International Journal on Data Science and Technology
2022; 8(4): 61-66
http://www.sciencepublishinggroup.com/j/ijdst
doi: 10.11648/j.ijdst.20220804.11
ISSN: 2472-2200 (Print); ISSN: 2472-2235 (Online)

Forecasting Wheat Production in India Using ARIMA and Radial Basis Function

Subbiah Selvakumar[1], Veluchamy Kasthuri[2]

[1]Department of Statistics, Government Arts and Science College, Nagercoil, India

[2]Department of Economics, Erode Arts and Science College, Erode, India

Email address:

selvastatmsu@gmail.com (Subbiah Selvakumar), Kasthuri.veluchamy@gmail.com (Veluchamy Kasthuri)

To cite this article:

Subbiah Selvakumar, Veluchamy Kasthuri. Forecasting Wheat Production in India Using ARIMA and Radial Basis Function. *International Journal on Data Science and Technology*. Vol. 8, No. 4, 2022, pp. 61-66. doi: 10.11648/j.ijdst.20220804.11

Received: September 9, 2022; **Accepted**: October 14, 2022; **Published**: November 29, 2022

Abstract: The time series is an arrangement of values in a specific order of time. Time series analysis, mostly used for forecasting. Prediction and analysis of wheat is a vital role in agricultural statistics. Indian wheat is largely a soft/medium hard, medium protein, white bread wheat, somewhat similar to U.S. hard white wheat. India is the second largest producer of wheat. The Agriculture Statistics System is very complete and provides data on a wide range of topics such as crop area and production, land use, water irrigation, land holdings, etc. Agricultural credit and subsidies are also considered important supporting factors for agriculture growth. Food grain production covers the dominant part of the cropped area (65%) in Indian agriculture. India is the world's largest producer of millets and second-largest producer of wheat, rice, and pulses. The present research work focused on the production of wheat in India using time series data ranging from 2001 to 2021. In this paper, Autoregressive Integrated Moving Average Model (ARIMA) and Radial Basis Function (RBF) for predicting wheat production of India was compared. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were compared. The outcomes were displayed numerically and graphically.

Keywords: ARIMA, RBF, MAE, MAPE, RMSE, Residual Analysis

1. Introduction

A time series is a collection of values based on time. Wheat is the leading cereal yield in India. The total area below the yield is about 29.8 million hectares in the country. Wheat crop has widespread manageability. It can be developed not only in the tropical and sub-tropical zones, but also in the pleasurable zone and the cold zones of the far north elsewhere even the 60 degrees north elevation. Uttar Pradesh is the leading producer of wheat and West Bengal is the prevalent producer of rice in India. China has the chief land area excited to wheat production. Punjab has appeared at the largest wheat producer state in India. Punjab is known as the wheat granary of India. Thasleema, Ahammad Basha Shaik [14] predicted production and yield of wheat in major growing states of India using ARIMA model. Arinjita Bhattacharyya [1] developed a hybrid forecasting model for COVID- 19. Sathish et al. [12] predicted discharge level per days using ARIMA and MLE models and also error estimates, RMSE, MAE were compared. Hiransha et al. [4] compared various model for stock pricing prediction. Athira et al. [2] predicted pollution and meteorological time series AirNet data using Neural Network. Sheetal et al. [13] forecasted time series data. Jayanthi Balaji et al. [5] predicted stock price movement data using Neural network. Yousif Alyousifi [15] indicated that the forecasting accuracy. Junling [6] focused exponential growth rate of an epidemic is an important measure of the severeness of the epidemic. Paarth Thadani [8] segregated noise data result into enhanced model show with marginal computational source and arrangement. Ben A. Smith [3] developed modified Incidence Decay and Exponential Adjustment model to predict the progression of transferrable disease occurrences. Rediat Takele [10] applied ARIMA model for prominent coronavirus (COVID-19) occurrence in selected East African countries. Rout et al. [11] applied low complexity recurrent neural network for stock market prediction. Rahim et al. [9] applied the sliding window technique for defining the suitable length of intervals.

Mao and Xiao [7] developed complex network analysis for fuzzy time series models. In this paper, Autoregressive Integrated Moving Average Model (ARIMA) and Radial Basics Function were used for wheat production prediction in India from 2001 to 2021. The enactment of these diverse models was assessed by foretelling truthfulness.

The chief unbiased of the learning is to estimate and forecasts the wheat production in India.

2. Time Series

Time series offerings for evaluating time series data. It is used to forecast future events based on well-known past events.

2.1. Autoregressive (AR) Model

The model (Y_t - δ) is

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + \ldots + \alpha_p (Y_{t-p} - \delta) + u_t \quad (1)$$

Where δ is the mean of Y and u_t is an uncorrelated random error term with zero mean and constant variance σ^2.

2.2. Moving Average (MA) Model

The moving average process is merely a linear combination.

The model Y_t is as follows,

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (2)$$

is an MA (q) process.

2.3. Autoregressive Moving Average (ARMA) Model

Both combinations of AR and MA and is therefore ARMA. Thus, Y_t follows an ARMA (1, 1) process if it can be written as,

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (3)$$

Because there is one autoregressive and one moving average term. θ Represents a constant term.

In general, in an ARMA (p, q) process, there will be p autoregressive and q moving average terms.

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} + \alpha_2 Y_{t-2} + \beta_2 u_{t-2} + \cdots + \alpha_p Y_{t-p} + \beta_q u_{t-q} \quad (4)$$

2.4. Autoregressive Integrated Moving Average Model (ARIMA)

ARIMA (p, d, q), where p is Autoregressive and q is the Moving Average Model and d is the differencing. If d =0, the data exhibits stationary and the order is denoted as (p, q), which is called ARMA process. If the data does not exhibit stationary, the first order differencing is carried out for converting it into stationary, hence the model is denoted as (p, d, q).

3. Radial Basis Function (RBF)

RBF networks, a discussion of feed forward networks have collective estimate abilities. It is recycled for curve fitting approximation difficult in a high dimensional space. It covers three layers that is input layer, hidden layer and output layer. Input layer is completed of source nodes that connect the network to its environment. Second is the hidden layer which put on a nonlinear transformation from the input space to the hidden space, which is of high dimensionality. Output layer is linear, providing the answer of the network to the activation configurations applied to the input layer. The RBF output layer consequences in a linear manner. The output y is calculated by:

$$y_i(x) = \sum_{k=1}^{J_2} W_{ki} \phi(\|X - C_k\|) \quad (5)$$

For i= 1,.., J_3 where $y_i(x)$ is the i[th] output of the RBFN. W_{ki} is the connection weight from the k[th] hidden to the i[th] output unit C_k is the prototype or center of the kth hidden unit, and $\|.\|$ denotes the Euclidean norms. The RBF $\phi(.)$ typically selected as the Gaussian function $\varphi(x - c_i) = e^{-\|x-c_i\|^2/2\sigma^2}$

Where $c_i = (c_{i1}, c_{i2}, \ldots, c_{in})$ is the center of the associated field, and σ is the width of the Gaussian function.

4. Residual Analysis

Thus, residuals denote the portion of the confirmation data not clarified by the model. Diverse kinds of error measurement namely.

1) Mean Absolute Error (MAE)
2) Mean Absolute Percentage Error (MAPE)
3) Root Mean Square Error (RMSE)

4.1. Mean Absolute Error (MAE)

Mean absolute error is a portion of inaccuracies among corresponding explanations stating the same occurrence.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|u_t| \quad (6)$$

Where n is the number of years and $|u_t| = Y_t - \hat{Y}_t$. Y_t is actual values at time t. $\hat{Y}_t$ is predicted values at time t.

4.2. Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error is defined as the difference between actual value and estimated values divided by actual values multiplied by 100 and taking summation and divided by number of observations is given by:

$$=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right| \tag{7}$$

4.3. Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is a usual manner to amount the inaccuracy of a model in forecasting numerical records.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}\left(Y_t-\hat{Y}_t\right)^2}{n}} \tag{8}$$

5. Results and Discussion

For the analysis data from 2001 – 2021 is measured. The data contain of the Wheat production of India. In this work we have considered wheat production of prediction for ARIMA and radial basis function. The data is taken from Http//:Agricoop.co.nic.in.

ACF and PACF of Wheat Production:

In order to make the points stationary, first order differencing carried out. Below the graphs give the details on the first order differencing.

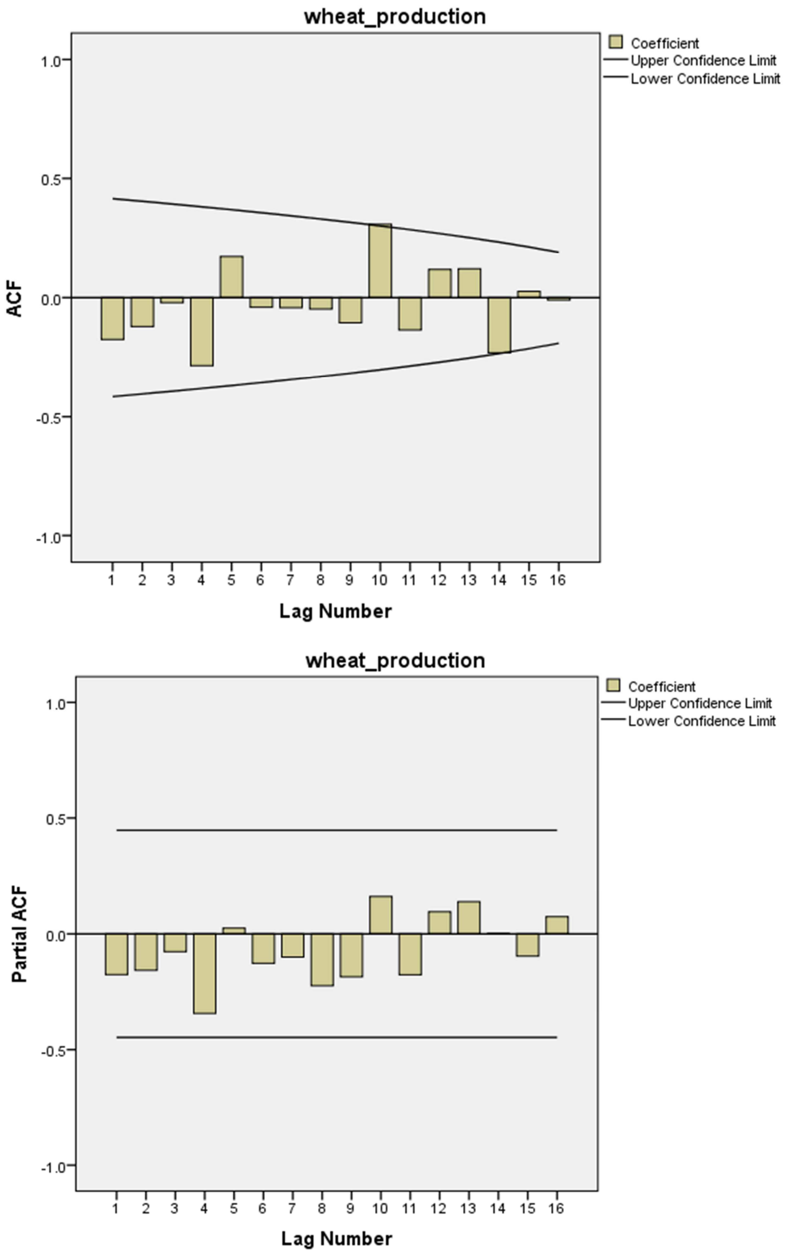

Figure 1. *Autocorrelation and Partial Autocorrelation graph for First Order Differencing.*

Figure 1 represents first order differencing using autocorrelation and partial autocorrelation function. All the data points within the control limits.

Table 1. *BIC Values of ARIMA (p, d, q).*

ARIMA (p, d, q)	Normalized BIC
ARIMA (1,1,0)	17.475
ARIMA (0,1,1)	17.479
ARIMA (1,1,1)	17.508
ARIMA (2,1,0)	17.694
ARIMA (2,1,1)	17.722
ARIMA (0,1,2)	17.507

When comparing with other models, smaller BIC statistic value indicates the better fitting model. The specified order is ARIMA (1, 1, 0) and hence the model is fitted and the forecasting is done.

Table 2 shows that forecasted values from 2022 is 113263 tonnes to 2025 124261 tonnes. forecasted values increased from 113263 to 124261 tonnes. And also upper control limit and lower control limit are calculated.

Table 2. *Forecasting using ARIMA Model.*

Year		2022	2023	2024	2025
Wheat	Forecast	113263	116744	120443	124261
Production	UCL	123764	130161	136483	142509
(tonnes)	LCL	102763	103328	104402	106012

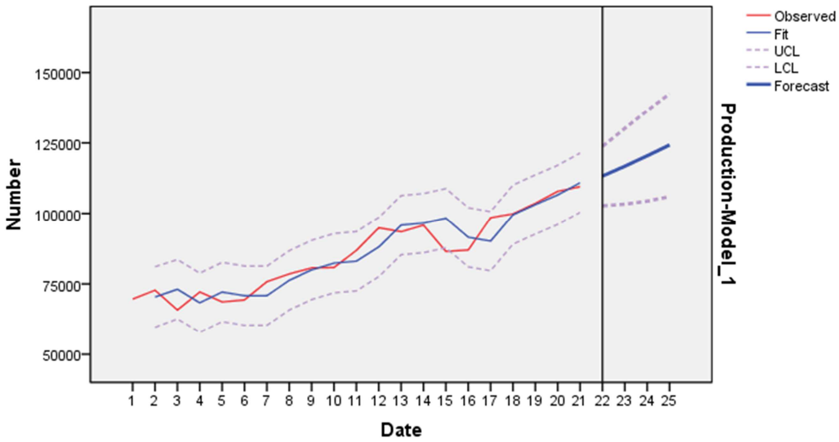

Figure 2. *Actual and Predicted Values of Wheat Production Using ARIMA Model.*

Figure 2 represents the actual and forecasted values using ARIMA model and also upper control limits and lower limits are shown.

ACF and PACF of Wheat Production (Residual Analysis)

The residual analysis is carried out using autocorrelation and partial autocorrelation function. Below graphs represents the residuals of ACF and PACF.

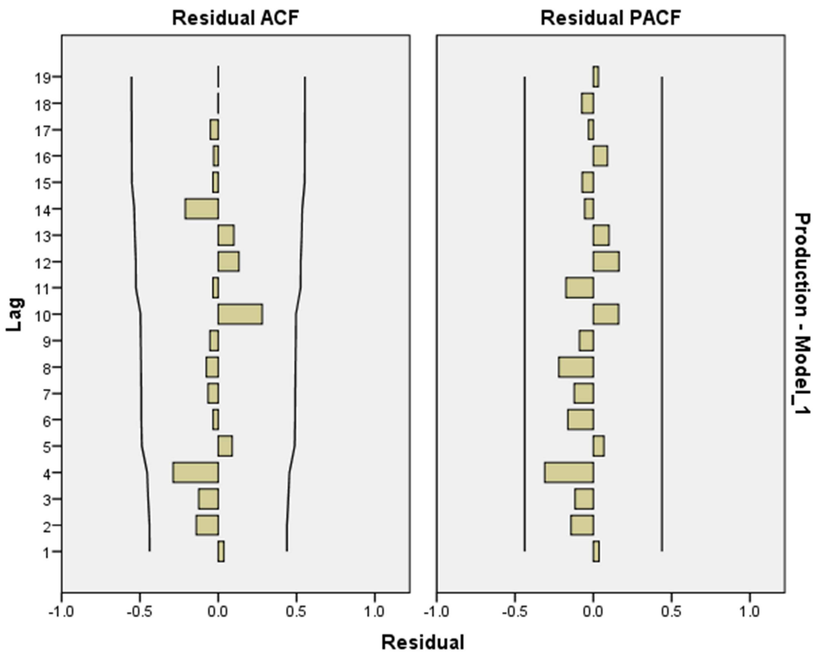

Figure 3. *Autocorrelation and Partial Autocorrelation graphs for Residuals.*

Figure 3 indicates residuals. All points within the control limits.

Table 3. *Forecasted Values using RBF Model.*

Year	**2022**	**2023**	**2024**	**2025**
Wheat Production (tonnes)	107047	107044	107024	106851

Table 3 shows that forecasted values from 2022 is 107047 to 2025 is 106851 tonnes. The forecasted values marginally decreased from 2022 to 2025.

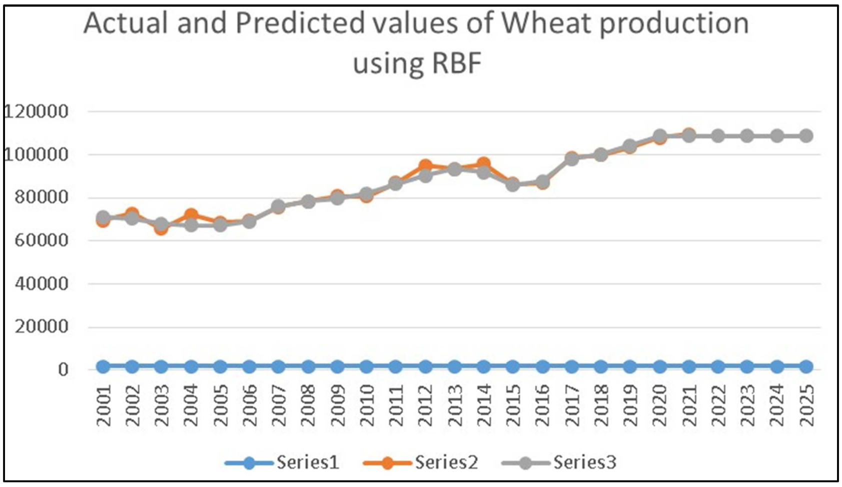

Figure 4. *Actual and Predicted of Wheat Production Using RBF.*

Figure 4 represents the actual and forecasted values using RBF.

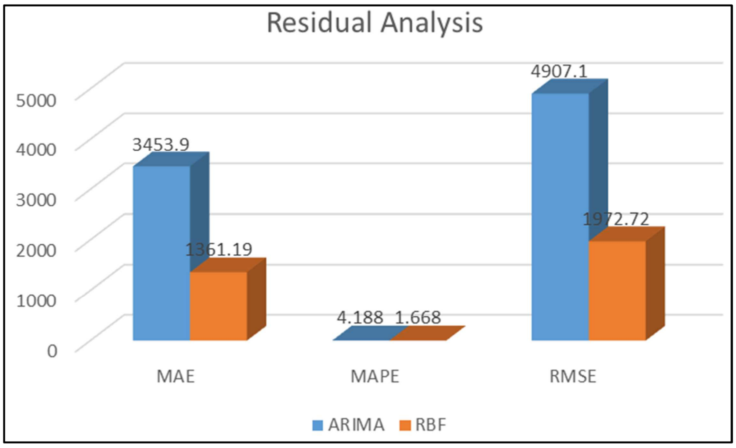

Figure 5. *MAE, MAPE and RMSE using ARIMA and RBF.*

Figure 5 shows that the mean absolute error, mean absolute percentage error and root mean square error obtained by using ARIMA and Radial Basics Function for wheat production prediction. Radial Basics Function is less values when compared to ARIMA. Radial Basics Function is better than that of ARIMA.

6. Conclusion

In this work, two models namely ARIMA and RBF were used for wheat production prediction in India. Autocorrelation function and partial autocorrelation function were used in ARIMA model. ARIMA (1,1,0) model is fitted and BIC =17.475 is minimum when compared to other order ARIMA model. Mean absolute error, mean absolute percentage error and Root Mean Square Error were compared using bar chart. Errors were minimum for RBF model when compared to ARIMA model. RBF is achieved superior than that ARIMA model. And also forecasting is done using radial basis function.

References

[1] Arinjita Bhattacharyya., Tanujit Chakraborty and Shesh N. Rai (2022): Stochastic Forecasting of COVID – 19 Daily New Cases Across Countries With a Novel Hybrid Time Series Model, Springer, vol. 107, pp. 3025-3040.

[2] Athira Va, Geetha Pb, Vinayakumar Rab, Soman K P (2018): DeepAirNet: Applying Recurrent Networks for Air Quality Prediction, International Conference on Computational Intelligence and Data Science, Procedia Computer Science, vol. 132 pp. 1394–1403.

[3] Ben A. Smith (2020): A Novel IDEA: The Impact of Serial Interval on a Modified- Incidence Decay and Exponential Adjustment Model for Projections of Daily COVID-19 Cases, Infectious Disease Modelling, vol 5, pp. 346-356.

[4] Hiransha M, Gopalakrishnan E. A, Vijay Krishna Menon, Soman K. P (2018): NSE Stock Market Prediction using Deep-Learning Models, International Conference on Computational Intelligence and Data Science, Procedia Computer Science, vol. 132, pp. 1351–1362.

[5] Jayanth Balaji A., D. S. Harish Ram, Binoy B. Nair (2018): Applicability of Deep Learning Models for Stock Price Forecasting An Empirical Study on BANKEX Data, Procedia Computer Science, vol. 143, pp. 947–953.

[6] Junling M (2020): Estimating Epidemic Exponential Growth Rate and Basic Reproduction Number, Infectious Disease Modelling, vol. 5, pp. 129-141.

[7] Mao. S and F. Xiao (2019): Time Series Forecasting Based on Complex Network Analysis, IEEE, vol. 7, pp. 40220–40229, 2019.

[8] Paarth Thadani (2021): Financial Forecasting using Stochastic Models: Reference from Multi-Commodity Exchange of India, Data Science in Finance and Economics, vol 1, pp. 196-214.

[9] Rahim, N. F., Othman, M Sokkalingam, R and E. A. Kadir (2018): Forecasting Crude Palm Oil Prices Using Fuzzy Rule-Based Time Series Method, *IEEE Access*, vol. 6, pp. 32216 - 32224.

[10] Rediat Takele (2020): Stochastic Modelling for Predicting COVID-19 Prevalence in East Africa Countries, Infectious Disease Modelling, vol. 5, pp 598-607.

[11] Rout A. K., Dash P. K., Dash R., and Bisoi R. (2017): Forecasting Financial Time Series Using a Low Complexity Recurrent Neural Network and Evolutionary Learning Approach., Journal of King Saud University-Computer and Information Sciences, vol. 4, pp. 536-552.

[12] Sathish, S, Khadar Babu SK. and C. U. Tripura Sundari (2019): Weather and Stochastic Forecasting Method for Generated Discharge Level Data at Sathanur Dam, International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, pp. 2278-3075.

[13] Sheetal S. Patil1, S. H. Patil, A. M. Pawar (2021): Study of Various Forecasting Models for Time Series Data Using Stochastic Processes, Journal of Computer Technology & Applications, vol. 12, pp. 26-32.

[14] Thasleema S. C, and Ahammad Shaik (2019): Stochastic Time Series Modeling for Rice Production in Andhra Pradesh, International Journal of Scientific Research in Mathematics and Statistics, Vol. 6, pp. 50 -55.

[15] Yousif Alyousifi, Mahmod Othman Andakram A and Almohammedi (2021): A Novel Stochastic Fuzzy Time Series Forecasting Model Based on a New Partition Method, IEEE, vol. 9, pp. 80236-80252.