Eco. Env. & Cons. 29 (August Suppl. Issue) : 2023; pp. (S24-S29)
Copyright@ EM International
ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2023.v29i04s.005

Comparison of Arima and Ann for Forecasting the Annual Rainfall of Nadia District, West Bengal, India

K. Sathees Kumar[1]*, T. Gowthaman[2] and Banjul Bhattacharyya[2]

[1]Department of Social Sciences, RVS Agricultural College, Thanjavur 613 402, Tamil Nadu, India
[2] Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, West Bengal

(Received 9 February, 2023; Accepted 10 April, 2023)

ABSTRACT

Every fluctuation in rainfall and temperature will affect crop yields since Indian agriculture is very susceptible to climate fluctuations, especially to temperature and rainfall. Planning and management of natural resources requires an understanding of the geographical and temporal distribution and changing trends in climatic factors. In order to better understand the variability pattern in climate data and perhaps even forecast short- and long-term changes in the series, time series analysis can be a very useful technique. The annual rainfall records for the Nadia district of west bengal from 1981 to 2021 have been examined in this study. The rainfall data were modelled using linear parametric technique Autoregressive Integrated Moving Average (ARIMA) and nonlinear nonparametric technique Artificial Neural Network (ANN). Model performance of ARIMA and ANN were compared. Result revealed that ARIMA was performed better than ANN for forecasting the rainfall of Nadia district. This forecasts from ARIMA are anticipated to assist decision-makers in the effective scheduling of agricultural management, urban planning, rainfall collection, and flood prediction.

Key words: *Annual rainfall, Linear parametric technique, Nonlinear nonparametric technique, ARIMA, ANN*

Introduction

Farmers, government officials, and other stakeholders-in particular, farmers-are expected to benefit from modelling climatic variables especially rainfall. When time series are modelled, it is possible to extract many of the intrinsic features that are present in the dataset and extrapolate the time series into the future using those features. Time series forecasting models have seen significant progress over the past several decades, and there are an endless number of stochastic processes that may be used to model and forecast for a given series. The amount and distribution of rainfall heavily influences a country's ability to produce food. As a result, precise rainfall modelling is essential for effective planning and policymaking. There have been many attempts in the literature to construct models that can describe climatic variables like rainfall and temperature.

Throughout past few decades, the analysis of time series data in statistical and stochastic models has mostly been dominated by the Box-Jenkins ARIMA approach. This model has become quite well-liked for simulating linear dynamics. The assumption of stationarity, nonetheless, underlies this model. Over the years, many people have used ARIMA to forecast rainfall trends (Ali, 2013; Bari *et al*., 2015; Graham and Mishra, 2017; Al Balasmeh *et al*., 2019; Lai and Dzombak, 2020; Gowthaman *et al.*, 2022). The approach offers a few unique qualities that make it more appealing to researchers. It simplifies forecasting by allowing researchers to utilise just one variable time series data while yet permitting numerous for more complicated instances. Time series rainfall data are modelled and forecasted using the statistical technique known as the autoregressive integrated moving average (ARIMA).

As the ARIMA model assumes linearity, it cannot detect any nonlinear patterns, which is a fundamental flaw. Time series can mostly have nonlinear components; in these cases, ARIMA models are insufficient for modelling and forecasting. There are several parametric nonlinear models like GARCH that can capture the nonlinear component to get around this problem. If the data generation process is very diverse, complex, and nonlinear in character, these parametric nonlinear models may occasionally fail. Artificial intelligence techniques are the only means to describe and anticipate such phenomena using such data. The most popular Artificial Intelligence (AI) method for modelling and predicting time series data is the Artificial Neural Network (ANN). Because there is no requirement to describe a specific model specification, neural networks have the flexibility of nonlinear modelling to handle complicated undefined data. Several researchers have employed ANN to predict rainfall trends throughout the years (Abbot and Marohasy, 2014; Sulaiman, 2013; Liu *et al*., 2019; Canchala *et al*., 2020)

Thus, this research objective was formed: the annual rainfall series for the Nadia district was predicted using the ARIMA model and ANN approaches, and the forecasting performances of these models were compared.

Materials and Methods

Study Area

The current study area is Nadia district of West Bengal (Fig. 1). Nadia is located between 22°53' and 24°11' North latitude and 88°09' and 88°48' East longitude. It has an area of around 390027 sq.km and is oriented North-South. The region lies Around 46 ft. above the mean sea level. The district is split in half by the cancerous tropic. Bangladesh is to the east of Nadia district, Bardhaman and Hugli are to the west, Murshidabad is to the north and north west, and North 24 Parganas is to the south and south east.

Auto-Regressive Integrated Moving Average

The Box-Jenkins modelling method, which is used in climatic time series analysis and is named after the statisticians Box and Jenkins (1970) describes stationary time series, and uses autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models to find the best fit of a time series to its past values in order to make forecasts.

Actual time series data that begins with the combination of autoregressive and moving average processes, known as ARMA, is more flexible when

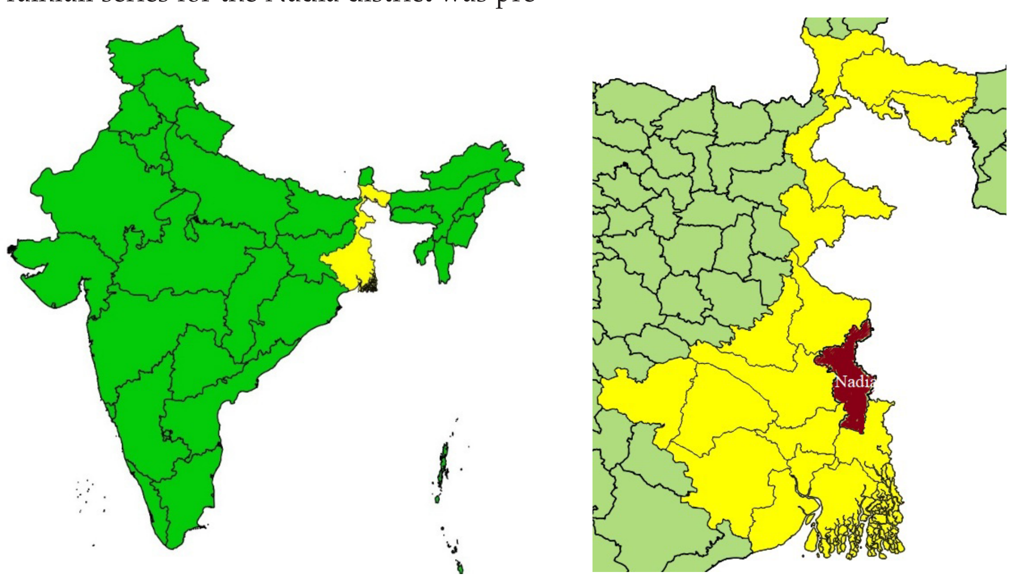

Fig. 1. Nadia district location in India

autoregressive and moving average processes are included (p,q).

ARMA (p,q) is indicated by

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

where

$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \emptyset_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

In which,

- B - the backshift operator express by $B(y_t) = y_{t-1}$.
- p – order of AR
- q – order of MA

By including "differencing" in the ARMA model, which is signified by ARIMA(p,d,q), the Box-Jenkins Autoregressive Integrated Moving Average [6] model was created.

$$\Delta^d Y_t = C + \phi_1 \Delta^d Y_{t-1} + \ldots + \phi_p \Delta^d Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

In which, $\varepsilon_t \sim N(0,\sigma^2)$.

Three steps constitute the methodology, and they are as follows:

Step 1: Identification

In order to identify a model, its suitable structure (p,d,q) and order must be given. Number of times differencing the series to make stationary series, autocorrelation function (ACF) and partial autocorrelation function (PACF) of stationary series plots can also be used to detect models (Box and Jenkins, 1970).

Step 2: Estimation

The models' coefficients can be calculated using non-linear least-squares estimation or maximum likelihood estimation. Parameter estimation for ARIMA models often demands a more difficult iteration process (Box and Jenkins, 1970).

Step 3: Model checking and forecasting

Testing the randomness of the model's residuals and the statistical significance of the estimated parameters are two key components of verication. Typically, the parsimony principle—according to which the best model is the simplest model—directs the fitting process.

Artificial Neural Network

A massively parallel distributed processor called an ANN has a natural tendency to store learned information from experiments and make it accessible for subsequent use. It is comparable to the human brain, whose speed and effectiveness have long fascinated experts. The ANN approach was created to better understand these processes and address the difficulties their cause. In essence, neural networks use a nonlinear modelling technique to approximate any function with reasonable accuracy. Its power derives from processing information from data in parallel. The procedure of constructing the model does not presume any knowledge of the model's shape. Instead, the data's properties play a major role in determining the network. The most often used model type for time series modelling and forecasting is the single hidden layer feedforward network.

Neural layers constitute an ANN. A network of three layers of interconnected basic processing units forms the model's defining feature. An input layer is the initial layer that receives input data. An output layer is the final layer that generates output data. There are hidden layers between the output and input layers. One or more hidden layers are possible. Using these connections between nodes in various layers, data can be transmitted.

To find misspecification in the above models, running a Ljung-Box test (Ljung and Box, 1978) on the residuals is helpful (Anderson, 1976).

Forecasts Evaluation Methods

In this research, the forecasting abilities of numerous models were examined using two common performance criteria. They are RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error). The subsequent formulae are used to calculate them.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}$$

$$\text{MAPE} = \frac{1}{n}\left(\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|\right) * 10 \quad 0$$

Where, Y_t is actual value and $\hat{Y}_t$ is predicted value.

Results and Discussion

For this investigation, annual rainfall series of Nadia district was collected from India Meteorological Department (https://mausam.imd.gov.in/). Data points obtained during 1981 to 2021. Last 5 data

Table 1. Descriptive statistics of Nadia district's rainfall series

Mean	Median	Minimum	Maximum	Standard Deviation	Skewness	Kurtosis	CV (%)
1145.85	1144.34	585.35	1766.60	308.80	0.01	-0.41	26.94%

Table 2. ADF test results

ADF test	Statistic	P value
Level	-2.61	0.34
Differenced	-3.97	0.02

points were used to assess the forecasting ability of the model, where other data points were utilised for model building. Descriptive statistics of Nadia district's rainfall series is given in Table 1. The rainfall series is symmetric and lower platykurtic. It means that outliers are less in the rainfall series. The coefficient of variation also confirms the low degree of chaos in the rainfall series.

Fitting of ARIMA model

In case of checking stationarity, ADF-Test was employed in the first step of the ARIMA model. It shows that the rainfall series is non-stationary (Table 2). So requires a first order difference to make it as a stationary series. To determine the values of p and q, respectively, the partial autocorrelation function (PACF) and autocorrelation function (ACF) of stationary series were computed. ACF (k) of the stationary series was cut off at the first spike and tailed off towards zero, whereas PACF (kk) of the stationary series also tailed off towards zero (Fig. 2). As a result, it was implied that the algebraic family of ARIMA on p=0, 1; d=1; and q=0, 1; could have been utilized. Table 3 contains the results. By using the Ljung-Box (Q) test to determine if the residuals' independence was verified, the residuals' assumptions were verified. Among the all fitted models, ARIMA (1,1,0) gave less RMSE and MAPE in the testing set, though other models were given less RMSE and MAPE in the training set. Parameter estimate of ARIMA (1,1,0) is given in Table 4.

Table 4. Parameter estimate of fitted ARIMA model

Fitted Model	AR (1)	S. E	Z value	P value
ARIMA (1,1,0)	-0.53	0.14	-3.74	<0.01

Fitting of ANN model

In case of ANN, the sigmoid (identity) activation function was used in the hidden (output) layer. One output node in the output layer was used and the iterative technique was used for multi-step-ahead forecasting. As a result, the model's uncertainty was only related to the number of tapping delays (p), which in this case represented the number of lagged observations and the number of hidden layer nodes (q). The number of tapped delays and hidden nodes were found via trial and error. For avoiding the local minima and finding the global minimum, p and q were varied from 1 to 6. I: Hs: Oi is a standard

Table 3. Performance of different ARIMA model in training set and testing set and residual diagnostics

ARIMA	Training set		Testing set		Ljung-Box test	
	RMSE	MAPE	RMSE	MAPE	Statistic	P value
(0,1,1)	247.63	18.44	280.35	18.94	5.24	0.51
(1,1,0)	252.99	18.59	259.84	17.12	3.52	0.74
(1,1,1)	247.15	18.21	279.74	18.89	4.78	0.57

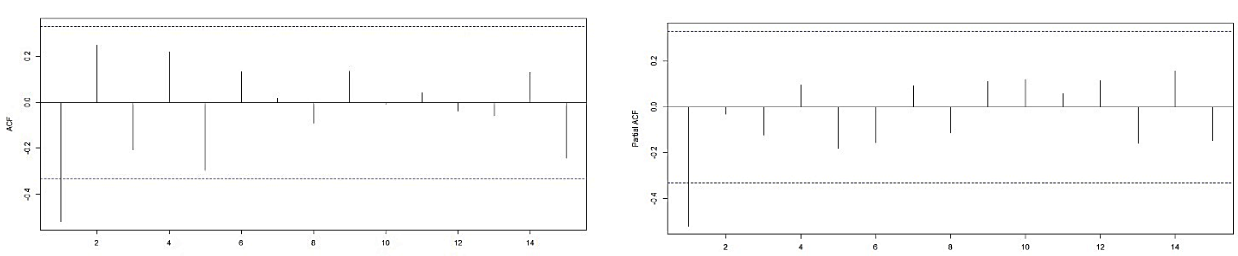

Fig. 2. ACF and PACF of differenced rainfall series

Table 5. Performance of different ANN model in training set and testing set and residual diagnostics

Model	Training set		Testing set		Ljung-Box test	
	RMSE	MAPE	RMSE	MAPE	Statistic	P value
(1:1s:1i)	207.71	14.84	358.88	17.08	12.27	0.42
(1:2s:1i)	201.75	13.90	358.83	17.09	10.46	0.58
(1:3s:1i)	194.18	13.24	356.53	17.14	10.08	0.61
(1:4s:1i)	184.73	12.76	350.29	17.15	11.26	0.51
(1:5s:1i)	180.96	12.58	346.00	17.22	11.20	0.51
(1:6s:1i)	181.62	12.46	355.78	17.33	14.89	0.25
(2:1s:1i)	182.24	12.78	331.08	16.98	12.73	0.39
(2:2s:1i)	151.46	11.48	356.59	17.31	7.35	0.83
(2:3s:1i)	127.07	9.06	351.38	17.72	5.60	0.93
(2:4s:1i)	116.90	8.59	358.30	20.12	6.43	0.89
(2:5s:1i)	108.40	7.31	362.25	20.49	4.48	0.97
(2:6s:1i)	100.43	6.85	695.32	40.89	3.84	0.99
(3:1s:1i)	177.80	12.56	304.97	16.87	18.65	0.10
(3:2s:1i)	162.14	11.12	325.93	16.98	18.27	0.11
(3:3s:1i)	120.65	8.47	342.70	17.97	15.30	0.23
(3:4s:1i)	92.47	5.88	388.61	19.75	11.83	0.46
(3:5s:1i)	78.79	5.12	670.28	39.95	14.58	0.27
(3:6s:1i)	52.06	3.56	502.20	24.85	9.46	0.66
(4:1s:1i)	178.20	12.50	304.94	16.89	17.71	0.12
(4:2s:1i)	143.37	9.66	328.86	17.12	20.58	0.06
(4:3s:1i)	99.45	6.99	351.52	17.62	18.64	0.10
(4:4s:1i)	68.75	4.64	333.95	15.20	17.39	0.14
(4:5s:1i)	37.85	2.28	343.52	16.62	11.38	0.50
(4:6s:1i)	25.94	1.59	989.48	39.89	9.29	0.68
(5:1s:1i)	171.67	11.66	303.85	16.87	16.39	0.17
(5:2s:1i)	119.50	7.89	311.53	16.56	15.71	0.20
(5:3s:1i)	67.91	4.63	364.59	17.79	14.71	0.26
(5:4s:1i)	35.79	2.53	403.31	20.96	17.47	0.13
(5:5s:1i)	18.48	1.37	395.08	19.36	8.90	0.71
(5:6s:1i)	12.09	0.58	397.13	20.50	7.80	0.80
(6:1s:1i)	165.61	10.97	308.05	17.91	15.67	0.21
(6:2s:1i)	108.90	7.22	269.57	17.98	16.81	0.16
(6:3s:1i)	69.82	4.61	334.43	18.53	15.57	0.21
(6:4s:1i)	28.17	1.96	392.91	18.72	22.69	0.03
(6:5s:1i)	9.84	0.66	380.69	18.42	7.07	0.85
(6:6s:1i)	3.13	0.16	374.70	21.12	7.88	0.79

ANN structure with a single hidden layer. Among the all-possible models, the ANN (6:2s:1i) model was given lower RMSE and MAPE than other models (Table 5).

Comparison of ARIMA and ANN

Forecasting ability of the both models was compared through the best ARIMA and ANN model performance in the testing set. According to Table 6, the RMSE and MAPE values for rainfall series are often lower in the ARIMA model than in the neural network model, indicating that the ARIMA model performs better than ANN. Similar result was found in Gao *et al.*(2017) and Nury*et al.*(2017). This comparison explains that ARIMA is sufficient for modelling the low-chaos series (Fig. 2). Though ANN performed better than ARIMA for high-chaos series (Jha and Sinha, 2014), it is not suitable for modelling low-chaos series. Forecast of next 5 years are ob-

Table 6. Forecasting ability comparison of ARIMA and ANN

Statistic	ARIMA	ANN
RMSE	259.84	269.57
MAPE	17.12	17.98

Table 7. Nadia district's rainfall series forecasts

Year	Forecast	Lower limit at 80%	Upper limit at 80%	Lower limit at 95%	Upper limit at 95%
2022	1404.46	1070.83	1738.08	894.22	1914.69
2023	1411.99	1044.01	1779.97	849.21	1974.77
2024	1407.96	962.74	1853.18	727.06	2088.87
2025	1410.12	922.19	1898.05	663.89	2156.34
2026	1408.97	870.86	1947.07	586.01	2231.93

tained from the fitted ARIMA model and are given in Table 7.

Conclusion

Using ANN and ARIMA approaches, the complexity of the yearly rainfall record's nature has been investigated. The models were developed and tested using yearly rainfall data for the Nadia district for the years 1981 to 2021. The research shows that the ARIMA model outperforms the ANN model and can be utilised as a suitable forecasting method to predict rainfall. For the long-term forecast, the model may benefit from further improvement collecting data independently from the various locations of the nation.

Acknowledgement

The authors are thankful to the India Meteorological Department.

Conflicts of Interest

The authors have sworn that there are no conflicting interests.

References

Abbot, J. and Marohasy, J. 2014. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmospheric Research*. 138: 166-178.

Al Balasmeh, O., Babbar, R. and Karmaker, T. 2019. Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan. *Arabian Journal of Geosciences*. 12: 1-19.

Ali, S.M. 2013. Time series analysis of Baghdad rainfall using ARIMA method. *Iraqi Journal of Science.* 54(4): 1136-1142.

Anderson, O. 1976. Time series analysis and forecasting: the Box-Jenkins approach. London, Butterworths.

Bari, S.H., Rahman, M.T., Hussain, M.M. and Ray, S. 2015. Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*. 7(1): 69-77.

Box, G.E.P. and Jenkins, G.M. 1970. Time series analysis: forecasting and control. San Francisco, Holden Day.

Canchala, T., Alfonso-Morales, W., Carvajal-Escobar, Y., Cerón, W.L. and Caicedo-Bravo, E. 2020. Monthly rainfall anomalies forecasting for southwestern Colombia using artificial neural networks approaches. *Water.* 12(9): 2628.

Gao, G., Lo, K. and Fan, F. 2017. Comparison of ARIMA and ANN models used in electricity price forecasting for power market. *Energy and Power Engineering*. 9(4B): 120-126.

Gowthaman, T., Kumar, K. S., Adarsh, V. S. and Bhattacharyya, B. 2022. Trend Analysis and ARIMA Models for Water Quality Parameters of Brahmani River, Odisha, India. *International Journal of Environment and Climate Change*. 12(12): 219-228.

Graham, A. and Mishra, E.P. 2017. Time series analysis model to forecast rainfall for Allahabad region. *Journal of Pharmacognosy and Phytochemistry*. 6(5): 1418-1421.

Jha, G. K. and Sinha, K. 2014. Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Computing and Applications*. 24: 563-571.

Lai, Y. and Dzombak, D.A. 2020. Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation. *Weather and Forecasting.* 35(3): 959-976.

Liu, Q., Zou, Y., Liu, X. and Linge, N. 2019. A survey on rainfall forecasting using artificial neural network. *International Journal of Embedded Systems*. 11(2): 240-249.

Ljung, G.M. and Box, G.E.P. 1978. On a Measure of Lack of fit in Time Series Models. *Biometrika*. 65: 297-303.

Nury, A. H., Hasan, K. and Alam, M.J.B. 2017. Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh. *Journal of King Saud University-Science.* 29(1): 47-61.

Sulaiman, J. and Wahab, S.H. 2013. Heavy rainfall forecasting model using artificial neural network for flood prone area. Springer Singapore.