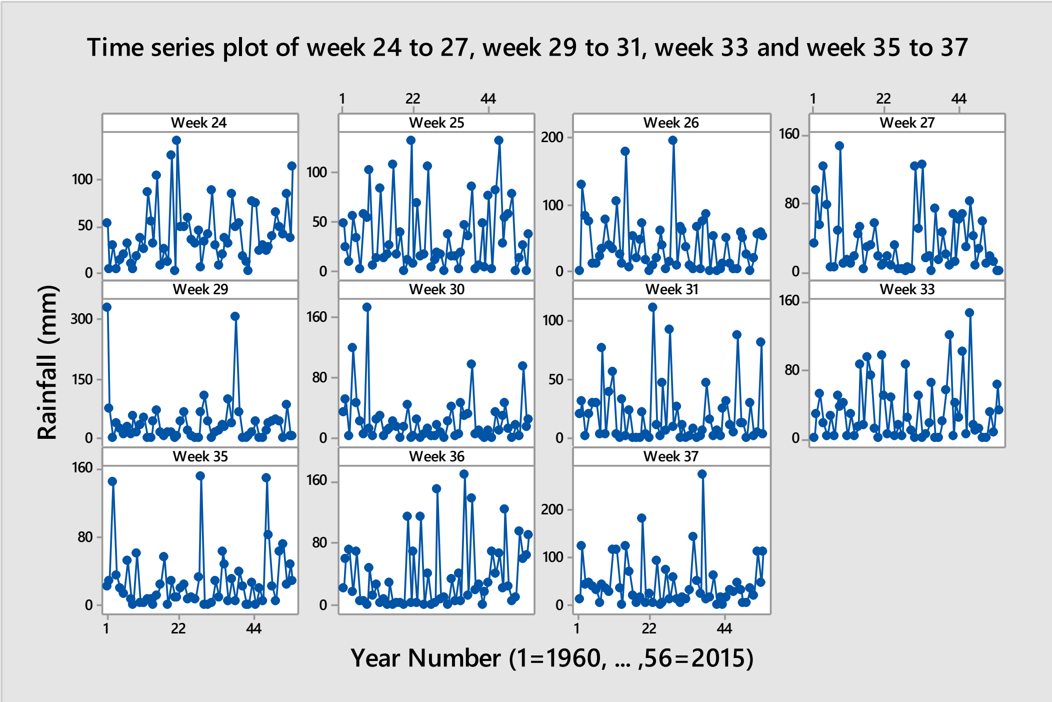

Figure A1.3: The time series plots of the weekly rainfall of weeks 24-27, 29-31, 33, 35-37

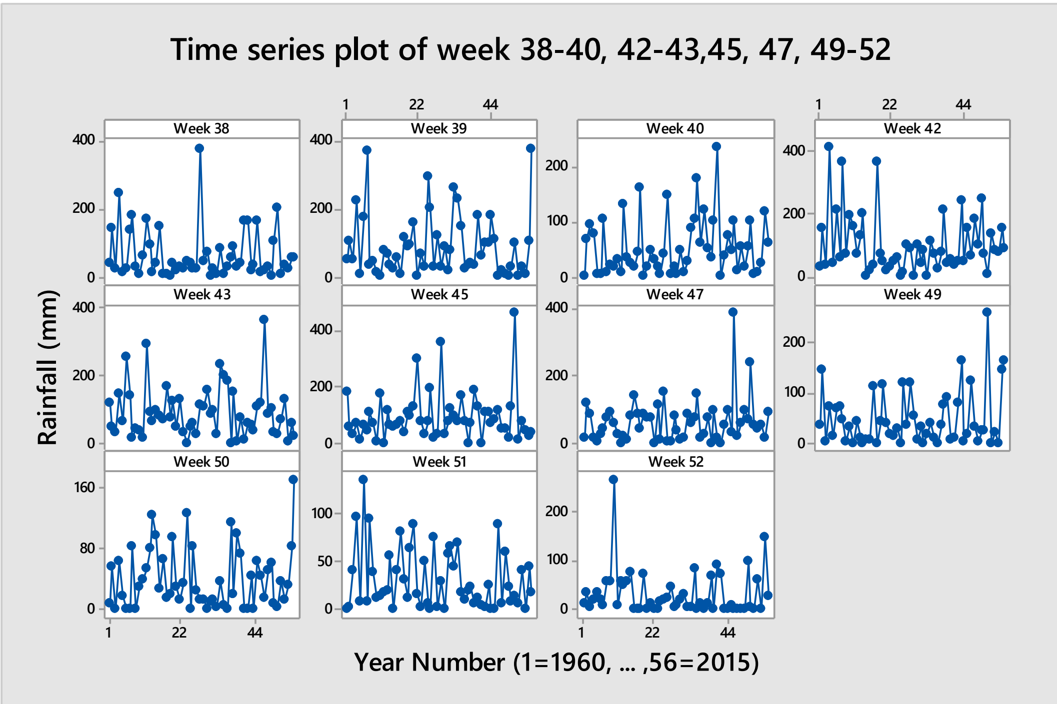

Figure A1.4: The time series plots of the weekly rainfall of weeks 38-40, 42-43, 45,47 and 49-52.

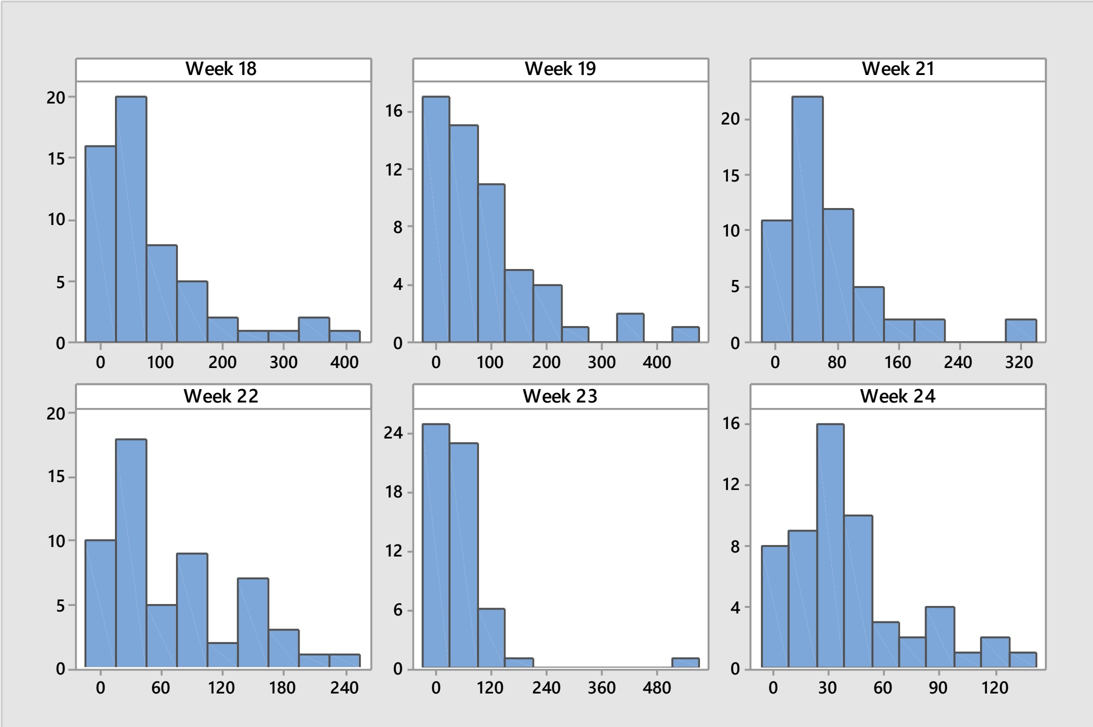

Figure A1.5: Histogram of the total weekly rainfall for week numbers: week 18,19 and 21-24 in SWM

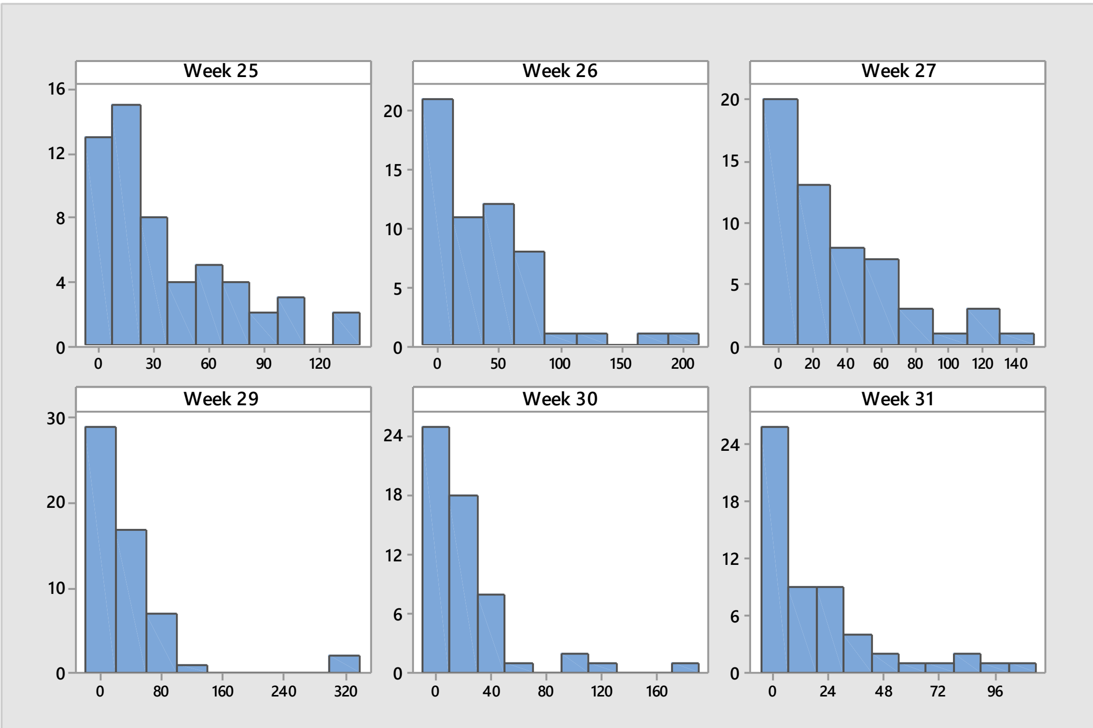

Figure A1.6: Histogram of the total weekly rainfall for week numbers: week 25-27 and 29-31 in SWM

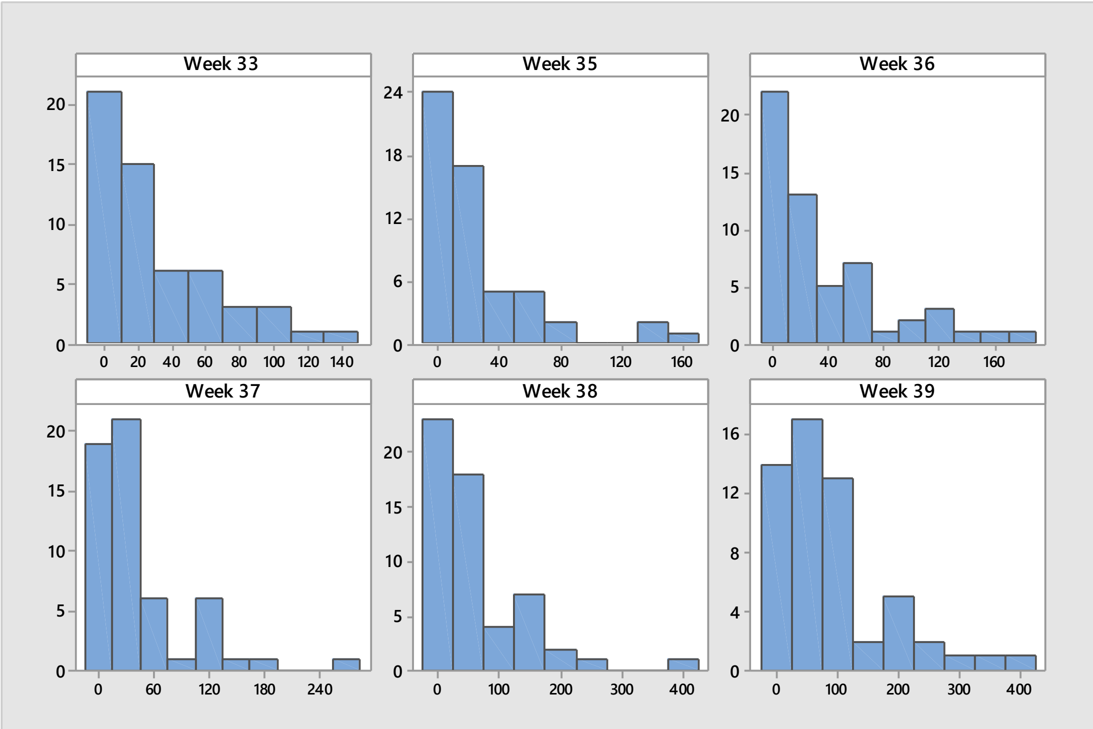

Figure A1.7: Histogram of the total weekly rainfall for week numbers: week 33 and 35-39 in SWM

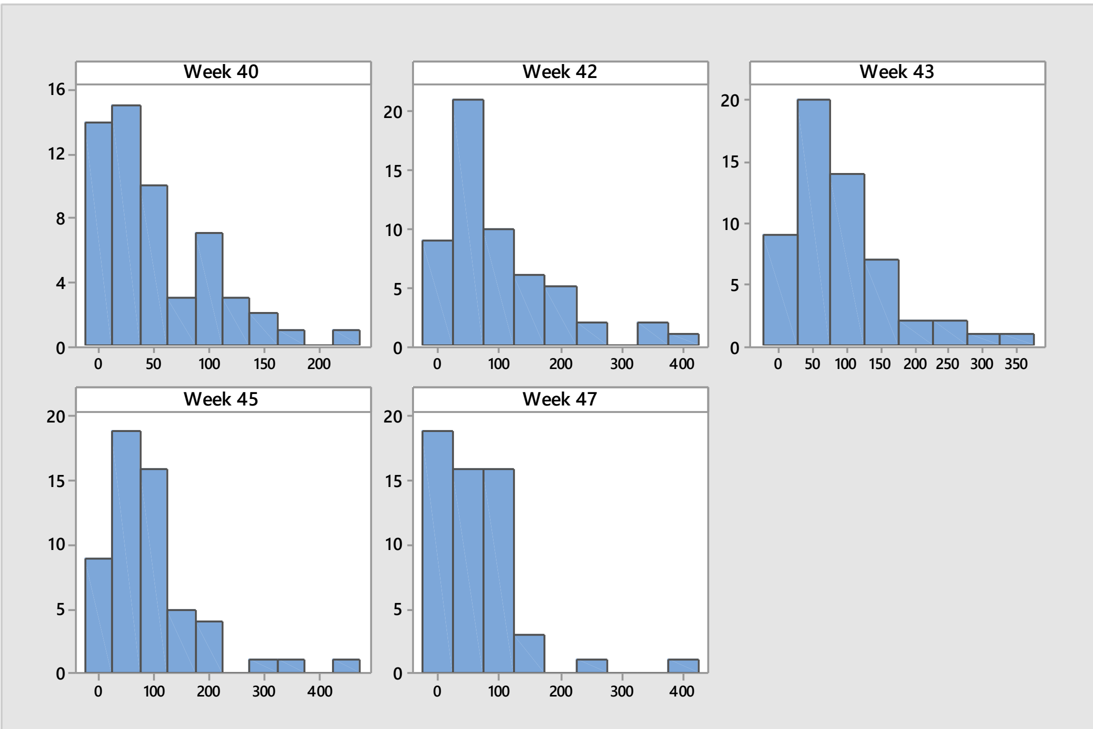

Figure A1.8: Histogram of the total weekly rainfall for week numbers: week 40, 42, 43, 45 and 47 in SIM

APPENDIX - 2

AUTO CORRELATION FUNCTIONS OF WEEKLY RAINFALL

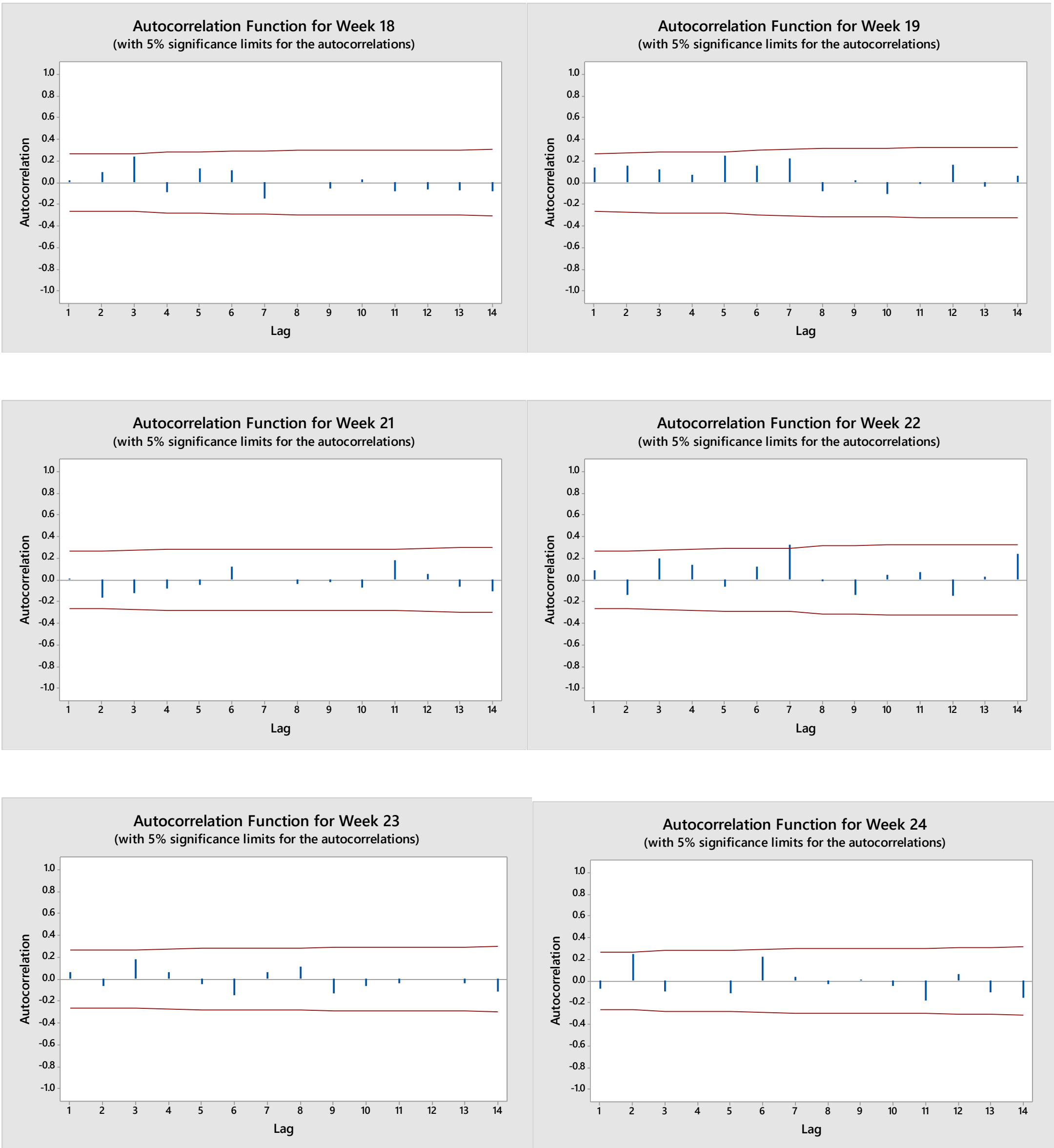

Figure A2.1: The auto correlation plots of the weeks 18,19 and 21-24 in SWM

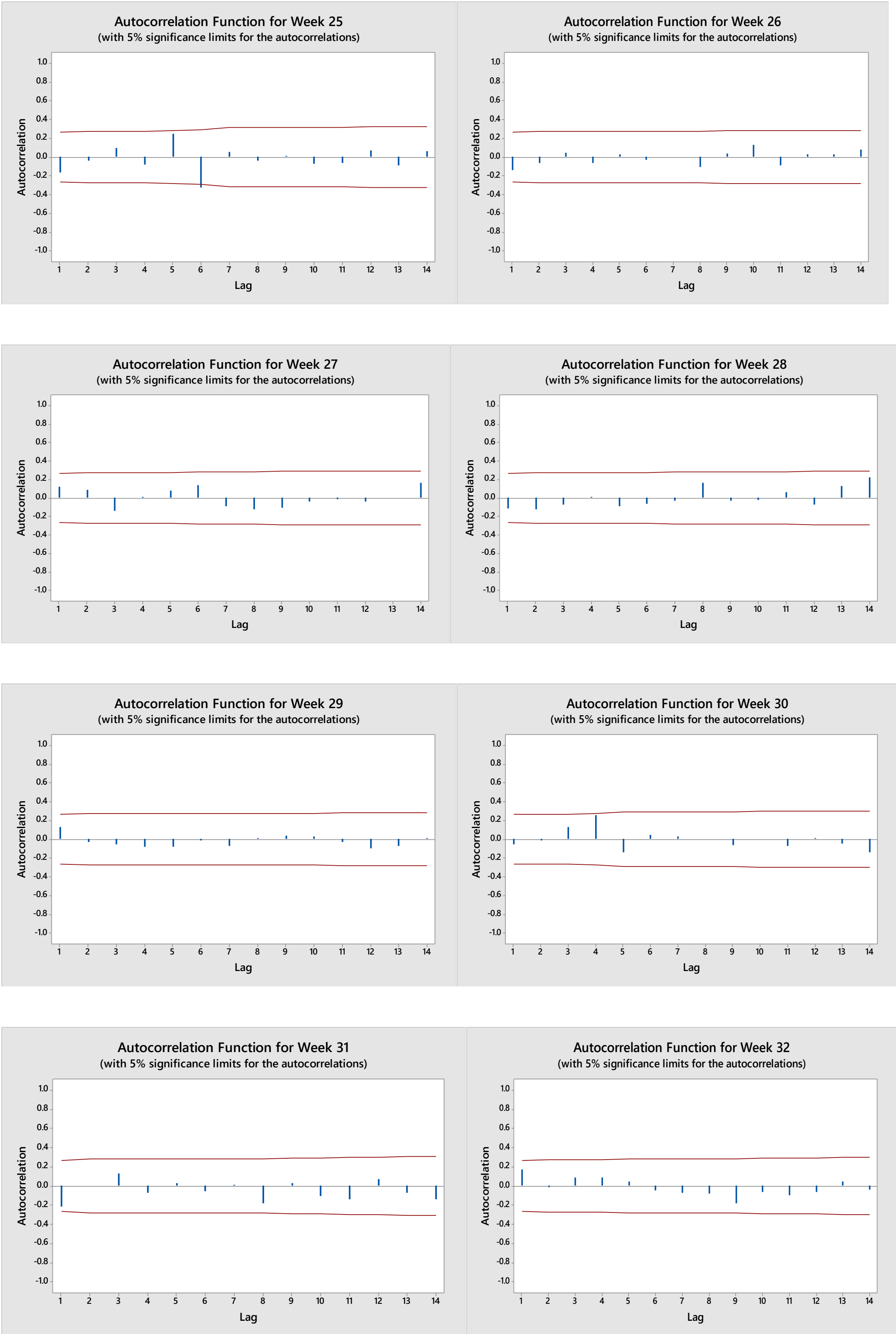

Figure A2.2: The auto correlation plots of the weeks: 25-32 in SWM

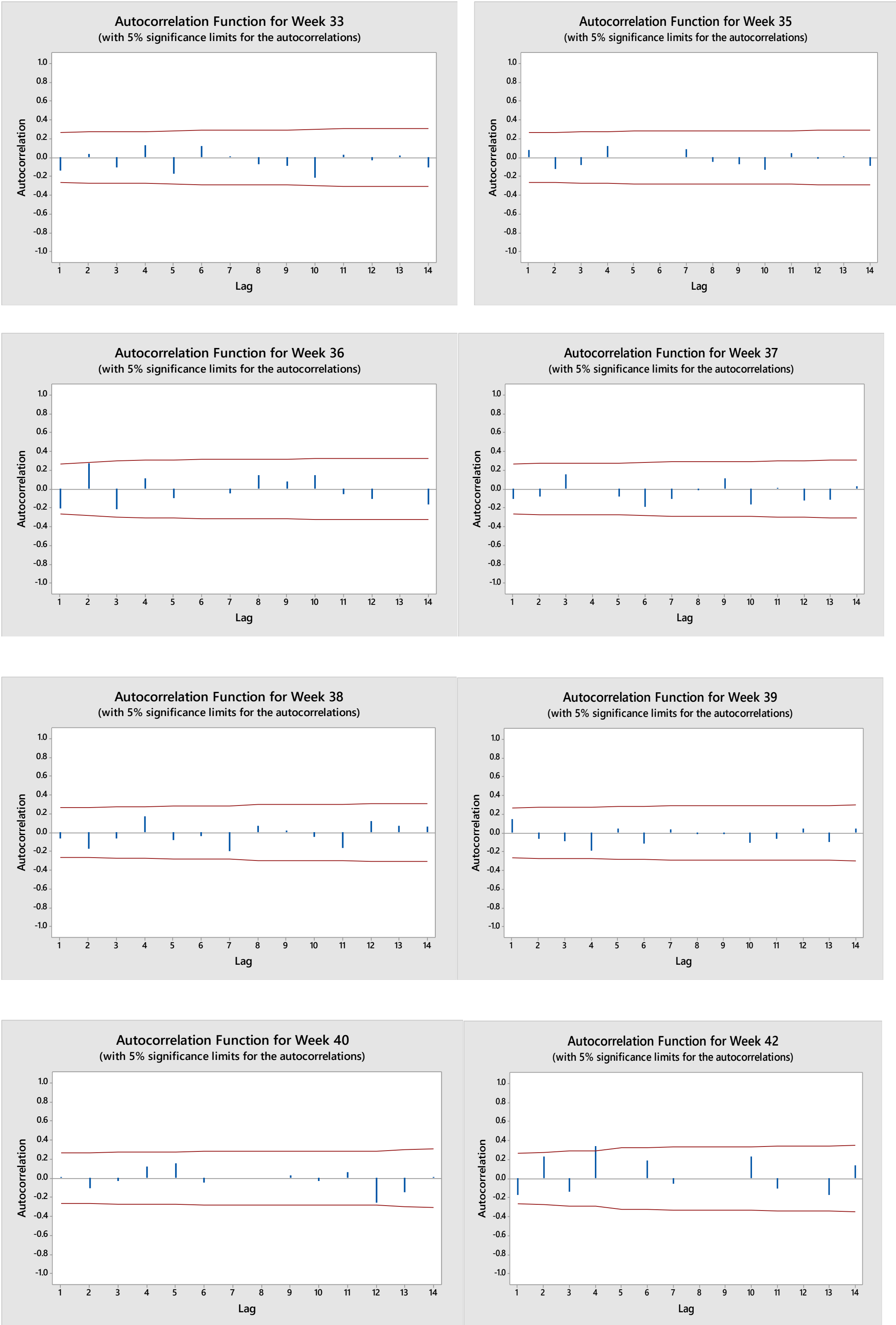

Figure A2.3: The auto correlation plots of the weeks: 33,35-39 in SWM & 40 and 42 in SIM

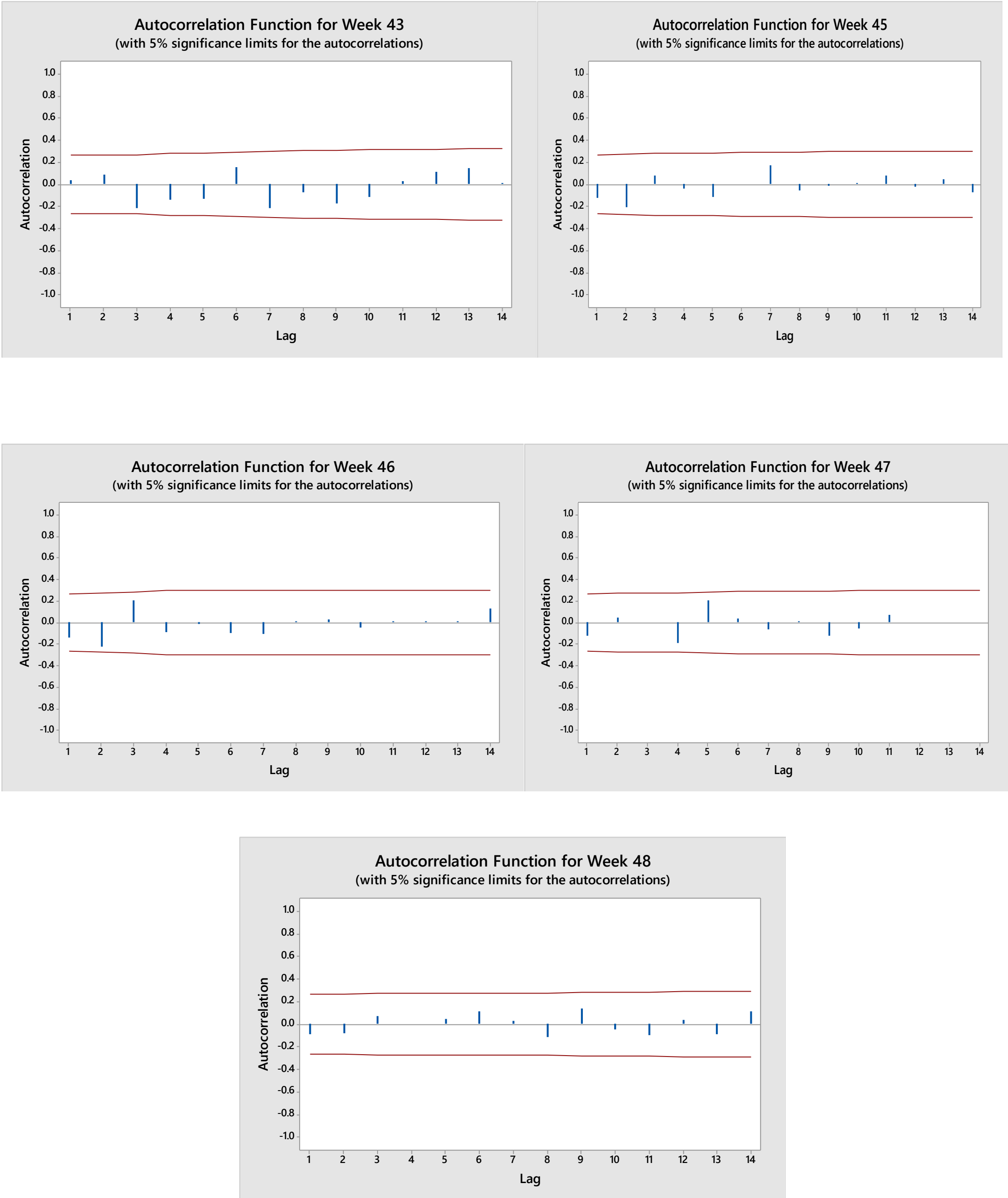

Figure A2.4: The auto correlation plots of the weeks: 43 and 45-48 in SIM

APPENDIX - 3

INDICES FORMATION

To evaluate the forecasting performance, an index was developed based on the absolute forecasting error and the corresponding calculations were listed below for each model separately.

Table A3.1: Indices for absolute error by Model 1 - ARFIMA (4,0.05792,4)

Absolute Forecasting Error in (mm)	Weights	The number of weeks				
		2015	**2016**	**2017**	**2015-2016**	**2015-2017**
0-10	9	10	7	10	17	27
11--15	7	6	8	4	14	18
16-20	5	6	4	7	10	17
21-25	3	4	8	8	12	20
26-30	1	6	2	2	8	10
31-35	-1	1	4	2	5	7
36-40	-3	4	1	1	5	6
41-45	-5	1	2	2	3	5
46-50	-7	2	4	3	6	9
More than 50	-9	12	12	13	24	37
	Index	**40**	**12**	**26**	**52**	**78**

Index (using 2015 data) = $\mathbf{9}\times10+\mathbf{7}\times6+\mathbf{5}\times6+\mathbf{3}\times4+\mathbf{1}\times6\mathbf{-1}\times1\mathbf{-3}\times4-5\times1\mathbf{-7}\times2\mathbf{-9}\times12 = \mathbf{40}$

Index (using 2016 data) = $\mathbf{9}\times7+\mathbf{7}\times8+\mathbf{5}\times4+\mathbf{3}\times8+\mathbf{1}\times2\mathbf{-1}\times4\mathbf{-3}\times1\mathbf{-5}\times2\mathbf{-7}\times4\mathbf{-9}\times12 = \mathbf{12}$

Index

(using 2015-2017 data) = $9\times27+\mathbf{7}\times18+\mathbf{5}\times17+\mathbf{3}\times20+\mathbf{1}\times10\mathbf{-1}\times7\mathbf{-3}\times6\mathbf{-5}\times5\mathbf{-7}\times9$

$\mathbf{-9}\times37$

$= 78$

Table A3.2: Indices for absolute error by Model 2 - ARFIMA (5,0.05999,5) for deseasonalized data

Absolute Forecasting Error in (mm)	Weights	The number of weeks				
		2015	2016	2017	2015-2016	2015-2017
0-10	9	17	9	11	26	37
11--15	7	2	5	7	7	14
16-20	5	6	3	6	9	15
21-25	3	2	7	3	9	12
26-30	1	2	7	6	9	15
31-35	-1	4	4	0	8	8
36-40	-3	1	2	1	3	4
41-45	-5	2	1	1	3	4
46-50	-7	3	3	6	6	12
More than 50	-9	13	12	11	25	36
	Index	**50**	**15**	**44**	**65**	**109**

Index (using 2015 data) = **9**×17+**7**×2+**5**×6+**3**×2+**1**×2-**1**×4-**3**×1-5×2-**7**×3-**9**×13 = **50**

Index (using 2017 data) = **9**×11+**7**×7+**5**×6+**3**×3+**1**×6-**1**×0-**3**×1-**5**×1-**7**×6-**9**×11 = **44**

Table A3.3: Indices for absolute error by Model 3 - ARFIMA (4,0.116577,6)-GARCH (1,1)

Absolute Forecasting Error in (mm)	Weights	The number of weeks				
		2015	2016	2017	2015-2016	2015-2017
0-10	9	10	8	10	18	28
11--15	7	3	5	4	8	12
16-20	5	6	5	4	11	15
21-25	3	6	5	6	11	17
26-30	1	7	7	7	14	21
31-35	-1	3	3	2	6	8
36-40	-3	1	2	1	3	4
41-45	-5	2	2	3	4	7
46-50	-7	2	3	2	5	7
More than 50	-9	12	12	13	24	37
	Index	**28**	**6**	**12**	**34**	**46**

Index (using 2016 data) = **9**×8+**7**×5+**5**×5+**3**×5+**1**×7-**1**×3-**3**×2-5×2-**7**×3-**9**×12 = **6**

Index (using 2015-2017 data) = **9**×28+**7**×12+**5**×15+**3**×17+**1**×21-**1**×8-**3**×4-**5**×7-**7**×7 -**9**×37

= **46**

Table A3.4: Indices for absolute error by Model 4 - ARFIMA (6,0.243588,5)-GARCH (1,1) for deseasonalized

Absolute Forecasting Error in (mm)	Weights	The number of weeks				
		2015	2016	2017	2015-2016	2015-2017
0-10	9	16	8	11	24	35
11--15	7	6	5	7	11	18
16-20	5	5	4	5	9	14
21-25	3	2	6	5	8	13
26-30	1	2	9	5	11	16
31-35	-1	1	3	1	4	5
36-40	-3	3	2	1	5	6
41-45	-5	1	1	3	2	5
46-50	-7	4	3	2	7	9
More than 50	-9	12	11	12	23	35
	Index	**68**	**20**	**52**	**88**	**140**

Index (using 2015 data)= **9**×16+**7**×6+**5**×5+**3**×2+**1**×2-**1**×1-**3**×3-5×1-**7**×4-**9**×12 = **68**

Index (using 2016 data)= **9**×8+**7**×5+**5**×4+**3**×6+**1**×9-**1**×3-**3**×2-**5**×1-**7**×3-**9**×11 = **20**

Table A3.5: Indices for absolute error by Model 5- Adjusted SARFIMA (1,0.115677,1) × (1,0.17075,0) -GARCH (1,1)

Absolute Forecasting Error in (mm)	Weights	The number of weeks				
		2015	2016	2017	2015-2016	2015-2017
0-10	9	12	8	8	20	28
11--15	7	4	5	6	9	15
16-20	5	4	4	6	8	14
21-25	3	5	8	5	13	18
26-30	1	6	4	3	10	13
31-35	-1	4	2	6	6	12
36-40	-3	3	4	2	7	9
41-45	-5	1	3	3	4	7
46-50	-7	3	5	3	8	11
More than 50	-9	10	9	10	19	29
	Index	**48**	**10**	**24**	**58**	**82**

Index (using 2015 data) = **9**×12+**7**×4+**5**×4+**3**×5+**1**×6-**1**×4-**3**×3-5×1-**7**×3-**9**×10 = **48**

Index (using 2017 data) = **9**×8+**7**×6+**5**×6+**3**×5+**1**×3-**1**×6-**3**×2-**5**×3-**7**×3-**9**×10 = **24**