A TIME SERIES ANALYSIS OF THE RAINFALL DATA IN MIZORAM DURING 1990-2020

Praveen Kumar Misra[1], Vanlalhraitsaka[2] and Lalpawimawha[2*]

[1]*Department of Mathematics and Statistics, DSMNR University, Lucknow, Uttar Pradesh*

[2]*Department of Statistics, Pachhunga University College, Mizoram University, AIzawl, Mizoram*

Abstract

Climate and rainfall are highly non-linear and complicated phenomena, which require classical, modern and details models to obtain accurate prediction. Rainfall is one of the climatological data which is widely analyzed for a long time. Analysis of Rainfall data is important as it facilitates policy decisions regarding the cropping pattern, showing date, construction of roads and providing drinking water to urban and rural areas. The amount of rainfall and its pattern is one of the important factors that affect the agriculture system. By precisely measuring precipitation with a rain gauge, we will know the right time to plant certain types of flowers, vegetables, fruits, or even trees. We will also when rainfall amounts are beneath optimal levels so that we can supplement with our own water. So, it is important to know the amount of rainfall for our future generation, so that we can have a better prepare for our future and we can give a better live for our country. The objective study of this study is to fit a time series model and forecast the Rainfall data of Mizoram, using time series data from 1990 – 2020.It is observe that ARIMA (0,1,1) Models are useful for forecasting the rainfall data of Mizoram. The findings of this study may be helpful for effective planning and better preparation for our future generation of our country and can also be helpful for the awareness of Afforestation.

Keywords: Mizoram, Rain fall, Time series analysis Forecasting

1. INTRODUCTION

India has a tropical monsoon climate and rainfall is an important element in Indian economy. The monsoon affects most parts of India where some regions receive a large amount of rainfall whereas the other receives almost none. Almost every year, some part of the country experiences a drought or a flood. Some of the countries are more prone to such events than other in the country. Usually, the Indian monsoon last between the end of May and the first week of October. The South-West Monsoon Rain which occurs between June and September every year and is responsible for nearly 76 percent of the annual precipitation in the country (www.indiaenvironmentportal.org.in>file). The rainfall seasons between the months of October and December is known as the North-East monsoon season and this effects south peninsular India. Northeast India has a subtropical climate that is influenced by its relief and influences from the southwest and northeast monsoons. The Himalayas to the north, the Meghalaya plateau to the south and the hills of Nagaland, Mizoram and Manipur to the east influences the climate. Since monsoon winds originating from the Bay of Bengal move northeast, these mountains force the moist winds upwards, causing them to cool adiabatically and condense into clouds, releasing heavy precipitation on these slopes. It is the rainiest region in the country, with many places receiving an average annual precipitation of 2,000 mm (79 in), which is mostly concentrated in summer during the monsoon season. Cherrapunji located on the Meghalaya plateau is one of the rainiest places in the world with an annual precipitation of 11,777 mm (463.7 in). Temperatures are moderate in the Brahmaputra and Barak valley river plains which decreases with altitude in the hilly areas. At the highest altitudes, there is permanent snow cover.Temperatures vary by altitude with the warmest places being in the Brahmaputra and Barak River plains and the coldest at the highest altitudes. It is also influenced by proximity to the sea with the valleys and western areas being close to the sea, which moderates temperatures. Generally, temperatures in the hilly and mountainous areas are generally lower than the plains which lie at a lower altitude. Summer temperatures tend to be more uniform than winter temperatures due to high cloud cover and humidity.

In the Brahmaputra and Barak valley river plains, mean winter temperatures vary between 16 to 17 °C (61 to 63 °F) while mean summer temperatures are around 28 °C (82 °F). The highest summer temperatures occur in the West Tripura plain with Agartala, the capital of Tripura having mean maximum summer temperatures ranging between 33 to 35 °C (91 to 95 °F) in April. The highest temperatures in summer occur before the arrival of monsoons and thus eastern areas have the highest temperatures in June and July where the monsoon arrives later than western areas. In the Cachar Plain, located south of the Brahmaputra plain, temperatures are higher than the Brahmaputra plain although the temperature range is smaller owing to higher cloud cover and the monsoons that moderate night temperatures year round.

In the mountainous areas of Arunachal Pradesh, the Himalayan ranges in the northern border with India and China experience the lowest temperatures with heavy snow during winter and temperatures that drop below freezing. Areas with altitudes exceeding 2,000 metres (6,562 ft) receive snowfall during winters and have cool summers. Below 2,000 metres (6,562 ft) above sea level, winter temperatures reach up to 15 °C (59 °F) during the day with nights dropping to zero while summers are cool, with a mean maximum of 25 °C (77 °F) and a mean minimum of 15 °C (59 °F). In the hilly areas of Meghalaya, Nagaland, Manipur and Mizoram, winters are cold while summers are cool.

The plains in Manipur have colder winter minimums than what is warranted by its elevation owing to being surrounded by hills on all sides. This is due to temperature inversions during winter nights when cold air descends from the hills into the valleys below and its geographic location which prevents winds that bring hot temperatures and humidity from coming into the Manipur plain.The southwest monsoon is responsible for bringing 90% of the annual rainfall to the region. April to late October is the months where most of the rainfall in Northeast India occurs with June and July being the rainiest months. Southern areas are the first to receive the monsoon (May or June) with the Brahmaputra valley and the mountainous north receiving later (later May or June). In the hilly parts of Mizoram, the closer proximity to the Bay of Bengal causes it to experience early monsoons with June being the wettest season. Mizoram (Mizo+Ram=Mizoram, ‘Mizo’ - a tribal community and ‘Ram’ means land), " province of Mizo tribe", is the twenty-4th largest state of India with an area of 21081 Sq Km which placed in the North-West part of India. According to the census of India 2011, approximately 1091014 people reside here. The state shares an international border to the west, east and south with its neighboring country Burma and its neighboring states are Tripura to the North-West, Manipur to the North-East and Assam to the North.The people of Mizoram mainly belong to Christian religion (near about 90.49% of total population. Some other religions of the state are Hinduism (3.61%), Buddhist (8.30%) and Muslim (1.10%).The climate of Mizoram usually is monsoon and humid type. Its temperatures varies from region to region with it elevation changes. Over all temperature of the state is quite pleasant and rainy throughout the year. Normally the summer season of Mizoram is continued from End of March to June. The temperatures varies during this season from 20°C to 30 °C.Normally the winter season of Mizoram is continued from mid of November to February and the average temperature during this season varies between 10°C to 22°C in Garo hill. Normally the monsoon season of Mizoram is continued for a long term. It starts from June and continue to October. Rain and cloud is very common to this state and it receives more or less rainfalls not only in monsoon but also the other season of the state. The average annual rainfall of the state is approximately 250 cm.Mizoram witnesses three seasons- summer, winter and monsoon. However, extreme temperatures are not observed in any of the seasons. Summer starts in March and then monsoon season starts taking over from May onwards. Winter is the best time to explore the state. The cool weather along with lovely landscapes will swoon you. The temperature during this period remains in perfect range for travelers.

2. Objective of the Study

The objective of the study is to propose models for Rainfall data of Mizoram from 1990 – 2020 using ARIMA techniques and forecasting the Rainfall data of Mizoram for the next 10 years.

3. Source of Data

To fulfill the objective of this present study, the data considered is secondary in nature. It has been collected from the report entitled “Meteorological Data of Mizoram 2017” and “Meteorological Data of Mizoram 2020” by Directorate of Economics and Statistics, Planning and Programme Implementation Department, Government of Mizoram. Website www.des.mizoram.gov.in

4. Methodology

Box-Jenkins Model is a mathematical model designed to forecast data ranges based on inputs from a specified time series. The Box-Jenkins Model can analyze several different types of time series data for forecasting purposes. The Box-Jenkins Model forecasts data using three principles: Autoregression, differencing, and moving average. These three principles are known as p, q and d respectively. Each principle is used in the Box- Jenkins analysis; together, they are collectively shown as ARIMA (p,d,q). Here p represents the amount of autoregression, d indicates the level of systematic change over time (trend) and q represents the moving part. Thus, Box-Jenkins Method is useful for finding out the best fit model.

One of the most important in time series analysis is Stationary. A stationary time series has constant mean, variance and autocorrelation structure. So, checking the raw data is whether stationary or non-stationary is very important to conduct time series series analysis by using ARIMA Model. So, to perform ARIMA Model, firstly we need to check whether the raw data is stationary or not. The following step is use to check the data is stationary or not as follows:-

1) Firstly we draw a diagram or plot for raw data, from that the stationary plot series will show the constant location and scale.

2) Conducting the ACF and PACF Plots. As well as looking at the time series plot of the data, the ACF plot is very useful for identifying non-stationary time series. For a stationary

time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{autocovariance\ at\ lag\ k}{variance\ of\ the\ time\ series}$$

3) Dickey-Fuller (ADF) unit root test.

So, if the time series is a stationary, then the time series has constant mean, variance and autocorrelation structure. But if the time series presents a trend, then the method of differencing can be use to remove the linear or curvilinear trend. After transforming non-stationary to stationary by using differencing method, we need to find the order of ARIMA(p,d,q), where p represents the amount of autoregression, d represents the level of systematic change over time (trend) and q represents the moving part. ACF and PACF plots of the series are necessary to determine the order of AR and MA terms. Though ACF and PACF do not directly dictate the order of the ARIMA model, the plots can facilitate understanding the order and provide an idea of which model can be a good fit for the time-series data.

After finding the parameters for ARIMA model from the ACF and PACF Plots, we get to know how many lags are significant. So we need to test various models upto those lags to find out the best model.AIC refers to Akaike Information Criteria and is a criterion for a selection among the finite set of models. The model with the lowest AIC is the best model. A good model is the one that has minimum AIC among all the other models.

$$\text{AIC} = -2\ln(L) + 2k$$

$$= n\ \ln(\sigma_a^2) + 2k$$

Where 'σ_a^2' is the M.L.E and k is the number of the parameters estimated in the model.

The formula for calculating Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) as follows;-

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

Where, $y_i = prediction$

$x_i = true\ value$

n = total number of data points.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \text{X}100\%$$

Where, $n = number\ of\ times\ the\ summation\ iteration\ happens.$

$A_t = Actual\ value.$

$F_t = Forecast\ value.$

RMSE = $\sqrt{\frac{\sum_{i=1}^{N}|y_i-z_i|^2}{N}}$

Where, N is the data points.

Yi = i^{th} *measurement.*

Zi= corresponding prediction.

Ljung – Box Test (L.B Test):-

LB Test is a type of statistical test of whether any of a group of autocorrelations of a time series is different from zero. It is widely applied in econometrics and other applications of time series analysis. It is applied to the residuals of a fitted ARIMA Model, not the original series, and in such applications the hypothesis actually being tested is that the residuals from the ARIMA Model have no autocorrelation. When testing the residuals of an estimated ARIMA Model, the degrees of freedom need to be adjusted to reflect the parameter estimation.

The test statistics is given by

$$LB = n(n+2)\sum_{k=1}^{h}\frac{\rho_k^2}{n-k} \sim \chi_m^2$$

Where n is the sample size, ρ^k is the sample autocorrelation at lag k and h is the number of lags being tested.

Since the calculated value of Chi-square is less than the observed value, then we reject the null hypothesis and there is no evidence of white noise between the lags.

5.1 Stationary Test

Figure 1 shows that the time series plots the rainfall data of Mizoram from 1990 – 2020. From the plots, we can see that the pattern is increasing from 1990 to 1991 and start decreasing to 1992 and again increasing and rapidly decreasing from 1993 to 1994 and again start highly increasing. So, we can observe that the raw data is non-stationary in mean and variance. We can also check the non-stationary pattern from observing ACF and PACF as in Figures 2 and 3.The ACF Plots tough the significant bound at lag 3 and start to decreasing slowly. From that the ACF Plots, we can conclude that the raw data is non-stationary in nature and for confirmation, we performed the ADF Test and from that Dickey-Fuller = -2.4039 and p-value = 0.4173. So, we can agree that the rainfall data of Mizoram from 1990-2020 is non-stationary. So, we can change that the nonstationary data to stationary, we should perform the differencing method.

Figure 1: Time series Plot

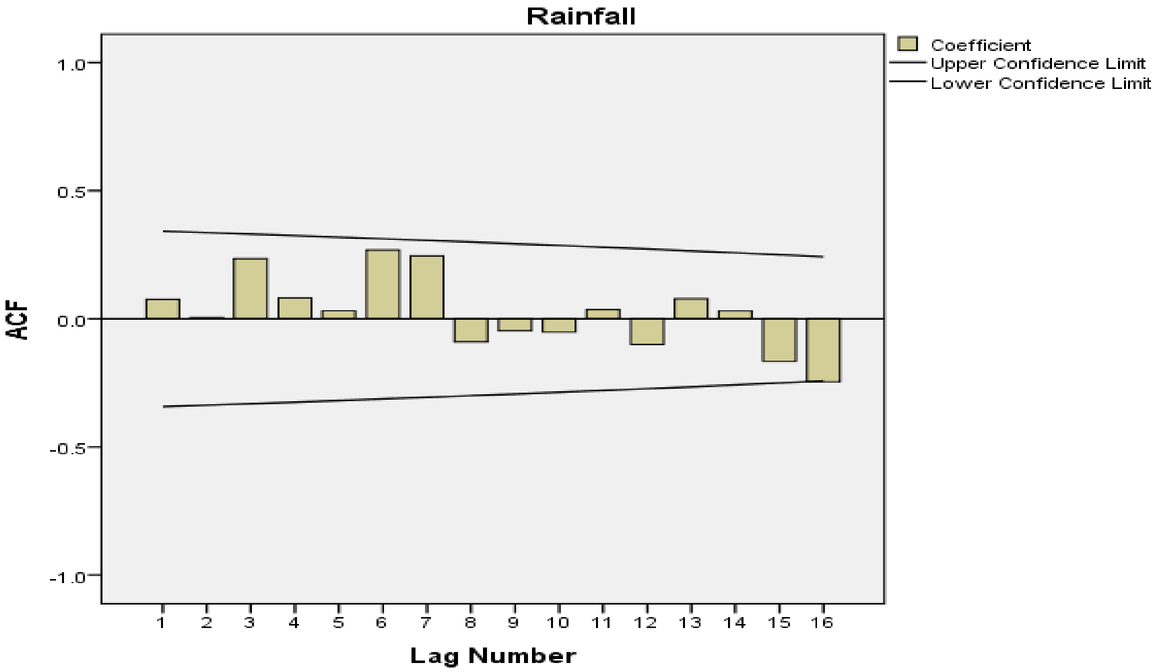


Figure 2: ACF Plot

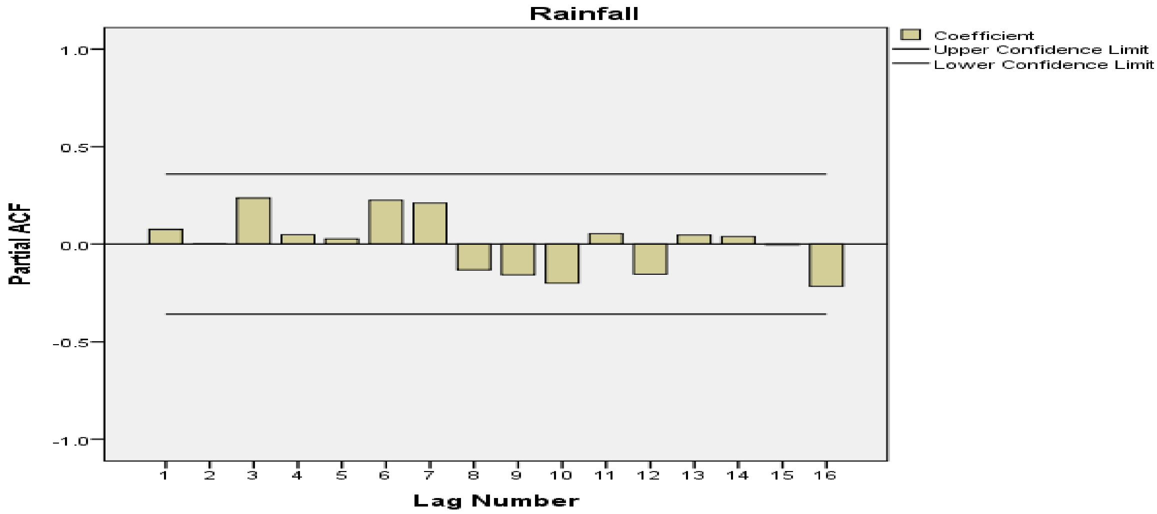


Figure 3: PACF Plot

Now, the first difference of the time series plots of the data of rainfall is given in Figure 4 and the first difference of ACF and PACF plots as in Figure 5 and Figure 6 respectively. From that the first difference plot, ACF plot and PACF plots, we can conclude that the non-stationary is now removed and become a stationary in mean. For confirmation of Stationary, we conduct the ADF Test and with ADF test, Dickey-Fuller = - 3.9444 and p-value = 0.02434. So, from ADF Test, we can say that the data is now stationary.

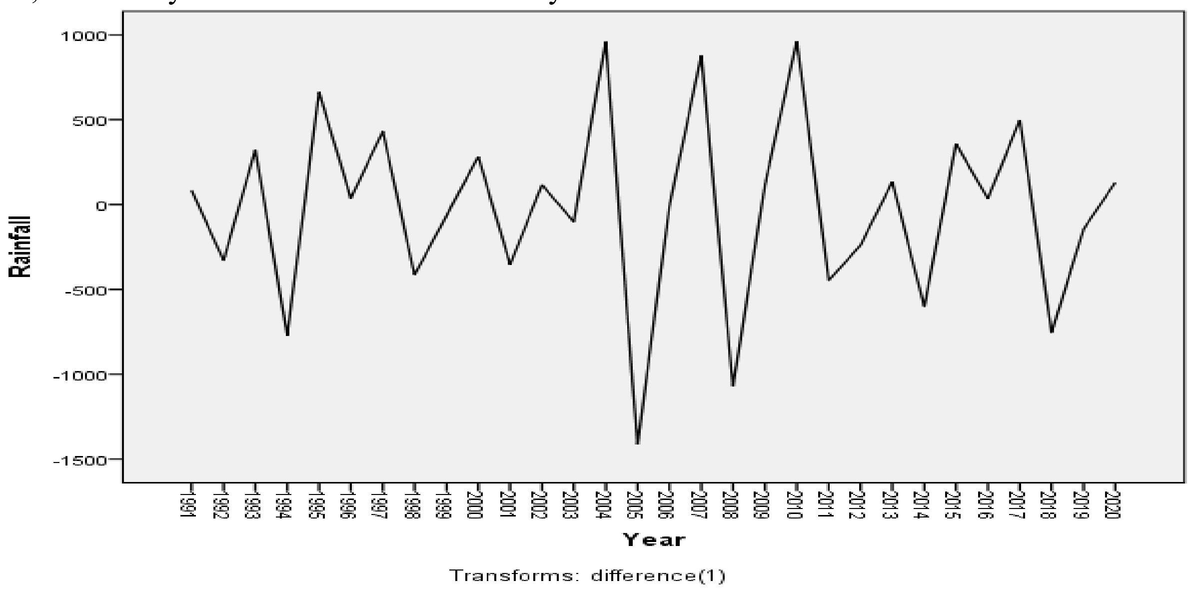


Figure 4: Time series plot after first differenced

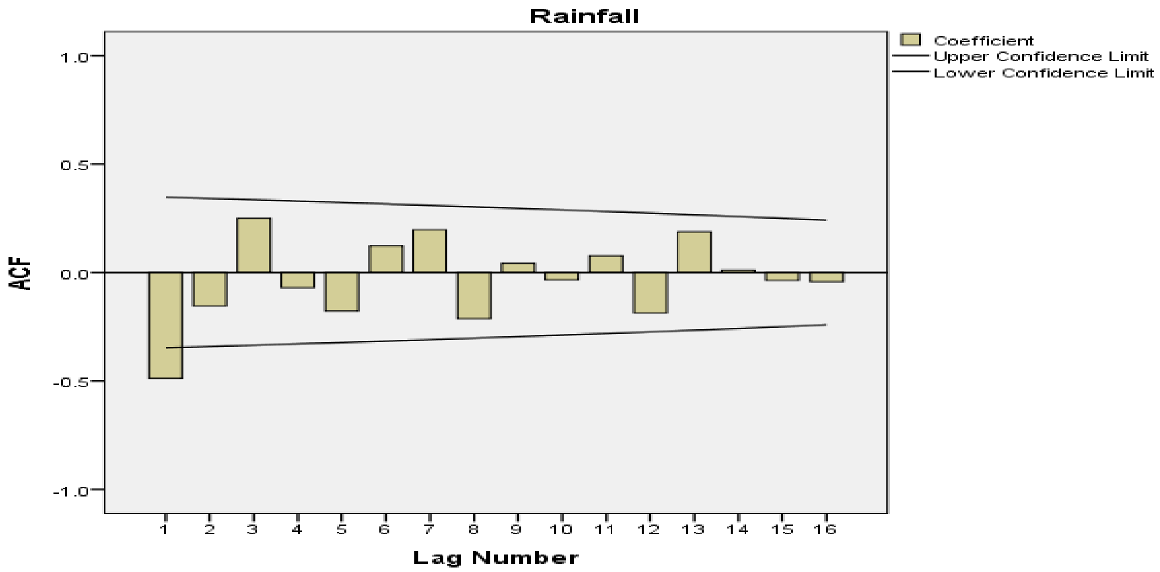


Figure 5: ACF plot after first differenced

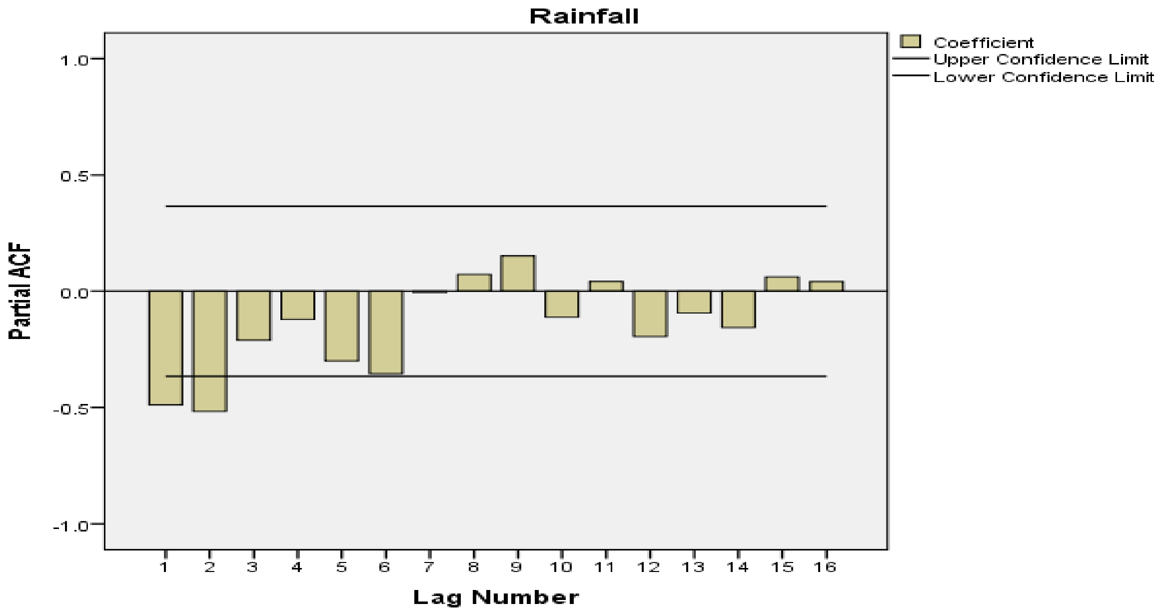


Figure 6: PACF plot after first differenced

So, from first differencing ACF Plots, we can see that the plots at lag 1 tough the significant bound and start drop quickly near to zero and all the other are under the significant bound. So, we can confirm that the time series is stationary.

5.2 Model Identification

The next step is to find out the order of ARIMA (p,d,q) from first difference of ACF and PACF Plots. In this present study, the data is once different and then the value of'd' is 1. As looking first difference of ACF plots from Figure 5, only lag 1 tough the significant bound and all the other were in under significant bound, so the possible value for 'q' is 1 and may be MA(1) is the most appropriate model and from the PACF Plots in Figure 6, lag 1 and 2 tough the significant bound and all the other were in under the significant bound. So, the possible value for 'p' is 2 and may be AR(2) is the most appropriate model. So, the possible best fit models are as follows:-

A) ARIMA (2,1,0)
B) ARIMA (3,1,1)
C) ARIMA (1,1,1)
D) ARIMA (2,1,1)
E) ARIMA (0,1,1)

The AIC, MAE and MAPE values for the different ARIMA Models are as follows:-

Model	AIC	RMSE	MAE	MAPE
ARIMA(2,1,0)	452.78	9.43	7.29	95.23
ARIMA(3,1,1)	454.89	9.62	8.36	96.45
ARIMA(1,1,1)	452.35	8.12	6.48	92.36
ARIMA(2,1,1)	452.78	7.58	7.23	95.46
ARIMA(0,1,1)	450.72	8.23	7.45	93.45

AIC refers to Akaike Information Criteria and is a criterion for a selection among the finite set of models. The model with the lowest AIC is the best model. A good model is the one that has minimum AIC among all the other models. Then from the above table, the best fit model is ARIMA(0,1,1) with the minimum AIC value.

5.3 Model Estimation

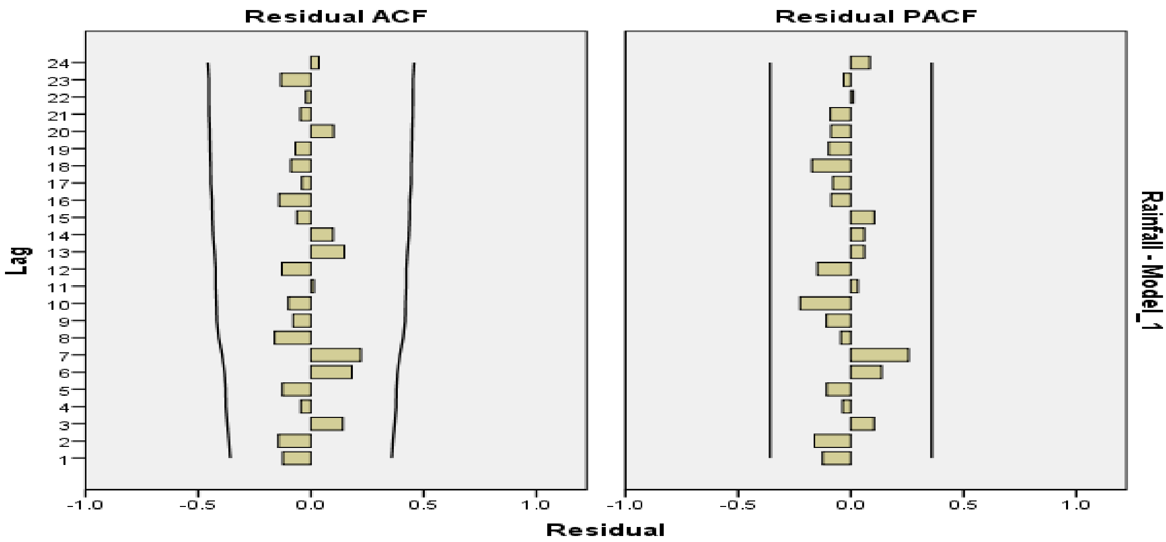
In the present study, the best fit model is ARIMA (0,1,1) with AIC value 450.72 . Then, the next step is to estimate the parameters and the coefficient of estimated parameters for ARIMA(0,1,1) Model is given by

Variables	Coefficients	S.E	σ^2	log likelihood
MA(1)	-0.818	0.09	29.78	-203.36

5.4 Diagnostic Check

Diagnostic Check performed when we want to make inferences with a model, in particular when the estimated standard errors of the parameters are used to make such inferences. To investigate the adequacy of the fitted model, we may examine the residuals, which should appear to be similar to observations from a white noise process, uncorrelated with each other and identically distributed. Diagnostic check that how our model is well fitted. Then after fitting the ARIMA(0,1,1) Model, the Ljung-Box and the residual of ACF and PACF Plots are given below. From residual ACF and PACF plots, we observed that all the plots were under the significant bound and no one tough the significant bound and hence, we can say that ARIMA(0,1,1) Model is the best model for the time series analysis of our rainfall data.

With Ljung Box Chi-Square Test, χ^2 =12.958 , p-value =0.05 and $\chi^2 = 27.59$ $(tabulated)$.



The above figure represents the Residual ACF and PACF Plot of the time series of rainfall.

5.5 Forecasting for the next 10 years.

After identifying the best fit model ARIMA(0,1,1) for our time series data, the next step is to forecast the amount of rainfall can be in Mizoram for the next 10 years. The prediction of rainfall is very important for all the people of Mizoram and especially for the people who live in rural areasand also for farmers.Now, the following table represents the forecast value of rainfall value in Mizoram with 95 prediction interval after fitting ARIMA(0,1,1) Model as follows:-

Year	Forecast	95% Prediction Interval	
		Lower	Upper
2021	2016.36	1350.77	2972.20
2022	2103.25	1337.56	2985.40
2023	2161.48	1324.57	2998.39
2024	2161.48	1311.78	3011.19
2025	2161.48	1299.17	3023.79
2026	2161.48	1286.75	3036.22
2027	2161.48	1274.50	3048.46
2028	2161.48	1262.42	3060.55
2029	2161.48	1250.49	3072.47
2030	2161.48	1238.73	3.84.24

6. Discussion

The Box-Jenkins ARIMA Model has been used to analysis of the rainfall data of Mizoram from 1990-2020. In this present study, we verified ARIMA (0,1,1) Model is the most appropriate model for our time series data. The raw data is getting differenced once to make non-stationary (with Dickey-Fuller = -2.4039 and p-value = 0.4173) to stationary(with Dickey-Fuller = -3.9444 and p-value = 0.02434) and also stationary verified by looking first differenced of ACF and PACF Plots. Then we reject the null hypothesis and the data is now becoming stationary. After that, we choose the parameters by observing the first differenced of ACF and PACF Plots and the best fit model is ARIMA (p,d,q) and the parameters is estimated and also performing the residual of Diagnostic Test. Now, the time series is stationary and we forecast the rainfall for the next 10 year with 95 prediction intervals by using IBM SPSS. From the forecast value, Mizoram will get more rainfall than the 2019 but less than the other years, that indicates that the amount of getting rainfall in Mizoram will decrease. So from our result, we can conclude that we need more awareness about afforestation and saving forest, so that we can make more rainfall in Mizoram and we can make Mizoram in better place. Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property, so a better forecasting model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms in better way.

References

1. *S.C. Gupta and V.K.Kapoor (2013). Fundamental of Mathematical Statistics, Sultan Chand and Sons, 23, Daryaganj, New Delhi-110002,* ***(2013).***
2. "Meteorological Data of Mizoram 2017 and 2020", Directorate Of Economics and Statistics, Planning and Programme Implementation Department, Government Of Mizoram. Website www.des.mizoram.gov.in .
3. *G.E.P. Box , G.M. Jenkin and G.C. Reinsel. Time Series Analysis and Forecasting and Control, Third Edition, Delhi, India Pearson Education,* ***(2003).***
4. *S. Chakraborty, L. Choudhury, R. Geeta and Sarmah. Results of Appeared Candidates in Matriculation / HSLC Examination in Assam (India): A Time Series Analysis (1951-2014),* ***(2016).*** *Pp. 4-15.*
5. *G.S. Shilpashree. Statistical Analysis of Rainfall Pattern in Mandya District, University of Agricultural Science, Bengaluru,* ***(2015).***
6. *S. Bisgaard and M. Kulahci. Time Series Analysis, Forecasting by Example, Hoboken, New Jersey, Wiley, A John Wiley & Sons, INC.,* ***(2011).***
7. *North Eastern Region Rainfall (url: https://en.Wikipedia.org/wiki/Northeast_India).*
8. *F. Jamal, E. Latifa, A. Zineb, E.L. Haj and L. Abdeslam. Forecasting of demand using ARIMA model,* ***(2018).***
9. *Analysis of Rainfall and Temperature Trends in Northeast India-2013 (url: https://rmets.onlinelibrary.wiley.com>joc).*
10. *R.K. Agrawal and R. Adhikari. An Introductory Study on Time Series Modelling and Forecasting,* ***(2013).***
11. *Analysis of Rainfall Data, Silchar, Assam (url: https://www.researchgate.net>publication)*
12. *S.S. Dash and H.V. Hema Kumar. "Statistical and Trend Analysis of Climate Data of Bapatla (A.P), India". (url: https://doi.org/10.20546/ijcmas.2017.610.470)*
13. *Time Series Analysis – ARIMA Model using R- solfware. (url: https://youtu.be/VphyVSJMbpA)*
14. *Time Series Forecasting ARMA and ARIMA Model. (url: https://youtu.be/TVTQNZLIfk)*
15. *Time Series Analysis. (url: https://youtu.be/Prpu_U5tKkE)*