Scan to know paper details and author's profile

Wheat Production Prediction in India using ARIMA, Neural Network and Fuzzy Time Series

S. Selvakumar & V. Kasthuri

ABSTRACT

A time series is a predetermination of data points that happen in repeated order of time. Forecasting productions play a necessary part in several fields such as, meteorological data, weather data, stock market data, rainfall data, agriculture data and so on. In recent years, fuzzy time series is used for forecasting. Song and Chissom (1993) proposed fuzzy time series for forecasting enrollments of data. In this paper, Autoregressive Integrated Moving Average Model (ARIMA), Neural networks for Radial Basis Function (RBF) and Multilayer Perceptran(MLP) and fuzzy time series for predicting wheat production of India were compared. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were compared. The results were displayed numerically and graphically.

Keywords: ARIMA, neural network, fuzzy time series, MAE, MAPE, RMSE, residual analysis and prediction.

Classification: FOR Code: 280212

Language: English

London Journals Press

LJP Copyright ID: 925652
Print ISSN: 2631-8490
Online ISSN: 2631-8504

London Journal of Research in Science: Natural and Formal

Volume 22 | Issue 15 | Compilation 1.0

Wheat Production Prediction in India using ARIMA, Neural Network and Fuzzy Time Series

S. Selvakumar[α] & V. Kasthuri[σ]

ABSTRACT

A time series is a predetermination of data points that happen in repeated order of time. Forecasting productions play a necessary part in several fields such as, meteorological data, weather data, stock market data, rainfall data, agriculture data and so on. In recent years, fuzzy time series is used for forecasting. Song and Chissom (1993) proposed fuzzy time series for forecasting enrollments of data. In this paper, Autoregressive Integrated Moving Average Model (ARIMA), Neural networks for Radial Basis Function (RBF) and Multilayer Perceptran(MLP) and fuzzy time series for predicting wheat production of India were compared. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were compared. The results were displayed numerically and graphically.

Keywords: ARIMA, neural network, fuzzy time series, MAE, MAPE, RMSE, residual analysis and prediction.

Author α: Govt Arts and Science College, Nagercoil, Tamil Nadu.
σ: Erode Arts and Science College, Erode, Tamil Nadu.

I. INTRODUCTION

The time series is a category of variables ordered in a specific order of time. Forecasting believes the future values of the time series. For the past decades, fuzzy time series has been widely used for predicting the historic data. Fuzzy time series is used to planning with forecasting difficulties in which the historical data are linguistic values. Song and Chissom (1993) proposed the fuzzy time series for the enrollments of university of Alabama. Wheat is the chief cereal crop in India. Wheat crop has well-known flexibility. Wheat is developed in a variety of soils of India. Yanpeng Zhang et.al(2020) proposed a novel fuzzy time series forecasting model by multiple linear regression and time series clustering for forecasting market prices. Singh, P. (2018) offered a new model to deal with four major issues of fuzzy time series (FTS) forecasting, viz., Wangren Qiu et.al(2015) proposed model was implemented in forecasting enrollment data at the University of Alabama. Ozge Cagcag Yolcu et.al(2016) proposed a novel high-order fuzzy time series approach that considers the membership values, where artificial neural networks are employed to identify the fuzzy relations. Adesh Kumar Pandey(2008) proposed fuzzy time series and neural network. The proposed method has been implemented in the historical data. Paarth Thadani(2021) presented non-linear forecasting models, including artificial neural networks, are popularly adopted in financial forecasting. Yousif Alyousifi et.al (2021) proposed Fuzzy Time Series Markov Chain – Transition Probability Matrix model is tested using two types of time series data, namely, air pollution index (API) data, and yearly enrollments for the University of Alabama.Wang et.al (2017) applied autoregressive integrated moving average model and artificial neural network for air pollution data. Alyousi et. al (2019) applied artificial neural network and Markov chain are applied for air pollution forecasting. In this work, ARIMA, Neural networks for Multilayer Perceptron and Radial Basics Function and fuzzy time series algorithms are used for wheat production prediction in India. Residual analysis for Mean

Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were compared.

II. METHODOLOGY

2.1 Autoregressive Integrated Moving Average Model

Autoregressive Integrated Moving Average Model (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 in converting it into a stationary, hence the model is denoted as (p, d, q).

2.2 Radial Basis Function (RBF)

RBF networks, a class of feed forward networks, called radial basis function that compute activations at the hidden neurons in a way that is different from what we have seen in the case of feed forward neural networks. Rather than employing an inner product between the input vector and the weight vector The RBF output layer results in a linear fashion. The output y is computed by

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

For i= 1,.., J_3 where $y_i(x)$ is the i^{th} output of the RBF. 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 $\|\cdot\|$ denotes the Euclidean norms. The RBF ϕ(.)typically selected as the Gaussian function

$$\phi(x - c_i) = e^{-\|x-c_i\|^2/2\sigma^2}$$

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.

2.3 Multi Layer Perceptron

Artificial Neural Network(ANN) is an operational model which consists of a large number of interconnected nodes(neurons). Each node contains a specific output function which is called an activation function. The connection between every two nodes represents a weighted value that passes through the connection signal, which is called weight. Weight is equivalent to the memory of ANN. Multi Layer Perceptron(MLP) has many layers, the first layer is the input layer, the last layer is the output layer, the middle layers are called hidden layers, each layer includes several neurons. This calculation process is called feed forward process of MLP. If there is an MLP, which contains m hidden layers, its input and output dimensions are respectively equal to n_1 and n_{m+2}. The number of nodes in each hidden layer is $n_2,n_3,\ldots,n_{m+1}$ respectively. In the feed forward process of this MLP, each node value is calculated using the following formula

$$X_{ij}= f(WX_{i-1}+b_{i-1})$$

fis the activation function.

Where X_{ij} represents the value of the j neuron i layer. W_i represents the weight vector of the j neuron in layer i-1 to layer i. X_{i-1} represents the value vector of all neurons in layer i-1. b_{i-1} represents the bias of the i-1 layer , and f is the activation function.

2.4 Fuzzy Time Series

It is the values of the observations of a special dynamic process are represented by linguistic values.

Computational Algorithm for Fuzzy Time Series

The step by step process is as follows:

Step1: Calculate the first order variation of the historical data.

Step2: Define the universe of discourse, U based on the range of available historical data.

$$U=[D_{min}-D_1,\ D_{max}+D_2]$$

Where D_{min} is the minimum value of the first order variation of the historical data,D_{max} is the maximum value of the historical data and D_1, D_2 are two positive integers.

Step3: Partition the universe of discourse U into equal length intervals: $u_1,u_2,\ldots,u_m$.

Step4: The number of intervals will be in accordance with the number of linguistic variables (fuzzy sets) $A_1,A_2,\ldots,A_m$ to be considered.

Step5: Fuzzify the variations of the historical data and establish the fuzzy logical relationship is represented by $A_i \rightarrow A_j$.

Step6: Rules for forecasting:

$[A_j]$ is corresponding interval u_j for which membership in A_j is supremum (i.e., 1)

$L[A_j]$ is the highest value of the interval u_j having supremum value in A_j.

$M[A_j]$ is the mid value of the interval u_j having supremum value in A_j.

For a fuzzy logical relationship $A_i \rightarrow A_j$..

A_i is the fuzzified wheat production of the current year n;

A_j is the fuzzified wheat production of the next year n+1;

D_i is the actual wheat production of the current year n;

D_{i-1} is the actual wheat production of the previous year n-1;

E_i is the variation wheat production of the current year n;

E_{i-1} is the variation wheat production of the previous year n-1;

F_j is the forecasted wheat production of the next year n+1;

Step7: Forecasting wheat production for the year n+1 is obtained from modified computational algorithm as follows;

Obtain the fuzzy logical relationship $A_i \rightarrow A_j$..

$$\text{If } E_i<M[A_j], \text{ then } F_j=D_{i-1}+(M[A_j]-1/6L[A_j])$$
$$\text{elseIf } E_i>M[A_j], \text{ then } F_j=D_{i-1}+(M[A_j]+1/6L[A_j])$$
$$\text{else } F_j=D_{i-1}+M[A_j]$$

Step8: Obtain the mean absolute error using actual values and forecasted values

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|u_t|$$

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.

Step9: Obtain the mean absolute percentage error using actual values and forecasted values

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

Step 10: Obtain the root mean square error using actual values and forecasted values

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

III. RESULTS AND DISCUSSIONS

Step1: Compute the first order variation of the historical data.

Step2: The universe of discourse U is defined as+

$$U=[D_{min}-D_1, D_{max}+D_2]$$

$$U=[-9323-77, 11510+90]= [-9400, 11600]$$

Where D_{min}= -9400 is the minimum value of the first order variation of the historical data.

D_{max}= 11600 is the maximum value of the first order variation of the historical data,

D_1=77 and D_2=90 are two positive integers. D_1, D_2 are choosing arbitrarily for the rounded off U value.

Step3: The universe of discourse U is partitioned into five equal length of intervals.

U_1=[-9400, -5200] U_2= [-5200, -1000] U_3= [-1000, 3200] U_4 = [3200, 7400] U_5 =[7400, 11600]

Step4: Define five fuzzy sets A_1,A_2,…,A_8 having some linguistic values on the universe of discourse U. The linguistic values are as follows:

A_1= very poor A_2= poor A_3= Moderate A_4= Good A_5= Very Good

Table 1: Fuzzified Wheat Production(tonnes) on Variations

Year	Wheat production(tonnes)	Variations	Fuzzified variations
2001	69681	-	-
2002	72766	3085	A3
2003	65761	-7005	A1
2004	72156	6395	A4
2005	68637	-3519	A2
2006	69355	718	A3
2007	75807	6452	A4
2008	78570	2763	A3
2009	80679	2109	A3

2010	80804	125	A3
2011	86874	6070	A4
2012	94882	8008	A5
2013	93506	-1376	A2
2014	95850	2344	A3
2015	86527	-9323	A1
2016	87000	473	A3
2017	98510	11510	A5
2018	99870	1360	A3
2019	103600	3730	A4
2020	107860	4260	A4
2021	109520	1660	A3

Step5: The membership of above mentioned linguistic variables is assigned through the Trapezoidal membership function by fixing the values arbitrarily. The memberships of Linguistic variables are as follows.

$$A_1= 1/u_1+0.5/u_2+0/u_3+0/u_4+0/u_5$$
$$A_2= 0.5/u_1+1/u_2+0.5/u_3+0/u_4+0/u_5$$
$$A_3= 0/u_1+0.5/u_2+1/u_3+0.5/u_4+0/u_5$$
$$A_4= 0/u_1+0/u_2+0.5/u_3+1/u_4+0.5/u_5$$
$$A_5= 0/u_1+0/u_2+0/u_3+0.5/u_4+1/u_5$$

Step6: The historical variations of the time series data are fuzzified in order to have the fuzzy logical relations obtained as follows: Variations in the fuzzy logic relationships

$$A_3\rightarrow A_1,\ A_1\rightarrow A_4,\ A_4\rightarrow A_2,\ A_2\rightarrow A_3,\ A_3\rightarrow A_4,\ A_4\rightarrow A_3$$
$$A_3\rightarrow A_3,\ A_3\rightarrow A_3,\ A_3\rightarrow A_4,\ A_4\rightarrow A_5,\ A_5\rightarrow A_2,\ A_2\rightarrow A_3$$
$$A_3\rightarrow A_1,\ A_1\rightarrow A_3,\ A_3\rightarrow A_5,\ A_5\rightarrow A_3,\ A_3\rightarrow A_4,\ A_4\rightarrow A_4$$
$$A_4\rightarrow A_3$$

Step-7: The forecasted values have been obtained by using the computational algorithm. Then the forecasted output while comparing with different models given as table 2.

Table 2: Forecasted Wheat Production(tonnes) by Different Models

Year	Actual Wheat Production	ARIMA	RBF	MLP	Fuzzy Time Series
2001	69681	70444	71136	68012	---
2002	72766	72919	70476	68511	71314
2003	65761	69333	68013	69209	66333
2004	72156	72296	67256	70171	72294
2005	68637	71247	67285	71465	69223
2006	69355	71458	69203	73150	69204
2007	75807	75788	76098	75258	74655

2008	78570	79249	78514	77761	77440
2009	80679	81977	79655	80561	80203
2010	80804	83173	82161	83488	81246
2011	86874	87615	86382	86339	87337
2012	94882	94466	90346	88932	94441
2013	93506	96192	93535	91147	91949
2014	95850	98445	91769	92941	95139
2015	86527	93432	85975	94331	87683
2016	87000	92056	87682	95372	87094
2017	98510	99045	98117	96132	98433
2018	99870	102568	100114	96676	100143
2019	103600	106355	104423	97061	103937
2020	107860	110576	108673	97331	107667
2021	109520	113292	108709	97518	109493

Table 2 shows that wheat production of actual values and forecasted values using different models, namely, ARIMA, RBF,MLP and fuzzy time series.

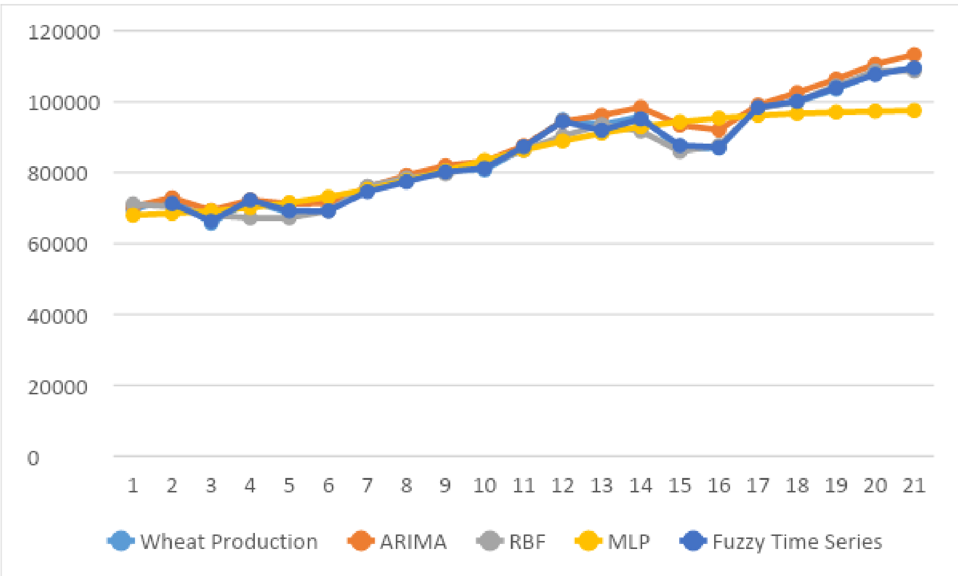

Figure 1 Shows that actual and forecasted values of wheat production of various models were compared by using line chart.

Figure 1: Actual and Forecasted Values of Wheat Production (tonnes)

Table 3: Residual Analysis by Different Models

Models	MAE	MAPE	RMSE
ARIMA	3453.9	4.188	4907.1
RBF	1361.19	1.668	1972.72
MLP	4033.857	4.5134	5171.001
Fuzzy Time Series	571.4	0.6976	733.569

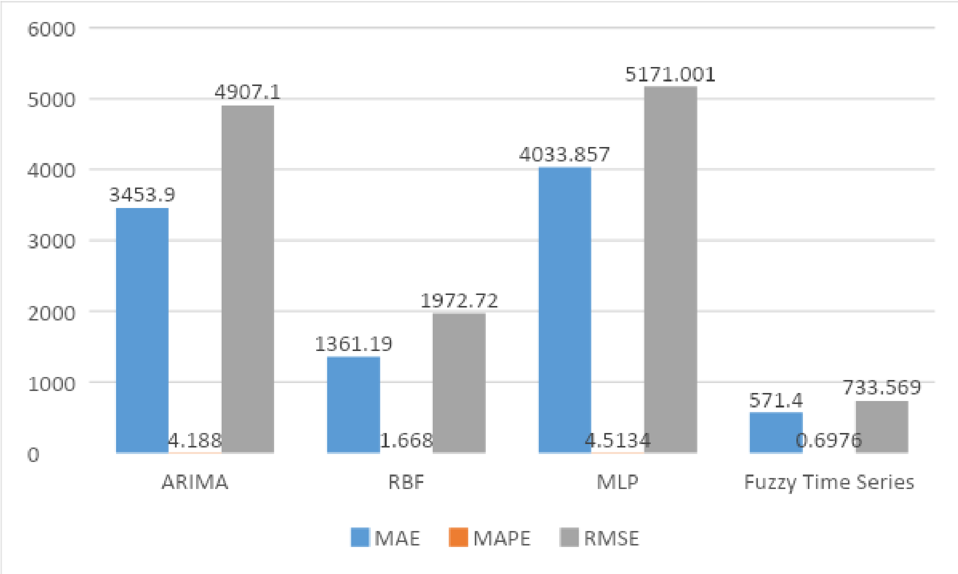

Figure 2: Residual Analysis by Using ARIMA, RBF, MLP and Fuzzy Time Series

Figure 2 shows that the mean absolute error, mean absolute percentage error and root mean square error obtained by using ARIMA and neural networks for Radial Basics Function and Multilayer Perceptron and fuzzy time series for wheat production prediction. Mean absolute error, mean absolute percentage error and root mean square error of fuzzy time series is less values when compared to ARIMA and neural networks. Mean absolute error, mean absolute percentage error and root mean square error show that the performance of fuzzy time series is better than that of ARIMA and neural networks.

III. CONCLUSION

In this work, three models, namely ARIMA, neural networks and fuzzy time series were used for wheat production prediction in India. Residual analysis for mean absolute error, mean absolute percentage error and root mean square error were compared using bar charts. Mean absolute error, mean absolute percentage error and root mean square error were minimum for fuzzy time series when compared to ARIMA and neural networks. Fuzzy time series is performed better than that of ARIMA and neural networks.

REFERENCES

1. Adesh Kumar Pandey, A.K Sinha, and V.K Srivastava(2008) : "A Comparative Study of Neural-Network & Fuzzy Time Series Forecasting Techniques - Case Study: Wheat Production Forecasting", International Journal of Computer Science and Network Security, vol.8,pp.382-387.

2. Alyousi, Y., K. Ibrahim, W. Kang, and W. Z. W. Zin(2019): “Markov Chain Modeling for Air Pollution Index Based on Maximum a Posteriori Method" , Air Qual., Atmos. Health, vol. 12, pp. 1521_1531.
3. Networks Based Model” , International Journal of Machine Learning & Cyber, vol.9,pp.491-506.
4. Ozge Cagcag Yolcu , Ufuk Yolcu , Erol Egrioglu, C. Hakan Aladag(2016): “ High Order Fuzzy Time Series Forecasting Method Based on an Intersection Operation”, Applied Mathematical Modelling, vol. 40,pp. 8750–8765.
5. Paarth Thadani(2021): “Financial Forecasting Using Stochastic Models: Reference From Multi-commodity Exchange of India”, Data Science in Finance and Economics, vol.3, pp.196-214.
6. Singh, P.(2018): “Rainfall and Financial Forecasting Using Fuzzy Time Series and Neural
7. Song. Q, B.S Chissom(1993):“Fuzzy Time Series and Its Models”, Fuzzy Sets and Systems, vol. 54, pp. 269-277.
8. Wang, P., H. Zhang, Z. Qin, and G. Zhang(2017): “A Novel Hybrid-GARCH Model Based on ARIMA and SVM for PM2:5 Concentrations Forecasting," Atmospheric Pollution Research, vol. 8, pp. 850-860.
9. Wangren Qiu, Ping Zhang, and Yan Hong Wang(2015):“Fuzzy Time Series Forecasting Model Based on Automatic Clustering Techniques and Generalized Fuzzy Logical Relationship” Mathematical Problems in Engineering ,vol.2, pp.1-7.
10. Yanpeng Zhang,Hua Qu,Weipeng Wang, and Jihong Zhao(2020): “ A Novel Fuzzy Time Series Forecasting Model Based on Multiple Linear Regression and Time Series Clustering”, Mathematical Problems in Engineering ,vol.20,pp.1-17.
11. Yousif Alyousifi, Mahmod Othman and Akram A. Almohammed(2021): “A Novel Stochastic Fuzzy Time Series Forecasting Model Based on a New Partition Method”, IEEE,vol 9,pp.80236-80252.