International Journal of Mathematical Archive-14(10), 2023, 13-21
Available online through www.ijma.info ISSN 2229 – 5046

RAINFALL PREDICTION USING ARIMA MODEL AND TREND ANALYSIS OF FUTURE RAINFALL OVER RANCHI DISTRICT, JHARKHAND, INDIA

CHANDAN KUMAR PANDIT*[1], ANAMOL KUMAR LAL[2]

[1]Research Scholar, [2]Assistant Professor
[1,2]University Department of Mathematics, Ranchi University, Ranchi, India.

(Received On: 13-09-23; Revised & Accepted On: 02-10-23)

ABSTRACT

Rainfall is getting heavily affected by climate change due to which pattern and intensity of rainfall of a region is getting affected. In this paper, rainfall prediction from year 2023 to 2050 is done for Ranchi by using ARIMA model. The historical rainfall data of 50 years is used to train the model. ARIMA(0,1,1) has the least normalized BIC value and is selected for future prediction. ARIMA(0,1,1) forecasted total annual rainfall for year 2022 with an accuracy of 87.35% which is very high accuracy for rainfall forecasting. Trend analysis of future rainfall from year 2023 to 2050 is done by using the predicted rainfall. Monthly future rainfall has significant downward trend for January, February, March, April, May, September and November. Whereas the monthly future rainfall of June, July, August, October and December has significantly upward trend. Seasonal future rainfall for Summer and Winter season has significantly downward trend whereas monsoon and post monsoon future rainfall have significantly upward trend. The future annual rainfall has significantly downward trend.

Keywords: *Mathematical model, ARIMA model, Rainfall forecasting, Trend analysis.*

INTRODUCTION

Rainfall is very important for sustainability of human kind as it is the only natural source of recharge ground water level. Human activities such as burning fossil fuels, deforestation and urbanization have led to increased global warming. Intergovernmental Panel on Climate Change (IPCC) reported that the warming of Earth will affect the frequency, magnitude and regional pattern of extreme events [1]. The increased global warming sets in new interaction among atmosphere, ocean and territorial land that reshape rainfall patterns on a global scale. There is a transformation in the intensity, frequency and distribution pattern of rainfall events.Researcher showed that the climate change had profound effect on rainfall variability and also on its intensity [2].The annual and monsoon rainfall of Ranchi has an upward trend [3]. Regions which used to receive reliable rainfall are now facing severe droughts, while the other regions which used to face less rainfall and even frequent droughts are getting frequent, intense and unpredictable rainfall.Therefore, it is very important to analyze the trend of future rainfall so that impact of extreme events can be minimized. Rainfall forecasting can be done by using mathematical modelling such Holt-Winter's method[4], ARIMA model, etc. ARIMA model was used for rainfall forecasting for Khordha district of Odisha which showed that the forecasted rainfall was very consistent with the observed rainfall [5]. Trend of future rainfall predicted by ARIMA was analyzed in Chattogram, Bangladesh which signified that monsoon and December month rainfall will increase for the period 2021-70[6]. The trend analysis of future weekly rainfall forecasted by ARIMA model in semi-arid Sinjar district of Iraq signified a decreasing trend [7]. ARIMA model was applied on water demand in Las Vegas Valley and found the forecasted value was very accurate [8]. The prediction of rainfall for Ranchi was done for the period 2023-30 using Holt-Winter's method which signified that Multiplicative Holt-Winter's method yields better forecasts than Additive Holt-Winter's method [9].

Therefore, in this paper, trend of future rainfall predicted by ARIMA model is analyzed.

Corresponding Author: Chandan Kumar Pandit*[1],
Research Scholar, University Department of Mathematics, Ranchi University, Ranchi, India.

STUDY AREA:
Ranchi is the capitol of state Jharkhand in India. Ranchi is situated on the Chota Nagpur plateau in eastern India. It is situated at average mean sea level of 629m with an area of 5097 Km^2. Ranchi has sub-tropical climate but due to surrounded by dense forest cover, it has very pleasant climate. Ranchi receives a unique topographical featured convectional rainfall during summer. The average annual rainfall is 1300mm out of which it receives 80.8% rainfall during south-west monsoon (June-September).

DATA AND METHODOLOGY:
The daily rainfall data of Ranchi is obtained from the website of National Centre For Environmental Information (https://www.ncdc.noaa.gov/cdo-web/datatools/findstation) for the period 1973-2022. Monthly, seasonal and annual rainfall are then calculated from the daily rainfall data. ARIMA model is used for the prediction of rainfall. ARIMA model requires stationary data and hence Dickey-Fuller Unit root test is applied on the data set to check stationarity of data. The better ARIMA model is selected by observing the lower value of normalized BIC. Trend analysis of future rainfall is analyzed by Mann-Kendall test and Sen's slope method. The methodologies of methods are described as follows:

ARIMA MODEL:
ARIMA means Autoregressive Integrated Moving Average. It is used in the time series analysis for the forecasting upcoming series points. ARIMA consists of three components: Autoregressive (AR), Integrated (I) and Moving Average (MA).The autoregressive (AR) component indicates that the evolving variable is regressed on its own lagged values. The Moving Average (MA) component shows that the regression error is essentially a linear combination of error components, the values of which happened simultaneously and at distinct points in the past. The Integrated (I) component indicates that the values of the data have been replaced by the difference between those values and the previous values .

ARIMA models are denoted by ARIMA (p, d, q) where p is the number of auto-regressive order, d is the order of differencing applied to the time series and q denotes the number of moving average order of the data series. The parameters p, d and q are non-negative integers.

ARIMA models require the data set to be stationary. It can also be applied on non-stationary data set by eliminating the non-stationarity of the mean function by introducing an initial differencing step one or more times. ARIMA models for different p, d and q values are developed and then the best model is selected by using the Akaike information criterion (AIC) and Bayesian information criterion (BIC).

ARIMA (p, d, q) is given as follows:

$$\bar{z}_t = \emptyset_1 \bar{z}_{t-1} + \emptyset_2 \bar{z}_{t-2} + \cdots + \emptyset_p \bar{z}_{t-p} + a_t - \theta_1 \bar{z}_{t-1} - \theta_2 \bar{z}_{t-2} - \cdots \theta_q \bar{z}_{t-q} \quad (1)$$

Where $\bar{z}_t = z_t - \mu$ and a_t is the shock.

Equation (1) can be applied after finding the backward shift operator (B) as follows:

$$\emptyset(B)(1-B)^d z_t = \theta(B) a_t \quad (2)$$

DICKEY-FULLER UNIT ROOT TEST:
The unit root test is used to check the stationarity of data set. It was developed by David Dickey and Wayne Fuller in 1979. When a data set has no unit root means that data set is stationary.
The unit root test in the time series y_t, the Dickey Fuller equation is given by

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \varepsilon_t$$

Where α represents the intercept and is constant, β represents the coefficient of trend and p is order of lag of autoregressive (AR) process. The unit root test is then carried out under the null hypothesis $\gamma=0$ against the alternative hypothesis of $\gamma < 0$. If $\gamma < 0$, then the series is stationary because there is no trend in the time series.

NORMALIZED BAYESIAN INFORMATION CRITERION (BIC):
It is used for the selection of model. It is developed by Gideon E. Schwarz in 1978. It is closely related to Akaike information criterion (AIC). It is possible to increase the maximum likelihood by adding parameters when fitting a model, but this can result in overfitting. BIC tries to resolve this problem by introducing a penalty term for the number of parameters in the model same as in AIC. However, the penalty term in BIC is larger than AIC.
The BIC is given as follows:

$$BIC = k \ln(n) - 2 \ln(\hat{L})$$

Where $\hat{L} = p(x \mid \hat{\theta}, M)$ is the maximize value of the model M and $\hat{\theta}$ is the parameter value that maximize the likelihood function; x is the observed value and n is the sample size of the data set; and k= the number of parameters estimated by the model

MANN-KENDALL TEST:
Mann-Kendall test is a non-parametric trend test which is used to determine whether or not there is a linear monotonic trend in a given time series data. It does not require the data to be normally distributed. The null hypothesis states that there is no monotonic trend, and this is tested against one of three possible alternative hypotheses: (i) there is an upward monotonic trend, (ii) there is a downward monotonic trend, or (iii) there is either an upward monotonic trend or a downward monotonic trend.

Let $x_1, x_2, x_3, \dots, x_n$ be data set of length n in a time series data. The indicator function Sgn$(x_i - x_j)$ is calculated as follows:

$$\text{Sgn}(x_i - x_j) = \begin{cases} 1, & x_i - x_j > 0 \\ 0, & x_i - x_j = 0 \\ -1, & x_i - x_j < 0 \end{cases}$$

Mean S and variance Var(S) of Sgn$(x_i - x_j)$ are calculated as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_i - x_j)$$

$$\text{And Var(S)} = \frac{[n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)]}{18}$$

Where t is the extent of any given tie.

Mann-Kendall test statistic Z is then computed as follows:

$$\mathbf{Z} = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & if\ S > 0 \\ 0 & if\ S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & if\ S < 0 \end{cases}$$

The value of Z>0 represents monotonic upward trend in data series whereas Z<0 represents monotonic downward trend.

SEN'S SLOPE METHOD:
It is a non-parametric test which is highly efficient to find out linear trend in univariate data series. This can be applied to a data set having missing value and outliers in the data series.

The Sen's estimator β of slope is calculated as follows:

$$\beta = \text{Median}\left(\frac{x_j - x_i}{j - i}\right)$$

Where x_j *and* x_i are the data values at time j and i (j>i) respectively.

β>0 indicates an upward trend whereas β<0 indicates downward trend in the data series.

RESULTS:
ARIMA model is used for forecasting rainfall by using monthly, seasonal and annual rainfall data for the period 1973-2022. The rainfall data is not stationary and hence differencing is introduced to make data stationary. The differencing of 1 makes the data stationary. Dickey-Fuller unit root test is applied to verify the stationarity of annual and seasonal rainfall data[Table 1]. The p value for annual and seasonal rainfall becomes 0 after applying differencing of 1, which emphasize that differencing of 1 make the data stationary. Annual rainfall plot and annual rainfall plot with differencing 1 are deployed in Figure 1.

Table-1: Dickey-Fuller unit root test statistics

Dickey-Fuller unit root test	t statistics	p value	Differencing =1	
			t statistics	**p value**
Annual	0.020	0.684	-7.574	0.00
Summer	-1.583	0.105	-6.122	0.00
Monsoon	0.057	0.69	-8.275	0.00
Post-Monsoon	-1.77	0.07	-8.143	0.00
Winter	-2.66	0.008	-7.037	0.00

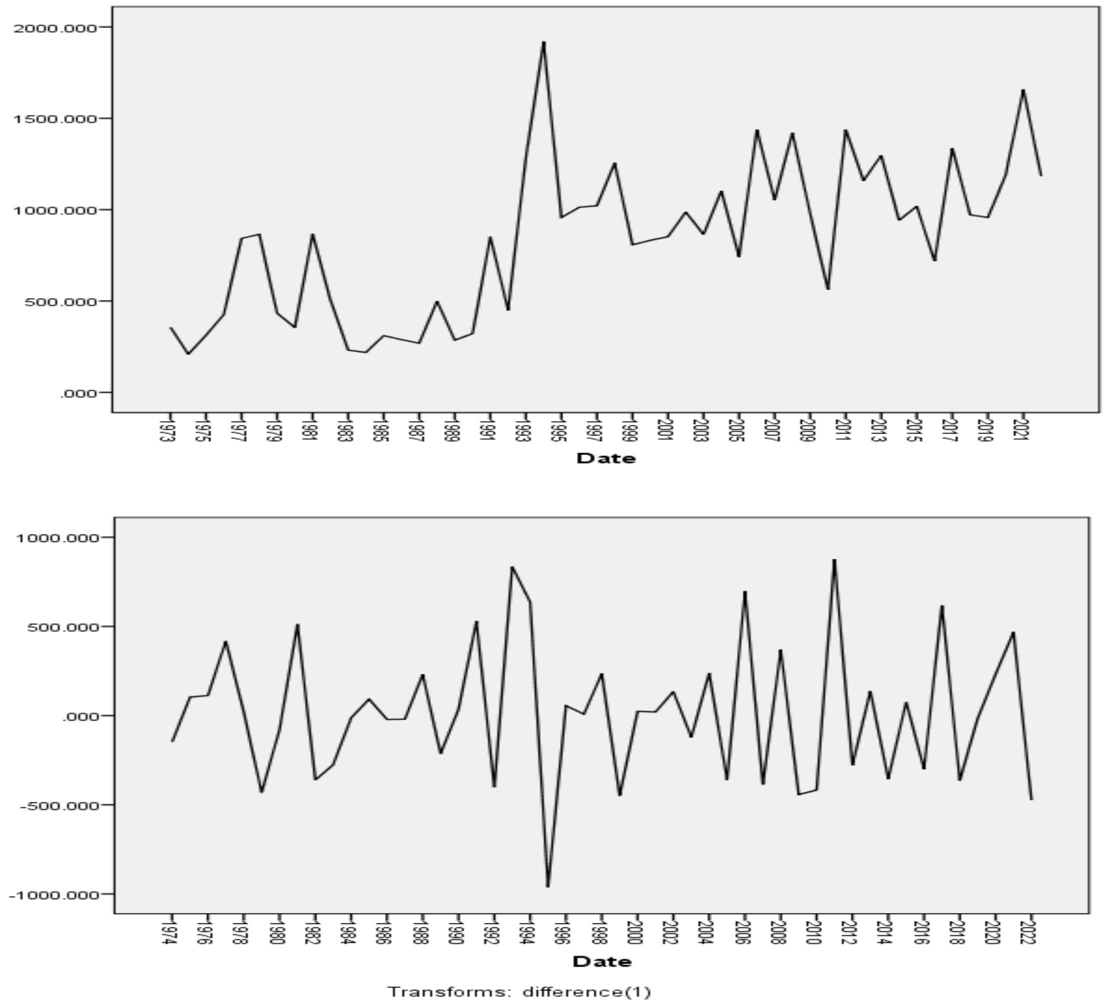

Figure-1: Annual rainfall graph (up) and Annual rainfall graph with differencing 1 (down)

Various ARIMA (p, d, q) is applied on the monthly, seasonal and annual rainfall data set for the period 1973-2022. Normalized BIC is calculated for all ARIMA models [Table 2]. The better ARIMA model is selected on the basis of lower value of normalized BIC. ARIMA (0,1,1) has normalized BIC of 8.405 which is the least among them.Hence, ARIMA (0,1,1) is selected for the future forecasting of rainfall from year 2023 to 2050. The forecasted value of year 2022 is then compared with the actual rainfall of year 2022[Table 3]. Forecasting error is calculated by subtracting the forecasted rainfall from the actual rainfall i.e.,

$$forecasting\ error = Acutal\ rainfall - forecasted\ rainfall$$

The plot of actual and forecasted rainfall is deployed together with the forecasting error [Figure 2].

The future rainfall forecast plots of seasonal, annual and monthly rainfall are deployed in Figure 3 and Figure 4 respectively. Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) of rainfall forecasts are also deployed. ACF and PACF plots of residuals clearly show that the forecasts are bounded within a region which means that forecasted values are close to the actual values.

Table-2: Normalized BIC value of various ARIMA models

ARIMA Model	Normalized BIC
ARIMA (0,1,1)	8.401
ARIMA (0,1,2)	8.490
ARIMA (0,1,3)	8.573
ARIMA (1,1,2)	8.581
ARIMA (1,1,1)	8.485

Table-3: Actual and forecasted rainfall of year 2022 with forecasting error

2022	Actual Rainfall	forecasted rainfall by ARIMA	Error
January	36.068	11.463	24.605
February	18.034	19.07	-1.036
March	0.254	20.947	-20.693
April	6.096	3.95	2.146
May	42.164	53.371	-11.207
June	96.52	198.585	-102.065
July	144.526	347.522	-202.996
August	556.26	315.396	240.864
September	255.016	236.304	18.712
October	29.21	85.532	-56.322
November	0.00001	13.449	-13.449
December	0.00001	28.417	-28.417

The annual rainfall of Ranchi for year 2022 is 1184.14 mm and the forecasted annual rainfall for year 2022 is1334 mm. The error % of forecasted annual rainfall is 12.65%. ARIMA(0,1,1) forecasted total annual rainfall for year 2022 with an accuracy of 87.35% which is very high accuracy for rainfall forecasting.

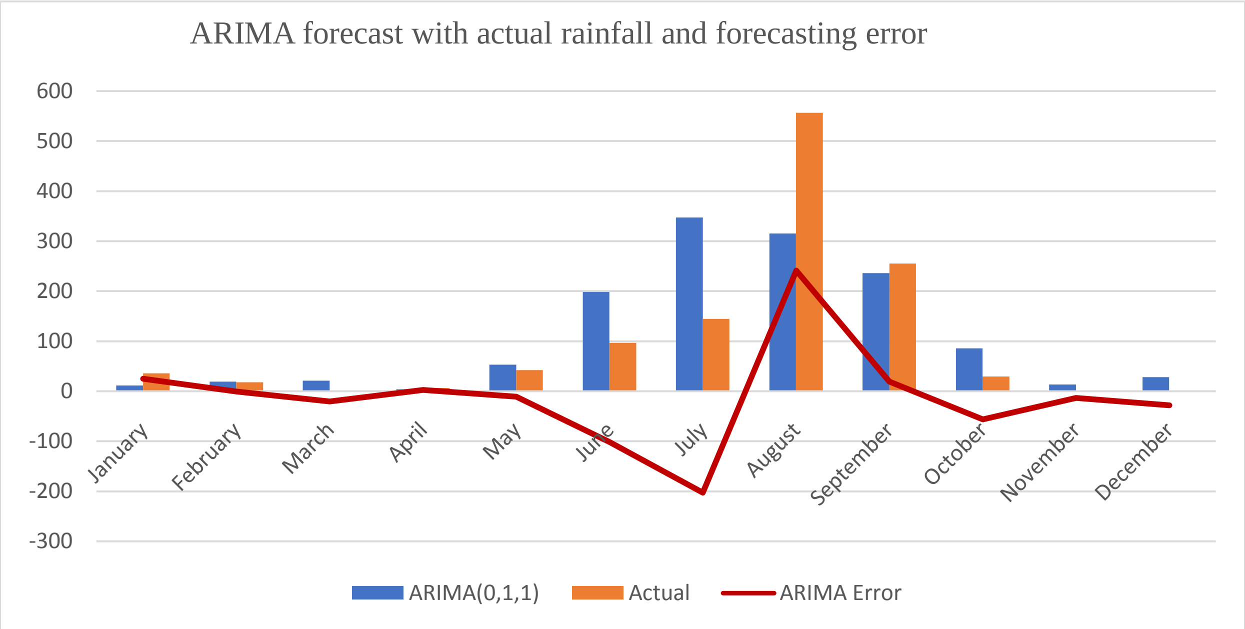

Figure-2: Actual and forecasted rainfall graph with forecasting error for year 2022

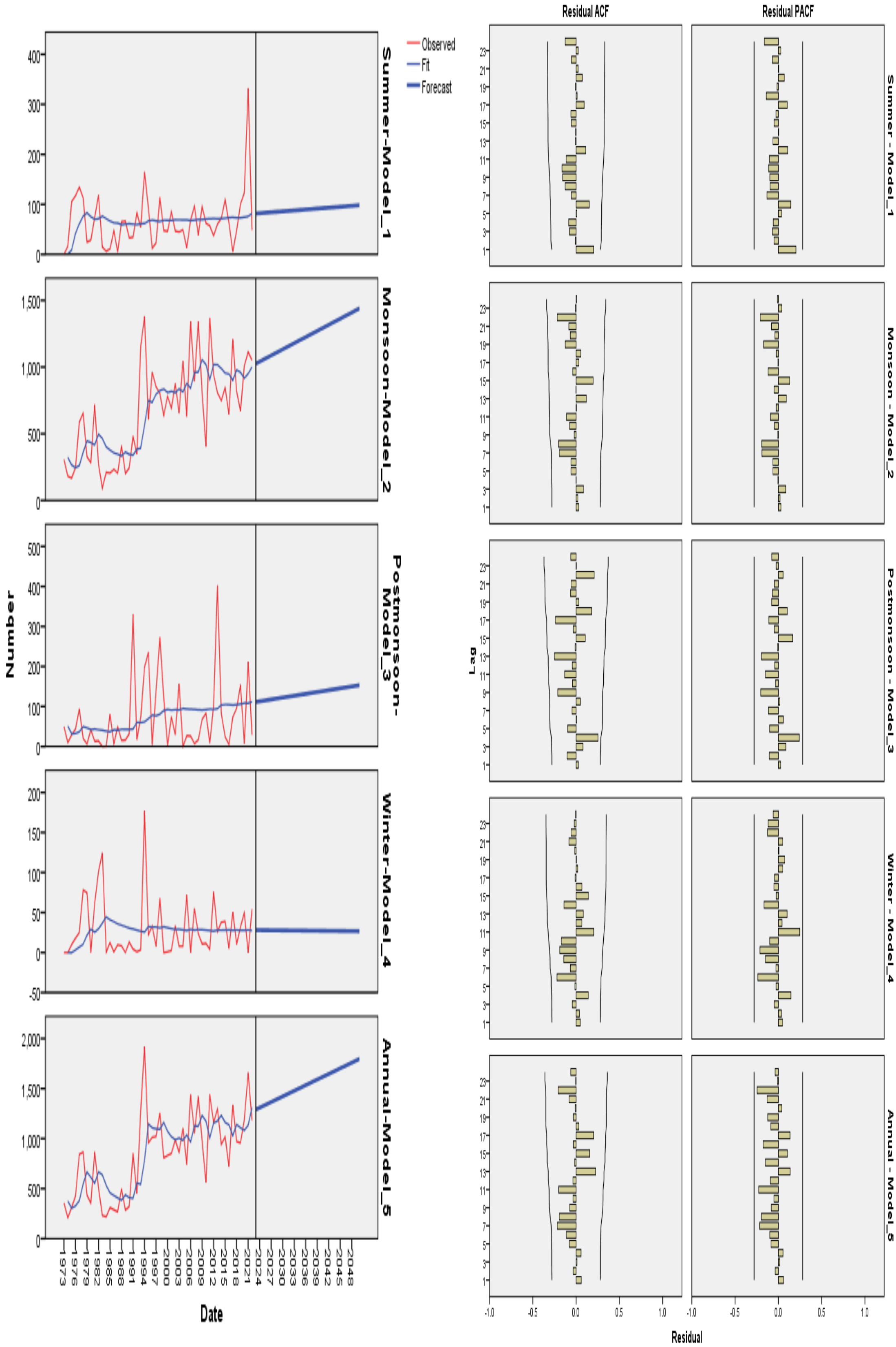

Figure-3: Seasonal forecasted rainfall graph (left) for 2023 to 2050 with ACF, PACF forecasting residuals graph (right)

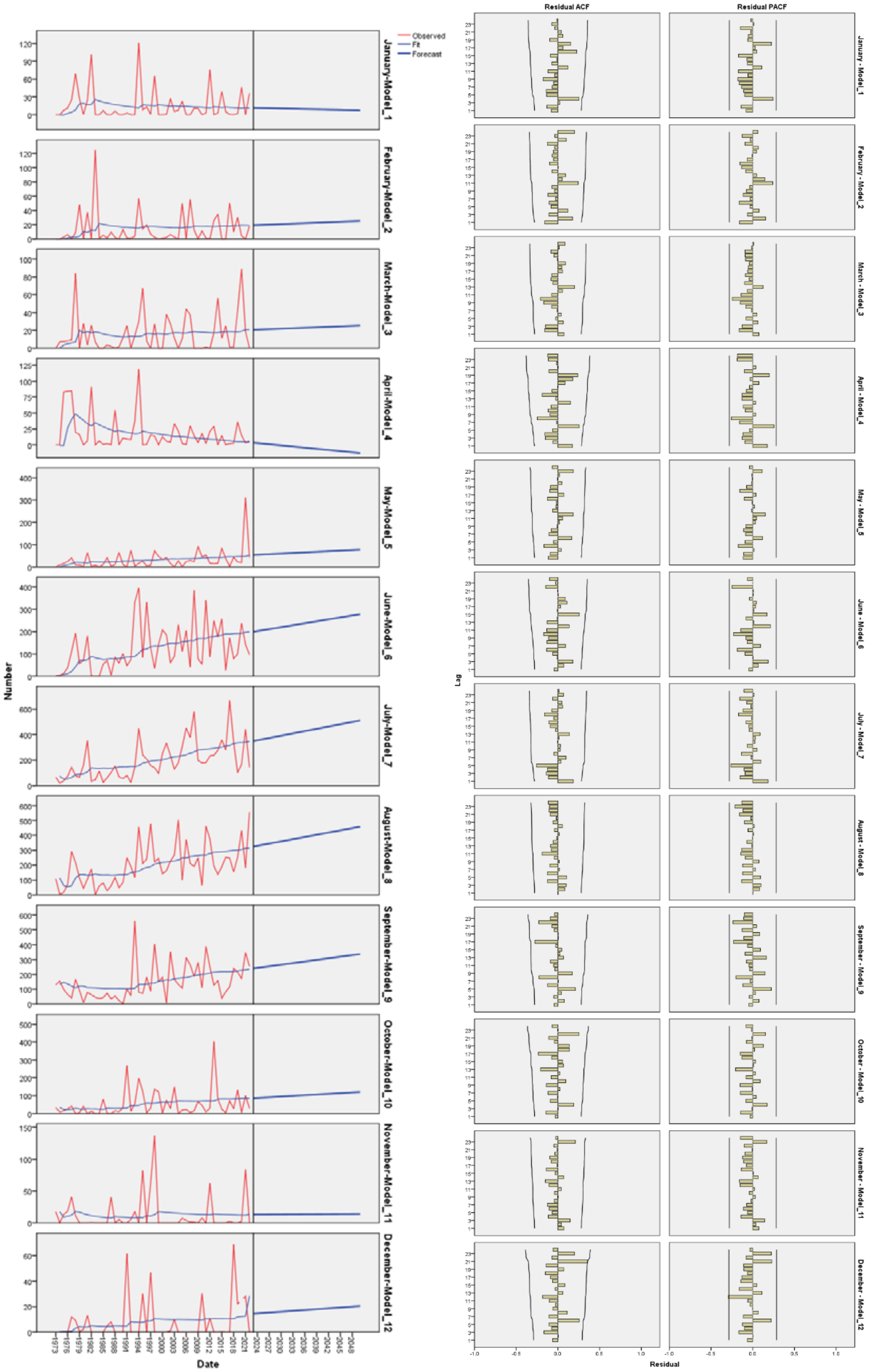

Figure-4: Monthly forecasted rainfall graph (left) for period 2023-50 and ACF,PACF forecasting residuals plot (right)

PREDICTED RAINFALL BY ARIMA MODEL:

Rainfall prediction for the period 2023-50 is done using ARIMA(0,1,1) [Table 4]. The predicted annual rainfall for the year 2050 is 1796 mm out of which 80.14% is received in the monsoon season. The summer, post monsoon and winter season contributes respectively 5.47%, 8.53%, 1.48%.

Table-4: Seasonal rainfall forecasts from year 2023 to 2050

Year	Summer	Monsoon	Post-Monsoon	Winter	Annual
2023	81.842	1026.037	111.541	28.07	1290.852
2024	82.452	1041.359	113.086	28.021	1309.573
2025	83.062	1056.682	114.63	27.971	1328.294
2026	83.672	1072.004	116.175	27.921	1347.015
2027	84.283	1087.327	117.719	27.872	1365.736
2028	84.893	1102.649	119.264	27.822	1384.457
2029	85.503	1117.972	120.809	27.772	1403.178
2030	86.113	1133.295	122.353	27.723	1421.899
2031	86.723	1148.617	123.898	27.673	1440.62
2032	87.333	1163.94	125.442	27.623	1459.341
2033	87.943	1179.262	126.987	27.574	1478.062
2034	88.553	1194.585	128.531	27.524	1496.783
2035	89.163	1209.907	130.076	27.474	1515.505
2036	89.773	1225.23	131.62	27.425	1534.226
2037	90.383	1240.552	133.165	27.375	1552.947
2038	90.993	1255.875	134.709	27.325	1571.668
2039	91.603	1271.198	136.254	27.276	1590.389
2040	92.213	1286.52	137.799	27.226	1609.11
2041	92.823	1301.843	139.343	27.176	1627.831
2042	93.433	1317.165	140.888	27.127	1646.552
2043	94.043	1332.488	142.432	27.077	1665.273
2044	94.653	1347.81	143.977	27.027	1683.994
2045	95.263	1363.133	145.521	26.978	1702.715
2046	95.873	1378.456	147.066	26.928	1721.436
2047	96.483	1393.778	148.61	26.878	1740.157
2048	97.093	1409.101	150.155	26.829	1758.878
2049	97.703	1424.423	151.699	26.779	1777.599
2050	98.313	1439.746	153.244	26.729	1796.32

TREND ANALYSIS OF FUTURE RAINFALL:

The future rainfall of 2023-2050 is analyzed for trend analysis. MK test statistics and Sen's slope estimator are calculated Table 5[Table 24]. The monthly future rainfall has significant downward trend for January, February, March, April, May, September and November. Whereas the monthly future rainfall of June, July, August, October and December has significantly upward trend. The seasonal future rainfall for Summer and Winter season has significantly downward trend whereas monsoon and post monsoon future rainfall have significantly upward trend. The future annual rainfall has significantly downward trend.

Table-5: MK test statistics and Sen's estimator for forecasted rainfall of period 2023-50

Months	Z	β	Trend
January	-3.53	-0.154	Downward
February	-3.48	0.223	Downward
March	-4.94	0.179	Downward
April	-6.46	-0.603	Downward
May	-5.42	0.889	Downward
June	4.28	2.930	Upward
July	7.14	5.985	Upward
August	4.89	4.936	Upward
September	-4.98	3.598	Downward
October	6.35	1.3	Upward
November	-3.79	0.032	Downward
December	5.91	0.223	Upward
Seasons			
Summer	-7.01	0.610	Downward
Winter	3.22	15.323	Upward
Post-Monsoon	3.04	1.545	Upward
Winter	-3.97	-0.050	Downward
Annual	-6.22	18.721	Downward

DISCUSSION:

Rainfall forecasting is done from year 2023 to 2050 by using ARIMA model. ARIMA (0, 1, 1) is selected for forecasting as it has the least normalized BIC value. Forecasting rainfall of year 2022 is compared with the actual rainfall of year 2022 and the error% of total annual rainfall for 2022 is 12.65%.ARIMA(0,1,1) forecasted total annual rainfall for year 2022 with an accuracy of 87.35% which is very high accuracy for rainfall forecasting. Trend analysis of future rainfall from year 2023 to 2050 is done by using the predicted rainfall. Monthly future rainfall has significant downward trend for January, February, March, April, May, September and November. Whereas the monthly future rainfall of June, July, August, October and December has significantly upward trend. Seasonal future rainfall for Summer and Winter season has significantly downward trend whereas monsoon and post monsoon future rainfall have significantly upward trend. The future annual rainfall has significantly downward trend.

REFERENCES

1. P.N. Lal, T. Mitchell, P. Aldunce, H. Auld, R. Mechler, A. Miyan, L.E. Romano, and S. Zakaria: National systems for managing the risks from climate extremes and disasters. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 339-392, 2012.
2. Saleh A. Wasimi. CLIMATE CHANGE TRENDS IN BANGLADESH. 2nd International Conference on Water & Flood Management (ICWFM-2009).
3. Chandan Kumar Pandit, Anamol Kumar Lal, Uma Shanker Singh; "ASSESSMENT OF TEMPORAL RAINFALL AND ITS CORRELATION WITH TEMPERATURE IN RANCHI, JHARKHAND", *International Journal of Current Research*, Vol. 15, Issue 08, pp. 25519-25522, August, 2023. DOI: 10.24941/ijcr.45782.08.2023
4. Charles C. Holt. "Forecasting seasonals and trends by exponentially weighted moving averages",*International journal of forecasting*, vol. 20, no. 1, pp. 5-10, 2004.
5. S. Swain, S. Nandi and P. Patel, "Development of an ARIMA Method for Monthly Rainfall Forecasting over Khordha District, Odisha, India",*In Recent findings in intelligent computing techniques*, pp. 325-331. Springer, Singapore, 2018. DOI:10.1007/978-981-10-8636-6_34
6. M. H. Masum, R. Islam, M. A. Hossen, A. A. Akhie,"Time Series Prediction of Rainfall and Temperature Trend using ARIMA Model",*J. Sci. Res*. 14 (1), 215-227, 2022.
7. S. Zakaria, N. Al-Ansari, S. Knutsson and T. Al-Badrany, "ARIMA models for weekly rainfall in the semi-arid Sinjar district at Iraq",*J. Earth Sci. Geotech*. Eng., 2(3),25–55, 2012.
8. P. Huntra and T.C. Keener, "Evaluating the impact of meteorological factors on water demand in the Las Vegas Valley using time-series analysis: 1990–2014"; *Int. J. Geo-Inf*. 6 249, 2017
9. Chandan Kumar Pandit, Anamol Kumar Lal, Uma Shanker Singh, "Rainfall Forecasts by Mathematical Model using Monthly Rainfall Data for Ranchi District of Jharkhand in India," *International Journal of Mathematics Trends and Technology*, vol. 69, no. 8, pp. 62-67, 2023. Crossref, https://doi.org/10.14445/22315373/IJMTT-V69I8P508

Source of support: Nil, Conflict of interest: None Declared.