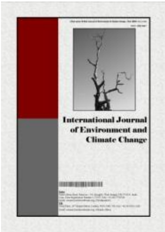

International Journal of Environment and Climate Change

12(11): 2879-2889, 2022; Article no.IJECC.91614
ISSN: 2581-8627
(Past name: British Journal of Environment & Climate Change, Past ISSN: 2231–4784)

Regional Time Series Forecasting of Chickpea using ARIMA and Neural Network Models in Central Plains of Uttar Pradesh (India)

Bukke Vennela [a*], Ekta Pathak Mishra [a], Shweta Gautam [a], Ashish Ratn Mishra [b] and Shraddha Rawat [a]

[a] *Department of Environmental Sciences and Natural Resources Management, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj - 211007, Uttar Pradesh, India.*
[b] *Centre for Geospatial Technologies, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj - 211007, Uttar Pradesh, India.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJECC/2022/v12i1131280

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/91614

Original Research Article

Received 14 July 2022
Accepted 24 September 2022
Published 01 October 2022

ABSTRACT

Climate and yield prediction are the most important and challenging tasks in modern agriculture during the climate change era. In general, climate and yield are mostly non-linear and highly complicated. India is an agricultural country and most of its economy depends upon agriculture therefore early prediction of climate and yield is necessary for the planned economic growth of our country. This research identifies superior forecasting models of Autoregressive Integrated Moving Average (ARIMA) as well as Artificial Neural Network (ANN) for predicting future climate and chickpea yield. Historical data for the climate and crop were used (1996-2020) and forecasting was done for the next 5 years (2020-2025). By using, RMSE and R^2 statistical tools simultaneously, the predictive accuracy of ARIMA and ANN models was compared. By comparing the R^2 values of ARIMA (0.591) and ANN (0.96), this study reveals ANN models can be used as more accurate forecasting tools to predict the future climate as well as yield, enabling timely agricultural management.

Keywords: ARIMA; ANN; MLP; RMSE; BIC; nodes; hidden layer; learning rate.

**Corresponding author: E-mail: bukkevennela24@gmail.com;*

1. INTRODUCTION

The yield of a crop is dependent on various technological, biological, and environmental aspects. Among various environmental aspects soil fertility, topography, water quality, and climate plays a important role in yield and growth of the plant. Climate change is majorly influenced by changes in temperature and rainfall. For assessing the impact of climate on crop yield various statistical models and methods are being used such as ARIMA, ANN, Seasonal Auto-Regressive Integrated Moving Averages (SARIMA) for interpreting the relation between climate and crop yield. ARIMA modeling was developed by Box-Jenkins which is used for predicting future values based on present trends. This is used in analyzing and forecasting weather and climate-based parameters in meteorological studies. ANN modeling utilizes to processing of the brain for developing algorithms that will be used to model complex patterns and predict problems. It allows the non-linear complex relations between the response variable and its predictors. It also compares the efficiency of models in fitting and future prediction. In present study, time series analysis on climate and yield parameters of chickpeas was analyzed and forecasted by using ARIMA and ANN models.

Chickpea (*Cicer arietinum*) is a member of the legume, pea, or pulse. It is a cool season crop, grows mainly in semiarid environments [1]. Chickpea is the second most important legume crop after the common bean . Estimating the yield parameters of chickpeas is very important, so in this connection, the present study was used for estimating the climate and yield parameters by using ARIMA and ANN. A several research has done using time series models such as Multi Linear Regression (MLR), Autoregressive Moving Average (ARMA), ARIMA, and SARIMA [2-4] (Muhamad Safiih et al. 2017). On the other hand, ANN is nonlinear and it is mainly depend upon the neurons. It can calculate the approximate value. Interest in the use of ANNs for developing climate change prediction models has increased in recent years due to ever-changing climate patterns in the world [5-8]. ANNs are computer systems inspired by biological neural networks to model relationship between independent and dependent variables. Hence, ARIMA and ANN are the most utilized technologies for data mining. The use of these models was capture the patterns of data and improving prediction accuracy [9]. Hence, we conducted our study with the following objectives:

- Forecasting of climatic parameters (maximum temperature, minimum temperature, and rainfall) using ARIMA.
- Time series (ARIMA) forecasting for the yield of chickpea in Prayagraj.
- To compare ARIMA and ANN models for chickpea yield prediction, based on climatic parameters as independent variables.

2. MATERIALS AND METHODS

In the present study, the climate prediction of the ARIMA model for time series analysis and ANN were used for forecasting algorithms based on the past values of the time series to predict the future values [10-15].

2.1 Study Area

Prayagraj is a district in the central plains of southeastern Uttar Pradesh. The city is situated at the confluence of the two rivers – Ganga and Yamuna. It lies between the parallels of 24º 47’ North latitude and 81º 19’ East longitudes. Due to unpredictable climate change and rapidly changing rainfall patterns, the city faces frequent floods.

2.2 Data Collection

The historical data for rainfall, minimum and maximum temperature were obtained from Indian Meteorological Department (IMD) for the years (1996-2020). The yield data was taken from ICRISAT.

2.3 Software Used

In this study, we used two software programs, SPSS for ARIMA and MATLAB for ANN.

2.4 Time Series Forecasting

It is the technique of data science to predict based on historical data. Time series forecasting isimportant in prediction of Machine Learning (ML). ML methods such as Support Vector Machines, Neural Networks, and Regression, be applied to it. Forecasting was done by historical data and using them to predict future observations.

2.5 Process of Time Series Forecasting

The Box-Jenkins Model forecasts data using mainly three principles: Auto Regression (AR), Differencing (I) and Moving Average (MA). These three principles are known as p, d, and q, respectively.

a) Ensuring stationarity (d)
b) Identification of AR(p) and MR(q)
c) Estimations using appropriate p, d. q values
d) Remove seasonality
e) Selecting the best seasonal ARIMA model

2.6 ANN

The ANN is a deep learning method that arises from the concept of the human brain's Biological Neural Networks. MATLAB (R2021a) was used to develop to predict the monthly minimum and maximum temperature and rainfall. In input variables of algorithms were used monthly rainfall and temperature and the target variable of the algorithm was used yearly yield data.

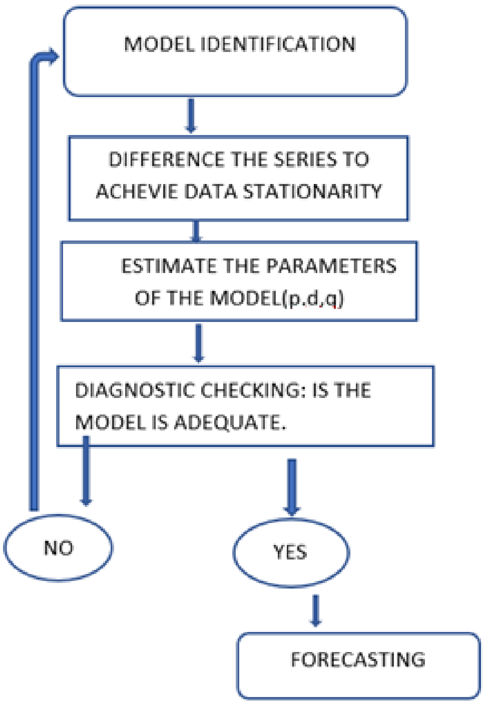

Fig. 1. Arima model flow chart

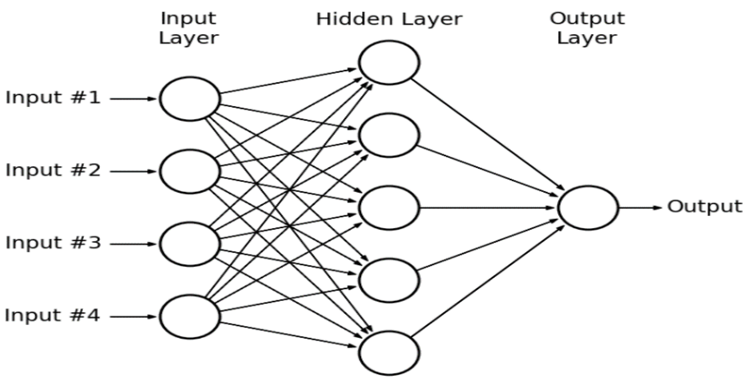

Fig. 2. Artificial Neural Network (ANN) architecture with input, hidden layer, and output

In Fig. 2, ANN models are the extreme simplification of human neural systems. An ANN comprises of computational units analogous to that of the neurons of the biological nervous system known as artificial neurons. Mainly, the ANN model constitutes of three layers, viz., input, hidden, and output. Each neuron in the nth layer is interconnected with the neurons of the (n + 1)th layer by some signal. Each connection is assigned a weight. The output may be calculated after multiplying each input with its corresponding weight.

$$Z = W_0 + W_1X_1 + W_2X_2 + \ldots + W_nX_n \qquad (1)$$

The creation of the ANN predictive model (in MATLAB software) involves the following aspects:

(i) Creating the network of ANN mainly consisit of input,hidden layer and output neurons.
(ii) Export the input and target to work shop. Use command as “nntool” or “nnstart” in command window.In data manager we import the input data and target data.
(iii) We created new neural network model i.e., feed backward propogation which was best fit for the collected data. Purelin function was used as it showed good and accurate results compare to other. The purpose of feedforward neural networks is to approximate functions. There is a classifier using the formula y = f* (x).
(iv) In the next step, we train the network, if the values are not good. We retrained the network for good accurate value in the regression graph for making the models as best fit for further validation.
(v) After that we use stimulation of network after stimulate the network then the output will came. The time-series data set for 25 years was divided into training (70%), testing (15%), and validation (15%).
(vi) Training , Validation , and testing are three main aspects of ANN.
(vii) The training set is used for the training of the algorithm. A validation set is used for accuarate of the model and to calculate the efficiency of the algorithm in terms of R-square and Root mean squared error (RMSE).

3. RESULTS AND DISCUSSION

3.1 ARIMA

The rainfall and temperature data were taken from 1996-2020. The Rainfall, max temperature and min temperature data were separately analyzed using different ARIMA models in SPSS statistical package**.** Then retrend the data to make it stationary, and then we choose possible values of p and q which we can adjust as model fitting progresses. And ACF (and PACF graphs are shown in Figs. 3-5.

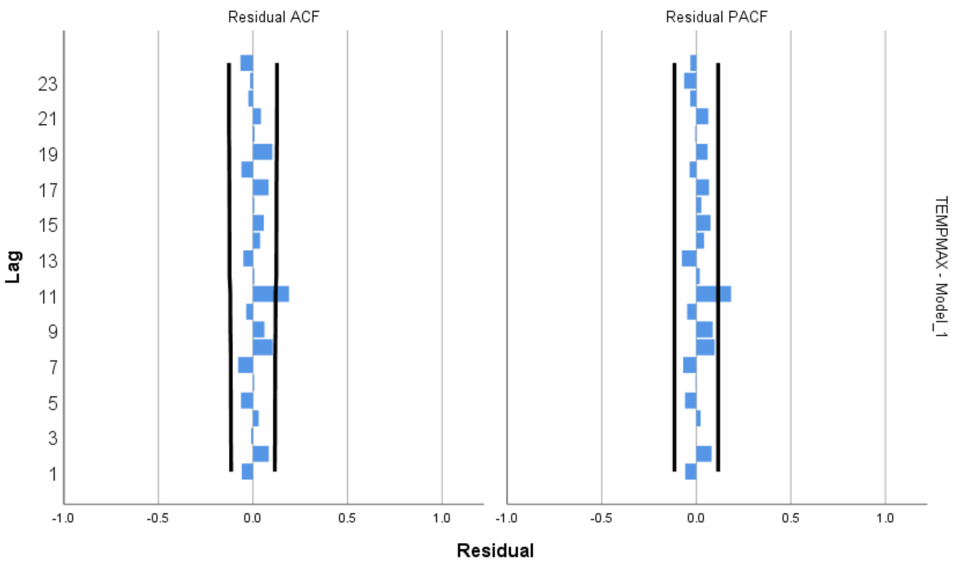

Fig. 3. ACF and PACF values of maximum temperature

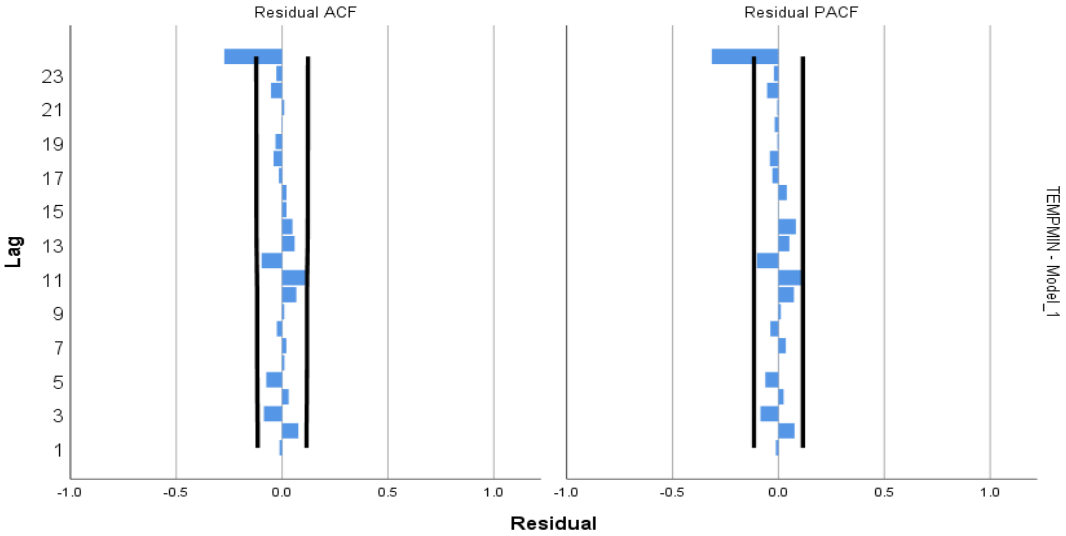

Fig. 4. ACF and PACF values of minimum temperature

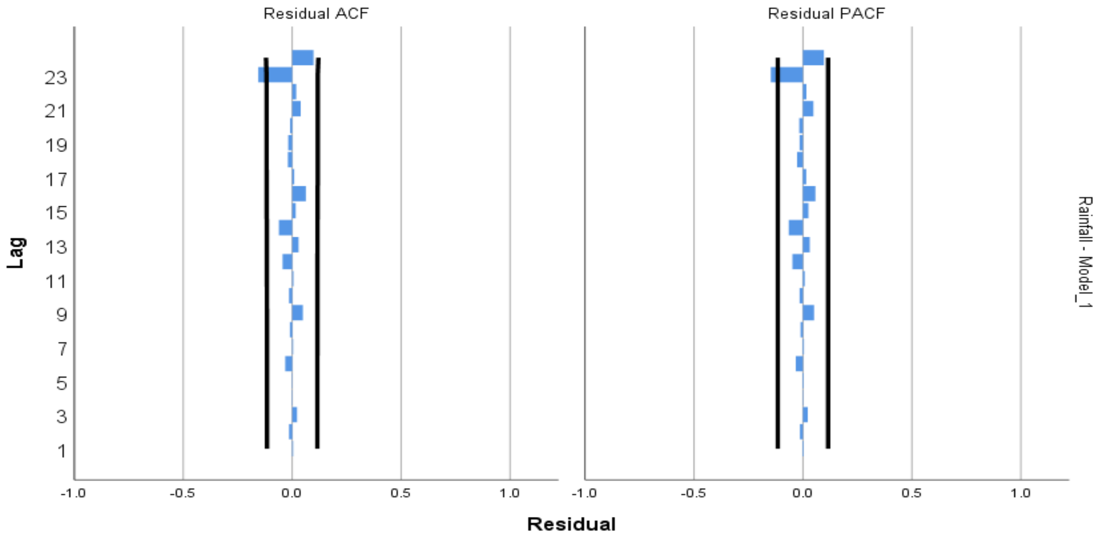

Fig. 5. ACF and PACF values of Rainfall

Table 1. Best fitted models of ARIMA

ARIMA	Best fitted Models
Rainfall	(0,0,0) (1,1,1)
Maximum Temperature	(1,0,0) (0,1,1)
Minimum Temperature	(1,0,1) (1,1,0)

After deciding the parameters p, d, and q evaluation of the model with fit statistics are required to quantify the performance. Some of the statistical measures are RMSE (Root Mean Square Error), MAPE (Mean Absolute percentage errors), and MAE (Mean Absolute Error). The values of these errors should be minimum for better performance of the model.

3.2 Model Validation and Forecasting

The best performed and predictive ability of the models are choosen the observed values, and predicted values were shown in Fig. 6. The graphs shows that the predicted values are well-fitted through the observed data. There is a slight over-prediction in March 2000, May 2004, and

November 2016 as compared with the original data. The rest of the predicted values are well-fitted through the original data. The ARIMA model is used to forecast rainfall in Prayagraj upcoming 5 years (2021-2025).

The best perfomed and predictive ability of the models are choosen, the observed values, predicted values, and fit values were plotted and displayed in Fig. 7. And the predicted values are well-fitted through the original data. This indicates that the models chosen for the maximum temperature series are the best-fitted ones for the Prayagraj. The ARIMA model is used to forecast rainfall in Prayagraj for the upcoming 5 years (2021-2025).

The best perfomed and predictive ability of the models are choosen, the observed values, predicted values, and fit values were plotted and displayed in Fig. 8 slight over can be seen in January 2017 as with the original data. Rests of the predicted values are well-fitted to the original data. This shows that the model chosen for the minimum temperature series are the best-fitted model for Prayagraj. The ARIMA model is used to forecast the minimum temperature in Prayagraj for the upcoming 5 years (2021-2025).

3.3 Diagnostic Analysis

After finding the three parameters p, d, and q calculate the model with fit statistics are required to quantify the performance of the forecast within its acceptable limits. Some of the statistical measures are RMSE (Root Mean Square Error), MAPE (Mean Absolute percentage errors), and MAE (Mean Absolute Error). The values of these errors should be minimum for better performance of the model.

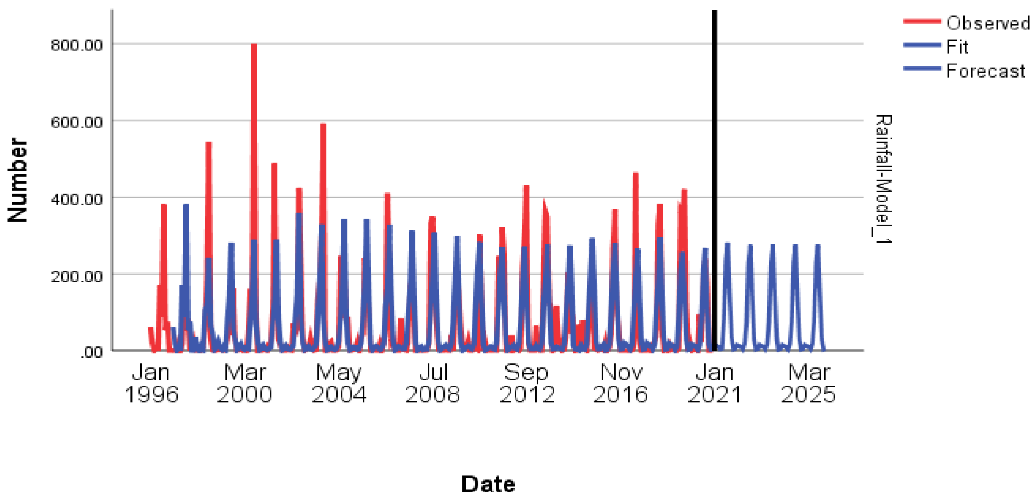

Fig. 6. Observed and fitted values of rainfall series

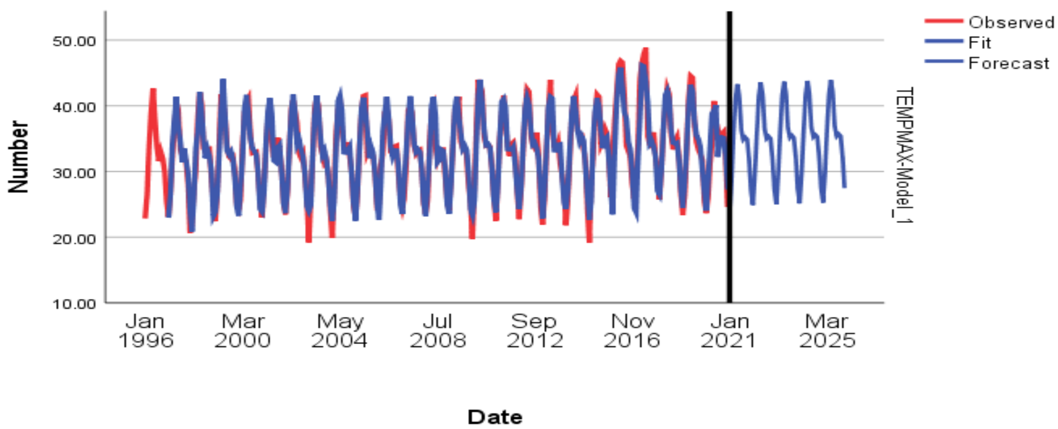

Fig. 7. Observed and fitted values of maximum temperature series

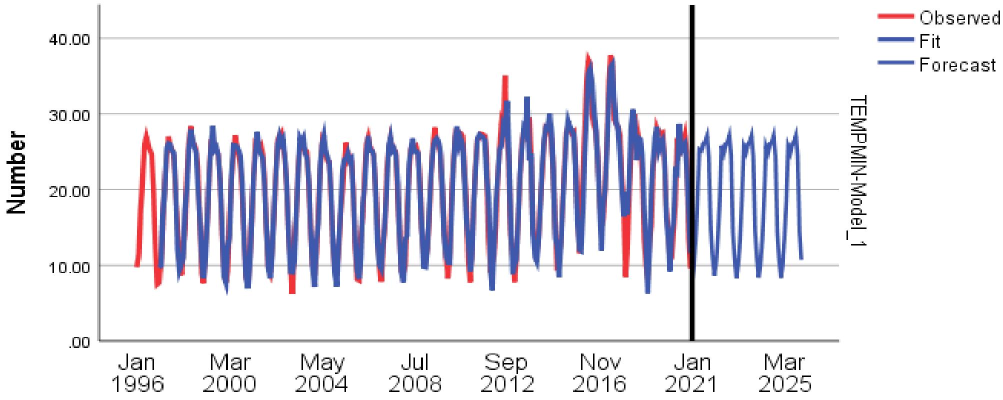

Fig. 8. observed and fitted values of minimum temperature series

Table 2. Forecasting accuracy statistics

Parameters	Rainfall model Arima (0,0,0)(1,1,1,)	Arima model Maximum temperature (1,0,0) (0,1,1)	Arima model Minimum temperature (1,0,1) (1,1,0)
Stationary R-squared	0.501	0.551	0.464
R- squared	0.497	0.908	0.927
RMSE	85.508	1.812	1.929
MAPE	337.466704	4.134	8.069
MAE	44.989	1.324	1.314

3.4 Forecasting of the Annual Yield of Chickpea by using the Arima Model

ARIMA model was used to forecast the yield of chickpea. Data used for model building is from the year 1996-2020. The data from 2021-2025 is used for cross-validation of the selected model. The yield data were separately analyzed using different ARIMA models in SPSS statistical package.To make data non stationary to stationary differencing was done. The ACF and PACF shows the values of q and p that would be suitable for a yield of chickpea are q=0 and p=0, Thus the Arima model that was found to be best fitted for a yield of chickpea is (0,1,0).

After finding the parameters p, d, and q evaluate the model with fit statistics are required.

3.5 Residual ACF and Residual PACF

The Fig. 9 are showing the residual ACF and PACF of yield. From the time plot of the residuals against time, we can see that there is no obvious pattern in the plot except for a possible outlier is independent. That is, ARIMA (0, 1, 0) is adequate for modeling the log-transformed yield data in PRAYAGRAJ.

The best fitted and predictive models are chosen, observed values, predicted values, and fit values were plotted in Fig. 10. slight over can be seen in January 2014 as with original data. Rests of the predicted values are most accurate through the original data. This indicates that the models chosen for yield data are the best-fitted ones for Prayagraj. The ARIMA model is used to forecast yield in Prayagraj for the upcoming 5 years (2021-2025).

3.6 Comparison of ARIMA and ANN

3.6.1 ARIMA

Based on yearly production data and monthly climate parameters were used to calculate ARIMA model. Here we take independent variables such as temperature, rainfall, and dependent variable as a yield. And the best fitted ARIMA model is (0,0,1) among all other ARIMA models. This has the R-Square value which is 0.591 with an RMSE value of 159.632. Among all models, ARIMA (0,0,1) is the best-fitted model for forecasting the future yield.

3.7 ANN (Artificial Neural Network)

In Prayagraj, timely sowing of chickpea crops was done in September month, hence monthly climate from September-November was taken for the study. And yearly wise 25 years (1996-2020) data was taken. Here independent variable (input) was taken as (maximum temperature, minimum temperature, rainfall), and the dependent variable (target) was taken as yield.

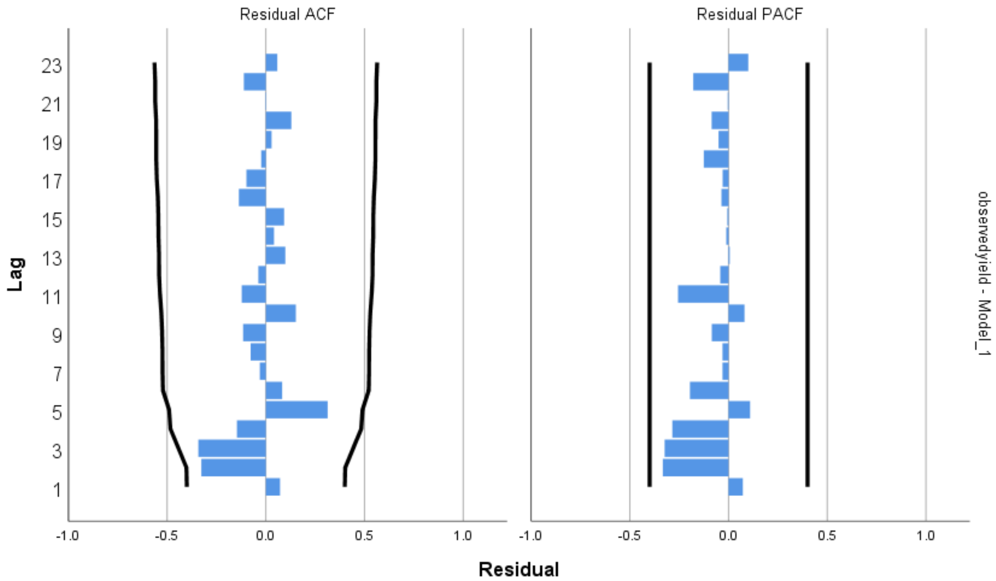

Fig. 9. Residual ACF and PACF of yield

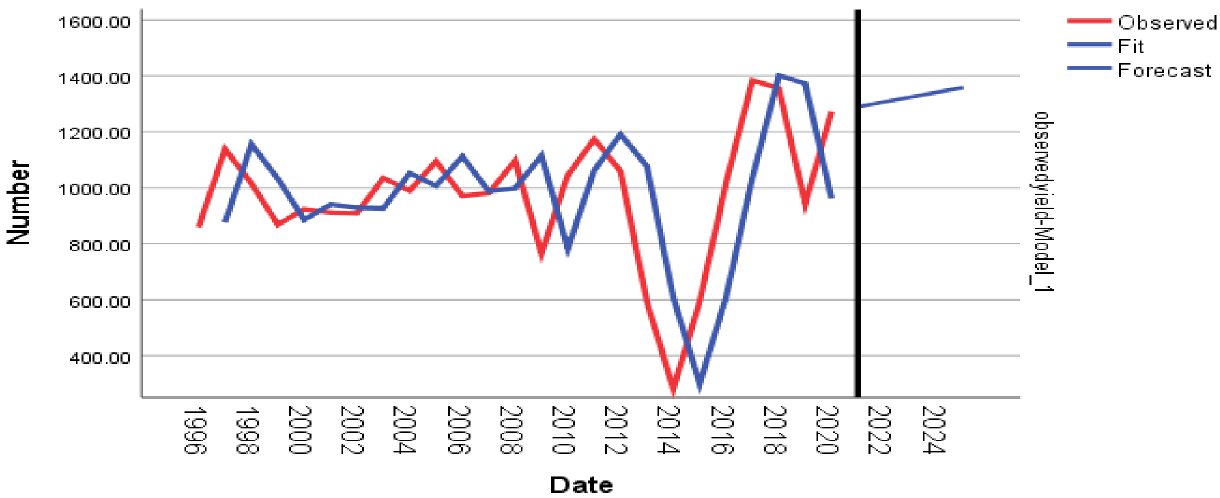

Fig. 10. Observed and fitted values of yield data

Table 3. Forecasting of yield parameters

Parameters	Forecasting of Yield Arima model (0,1,0)
Stationary R-squared	-2.220E-16
R- squared	-.025
RMSE	246.079
MAPE	24.748
MAE	194.370

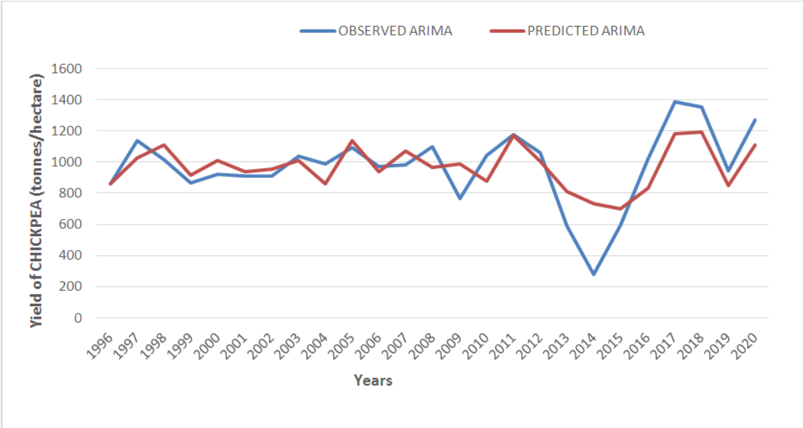

Fig. 11. Observed and fitted values of ARIMA

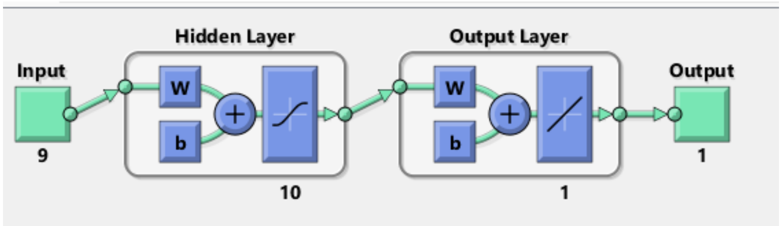

Fig. 12. Observed and fitted values of ANN

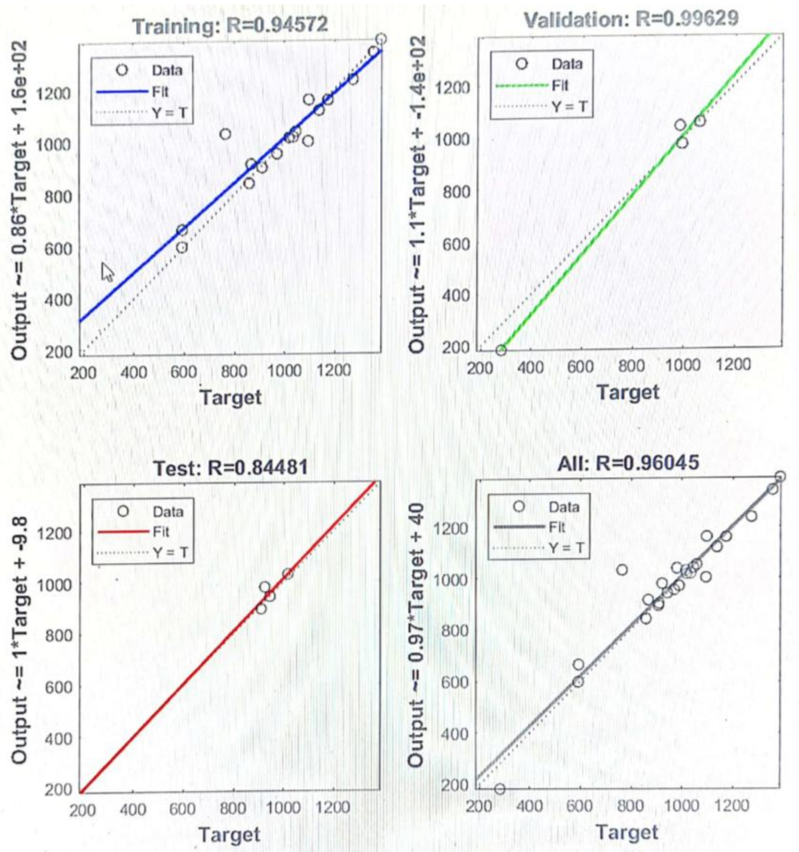

Fig. 13. Neural network training regression graph

Table 4. Forecasting parameters for ARIMA and ANN

Model	RMSE	R^2 value
ARIMA	159.632	0.591
ANN	66.716	0.96

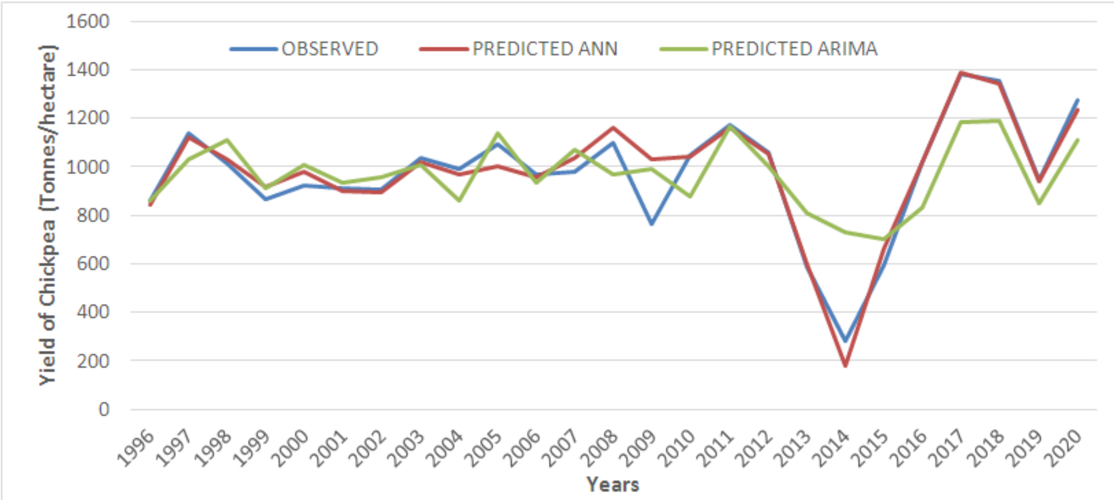

Fig. 14. Observed and forecast values of Chickpea yield by ARIMA and ANN

The ANN was trained with 70% of target data and 15% of data were used to validate and 15% of data was used for testing. In addition, 10 hidden neurons and 2 delays were used in the network. The details are shown in Fig. 10. Performance of each training algorithm in predicting the atmospheric temperatures was using the Mean Squared Error (MSE) and the correlation coefficient (R).

On similar data set, the ANN model was used. After a lot of training the best fit model of the ANN is with the R - Square value of 0.960 and RMSE value of 66.716.

3.8 Comparison of ARIMA and ANN

For comparison purposes, the training and testing performance of the ANN and ARIMA model were used. The ARIMA and ANN forecasts are close to actual values. It shows that both approaches work well for chickpea forecasting. Empirical results of two different models clearly reveal the efficiency of the ANN model. It shows ANN models are the best fit when compared to the ARIMA model.

In the work, it is clearly reveals that ANN is the best fit model for forecasting the impact of climate change in the Prayagraj district. ANN models always have the highest R-Square value and low RMSE than all other models so ANN is the best model to fit. From the graph, it could be observed that the yield parameter was more accurate with the ARIMA model when compared with the ANN model.

4. CONCLUSION

Rainfall and temperature are the main factors for climate prediction. The ARIMA and ANN forecasts are close to the actual value. It shows that both of the approaches work well for chickpea forecasting. Based on the results we concluded ARIMA models are not always adequate for the time series data. It is mainly seen from the comparison of the R – Square value and RMSE of the models. It reveals that ANN is the best fit model for forecasting the impact of climate change in the Prayagraj district. ANN models always have the highest R-Square value and low RMSE than all other models so ANN is the best model to fit. Various models of ARIMA was used for predicting the climate change and crop yield of Chickpea and the best predicted model was 0.591. When Compared with the results of ARIMA and ANN for climate change and crop yield. ANN is accurate to predict the future values. From the graph, it could be observed that the yield parameter was more accurate with the ARIMA model when compared with the ANN model.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Saxena MC. Problems and potential of chickpea production in nineties. In: Chickpea in the nineties. Proceedings of the second international workshop on chickpea improvement, 4–8 Dec 1989, ICRISAT Center Patancheruvu India; 1990.
2. Cornillon P-A, Imam W, Matzner-Løber E. Forecasting time series using principal component analysis with respect to instrumental variables. Comp Stat Data Anal. 2008;52(3):1269-80.
3. Ibrahim MZ. Time-series analysis of pollutants in East Coast peninsular Malaysia. J Sustainability Sci Manag. 2010;5(1):57-65.
4. Samsuri A. Multiple Linear Regression (MLR) models for long term Pm10 concentration forecasting during different monsoon seasons. J Sustainability Sci Manag. 2017;12(1):60-9.
5. Acharya N. Development of an artificial neural network based multimodel ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Clim Dynam. 2014;43(5-6): 303-1310.
6. Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. Journal of Hydrology. 2014;508:418-29.
7. Hashim FR, Nik Daud NG, Ahmad KA, Adnan J, Rizman ZI. Prediction of rainfall based on weather parameter using artificial neural network. J Fundam Appl Sci. 2017;9(3S):493-502.
8. Mishra N, Soni HK, Sharma S, Upadhyay AK. Development and analysis of artificial neural network models for rainfall prediction by using time-series Data. IJISA. 2018;10(1):16-23.
9. Qiu M, Song Y. Predicting the direction of stock market index movement using an optimized artificial neural network model. PLOS ONE. 2016;11(5):e0155133.
10. Box GEP, Jenkins GM. Time series analysis: forecasting and control. rev ed. San Francisco: Hoden-Day; 1976.
11. Mislan, Haviluddin, Hardwinarto S, Sumaryono, Aipassa M. Rainfall monthly prediction based on artificial neural network: A case study in tenggarong station East Kalimantan-Indonesia. Procedia Comput Sci. 2015;59: 142-51.
12. Moghim S, Bras RL. Bias correction of climate modeled temperature and precipitation using artificial neural networks. J Hydrol Meteorol. 2017;18(7):1867-84.
13. Mohapatra. Rainfall prediction based on 100 years of meteorological data, Computing and Communication Technologies for Smart Nation (IC3TSN) International Conference on. 2017. IEEE Publications; 2017;162-6.
14. Sachan, Abhishek. Forecasting of rainfall using ANN, GPS and meteorological data, Convergence of Technology (I2CT) International Conference for 1-4 IEEE. Vol. 2014; 2014.
15. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants features and applications. Trends Biotechnol. 2005;23(1):48-55.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/91614