International Journal of Water Resources and Arid Environments 11(1): 62-80, 2022
ISSN 2079-7079

Application of ARIMA Models in Forecasting Average Monthly Rainfall in Birzeit, Palestine

[1]*Deeb Aborass*, [2]*Hassan Abu Hassan*, [3]*Ibrahim Sahalash and* [4]*Hussein Al-Rimmawi*

[1]MA degree in Applied Statistics and Data Mining, Birzeit University, Palestine
[2]Department of Mathematics, Birzeit University, Palestine
[3]Chemistry Department, Birzeit University, Palestine
[4]Geography Department, Birzeit University, Palestine

Abstract: Town of Birzeit is located on top mountains of the district of Ramallah and Al-Bireh, Palestine. Its elevation is about 800 meters above sea level and enjoys Mediterranean climate. The whole West Bank of Palestine in general and Birzeit in specific suffers from lack of water supplies especially in summer. It becomes necessary to analyze the present rainfall pattern and to predict its future pattern too since it has not been tackled before. In this study authors have reported the Box-Jenkins ARIMA methodology and comparative study of ETS model for rainfall forecasting at Birzeit for the period which extendsfrom September -2003 to August-2021. This study aimed to use two types of rain forecasting models and to select best one. The study concluded that the seasonal ARIMA model (11.0.2) x (11.0.2) is the best in predicting rainfall for the next four years, from September -2021 to August -2025. In Birzeit, rainfall starts in October and gradually increases in December, January and February. However, it starts declining duringmonths of March, April and May.

Key words:

INTRODUCTION

Water is the backbone of life and is necessary for any living being.God says

"And made from water every living thing? Then will they not believe?"Al-Anbya 30.

Rainfall is the main source of nourishment in Palestine, in addition to other importantsources such as rivers, lakes, underground, springs and dew. Despite these sources, Palestine suffers from limited water availability since all water sources are controlled by the Israeli occupation authorities. Hence comes the important of this study which tackles the present and future patterns of Birzeit rainfall. This town is famouswith olive groves and olive oil production. And it is place of "Birzeit University"with has about 15000 students.

Birzeit depends on annual rainfall and springs for its various uses. In fact, there are five springs - Al-Hamam, Filifila, Al-Qaws, Uqban and Al-Marj. However, this town and its university rely on network water which is delivered by Water Authority in Ramallah. In winter, the town is exposed to rainy southwesterly winds and sometimes to relatively cold dry northeasterly winds. The average recorded rainfall is about 651millimeters. In general, the average temperature in winter rarely reaches zero Celsius and in summer it rarely exceeds 35 Celsius, The annual average temperature ranges between 5 in winter and 25 Celsius in summer. However, at present Birzeit witness's high temperature during summer and this is related to global climate change.

Rainfall forecast helps authorities and farmers to prepare well for the agricultural season. Forecasting is one of the most important decision-making tools and the most important elementin the planning process for future. In order to make the right decision, past and present available rainfall data may be analyzed and predicted for years to come. For that, future forecasting leads to raising degree of confidence in the decision making. There are many methods that are used in building forecast models and different time series analysis methods are among the most used such asAutoregressive Moving Average ARMA, Autoregressive Integrate Moving Average

Corresponding Author: Deeb Aborass, MA Degree in Applied Statistics and Data Mining, Birzeit University, Palestine.
E-mail: deshataweew@gmail.com.

ARIMA. In this research, time series forecasting is addressed using method represented by Box Jenkins models.

Problem Statement: Since Birzeit Suffers from insufficient water resources, Therefore, Rainfall Prediction process becomes necessary. In fact, time series is directly affected by selection of the appropriate model for data analysis, as this step directly affects the accuracy of prediction results. Time series data in different sectors are mostly non-linear and suffer, sometimes for its being random and turbulent.In order to obtain prediction models for time series data that have the ability to depict reality and high accuracy in future predictions, they must take all considerations related to data from being linear or non-linear, data quality and other influences related to data. Therefore, the research problem can be represented in the following questions:

- To what extent can Box Jenkins models deal with the reality of Birzeit rainfall time series data in terms of linearity and non-linearity?
- What are the advantages and failures of models built with these methods?
- Is it possible to set criteria through which a specific model that is compatible with the nature of the data?

Importance of the Study: This research explores prediction of recorded rainfalldata in Birzeit. Such prediction helpsauthorities and concerned people, to prepare for the coming rain seasons. In addition, the importance of the research lies in dealing with different time series methods, such as Box-Jenkins and its efficiency for time series prediction of recorded rainfall at Birzeit weather station. The importance of this study is because it is a kind of trade-off between Box-Jenkins method and Exponential Smoothing State Space (ETS).

Research Hypothesis:

- Recorded rainfall at Birzeit Weather station from 2003 to 2021 may be Predicted using Box-Jenkins models for few years to come.
- Box-Jenkins models and Exponential Smoothing State Space (ETS)models may be used together to increase prediction accuracy. However, ETS models are more efficient and more accurate than Box Jenkins models.

Research Methodology and Tools Used: In this research descriptive and analytical will be tackled and explained. Therefore, the research is divided into two aspects. Theoretical aspect addressees the theoretical foundations of time series models in terms of general form and stages. As for the applied side, an applied study was conducted on recorded rainfalldata in Birzeit weather station from Sep 2003 to Aug 2021. A model will be built model to predict rainfall in Birzeit for coming few years. The last part includes the most important conclusions, recommendations. and tools used R-studio software.

Study Area: Birzeit enjoys Mediterranean climate. In winter, the town is exposed to rainy southwesterly winds and sometimes to relatively cold dry northeasterly winds. Average rainfall is about 500 millimeters per year. In general, average temperature in winter rarely reaches zero degrees Celsius and in summer it rarely exceeds 35 degrees Celsius. Annual average temperature ranges between 5 to 25 Celsius. In early April, Khamasin dry winds starts to blow and it usually carries dust. In fact, they are originated from African Sahara. Furthermore, Palestine receives another dry winds which blow from Jordan and affect plantations in case they blow in grow season.

In general, Birzeit climate is pleasant and refreshing in summer. Table (1) illustrates summary of monthly temperature Ramallah and Al-Bireh Governorate and Figure (1) shows location of Ramallah and Al-Bireh within Palestine.

Literature Review: Auto Regression Integrated Moving Average (ARIMA) and Seasonal Auto Regression Integrated Moving Average (SARIMA) models have largely been used to analyze weather variables such as precipitation and temperature [2]. Hasan [3] used time series, regression and ARIMA methods to statistically analyze rainfall in Sudan for the period 1965-2005. The study concluded that Sudanese rainfall has gradient pattern in which it more in the south and less in the north. The revealed result showed that exponential model is the best to represent any change in rainfall patterns in Sudan. Abd Elseed [4] employed Box – Jenkins time series models (ARIMA) in order to select and examine the most appropriate model for prediction of rain falls in Wad Medani city in Sudan. The study revealed that the series was stationary at the raw data level and the results showed decrease in rainfall inWad Medani.In Al-Ghab region, Syria Khadam, Abd Allah and Hana [5] used ARIMA to analyze climate data and to predict their values

Table 1: Ramallah and Al-Bireh Governorate Monthly Temperature

Climatic data for Ramallah													
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	annual rate
Max degree °C	23.4	25.3	27.6	35.3	37.2	36.8	40.6	44.4	37.8	33.8	29.4	26	44.4
Average maximum temperature °C	11.8	12.6	15.4	21.5	25.3	27.6	29	29.4	28.2	24.7	18.8	14	21.5
Average minimum temperature °C	6.4	6.4	8.4	12.6	15.7	17.8	19.4	19.5	18.6	16.6	12.3	8.4	13.5
Min degree °C	-6.7	-2.4	-0.3	0.8	7.6	11	14.6	15.5	13.2	9.8	1.8	0.2	-6.7
Rainfall	133.2	118.3	92.7	24.5	3.2	0	0	0	0.3	15.4	60.8	105.7	554.1

Fig. 1: Ramallah and Al-Bireh Governorate
Source: Palestineremembered, 2022 [1]

for next five years' period. The study revealed decline in February and March rainfall which affects plant production. ARIMA model also showed that average recorded temperature will remain in about the same levels. Hussein [6] used time series for rainfall averages are not seasonal and random which meant that is not affected by time and has steady mean and variance. The study also showed that ARIMA is the best model to analyze Iraqi rainfall. Hayek and Hamad [7] carried out a study using Box-Jenkins models in order to analyze and predict the incoming monthly water volumes for fifteen years to Al-Aroos River which is located in Syrian coast. They found that the best model SARIMA (0, 1, 2) (1, 2, 1)12. which gives the nearest prediction of measured actual data. Hayek and Ammar [8]employed Box – Jenkins time series models (ARIMA) in order to predict future rainfall patterns of Tartous weather station, Syria. The study showed that ARIMA model of (3, 0, 4) is the best for such prediction to cover 20 years' period. Afrifa-Yamoah [9] found thatSARIMA (1, 0, 0)×(0, 1, 2)12 was an appropriate model for predicting monthly average temperature for the Brong Ahafo Region of Ghana from 1975 to 2009. Author obtained data from theDepartment of Meteorology and Climatology in the BA Region.Graham & Mishra [10] conducted a study using ARIMA to predict rainfall parameters on monthly scale for the period 1985-2015 at Allahabad, Uttar Pradesh, India. The study revealed that Box-Jenkins methodology was the appropriate tool to forecast rainfall in Allahabad for five years' period despite of less accuracy because of some missing values.Akinboblal, Okogbue; & Ayansolal [11] used Seasonal Autoregressive Integrated Moving Average (SARIMA) model to predict the monthly rainfall spanning a period of 30 years (1981-2010) for fourteen Nigerian weather stations. Data were collected from the Nigerian Meteorological Agency. Authors report that the results were useful for forecasting rainfall patterns and provided great help for decision makers to formulate policies to mitigate the problems of water resources management, soil erosion, flooding and drought. Balibey & Turkeyilmaz [12] believe that Turkey is very sensible country for climate variation. For that they obtained data from Turkish State Meteorological Service (TSMS) for 1999-2010 period and carried out a study on historical data of total monthly precipitation. In their study they applied Box-Jenkins ARIMA approach and found out that the sinusoidal model predictions are more appropriate than ARIMA technique for monthly averages and contribute to decision makers about agricultural politics and environmental politics to take preventive measures against drought. Kibunja; Kihoro; Orwa and Yodah [13] carried out a study on Mount Kenya region in order to determine the forecasted values of precipitation using SARIMA model. Authors found out that SARIMA model is a good model for forecasting precipitation in the area of study. In Jordan, by using ARIMA model for monthly precipitation data (1922-1999) in Amman Airport, Momani [14] found that time series analysis could also to be applied to forecast peak values of rainfall data. Papalaaskaris; Panagiotidis andPantrakis [15] used Seasonal Autoregressive integrated moving average (SARIMA) to discover the most appropriate match of a time-series to find out future prediction and forecast of rainfall potential patterns of flood and drought cycles in Kaval city, Greece, North-Eastern Mediterranean Basin. Authors concluded that SARIMA model fit for the total recorded rainfall data od Kavala city, Greece for the period 2006-2014.Ramli; Rusdiana; Yulianur and Achmad [16]. Compared the predicted rainfall results of three regions (Sabang, Aceh Besar & Tengah) in Indonesia using ARIMA to analyze the 1988-2015 rainfall data. Rainfall best predicting model was ARIMA for the short-term. However, accuracy of forecasting for the long-term was not good because the trends of rainfall was flat. Zakaria; Al-Ansari, & Al-Badrany [17] used ARIMA model to predict rainfall in Sinjar, Iraq for five years, in which they used data extended from 1990 to 2011. They used ACF and PACF to determine the best model which that gave the minimum AIC is ARIMA (3, 0, 2) x (2, 1, 1)30, (1, 0, 1) x (1, 1, 3)30. Box-Jenkins methodology were also used to build Autoregressive Integrated Moving Average (ARIMA) models for weekly rainfall data from four rainfall stations in the North West of Iraq: Sinjar, Mosul, Rabeaa and Talafar for the period 1990-2011. Four ARIMA models were developed for the above stations as follow: (3, 0, 2) x (2, 1, 1)30, (1, 0, 1) x (1, 1, 3)30, (1, 1, 2) x (3, 0, 1)30 and (1, 1, 1) x (0, 0, 1)30. These models were used to forecast the weekly rainfall data for the years of (2012 to 2016). Mahdi; Provost; Salha and Nashwan, [18] carried out a study to build a multivariate model to predict recorded rainfall (from 1973 to 2014) infive districts of Gaza strip, Palestine. Recorded Rainfall were aggregated to monthly data using the Thiessen polygons method and multivariate seasonal vector integrated autoregressive moving average models (sVARIMA) which they confirm that it provide a useful approach for forecasting rainfall data as a preliminary guideline for short and long-term sustainable water resources management. The model they got was sVARIMA (0; 0; 1) (1; 1; 0). Shiban, Lama; Alasaad, Ali and Abdelrahman and Abbas [19] carried out a research in which they tackled rainfall in ALHWAIZ basin in Syria in order to explain time series behavior and by applying one model of Box-Jenkins they predicted rainfall in future. Recorded rainfall data from 1959 to 207 were used to build forecasting ARIMA model for the coming six years. They concluded thatthat Autoregressive Integrated Moving Average Model (ARIMA Model) ARIMA (1, 1, 3) model to represent data and ARIMA (2, 1, 0) model is the right model to forecast future rainfall. Mithiya; Mandal and Bandyopadhyay [20] attempted to forecast monthly rainfall in India with the help of time series analysis using monthly rainfall data trough linear and non-linear models. The non-linear model of Artificial Neural Networks (ANNs) showed better results than linear model namely simple seasonal exponential smoothing and seasonal Auto-Regressive Intergrated Moving Averages to forecast rainfall which will help to identify proper cropping pattern. They concluded that forecasting of rainfall with ANN is more efficient than that using SARIMA model. Alhashimi [21] Predicted Monthly Rainfall (from 1970 to 2008) in Kirkuk, Iraq, using Artificial Neural Network (ANN) and Time Series Models. She developed and implemented three rainfall prediction models based on past observations such as time series models based on autoregressive integrated moving average (ARIMA), Artificial Neural Network ANN model and Multi Linear Regression MLR model. The study revealed that ANN model can be used as an appropriate forecasting tool to predict the monthly rainfall, which is preferable over the ARIMA model and MLR model.

Time Series: Time series is a data set that tracks a sample over time and allows one to see what factors influence certain variables from period and has various influences such as national income, unemployment, industrial production, annual sales of commercial and industrial companies during a certain time. This does not mean that the chainsof time series are limited to economic and commercial fields, but also extends to other areas such as quantitative measurementof rainfall in a given area. Furthermore, timeseries analysis plays a significant role in modelling meteorological data such ashumidity, temperature, rainfall and other environmental variables Graham and Mishra [10] There are two types of time series, first is continuous time series which measure values of changes throughout certain time period. Second, is discrete time series which refers to values ofchanging phenomenon at a certain time. Time Series analysismethod isdeveloped for the purpose of forecasting future which is based on the different values. Furthermore, there are four components for time series:

- Trend component (T): It is the regular and continuous change that occurs in values of the studied phenomenon due to effectsof certain factors which may result in a positive or negative trends.
- Seasonal component (S): changes that occur at specific times of each year, such as holidays or the beginning of the school year. However, Weather, traditions and religious occasions may play a role in affecting the seasonal change.
- Cyclic component (C): It refers to changes that recur in specific years and are similar to seasonal component, except that the cyclic component has a periodicity that is much larger than the seasonal component. Cyclic componentsneed many years to be discovered, so they are called long-term variables.
- Stochastic Component (I): A group of factors that cannot be explained through time series, since they occur suddenly and randomly and do not recur at specific times, such as earthquakes and volcanoes.

Types of time series analysis in terms of Stationary:Theyrefer to patterns of studied phenomenon and their distribution. There are two types of time series:

- Stationary Time Series which are characterized by the fact that their data fluctuates around a fixed arithmetic mean and constant variance.
- Non-Stationary Time Series: This series do not have a mean or a variance and their data tend to increase or decrease.

Autocorrelation Function (ACF): This is the measure of the degree of relationship between values of the same variable at a specific time and is used to findout the suitability of the models used in forecasting.

Partial Autocorrelation Function (PACF): The partial correlation is used as to determine the order of the autoregressive model and it is used to determine the rank of the ARMA model.

Box-Jenkins Approach in Time Series: This is a method that is based on a set of probabilistic models called Box Jenkins model and is used to represent time series data for a particular phenomenon. It is considered the most powerful and effective in many cases.

Box Jenkins models may come as non-seasonal or seasonal model which are used to represent two types of chains:

Stationary Time-Series Models: These models have a stable property and do not have a general trend. They have a fixed arithmetic mean that fluctuates around it. These models include three types:

Autoregressive Models (AR): It is symbolized by AR (P), where P represents the order of the model, which is a positive integer. The general form of this model is as follows;

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p}$$

where:
Yi: value of the series at time I
φ_i: Denotes parameters of the autoregressive model factors
α: constant value

Moving Average (MA): It is usually denoted by MA (q) where q is a positive integer indicating the order of the model and the general form of this model is as follows,

$$Y_t = \alpha + \theta_1 \varepsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

where:
Y_i: value of the series at time i
θ_{ii}: Denotes parameters of the autoregressive model factors
ϵ_i: indicates errors over a specific period of time

Mixed Model ARMA: These models consist of integrating the previous two models, the autoregressive model and the moving average model. This model is the most widely used Jenkins-box model for its flexibility and suitability for different types of data. This model is denoted by the symbol ARMA (p, q), where (p, q) represents the two degrees of the model and the general form of this model can be written in the following formula;

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

Non-Stationary Time-Series Models: These models are usually used to represent time series in which the trend is one of its components. They haveseveral mediums around which data fluctuates. These models can be represented by models of first type, but after removing the instability from the original series using the method of differences (d). These models are similar to stationary time series, but it adds integration in order to indicate the use of stationary time series models on non- stationary time series after converting them to time series. The model code becomes as follows (pdq) ARIMA.

To build Jenkins model, they are four main stages: Identification, estimation, diagnostic checking and forecasting.

Identification: The identification stage of the model is one of the most important stages in the analysis of time series since it is based on the data set and on understanding the basic characteristics of time series, especially autocorrelation and partial autocorrelation functions. If it is stationary, researchers my move on to studying and determining the appropriate model, but if it is the opposite, it must be made stationary either by using differences if the stationary is in the average or by using transformations.The identification stage includes the following steps:

- First step is plotting data to find out if it has seasonality or non-seasonality.
- Second step is the calculation of Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) of the sample to determine the degree of differences (in the case of non-stationary). where d>0 (and often d=0, 1, 2)
- Third step is the calculation checking the ACF and PACE of the sample in order to diagnose the model, to find out whether it contains duality between ARMA(1, 0) or AR(1) models and ARMA(0, 1) or MA(1) models according to the three functions.

Estimation: After defining the proposed model, estimation of parameters comes through Least square method, Maximum likelihood, linear estimation method or root mean square error RMSE

Diagnostic Checking: At this stage, the model is tested to determine its suitability to represent the studied phenomenon and to use it in obtaining future predictions. This is can be done by calculation of residuals to detect whether there is any factor other than randomnessby tests such as box-pierce (q), which relies on zero autocorrelation. This means testing the null hypothesis

$$H_0: r1=r2=r3....rk=0$$

In this test, the calculated value of (q) is compared with the tabular value of $\times^2$ at a degree of freedom (h-m) and a certain level of significance.If the calculated value of (q) is less than the tabular $\times^2$, then we will accept the

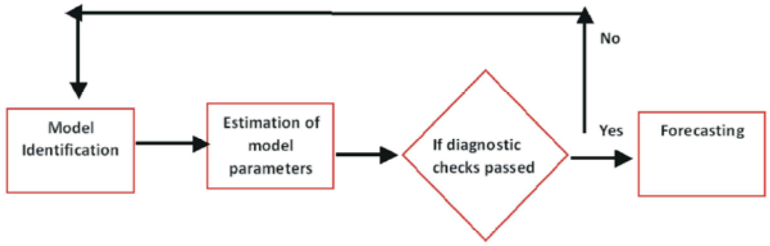

Fig. 2: flowchart of the Box-Jenkins Method

null hypothesis h0 and that the model is suitable to represent this phenomenon. Another test is BOX-Liun and is called the (Portmantean) and takes thesymble (Q) and in general it is the autocorrelation function ACF with respect to stationary time series ofa special form. Calculated value of Q is compared to tabular value of $\times^2$.

Forecasting: Prediction is the main goal of time series analysis, but after the testing of initial model. Time series has become the most widespread and widely used method because of its many advantages which leads to obtaining an excellent system and reliable prediction. The following figure illustratesflowchart of the Box-Jenkins Method.

Exponential Smoothing: It is a time series forecasting method for univariate data and it may be considered an alternative to the popular Box-Jenkins ARIMA class of methods for time series forecasting. This method is sometimes referred to as ETS model.

Methodology: The aim of this research is to predict rainfall in Birzeit for next four years: from2021- to 2025 using to different method such as Box-Jenkins and ETS methods. Descriptive research method was used as well as the time-series analysis technique in order to analyze the current rainfall records and to predict rainfall for the next four years. Rainfall records for the past 19 years (2003 to 2021) was collected from Birzeit university weather station.Required preprocessing tasks were applied to enhance data efficiency before applying the forecasting algorithms. Preprocessing includes several techniques such as cleaning, reduction, transformation and integration. Microsoft Excel 2016 version was employed to perform the following data preprocessing:

- Data integrationwhich means combining data from multiple datastores into one consistent dataset.
- Data reduction: Dataset was reduced into smaller volumes in whichirrelevant attributes were removed.
- Missing values: Descriptive statistics showed that there were no missing values.
- Values of Outliers: There were three outlier values in data (using cooks distance and R studioversion 4.1.0 was used to find best fit model to predict rainfall for next five years.

Figure 3 represent the outlier values and table 2 shows outlier values too in which old recorded rainfall was replaced by new recorded ones (mean of annual rainfall for outlier value).

- Decomposition of time series into its basic components by means of dynamic series model:

Y (t)=u(t)+ v (t)+ s (t) +e (t)

where:

U (t) = trend due to the action of \ permanent factors.
V (t) = cyclic variations due to the action of rhythmical factors.
s (t) = stochastic components assumed to follow an autoregressive moving average Process.
e (t) = random component due to the action of random factors[2].

Analysisof Recorded Birzeit Rainfall: Table3 illustrates monthly rainfall descriptive statistics from Aug-2003 till Sep-2021. The highest recorded rainfall was 447mm which was recorded in January of 2020. However, the lowest was 0.50mm and was recorded in May of 2005. Standard deviation was 92.42 and illustrates large variances of recorded rainfall at Birzeit. In addition, figure 4 illustrates curve of recorded rainfall from 2003 to 2021. Timeseries plot shows us that rainfall have seasonality pattern without any occurring trends. Rainfall reach its highest in January and its lowest in September.

Regarding annual rainfall, table 4 shows such records. The highest rainfall was 864.5mm in 2018- 2019 season and the lowest was 464.1 mm in season of 2013- 2014. Furthermore, Figure 5 shows curve of annual rainfall and it is clear that rainfall fluctuate between seasons.

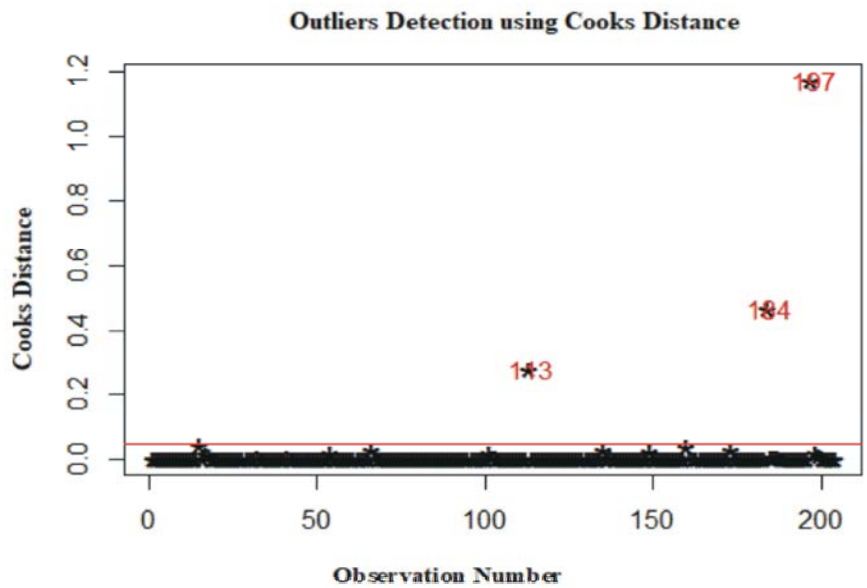

Fig. 3: Outlier values using Cooks Distance

Table 1: Outlier values

Month	Old-recordedrainfall (mm)	New-recorded rainfall (mm)
Jan-20	447	202
Dec-18	373	200
Jan-13	359	206

Table 2: Rainfall descriptive statistics

	Maximum value	Minimum value	Mean	Standard deviation
Rain Fall amount (mm)	447	0.50	101.75	92.42
Year	2020	2005		
Month	January	May		

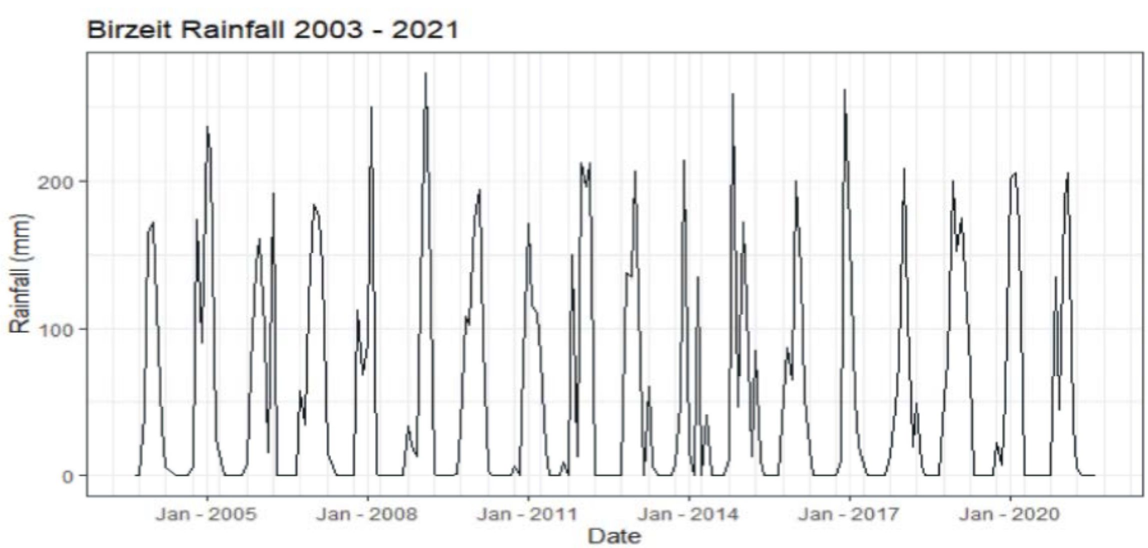

Fig. 4: Represent total annual rainfall 2003-2021

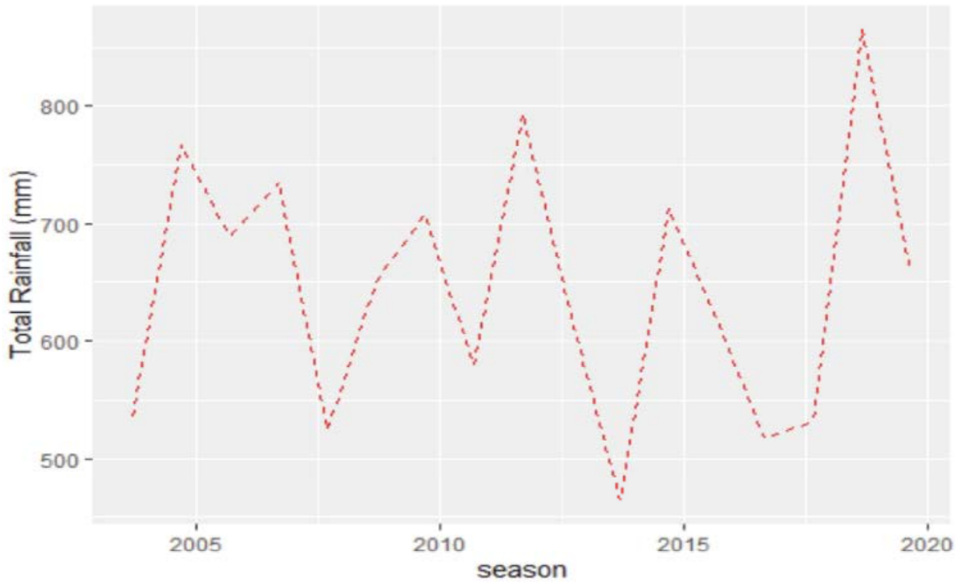

Fig. 5: Total Rainfall Per Season

Table 3: Total Rainfall amount per season

Season	Total Rainfall (mm)
2003 – 2004	535.4
2004 – 2005	766.8
2005 – 2006	689.6
2006 – 2007	734.5
2007 – 2008	523.5
2008 – 2009	650.2
2009 – 2010	706.9
2010 – 2011	580.0
2011 – 2012	793.0
2012 – 2013	621.5
2013 – 2014	464.1
2014 – 2015	713.0
2015 – 2016	617.0
2016 – 2017	517.0
2017 – 2018	531.2
2018 – 2019	864.5
2019 – 2020	658.0
2020– 2021	613.9

Table 5 shows percentage distribution of recorded rainfall over each season. It illustrates that rainfall are mainly concentrated over months of December, January, February and March. In addition, months of May, June, July, August and September are free of Rainfall.

Model Estimation: Aggregation model considers that the value ofmonthly rainfall is the sum of different effect components such as general component trend, seasonal component and random component. When adding the effect of smooth compounds at a particular month the result will be the value of rainfall in that month and therefore the challenge lies in the decompose and analysis of recorded rainfall series in Birzeit. Time seriesof data will be decomposed into more detail based on trend, seasonality and remainder components. Therefore, more insight about rainfall properties during 2003-2021will be achieved. Figure6 shows series of each component.

First row of figure 6represents data without any decomposition, second row represents random component effect, third-row seasonal component effect and last row represents trend component effect. Moreover, trend cycle and seasonal plots showed there are seasonal fluctuations occurred with no specific trend and random remainder/residual. Calculationformulas to measure trend and seasonality strength are:

Tt: Trend component
St: Seasonality component
Rt: Reminder component
F_T: Trend Strength

$$F_T = Max(0.1 - \frac{VAR(Rt)}{VAR(Tt + Rt)})$$

F_S: Seasonal Strength

$$F_S = Max(0.1 - \frac{VAR(Rt)}{VAR(St + Rt)})$$

Value of measured trend and seasonal strength lies between 0 and 1, while "1" means strong trend and seasonal. In this case trend strength is 0.1 and seasonal

Table 5: Rainfall distribution over the season 0% of rainfall high % of rainfall

Season	January	February	March	April	May	June	July	August	September	October	November	December	Total
2003 - 2004	32%	22%	6%	1%	1%	0%	0%	0%	0%	0%	8%	31%	100%
2004 - 2005	31%	28%	3%	2%	0%	0%	0%	0%	0%	1%	23%	12%	100%
2005 - 2006	23%	14%	2%	28%	0%	0%	0%	0%	0%	1%	11%	21%	100%
2006 - 2007	25%	24%	19%	2%	1%	0%	0%	0%	0%	8%	5%	16%	100%
2007 - 2008	17%	48%	0%	0%	0%	0%	0%	0%	0%	0%	22%	13%	100%
2008 - 2009	19%	42%	28%	0%	0%	0%	0%	0%	0%	5%	3%	2%	100%
2009 - 2010	25%	27%	13%	1%	0%	0%	0%	0%	0%	5%	15%	14%	100%
2010 - 2011	29%	20%	19%	14%	3%	0%	0%	0%	0%	1%	0%	13%	100%
2011 - 2012	27%	25%	27%	0%	0%	0%	0%	0%	1%	0%	19%	2%	100%
2012 - 2013	33%	12%	0%	10%	1%	0%	0%	0%	0%	0%	22%	22%	100%
2013 - 2014	4%	0%	29%	0%	9%	0%	0%	0%	0%	2%	11%	46%	100%
2014 - 2015	24%	16%	2%	12%	2%	0%	0%	0%	0%	2%	36%	7%	100%
2015 - 2016	32%	22%	9%	3%	0%	0%	0%	0%	0%	9%	14%	11%	100%
2016 - 2017	32%	10%	4%	1%	0%	0%	0%	0%	0%	0%	2%	51%	100%
2017 - 2018	39%	21%	4%	9%	1%	0%	0%	0%	0%	3%	9%	15%	100%
2018 - 2019	18%	20%	16%	8%	0%	0%	0%	0%	0%	6%	9%	23%	100%
2019 - 2020	31%	31%	25%	0%	0%	0%	0%	0%	0%	3%	1%	8%	100%
2020 - 2021	31%	33%	6%	1%	0%	0%	0%	0%	0%	0%	22%	7%	100%
2021 - 2022	22%	24%	15%	7%	1%	0%	0%	0%	0%	5%	11%	15%	100%
2022 - 2023	26%	32%	18%	1%	1%	0%	0%	0%	0%	2%	18%	2%	100%
2023 - 2024	26%	22%	7%	8%	1%	0%	0%	0%	0%	2%	16%	18%	100%
2024 - 2025	23%	21%	21%	4%	1%	0%	0%	0%	0%	2%	11%	18%	100%
Average	26%	23%	12%	5%	1%	0%	0%	0%	0%	3%	13%	17%	100%

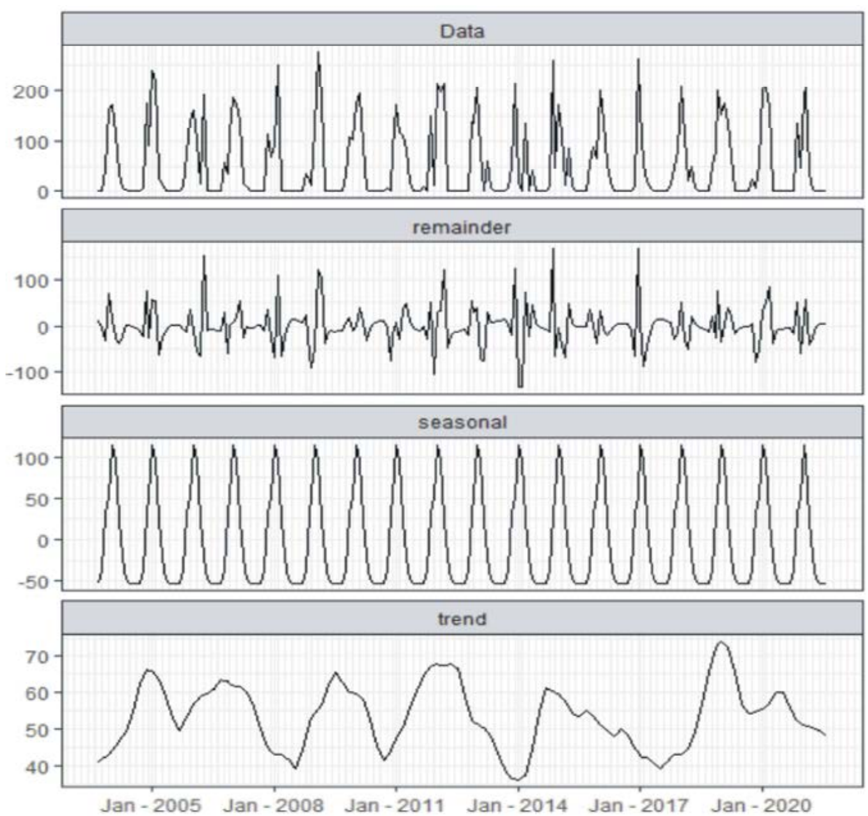

Fig. 6: Series for Each Component

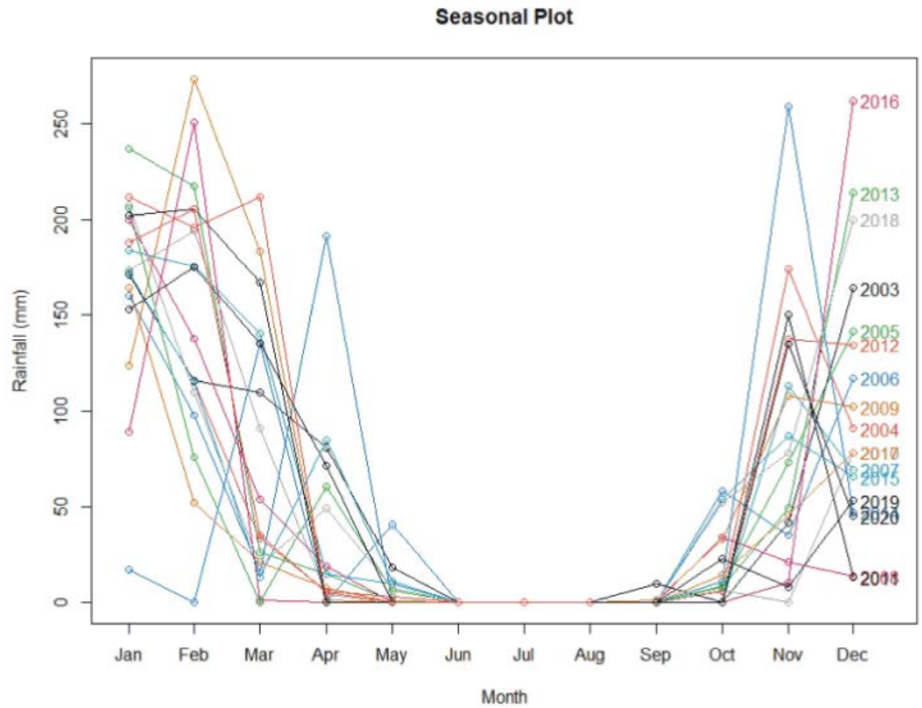

Fig. 7: Seasonal plot for total period

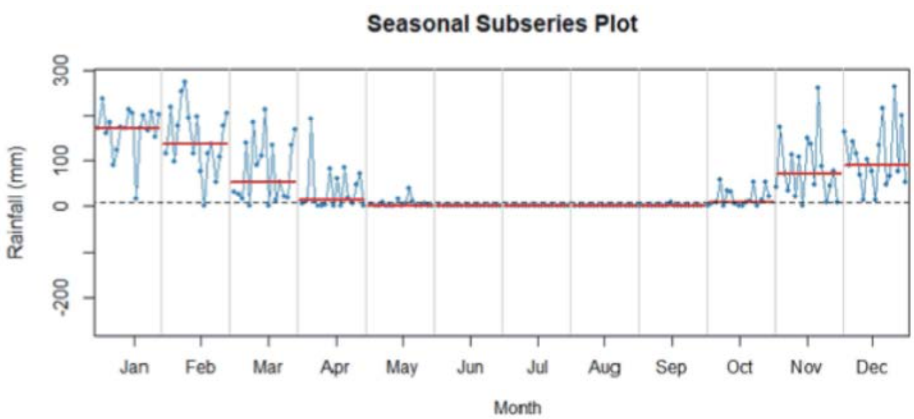

Fig. 8: Seasonal subseries Plot

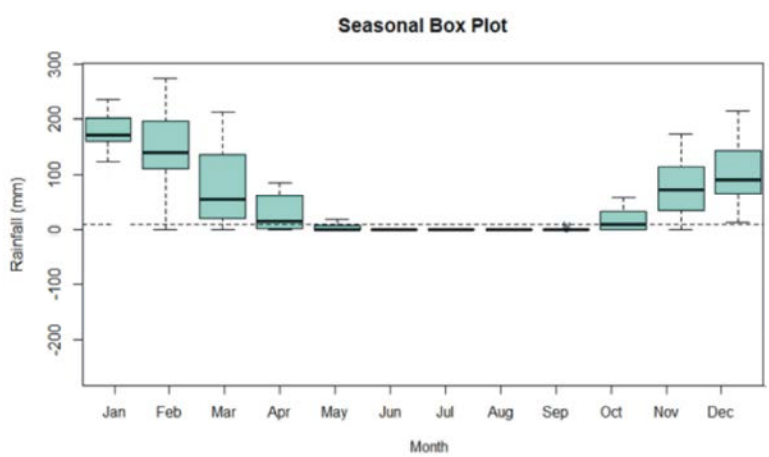

Fig. 92: Boxplot series

strength is 0.6. This result proves that our data has seasonality pattern. To explore more about rainfall data seasonality pattern; seasonal plot and seasonal-subseries plot, seasonal boxplot is used to provide much more insight explanation of our data. Figure 7 shows seasonal plot for total period and for each year. seasonal plot indeed shows a seasonal pattern occurred each year, but due to variances on several years, a pattern can't be clearly seen. Therefore, Seasonal Boxplot and subseries plot is used (Figures 8 and 9).

Using seasonal boxplot and subseries plot, patterns can clearly be seen. The horizontal lines indicate means of every month and indicates rainfall start to decrease from March and reach its lowest point in September. This period is called dry season. Rainfall start to rise again after October and reach its peak in January. This period is called rainfall season.

Box-Jenkins models is used to analyze time series to find the equation that predicts the reminder component, but before that it must be checked whether time series is stationary or not. Time series is said to be stationary if there is no deviation in data (absence of a general trend). If data is scattered horizontally around a fixed average and fluctuates around a fixed arithmetic mean independent of time, as well as the variance is constant over time. This makes sure that the series is stationary. Therefore there is a need to useDickey Fuller Test. Dickey-Fuller value is (-10.41) and the significant (p-value)is 0.01 which is less than p-value of 0.05. This result means that series is stationary.

ARIMA models and ETS model: First step on forecasting is to choose the right model which implies splitting timeseries dataset into train and test sets. Train setis used to train several models and used all data until August-2020 as training set. However, test set starts from September-2021to the end of series.

Model Assignment: Forecasting is carried out by using both ARIMA and ETS model and they are comparedwith each other and evaluated using some parameter against test set.

ARIMA Model: To build ARIMA model an autocorrelation plot on stationary time series was carried out and there was a need to do differencing. Therefore, number of differences (d, D) on this model was set as "zero". It is clear that dataset has seasonality and there was need to build ARIMA (p, d, q)(P, D, Q)m , to get (p, P, q, Q). ACF Plot is used to get MA parameter (q, Q). there wasa significant spike at lag 1 and sinusoidal curve indicated annual seasonality. PACF Plot is used to get AR parameter (p, P), there was a significant spike at lag 1 for AR parameter.

ACF/PACF (Figure 10) was used to identify possible ARIMA model and to figure matrix between AR and MA (Figure 11) was used to identify ARIMA model.

To identify AR MA from previous matrix, each triangle represented AR MA. For example, red color triangle the AR is 5 and MA is 11 so ARIMA is (5, 0, 11) and so on.Table 6illustrates the suggested ARIMA models.

The model with minimum AIC often is a best model for forecasting.AIC value of model number 1 SARIMA(11, 0, 2)x(11, 0, 2)is the lowest among other models and this ARIMA model is best for forecasting. Butthere is a need to have residuals checked for this model to make sure that this model is appropriate for timeseries forecasting. Figure 12 shows the result of fit SARIMA(11, 0, 2)x(11, 0, 2)

Based on the Ljung-Boxtest (figure 13) and ACF plot of model residuals, null hypothesis is accepted since the p-value is greater than 0.05 and this means that residuals are independently distributed. It can be concluded that this model is appropriate for forecasting since its residuals shows white noise properties and they are not correlated with each other.

ETS Model: There was a need to build ETS model and compare it with the chosen ARIMA model for the purpose of finding which model is better against Test Set. Similar to ARIMA model, there is a need to check its residuals behavior and to make sure that this model is working well for forecasting, Figure (14).

Based on the Ljung-Boxtest (Figure 15) and ACF plot of model residuals, the null hypothesis is rejected since the p-value is greater than 0.05 which means that residuals are independently distributed).

After checking residuals, ACF Plot shows ETS Model residuals has little correlation between with each other on several lag, but most of residuals are still withinand it may be said that using this model as compared with chosen SARIMA model.

ARIMA vs ETS Model on Test Set: Better model for this data time series can be checked using test set. Both SARIMA and ETS modelswere employed to predict and see their accuracy against test set (September 2020 to August 2021). First step, there is a need to plot visualization between ARIMA Model, ETS Model and our actual test set. Figure 16 shows the ARIMA and ETS forecasting for one year compared to actual data.

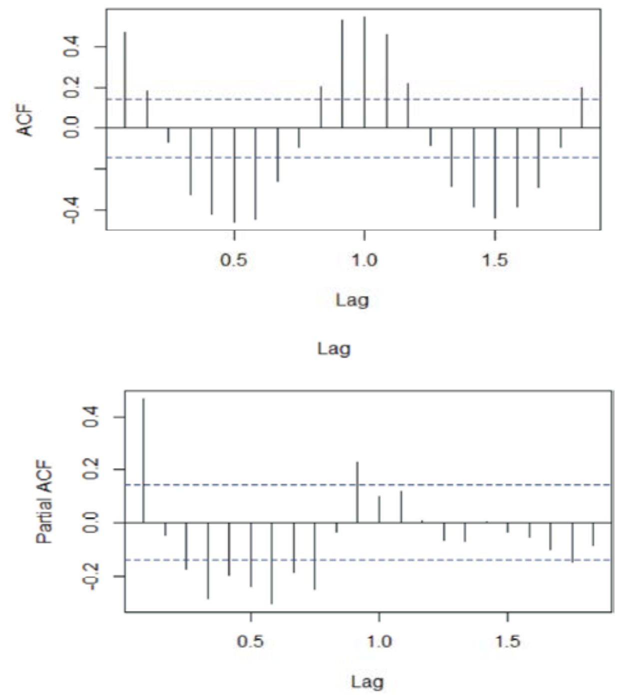

Fig. 10: ACF and PACF

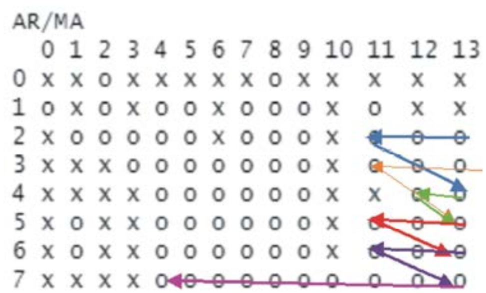

Fig. 11: AR and MA matrix

Table 6: Suggested ARIMA Models

Model Number	Model	AIC
Model 1	SARIMA (11, 0, 2)x(11, 0, 2) $_{12}$	2185.81
Model 2	SARIMA (11, 0, 3)x(11, 0, 3) $_{12}$	2192.91
Model 3	SARIMA (12, 0, 4)x(12, 0, 4) $_{12}$	2192.83
Model 4	SARIMA (11, 0, 5)x(11, 0, 5) $_{12}$	2191.83
Model 5	SARIMA (11, 0, 6)x(11, 0, 6) $_{12}$	2177.31
Model 6	SARIMA (4, 0, 7)x(4, 0, 7) $_{12}$	2188.08

Eventhough both ARIMA and ETS models are not exactly fit same value with actual data, but surely both of them plotting quite similar. Accuracy between these models and actual data is calculated and decided the best model and it was found that ARIMA Model doing better in test set compared to ETS Model, Table 7.

2021-2020 Rainfall Forecasting: Using the same parameter which was employed in train set, rainfall for the four years is forecasted. there researcher will forecast rainfall for next 4years, from October 2021 to September 2025. Table 8 illustrates forecasted points for both models Figure 17 shows SARIMA forecasted with observed data and it is clear that is no big differences between them. The points that were predicted kept the original data pace which means that rainfall season is still the same. Furthermore, Figure 18 illustrates ETS forecasted rainfall with actual data and it is as well as SARIMA forecast.

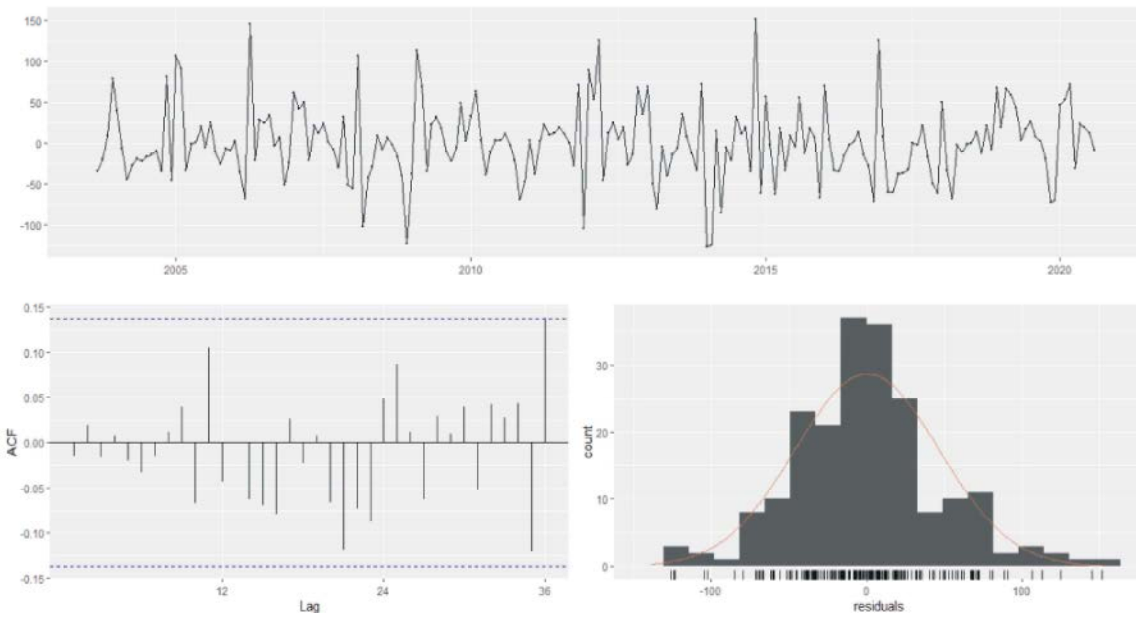

Fig. 12: Residuals from SARIMA (11, 0, 2) (11, 0, 2)$_{12}$

```
        Ljung-Box test

data:  Residuals from ARIMA(11,0,2) with non-zero mean
Q* = 16.183, df = 10, p-value = 0.09451

Model df: 14.   Total lags used: 24
```

Fig. 13: Ljung-Box test resultSARIMA model

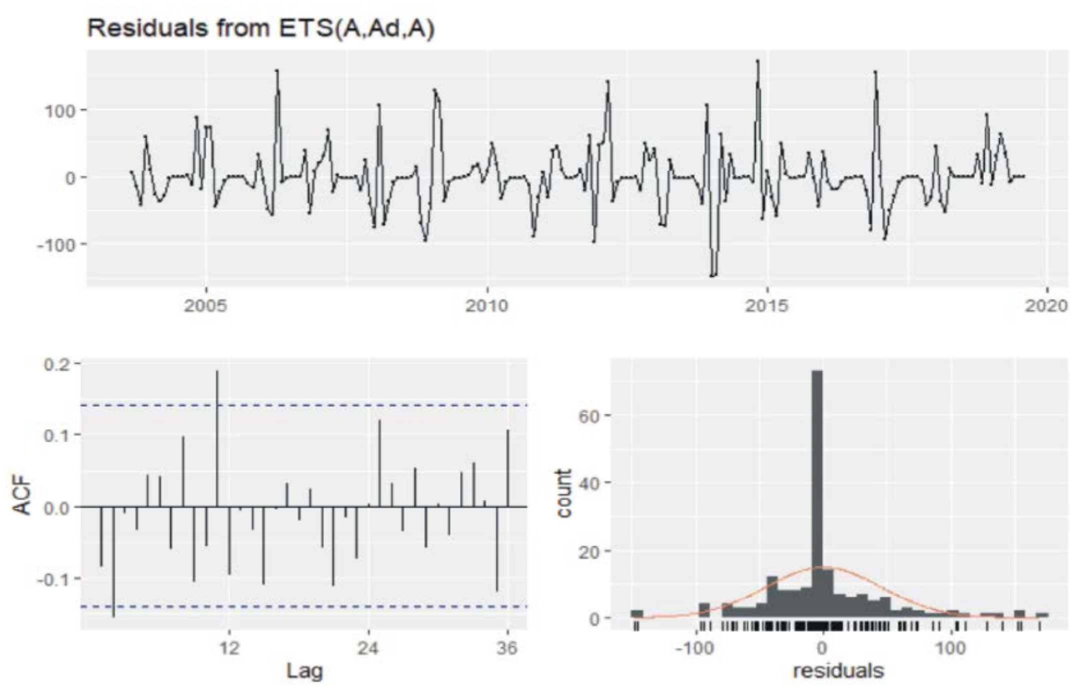

Fig. 14: Residuals from ETS model

```
        Ljung-Box test

data:  Residuals from ETS(A,Ad,A)
Q* = 29.985, df = 7, p-value = 9.555e-05

Model df: 17.   Total lags used: 24
```

Fig. 15: Ljung-Box test result ETS model

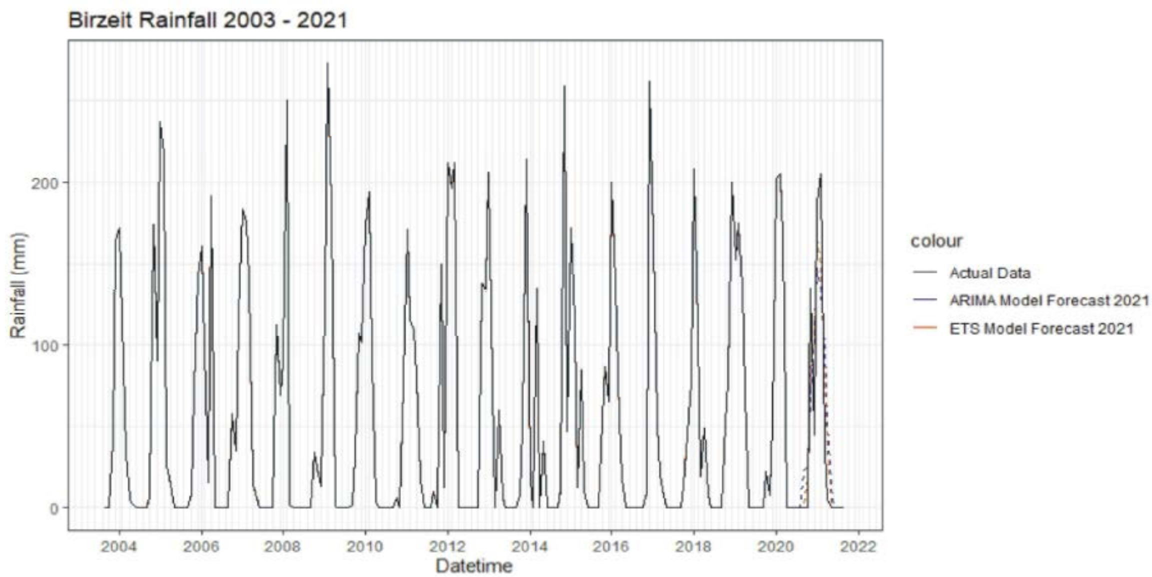

Fig. 16: ARIMA and ETS compared to actual data

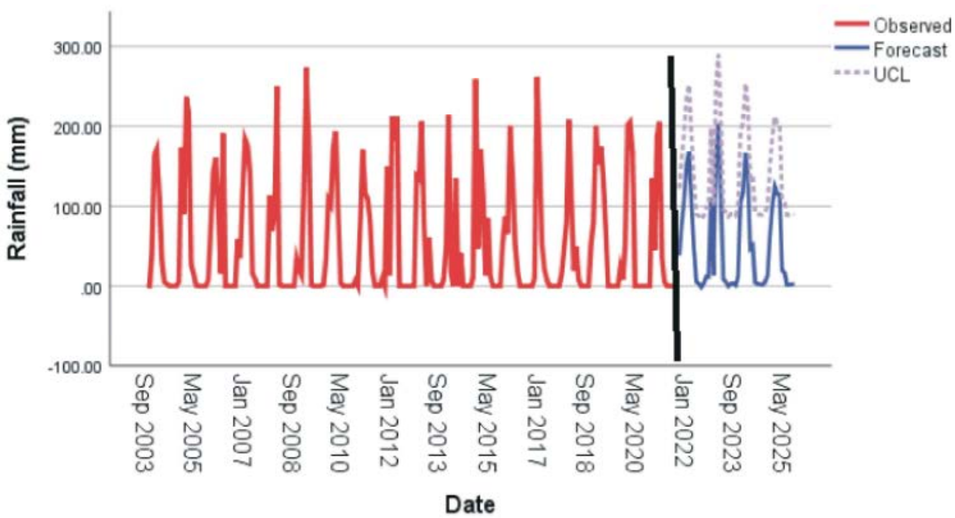

Fig. 17: SARIMA Forecasted vs Actual Data

Table 7: Accuarcy Result for ARIMA and ETS Model

Model	RMSE
ARIMA	33.05
ETS	42.90

Table 8: Forecasted points for both models

	SARIMA		ETS	
Period forecast	Forecasted points	Upper limit	Forecasted points	Upper limit
Oct 2021	16.64	107.58	38.09	122.11
Nov 2021	84.28	175.53	78.03	162.05
Dec 2021	104.45	196.01	107.24	191.56
Jan 2022	168.07	259.93	153.08	237.40
Feb 2022	150.27	242.43	167.88	252.36
Mar 2022	76.97	169.44	101.76	186.24
Apr 2022	33.42	126.20	47.17	131.68
May 2022	4.92	98.00	4.96	90.04
Jun 2022	0.00	12.90	0.00	11.00
Jul 2022	0.00	16.12	0.00	13.75
Aug 2022	0.00	20.15	0.00	17.19
Sep 2022	0.00	25.19	0.00	21.48
Oct 2022	16.55	111.14	11.16	98.66
Nov 2022	84.19	179.07	110.35	197.85
Dec 2022	104.36	199.54	12.98	100.48
Jan 2023	167.98	263.45	164.63	252.16
Feb 2023	150.18	245.94	202.26	289.80
Mar 2023	76.88	172.94	111.83	199.37
Apr 2023	33.33	129.68	8.12	95.79
May 2023	4.83	101.47	5.35	93.01
Jun 2023	0.00	14.00	0.00	12.00
Jul 2023	0.00	17.50	0.00	15.00
Aug 2023	0.00	21.88	0.00	18.75
Sep 2023	0.00	27.34	0.00	23.44
Oct 2023	16.47	114.55	10.94	98.65
Nov 2023	84.10	182.48	105.52	193.23
Dec 2023	104.27	202.94	116.17	203.89
Jan 2024	167.89	266.83	166.30	254.02
Feb 2024	150.09	249.32	140.78	228.50
Mar 2024	76.79	176.30	46.75	134.48
Apr 2024	33.24	133.04	50.37	138.10
May 2024	4.74	104.82	4.26	91.99
Jun 2024	0.00	11.00	0.00	10.00
Jul 2024	0.00	13.75	0.00	12.50
Aug 2024	0.00	17.19	0.00	15.63
Sep 2024	0.00	21.48	0.00	19.53
Oct 2024	16.38	117.85	12.27	99.92
Nov 2024	84.02	185.77	61.00	148.64
Dec 2024	104.18	206.21	100.75	188.39
Jan 2025	167.80	270.10	124.45	212.10
Feb 2025	150.00	252.58	115.23	202.88
Mar 2025	76.70	179.55	114.83	202.48
Apr 2025	33.15	136.28	19.44	107.09
May 2025	4.65	108.05	3.22	104.42
Jun 2025	0.00	9.00	0.00	8.00
Jul 2025	0.00	11.25	0.00	10.00
Aug 2025	0.00	14.06	0.00	12.50
Sep 2025	0.00	17.58	0.00	15.63

Table 9 shows that average rainfall for whole period is close to forecasted by SARIMAand ETS but results of forecasted are slightly less than the actual average. However, SARIMA gave better forecast than ETS since its result (637.9) is closer to actual average (655.1).

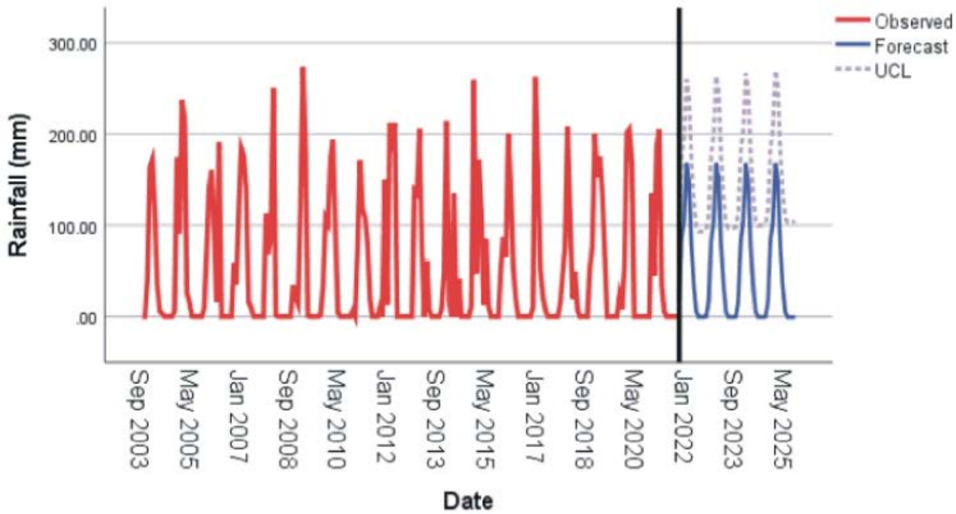

Fig. 18: ETS Forecasted vs Actual Data

Table 9: Oberseved data and forcasted points

Period	Average rainfall	Difference %
Observed Sep-2003 to August -2021	655.1	
SARIMA Sep-2021 to August 2025	637.9	-2.6%
ETS Sep-2021 to August 2025	629.3	-3.9%

Table 10: Winter Months Observed vs. Forecasted

		Forecast period			
Month	Observed (Average)	SARIMA	ETS	Observed vs. SARIMA	Observed vs. ETS
October	17.10	16.51	18.12	-3.5%	5.9%
November	84.70	84.15	88.73	-0.7%	4.8%
December	104.90	104.32	84.28	-0.6%	-19.7%
January	168.50	167.93	152.11	-0.3%	-9.7%
February	150.70	150.13	156.54	-0.4%	3.9%
March	77.40	76.83	93.79	-0.7%	21.2%

Table 11: Rainfall distribution over the season observed vs. forecast 0% of rainfall high % of rainfall

Season	January	February	March	April	May	June	July	August	September	October	November	December
2003 - 2004	32%	22%	6%	1%	1%	0%	0%	0%	0%	0%	8%	31%
2004 - 2005	31%	28%	3%	2%	0%	0%	0%	0%	0%	1%	23%	12%
2005 - 2006	23%	14%	2%	28%	0%	0%	0%	0%	0%	1%	11%	21%
2006 - 2007	25%	24%	19%	2%	1%	0%	0%	0%	0%	8%	5%	16%
2007 - 2008	17%	48%	0%	0%	0%	0%	0%	0%	0%	0%	22%	13%
2008 - 2009	19%	42%	28%	0%	0%	0%	0%	0%	0%	5%	3%	2%
2009 - 2010	25%	27%	13%	1%	0%	0%	0%	0%	0%	5%	15%	14%
2010 - 2011	29%	20%	19%	14%	3%	0%	0%	0%	0%	1%	0%	13%
2011 - 2012	27%	25%	27%	0%	0%	0%	0%	0%	1%	0%	19%	2%
2012 - 2013	33%	12%	0%	10%	1%	0%	0%	0%	0%	0%	22%	22%
2013 - 2014	4%	0%	29%	0%	9%	0%	0%	0%	0%	2%	11%	46%
2014 - 2015	24%	16%	2%	12%	2%	0%	0%	0%	0%	2%	36%	7%
2015 - 2016	32%	22%	9%	3%	0%	0%	0%	0%	0%	9%	14%	11%
2016 - 2017	32%	10%	4%	1%	0%	0%	0%	0%	0%	0%	2%	51%
2017 - 2018	39%	21%	4%	9%	1%	0%	0%	0%	0%	3%	9%	15%
2018 - 2019	18%	20%	16%	8%	0%	0%	0%	0%	0%	6%	9%	23%
2019 - 2020	31%	31%	25%	0%	0%	0%	0%	0%	0%	3%	1%	8%
2020 - 2021	31%	33%	6%	1%	0%	0%	0%	0%	0%	0%	22%	7%
2021 - 2022	22%	24%	15%	7%	1%	0%	0%	0%	0%	5%	11%	15%
2022 - 2023	26%	32%	18%	1%	1%	0%	0%	0%	0%	2%	18%	2%
2023 - 2024	26%	22%	7%	8%	1%	0%	0%	0%	0%	2%	16%	18%
2024 - 2025	23%	21%	21%	4%	1%	0%	0%	0%	0%	2%	11%	18%

Table 10 shows the difference between average monthly observed data and forecasted points for SARIMA and ETS. It's that SARIMA results are closer to actual data than ETS.

Regarding to rainfall distribution table 11 show the rainfall distribution for forecast overall months compared to observed data. From table 11 there is little change in rainfall distribution, since the December had low percentage compared to previous while there in increase in November

CONCLUSION

In this research researchers found out that Box-Jenkins, ARIMA methodology and comparative study of ETS models are good to forecast Rainfall at Birzeit weather station. Forecasting of rainfall at Birzeitis carried outby building ETS and seasonal ARIMA models.The performance of the model is evaluated with reference toRoot Mean Squared Error (RMSE) and model fit. Thecomparative analysis revealed that seasonal ARIMA modelaccurately forecasts rainfallwith less error. Thus, derivedmodel could be used to forecast monthly rainfall for theupcoming years. Researchers suggest thatseasonal ARIMA (11; 0; 2) - (11; 0; 2) model is the most appropriate for forecastingrainfall at Birzeit. It canbe concluded that seasonalARIMAmodel provides a useful method to forecast rainfall data. It could prove very helpful toenvironmental scientists, especially those based in Middle Eastern countries, to utilizesuch model to forecast rainfall as a preliminary guideline toward short andlong-term sustainable water resources management.Researchers recommends using seasonal ARIMA for rainfall predication in other parts of Palestine and other Arab countries.

REFERENCES

1. Palestine Remembered, 2022. Palestine Maps Before and After Nakba, 1948. https://www.palestineremembered.com/Maps/index.html. Retrieved 4-3-2022.
2. Dimri, Tripti, Ahmad, Shanshad and Sharif, Mohammad, 2020. Time series analysis of climate variables using seasonal ARIMA approach. Journal of Earth Systems and Sciences, pp: 129-149.
3. Hasan, Balqees, 2011. Rainfall Statistical Analysis in Sudan: 1965-2005 using Time Series and regression methods. PhD. Dissertation, Islamic University of Um-Durman, Sudan. In Arabic.
4. Abd Elseed, Haytham, 2015. Using of ARIMA Models for Forecasting of rains in Wad Medani city Sudan (1981 – 2010). MA Degree in Applied Statistics. Al-Jazeera University, Sudan. In Arabic.
5. Khadam, Munzer, Abd Allah, Abraham and Hana, Maisa, 2019. Statistical analysis of some climatic variables and prediction of their future in the region of the Alghab. Tishreen University Journal - Biological Sciences Series, 41(4): 147-161. http://journal.tishreen.edu.sy/index.php/bioscnc/article/view/9042.
6. Hussein, Ali, 2017. Use the Time Series method to predict rainfall rates in Iraq based on time series data for rainfall in Iraq and for the period (2006-2016). International Journal of Statistics and Applications, 5(5): 237-246. Journal of Economic Sciences, 12(47): 102-121. In Arabic.
7. Hayek, Sharif and Hammad Munzer, 2016. Predict of the monthly water volumes incoming in AL-ROOS River in the Syrian Coast by using the time series analysis. Tishreen University Journal for Research and Scientific Studies Engineering Sciences Series, 38(5): 45-65. In Arabic. https://search.emarefa.net/detail/BIM-338865.
8. Hayek, Sharif and Ammar Ghatfan, 2015. Rainfall Prediction in Tartous Station Located in the Southern Part of the Syrian Coast. Tishreen University Journal for Research and Scientific Studies: Engineering Sciences, 73(2).
9. Afrifa-Yamoah, Ebenezer, 2015. Application of ARIMA Models in Forecasting Monthly Average Surface Temperature of Brong Ahafo Region of Ghana.International Journal of Statistics and Applications, 5(5): 237-246. https://www.researchgate.net/publication/282848214_Application_of_ARIMA_Models_in_Forecasting_Monthly_Average_Surface_temperature_of_Brong_Ahafo_Region_of_Ghana/link/561e2bd608aec7945a2541b6/download.
10. Graham, Anosh and Mishra Pathak, 2017. Time series analysis model to forecast rainfall for Allahabad region. Journal of Pharmacognosy and Phytochemistry, 6(5): 1418-1421.
11. Akinbobola, A., E.C. Okogbue and A.K. Ayansola, 2018. Statistical Modeling of Monthly Rainfall in Selected Stations in Forest and Savannah Eco-Climatic Regions of Nigeria. J Climatol Weather Forecasting, 6(S1): 1-9. https://www.iomcworld.com/open-access/statistical-modeling-of-monthly-rainfall-in-selected-stations-in-forest-and-savannah-ecoclimatic-regions-of-nigeria-2332-2594-1000226.pdf.

12. Balibey, Mesut and Turkeyilmaz, Serpil, 2015. A time series approach for precipitation in Turkey. Gazi University Journal of Science, 28(4): 549-559.
13. Kibunjia, H., J. Kihoro, G. Orwa and W. Yodah, 2014, 50-58. Forecasting Precipitation Using SARIMA Model: A Case Study of Mt. Kenya Region. Mathematical Theory and Modeling, 4(11): 50-58.
14. Momani, Nail, 1999. Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis. American Journal of Environmental Sciences, 5(5): 599-604.
15. Papalaaskaris, Thomas, Panagiotidis, Theologos and Pantrakis, Athanasios, 2016. Stochastic Monthly Rainfall Time Series Analysis, Modeling and Forecasting in Kaval City, Greece, North-Eastern Mediterranean Basin. Procedia Engineering, 162: 254-263.
16. Ramli, I., S. Rusdiana, A. Yulianur and A. Achmad, 2019. Comparisons among rainfall prediction of monthly rainfall basis data in Aceh using an autoregressive moving average. Earth and Environmental Science.
17. Zakaria, Saleh, Al-Ansari, Nadhir, Knutsson, Sven and Al-Badrany, Thafer, 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.
18. Mahdi, Esam, Provost, Serege, Salha, Raid and Nashwan, Imad, 2017. Multivariate Time Series Modeling of Monthly Rainfall Amounts. Electronical Journal of Applied Statistical Analysis, 10(10): 65-81.
19. Shiban, Lama, Alasaad, Ali and Abdelrahman, Abbas, 2019. Building a Forecasting Model for Annual Rainfall in Alhwaiz Basin Using Box-Jenkins Methodology. American Journal of innovative and Applied Sciences, 8(1): 18-23.
20. Mithiya, Debasis, Mandal, Kumar and Bandyopadhyay, Simanti, 2020. Time series Analysis and Forecasting of Rainfall for agricultural Crops in India: An application of Artificial Neural Network. Research in Applied Economics, 12(4): 1-21.
21. Alhashimi, Shaymaa, 2014. Prediction of Monthly Rainfall in Kirkuk Using Artificial Neural Network and Time Series Models. Journal of Engineering and Development, 18(1): 129-143.
22. Haidu, I., P. Serbrn and M. Simota, 1987. Fourier-ARIMA Modelling of Multiannual FlowVariation. In: Proceedings of Vancouver Symposium. August, 1987, IAHS Publication No. 68, The Influence of Climate Changeand Climate Variability on the HydrologicRegime and Water Resources. Proceedings ofthe Vancouver Symposium, August, 1987, IAHS Publications, 168: 281-287.