The Pharma Innovation Journal 2022; SP-11(4): 1060-1066

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(4): 1060-1066

www.thepharmajournal.com
Received: 28-02-2022
Accepted: 30-03-2022

Nishtha Pradarshika Rai
Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal, India

Arunava Ghosh
Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal, India

Naffees Gowsar SR
Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal, India

Chowa Ram Sahu
Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal, India

Corresponding Author
Nishtha Pradarshika Rai
Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Pundibari, Cooch Behar, West Bengal, India

Modelling on annual and monsoon rainfall of Cooch Behar district in West Bengal: An ARIMA approach

Nishtha Pradarshika Rai, Arunava Ghosh, Naffees Gowsar SR and Chowa Ram Sahu

DOI: https://doi.org/10.22271/tpi.2022.v11.i4So.12091

Abstract
The present study was conducted to identify the best fitted time series model to describe the annual and monsoon rainfall in Cooch Behar District of West Bengal. Rainfall data for 40 years, i.e. from 1978 to 2017 was collected from Indian Meteorological Department, Pune. The three stage Box Jenkins procedure for selection of Auto Regressive Integrated Moving Average model was performed. From the evaluation of model combinations, it was found that ARIMA (0,0,1) is the most appropriate model for fitting both monsoon as well as annual rainfall data. ARIMA (0,0,1) possessed lowest Bayesian Information Criteria and Root Mean Square Error. The Absolute Relative Percent Forecast Error was utilized for confirming the predictive power of the identified best fitted model. The same model was used for predicting the annual and monsoon rainfall for the next three years.

Keywords: Rainfall, ARIMA, ACF, PACF, BIC, RMSE, ARPFE

1. Introduction
Rainfall is a spectacular feature and the most influential climate variable of the Indian subcontinent. The distribution of rainfall varies widely in both space and time across the nation due to its heterogeneous nature of distribution (Nirmala and Sundaram, 2010) [6]. The Indian Subcontinent is primarily fed by the South West Monsoon (SWM) or Indian Summer Monsoon Rainfall (ISMR) occurring primarily from June to September and is also called JJAS Rainfall. The ISMR is not only responsible for climatic variability but also for economic stability in the country. Rainfall Modelling is imperative for the general understanding of hydrology and in comprehending rainfall distribution in specific areas.
In West Bengal, South West Monsoon is accompanied by thunderstorms in the Northern and Terai regions of the state. Cooch Behar, also known as Koch Behar is categorized under Eastern Himalayan Region (Agro Climatic Zone) by the Planning Commission of India. The climate of Cooch Behar is described by substantial precipitation (98%) occurring from the first week of June to the last week of September and slight precipitation from October to mid-November. The annual average rainfall in the region is between 3200 mm to 3300 mm.
A handful of investigations related to rainfall pattern are reported of being conducted in Cooch Behar. Understanding the annual and monsoon inconsistency is significant for the improvement of the likely effect of changes in the precipitation pattern. The ultimate goal of rainfall modelling studies is to build an estimate of the prevailing rainfall pattern.
With the passage of time, several kinds of models were constructed and of which the Time Series model is most frequently used. A time series model is an empirical model that uses past values in order to predict future values. In West Bengal, Pal *et al.* (2015) [7], Nandargi and Barman (2018) [5] have made use of the Time Series model to study ISMR. Jeet *et al.* (2021) [9], Dwivedi *et al.* (2019) [3], Shivhare *et al.* (2019) [9] have explained rainfall using ARIMA Models to forecast rainfall in Ranchi, Gujarat and Varanasi, respectively.

2. Materials and Methods
2.1 Data
Annual rainfall and monsoon rainfall data of Cooch Behar station over a span of 40 years, i.e. from 1978 to 2017 was collected from IMD, Pune. It is reported by the IMD that four years data viz. 1993, 1994, 1999, and 2001 were missing. This data was substituted by data obtained from the State Govt. of West Bengal.

2.2 Description of the Model
The time-series data of Cooch Behar on annual rainfall and monsoon rainfall was utilized for identification of the best ARIMA model by performing the three-stage (identification, estimation, and diagnostic checking) iterative procedure (Box and Jenkins, 1976) for forecasting rainfall. The general form of the ARIMA model is written as:

$$(1-\varphi_1 B-\varphi_1 B^2-\cdots-\varphi_p B^p)(1-B)^d y_t=(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)a_t$$

Where, φ's and θ's are the AR (autoregressive) and MA (moving average) coefficients; B is the backshift operator; d is the order of differencing, p is the AR operator order, q is the MA operator order, y_t is the t-th observed time series value; and a_t is the t-th white noise.

a. Identification of models

A time series variable should be strictly stationary under normality assumption. A non-stationary variable is to be converted to a stationary one by suitable transformations. Normality assumption can be checked graphically by quantile-quantile (Q-Q) plot and independence by employing Ljung-Box (Ljung and Box, 1978) [4] statistic. The theoretical autocorrelation functions (ACF's) and partial autocorrelation functions (PACF's) are compared with their estimated ACF's and PACF's for determining the order of the (AR) and (MA) processes.

b. Estimation of parameters

In the ARIMA (p,d,q) model, p, d, q are the parameters and represents the order of AR process, differencing, and MA process respectively. The iterative finite-unconditional-least-squares method is applied for obtaining the precise estimates of parameters. Each coefficient is estimated and tested for its significance. Finally, stationarity and invertibility are checked for these estimated coefficients.

c. Diagnostic checking of models

This stage is important for screening the models for their goodness-of-fit, i.e., the model adequacy for predicting the future values. A set of estimated residuals are evolved after undergoing for estimation of parameters. Independence (p value of Q statistic > 0.05) and normality should be satisfied by these residuals for obtaining adequate models.

2.3 Model evaluation techniques
The Mean Absolute Error (MAE), Mean Absolute Per cent Error (MAPE), Bayesian Information Criteria (BIC), and estimated residual mean square (RMS) for each model are computed to check the accuracy of selected model.

- Mean Absolute Error (MAE) is the arithmetic average of the differences between predicted and actual absolute error values.

MAE = $T^{-1} \times \sum_{t=1}^{T} |p_t - y_t|$

- Percentage of error present in the forecasting model is measured by the Mean Absolute Per cent Error (MAPE) and is computed as:

MAPE = $100 \times T^{-1} \sum_{t=1}^{T} | \hat{a_t} / y_t|$

- Bayesian Information Criteria (BIC) is derived from Bayesian Probability and Inference. It uses maximum likelihood function for model fitting.

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

- Residual Mean Square (RMS) or Mean Square Error (MSE) is the average of the square of errors.

RMS = $T^{-1} \sum_{t=1}^{T} \hat{a}_t^{\,2}$

Where, T = no. of observations, p_t = predicted value at time 't', y_t = observed value at time 't', $\hat{a}_t$ = estimated value of the white noise at time 't', $\hat{a}_t = y_t - p_t$, k = no. of parameters, L = maximum likelihood function of the estimated model and $\hat{a}_t^{\,2}$= squared value of the estimated white noise at time 't'.

Absolute Relative Percent Forecast Errors: The first 37 years observations (out of 40 years) are re-fitted for confirming the predictive power of an identified best fitted model for forecasting the future values. ARPFE's for the rest observed values are computed as:

$$ARPFE = |\{y_{t+m} - \widehat{y_{t+m}}\}/y_{t+m}| * 100$$

Where, y_{t+m}= (t+m)th observed value, $\widehat{y_{t+m}}$= (t+m)th estimated value from the re-fitted ARIMA model based on t = 37 years. Here (t+1), (t+2), (t+3) = 38th, 39th and 40th observation.

3. Results and Discussion
On the basis of visualization of plots of observed time-series variables (Fig. 1) at the identification stage, annual rainfall and monsoon rainfall are found to be stationary with respect to mean and variance. It is also found that both the time-series variables show normal distribution from the observation of Q-Q plots (Fig. 4 and Fig. 5) and Shapiro Wilk's Test (Table 2) and independence (from the value of estimated autocorrelations and Ljung Box Test statistic) as presented in Table 2. As both the variables show stationarity, no transformation/differencing are required before performing ARIMA modeling.

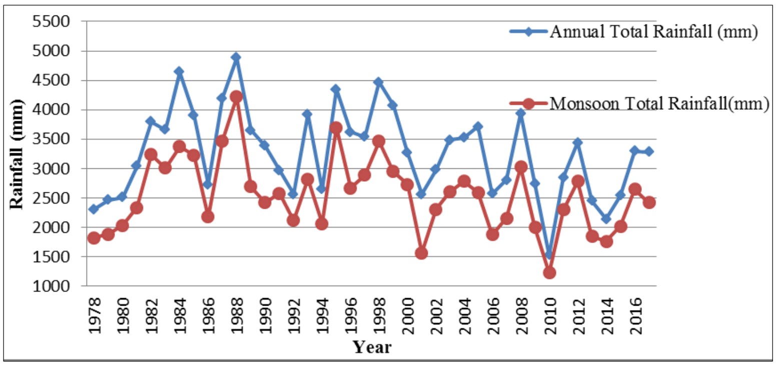

Fig 1: Yearly distribution of Annual Rainfall and Monsoon Rainfall

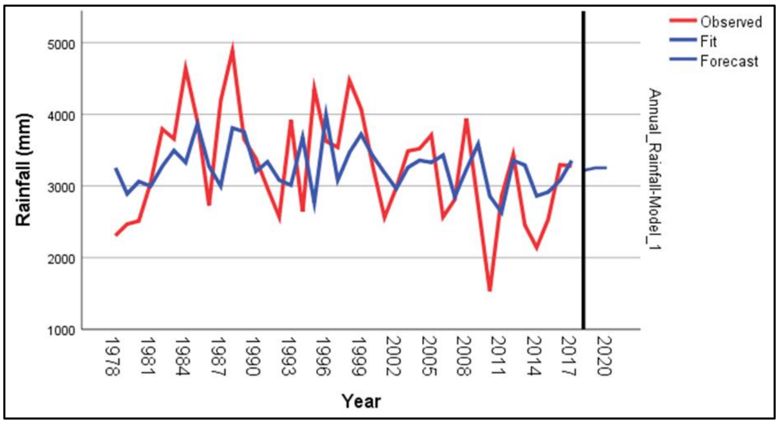

Fig 2: Observed value, corresponding fitted value and predicted value of Annual Rainfall in MA (1) model

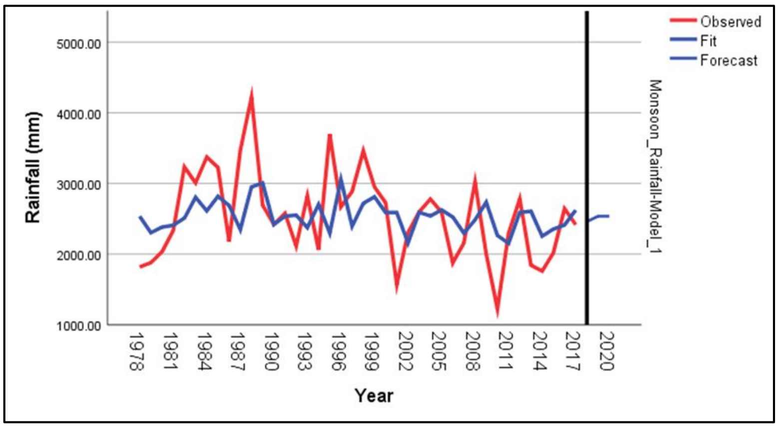

Fig 3: Observed value, corresponding fitted value and predicted value of Monsoon Rainfall in MA (1) model

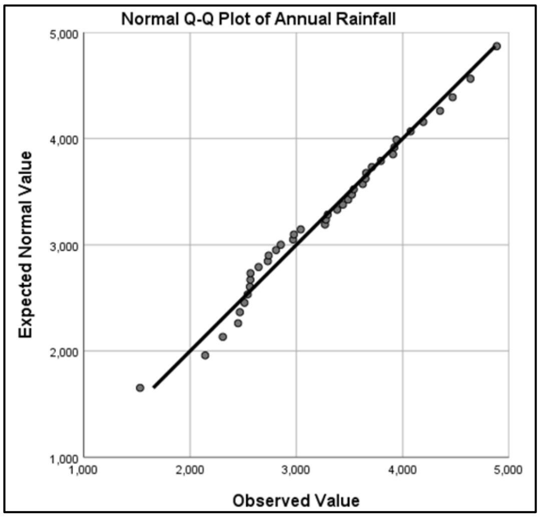

Fig 4: Quantile Quantile Plot of the observed values of Annual Rainfall

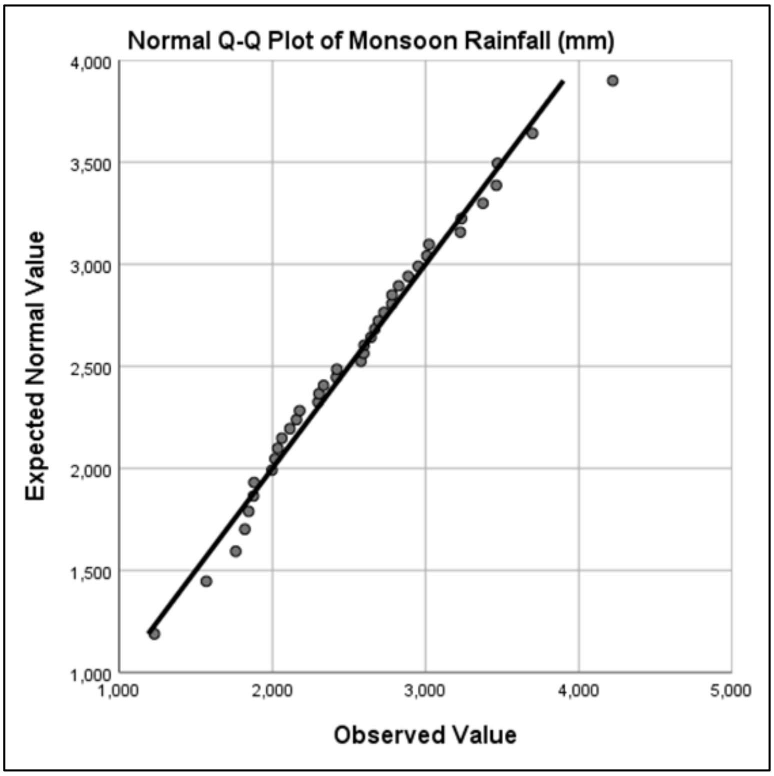

Fig 5: Quantile Quantile Plot of the observed values of Monsoon Rainfall

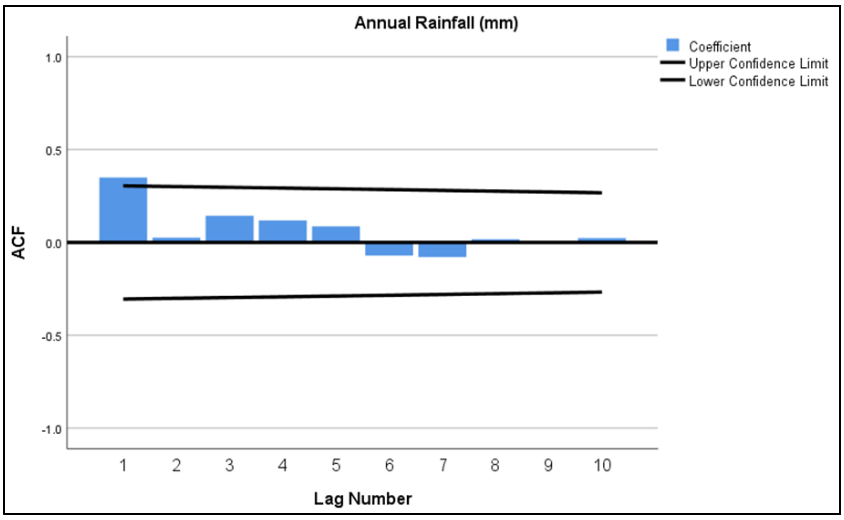

Fig 6: Estimated ACF of Annual Rainfall

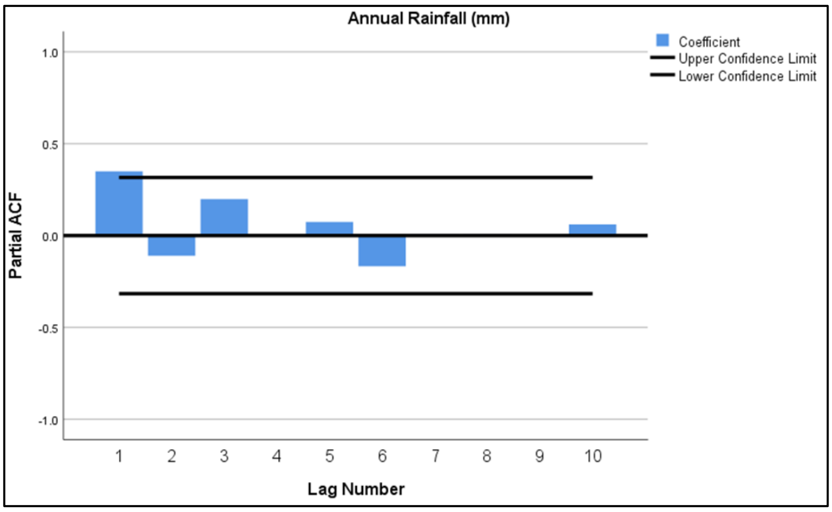

Fig 7: Estimated PACF of Annual Rainfall

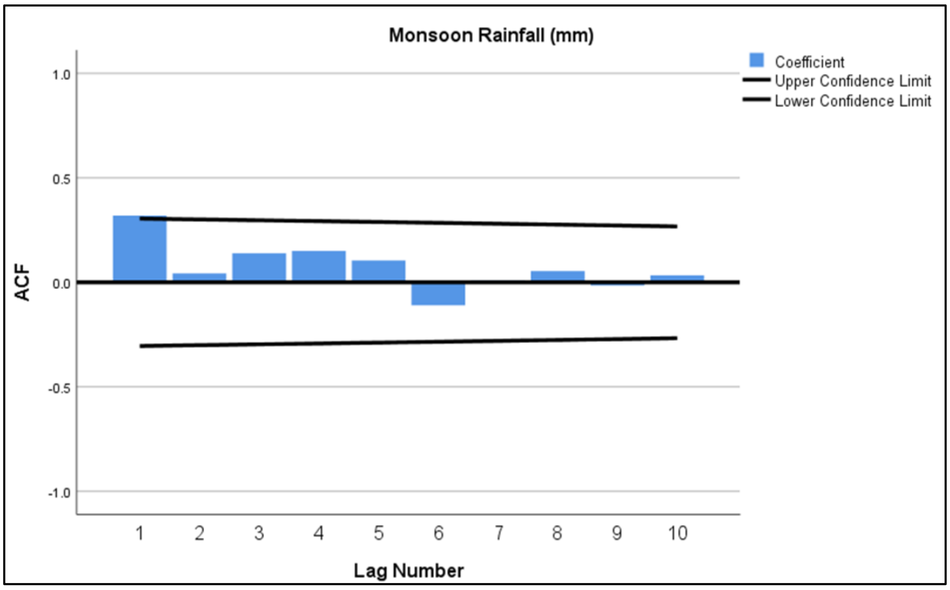

Fig 8: Estimated ACF of Monsoon Rainfall

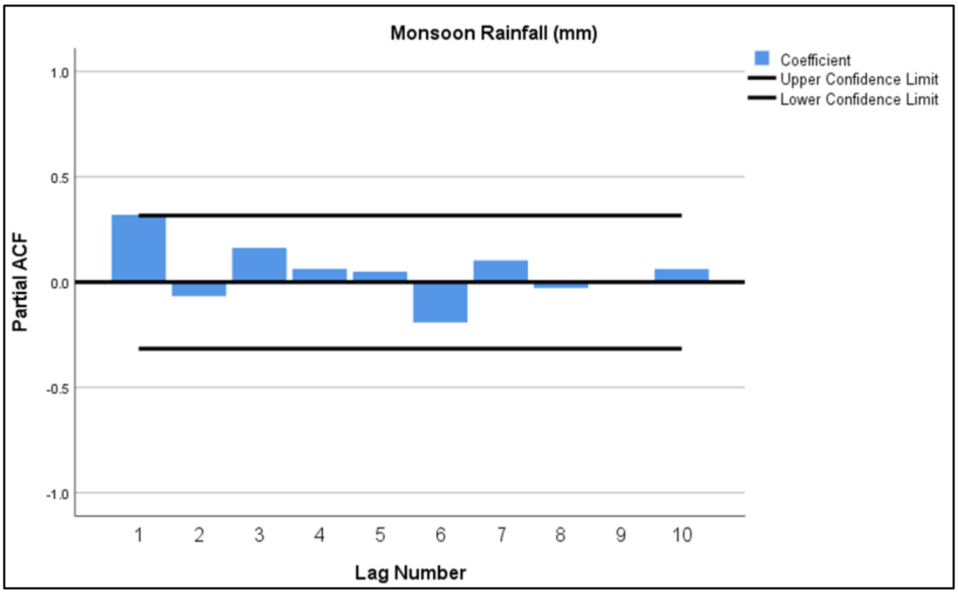

Fig 9: Estimated PACF of Monsoon Rainfall

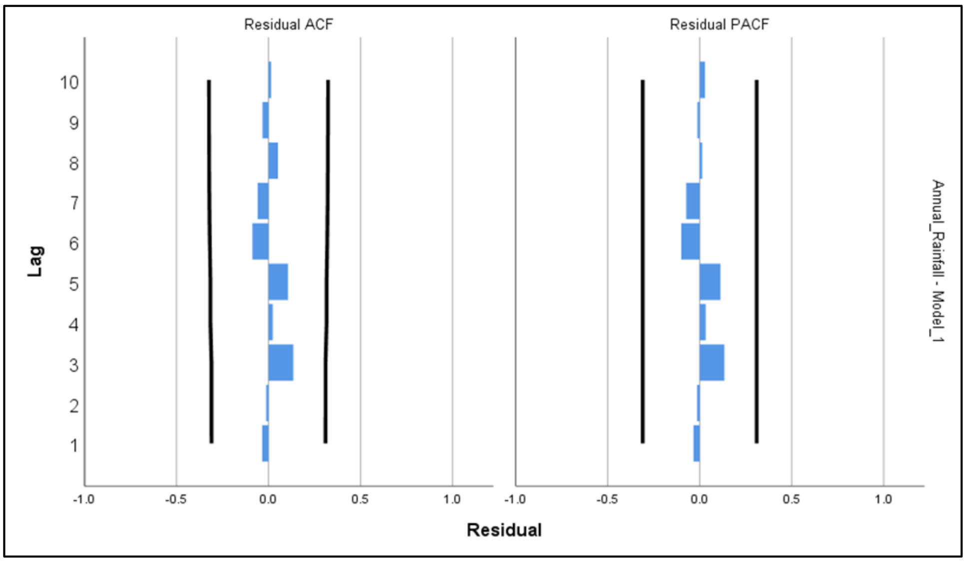

Fig 10: Estimated ACF and PACF on residuals of the fitted MA(1) model of Annual Rainfall

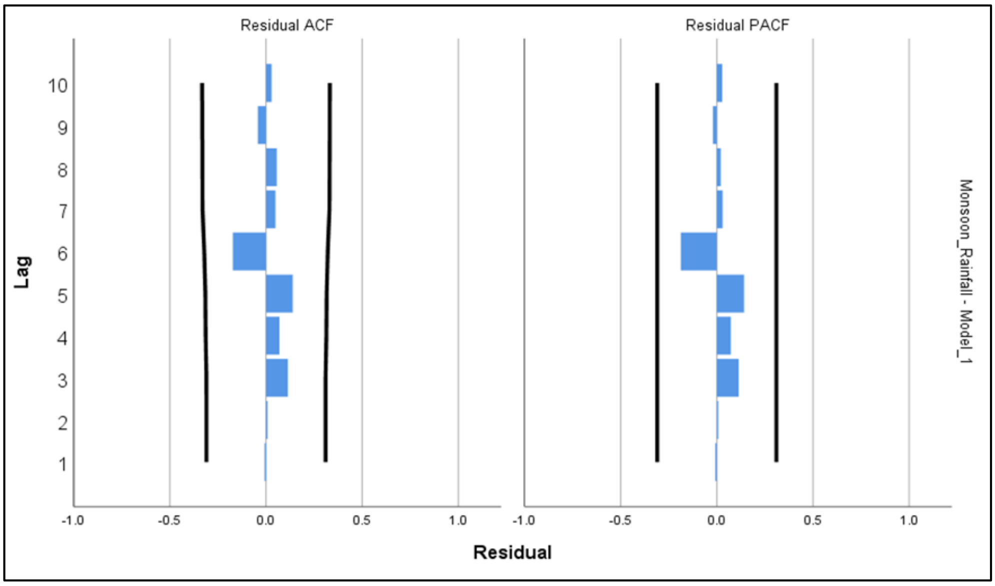

Fig 11: Estimated ACF and PACF on residuals of the fitted MA(1) model of Monsoon Rainfall

Table 1: Estimated parameters and model evaluation statistics of selected model

Variable	No. of observation	Estimated coefficients		Bayesian Information Criteria (BIC)	Root Mean Square Error (RMSE)	Mean Absolute Percent Error (MAPE)	Mean Absolute Error (MAE)	ARPFE	
		Constant	MA(1)					No. of observation	Value
Annual Rainfall (0,0,1)	40	3252.176** (159.412)	-0.467** (0.145)	13.275	695.936	17.886	538.678	-	-
	37	3261.407** (171.003)	-0.461** (0.152)	13.357	721.346	18.731	563.975	38	15.092
								39	1.002
								40	0.512
Monsoon Rainfall (0,0,1)	40	2536.796** (128.698)	-0.370** (0.152)	12.977	599.606	19.803	466.271	-	-
	37	2857.701** (268.863)	-0.330** (0.163)	13.054	619.936	20.551	483.058	38	17.138
								39	3.609
								40	5.437

Figure in the parentheses are standard errors of the estimated coefficient
* Significant at P = 0.05 ** Significant at P = 0.01

Table 2: Tests of Normality and Independence

ARIMA Model	Shapiro Wilk's Test		Ljung- Box (Q)	
	Statistic	Sig.	Statistic	Sig.
Annual Rainfall (0,0,1)	0.983	0.806	5.431	0.996
Noise residual from Annual Rainfall (0,0,1)	0.979	0.648	5.399	0.996
Monsoon Rainfall (0,0,1)	0.985	0.873	6.577	0.988
Noise residual from Monsoon Rainfall (0,0,1)	0.974	0.488	6.597	0.988

Table 3: Forecasted values of annual and monsoon rainfall in Cooch Behar District of West Bengal

Year	Annual Rainfall (mm)	Year	Monsoon Rainfall (mm)
2018	3216.9	2018	2460.7
2019	3252.2	2019	2536.8
2020	3252.2	2020	2536.8

Different ARIMA models with combinations of orders (p, 0, q) are tried to fit the data on both the variables. The characteristics of estimated ACF's and PACF's are much similar to the theoretical ACF's and PACF's of the moving average (MA) process. The estimated ACF has one spike at lag1 and decays towards zero (Fig. 6 and Fig.8) and PACF alternates in sign starting from the positive side (Fig. 7 and Fig. 9), the moving average of order one [MA (1)] is found appropriate. The coefficient θ for MA (1) of these time-series variables are found to be significant and also with a negative sign (Table 1) because the spike is on the positive side. The invertibility condition is satisfied since $|\widehat{\theta_1}| < 1$ (Table 1).

With regard to the adequacy of a model for predicting the future values, the satisfaction of goodness of fit criteria is very much essential. The MA(1) model for both the variables satisfies essentially the goodness of fit criteria as MA (1) has lowest BIC and RMSE with a comparatively smaller MAPE and MAE among all competitive models (Table 1).

As shown in Table 2, the residuals obtained from the time series models satisfied the normality conditions (from Q-Q plots and Shapiro Wilk's Test) and independence criteria (from the value of Q statistic and observation of residual plots). The residual plot of ACF and PACF (Fig. 10- Fig. 11) ensured that the residuals obtained from the best fitted model

[MA(1)] possess all the desired properties of white noise.
The model MA(1) or ARIMA (0,0,1) is founded to be a best fitted and parsimonious model with good predictive power based on the ARPFE value as depicted in Table 2. The model is used for forecasting annual rainfall and monsoon rainfall for the next three consecutive years (Fig 2 - 3) and the predicted values is given in Table 3.

Conclusion
The investigation reveals that annual total rainfall and monsoon rainfall data of Cooch Behar district of West Bengal fits well to MA(1) as shown in Fig.2 and Fig.3. These models have been algebraically represented below for describing their intrinsic behaviour and also predicting near-future values.

- **MA (1) model for Total Annual Rainfall**

$$\widehat{Z_t} = 3252.176 + 0.467\widehat{a_{t-1}} + a_t$$

- **MA (1) model for Monsoon Rainfall**

$$\widehat{Z_t} = 2536.796 + 0.370\widehat{a_{t-1}} + a_t$$

Where, $\widehat{Z_t}$ = estimated value of Z_t, $\widehat{a_{t-1}}$ = estimated model residual at time 't-1' and a_t = white noise term at time 't.'

References

1. Box GEP and Jenkins GM. Time Series Analysis: Forecasting and Control. 2nd ed., Holden-Day, San Francisco, 1976.
2. Dwivedi DK, Kelaiya JH, Sharma GR. Forecasting monthly rainfall using autoregressive integrated moving average model (ARIMA) and artificial neural network (ANN) model: A case study of Junagadh, Gujarat, India. Journal of Applied and Natural Science. 2019;11(1):35-41.
3. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978;65:297-303.
4. Nandargi SS, Barman K. Analysis of Trends and Variability in Rainfall over West Bengal. International Journal of Current Advanced Research. 2018;077(7):14223-14229.
5. Nirmala M, Sundaram SM. A seasonal arima model for forecasting monthly rainfall in Tamil Nadu. National Journal on Advances in Building Sciences and Mechanics. 2010;1(2):43-47.
6. Pal S, Majumder D, Chakraborty P. District-wise trend analysis of rainfall pattern in last century (1901-2000) over Gangetic region in West Bengal, India. Journal of Applied and Natural Science. 2015;7:750-757.
7. Shivhare N, Rahul AK, Dwivedi SB, Dikshit PKS. ARIMA based daily weather forecasting tool: A case study for Varanasi. Mausam. 2019;70(1):133-140.
8. Jeet P, Singh KN, Ranjan Kumar R, Gurung B, Singh AK, Upadhyaya A. Modeling and Trend Analysis of Climatic Variables of Ranchi District, Jharkhand. Journal of Agri Search. 2021;8(2):120-128.