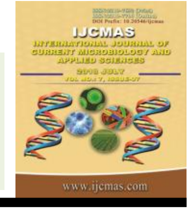
Int.J.Curr.Microbiol.App.Sci (2018) *7*(7): 1634-1642

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 7 Number 07 (2018)**
Journal homepage: http://www.ijcmas.com



Original Research Article https://doi.org/10.20546/ijcmas.2018.707.191

Forecasting of Nitrogen Content in the Soil by Hybrid Time Series Model

Neelam Chouksey*, G.C. Mishra, S.K. Singh, Munneswar Singh and Rajesh Chouksey

RMD College of Agriculture and Research Station, Ambikapur (C.G.), India
Institute of Agricultural Science, Banaras Hindu University, Varanasi (U.P.), India
Indian Institute of soil science, Bhopal (M.P.) and Krishi vigyan Kendra, surguja, India

**Corresponding author*

Keywords

ARIMA, ARIMAX, Hybrid model, Nitrogen content

Article Info

Accepted:
10 June 2018
Available Online:
10 July 2018

A B S T R A C T

There are several linear time-series forecasting models available in literature. One of the important and widely used technique for analysis of univariate time-series data is Box Jenkins' Autoregressive integrated moving average (ARIMA) methodology (Box et al., 2007). Sometimes addition of the other exogenous variables increases the prediction accuracy of ARIMA model (ARIMAX). For this aspect we applied different p and q order ARIMAX model for five nutrient combinations of nitrogen content which is further developed by including organic carbons an input (exogenous) variable. Among the linear models the ARIMAX model performed better as compare to ARIMA model. But the performance of machine intelligence techniques like Hybrid of linear and nonlinear model is better as compared to linear time series models. The variations in nitrogen content data for all treatments are large. This could be the reason that nonlinear machine learning techniques can capture the heterogeneous trend in the data set and performed well as compare to ARIMA and ARIMAX. Further the highest forecasted value by hybrid model for nitrogen content for the year 2018, was found to be 270.39 kg/ha by using 100%NPK+FYM treatment combination. On the basis of forecasted value we can say that the combination of recommended dose with farm yard manure might be useful as the best combination for establishing higher nitrogen in the soil.

Introduction

For the last three decades all over the developed world fingers have been raised on fertilizer, particularly nitrogen, as the number one enemy of sustainable agriculture (Nolan and Ruddy, 1997). Declining land fertility and crop productivity with negative balance is the main concerns against the food security problems in the country. Fertilization is one of the most notable measures that help to increase agricultural production, so application of adequate amount of mineral nutrient to crop and significant combination of fertilizer is one of the important factors for achieving higher productivity. Accurate estimation of nitrogen and forecasting for best fertilizer dose might be helpful for enhancing the productivity of crops.

There are several linear time-series forecasting models available in literature. One of the important and widely used technique for analysis of univariate time-series data is Box Jenkins' Autoregressive integrated moving average (ARIMA) methodology (Box et al., 2007). The ARIMA model is so popular due to its statistical properties. ARIMA is a flexible class of models including pure autoregressive (AR) models, pure moving average (MA) models, combined AR and MA (ARMA) models. Many researchers have been used ARIMA model to forecast. Mishra *et al.,* (2011) used ARIMA for the estimation of fertilizer for coming years. The ARIMA models were fitted for pre-harvest sugarcane yield estimation and the level of accuracy achieved by ARIMA model(s) with weather as input series was considered adequate for estimating the sugarcane yield(s) i.e. the ARIMAX models consistently showed the superiority over ARIMA models in capturing the percent relative deviations pertaining to sugarcane yield forecasts. (Sanjeev, Verma, Urmil, 2015). In addition to ARIMA, various exponential models can also be used to forecast a linear time-series process. One of the major limitations of these models is the pre-assumed linear form of the models. But if the data is nonlinear in the situation, performance of machine intelligence techniques like ANN and Hybrid of linear and nonlinear model is better as compared to linear time series models. Mitra and Paul (2017) was adopted hybrid methodology namely ARIMA-GARCH and ARIMA-ANN have been applied for modeling and forecasting of wholesale potato price in Agra market of India and observed that ARIMA-ANN hybrid model outperforms the other combinations and individual counterpart for the data under consideration.

Keeping the above points in view, the present study was undertaken to find better combination of nutrients on the basis of nitrogen content in the soil and found the best prediction model for forecasting of nitrogen content.

Materials and Methods

In the present study, yearly data from 1993-2017 on nitrogen content in the soil and organic carbon (kg/ha) was collected over the time from the long term fertilizer effect (LTFE) and trail data of IGKVV Raipur, Chhattisgarh. In the study time series analysis technique has been used by the help of past behavior of a time series data to make inferences about its future behavior.

There are four forecasting method used in nitrogen content in the soil. First method is ARIMA, next are ARIMAX, ANN modeling, and the error from ARIMAX model are used for hybrid ARIMAX-ANN modeling. The best model and method are selected based on the out-of-sample root mean squared error (RMSE).

Autoregressive Integrated Moving Average (ARIMA) Model

It is linear time series model. In the model to achieve greater flexibility in fitting of actual time-series data, include both autoregressive and moving average processes. This leads to the mixed autoregressive-moving average model. Let us denote the values of a process at equally spaced time epochs *t*, *t-1*, *t-2*, . . . by y_t, y_{t-1}, y_{t-2}, . . .

$$y = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + ...+ \varphi_p y_{t-p} + \varepsilon_1 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} + ... - \theta_q \varepsilon_{t-q} \quad .... (1)$$

where ϕ is autoregressive coefficient; θ is moving average coefficient and ε_t is white noise This is written as ARMA (*p*, *q*) model. In practice, it is frequently true that adequate representation of actually occurring stationary time-series can be obtained with autoregressive, moving average, or mixed models.

A generalization of ARMA models which incorporates a wide class of nonstationary time-series is obtained by introducing the differencing into the model. Box and Jenkins proposed a practical four-stage procedure for finding a good model which is denoted by ARIMA (*p*, *d*, *q*). The ARIMA methodology is carried out in four stages, viz. identification, estimation and diagnostic checking and forecasting. Parameters of the tentatively selected ARIMA model at the identification stage are estimated at the estimation stage and adequacy of tentatively selected model is tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration.

Autoregressive Integrated Moving Average with explanatory variable (ARIMAX) – transfer function model

ARIMAX model is an ARIMA model with the addition of exogenous variables. Organic carbon has been used as exogenous (independent) variable and there effects on the nitrogen content in the soil for different treatments for the period of 1993 to 2017 were studied by the ARIMAX models. The ARIMAX model written as the following model:

Assume two time series denoted Y_t and X_t, which are both stationary. Then, the ARIMAX model can be written as follows:

$$Y_t = C + V(B) X_t + N_t \qquad (2)$$

where: Yt is the output series (dependent variable), Xt is the input series (independent variable), C is constant term, N_t is the stochastic disturbance, i.e. the noise series of the system that is independent of the input series. $V(B)X_t$ is the transfer function (or impulse response function), which allows X to influence Y via a distributed lag. B is backshift operator, This ARMAX model is quite different from ARIMA model, because we work with two different series Xt and Yt , there are output series Yt is related to input series Xt. Then, we have ARIMAX model formula with maximum lag denoted by p (free-form distributed lag model):

$$y = C + v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + v_3 X_{t-3} + \dots\dots + \varphi_p y_{t-p} + \theta_q y_{t-q} + N_\tau \qquad \dots(3)$$

Where ϕ is autoregressive coefficient with p order and θ is moving average coefficient with q order. Construction of ARIMAX is similar iterative process as construction of univariate Box-Jenkins ARIMA model, i.e. identification, estimation and diagnostic checking.

Artificial neural network (ANN)

An artificial neural network (ANN) is an information-processing system that has certain performance characteristics in common with biological neural networks. Artificial neural networks have been developed as generalizations of mathematical models of human cognition or neural biology. A neural network is characterized by (1) its pattern of connections between the neurons (called its architecture), (2) its method of determining the weights on the connections (called its training, or learning, algorithm) and (3) its activation function (Fauset, 1994). The multilayer perception has a simpler structure and a much faster training process. The hidden layer performs a nonlinear transform of the input, and the output layer is a linear combiner mapping the nonlinearity into the new space. In mathematical terms, artificial neuron may be described by the following equations.

$$v_k = \sum_{i=0}^{m} w_{kj} \times x_{(j)} \qquad \dots(4)$$

$$y_k = \emptyset(v_k) \quad \ldots(5)$$

where $x_0 = 1$ and $x_1, x_2, \ldots, x_m$ are the input signals; $w_{k1}, w_{k2}, \ldots, w_{km}$ are the respective synaptic weights of neuron k; b_k is the bias; v_k is the "activation potential"; φ is the activation function; y_k is the output signal of the neuron.

Hybrid Model (ARIMAX –ANN)

Zhang (2003) proposed a hybrid approach that decomposes a time-series process into its linear and nonlinear component. The hybrid model considers the time-series y_t as a combination of both linear and nonlinear components.

That is, $$y_t = L_t + N_t \quad \ldots(6)$$

Where L_t and N_t represent the linear and nonlinear component present in the given data respectively. These two components are to be estimated from the data. This hybrid method of combining forecasting has following steps:

1) First, a linear time-series model, ARIMAX is fitted to the data.

2) At the next step residuals are obtained from the fitted linear model. The residuals will contain only the nonlinear components. Let e_t denotes the residual at the time t from the linear model, Then

$$e_t = y_t - \widehat{L_t} \quad \ldots(7)$$

Where $\widehat{L_t}$ is the forecast value for the time t from the estimated linear model.

3) Diagnosis of residuals is done to check if there is still linear correlation structures left in the residuals. The residuals are tested for nonlinearity by using BDS test.

4) Once the residuals confirm the nonlinearity, then the residuals modelled using a nonlinear model ANN.

5) Finally the forecasted linear and nonlinear components are combined to obtain the aggregated forecast values as

$$\widehat{y_t} = \widehat{L_t} + \widehat{N_t} \quad \ldots(8)$$

The graphical representation of hybrid methodology is expressed in following Figure 1.

Accuracy of models

Models are compared according to the minimum values of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and maximum value of Coefficient of determination (R^2).

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{X}_i - \bar{X}_i)^2}{\sum_{i=1}^{n} (X_i - \bar{X}_i)^2} \quad \ldots(9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{X}_i - \bar{X}_i)^2}{n}} \quad \ldots(10)$$

Results and Discussion

Time series plot (Fig. 2) shows that treatment NPK + FYM, performs consistently higher nitrogen content and increases at maximum time period than others. Control and 100% N Treatment gave the decreasing trend at maximum time period. Nitrogen content average are not similar between combination of fertilizer dosage (Table 1). However, much variation exists in the nitrogen content between different treatments and the pattern is not systematic among the all treatments. The summary or descriptive of treatment wise nitrogen content is given in Table 1, which indicated the data under consideration is heterogeneous figure 2 indicates the all most series are non stationary in nature.

After the evaluation of trend of every series, our next task was to forecast the series for the coming year and for this purpose ARIMA model was adopted in the time Series data on nitrogen content. Sometimes addition of the other exogenous variables increases the prediction accuracy of ARIMA model (ARIMAX). For this aspect we applied different p and q order ARIMAX model for five nutrient combinations of nitrogen content which is further developed by including organic carbon an input (exogenous) variable. In this study firstly we identified the model by ACF and PACF, which was non autocorrelated except control series. So there has been applied differencing in all nonautocorrelated series , thereafter we checked stationarity for all treatment series and observed that all the series were stationary in nature then its adequacy was checked by Ljung Box test and then it has been used to forecast the nitrogen content for different level of fertilizer dosage for coming years.

Table.1 Summary statistics of nitrogen content for different treatments

Treatment	Mean	Coefficient of variation	Skewness	Kurtosis
Control	189.12	14.42	0.06	-1.18
100%NPK	221	18.98	-1.03	0.24
100%NPK+Zn	215.37	20.85	-1.31	-0.64
100%NPK+FYM	229.23	17.79	-1.11	0.23
50%N+100%PK+GM	215.73	22.14	-1.77	1.08

Table.2 Comparison of forecasting performance of all models for nitrogen content time series

Treatments	Model	RMSE	R-squared
Control	ARIMA	11.86	0.71
	ARIMAX	11.35	0.75
	ANN	6.53	0.98
	ARIMAX- ANN	8.34	.52
100%NPK	ARIMA	11.54	0.66
	ARIMAX	11.27	0.63
	ANN	9.18	0.99
	ARIMAX- ANN	5.23	0.85
100%NPK+Zn	ARIMA	8.93	0.71
	ARIMAX	9.79	0.56
	ANN	6.34	0.92
	ARIMAX- ANN	5.39	0.71
100%NPK+FYM	ARIMA	12.83	0.05
	ARIMAX	14.37	0.50
	ANN	9.91	0.98
	ARIMAX- ANN	6.64	0.62
50%N+100%PK+GM	ARIMA	7.50	0.29
	ARIMAX	8.35	0.22
	ANN	5.83	0.99
	ARIMAX- ANN	4.60	0.86

Fig.1 Schematic representation of ARIMAX – ANN Hybrid methodology

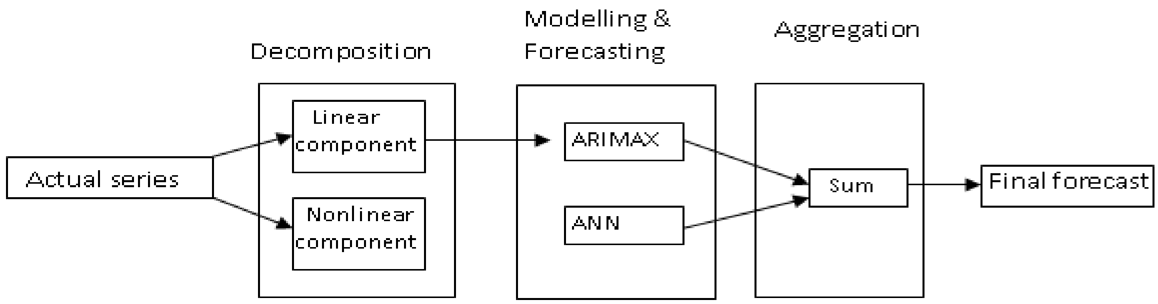


Fig.2 the Time series plot of nitrogen content for different treatments

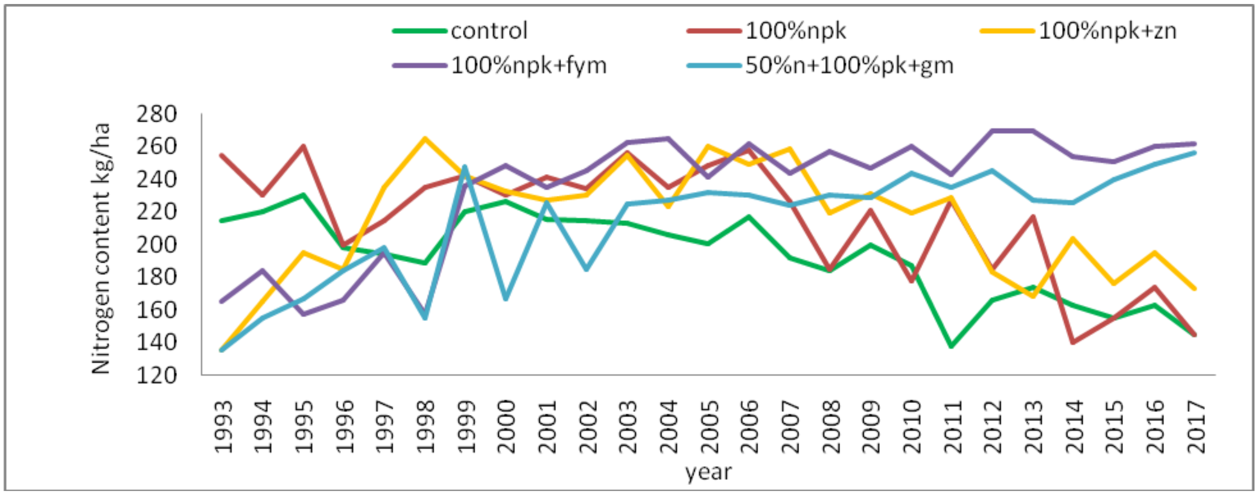


Fig.3 Time series plot for the control series

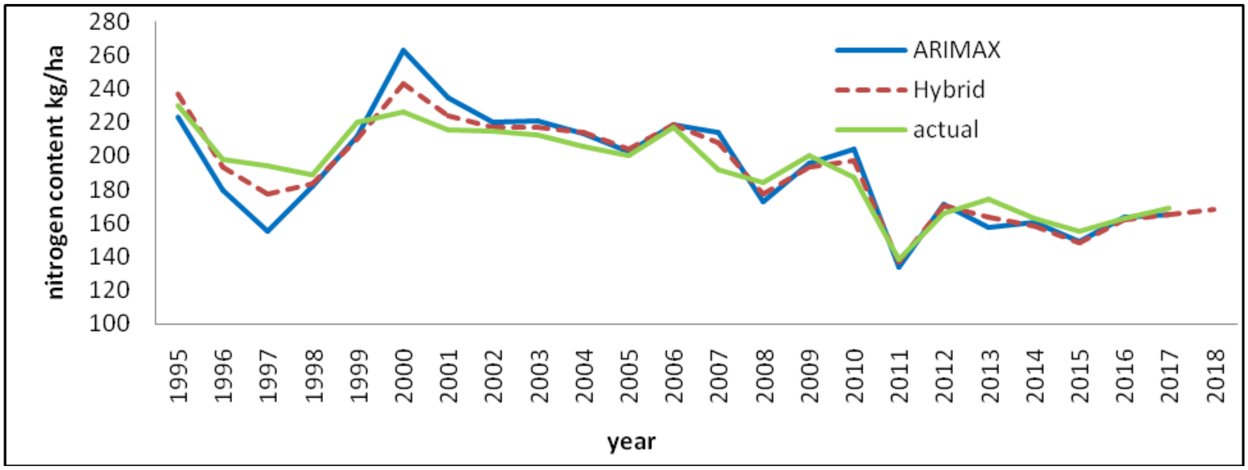


Fig.4 Time series plot for the 100%NPK series

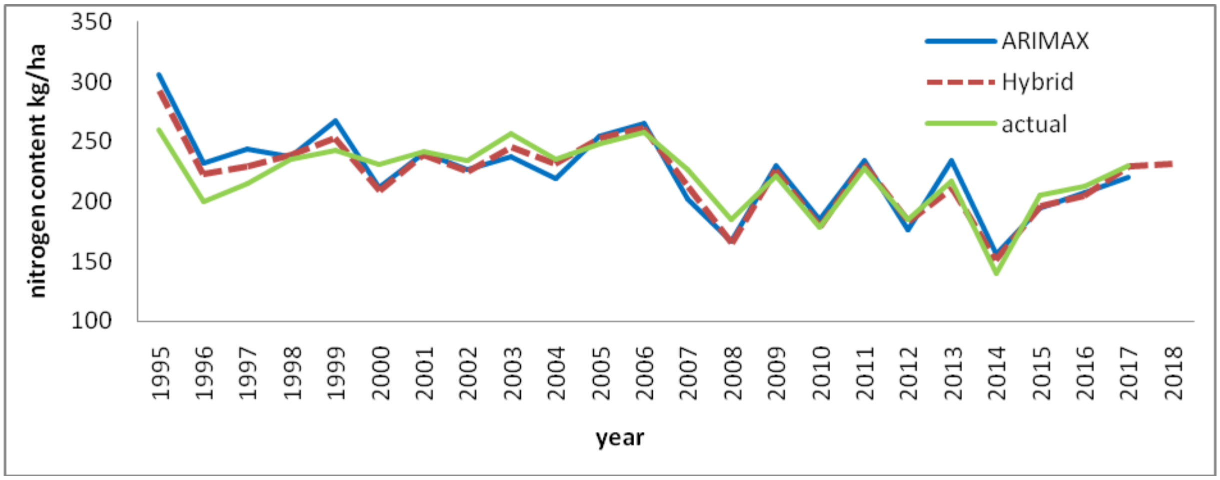


Fig.5 Time series plot for the 100%NPK+Zn series

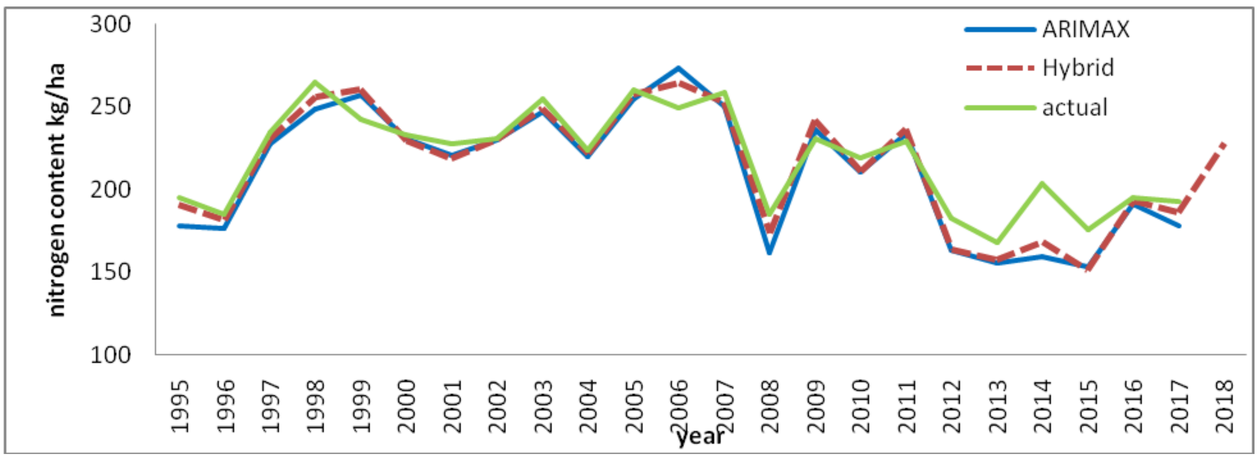


Fig.6 Time series plot for the 100%NPK+FYM series

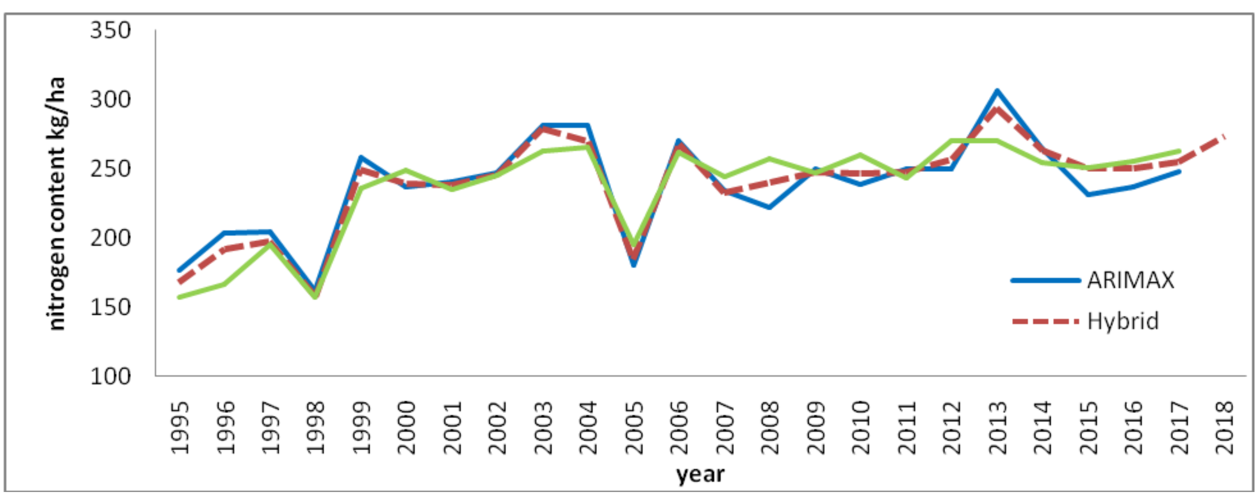
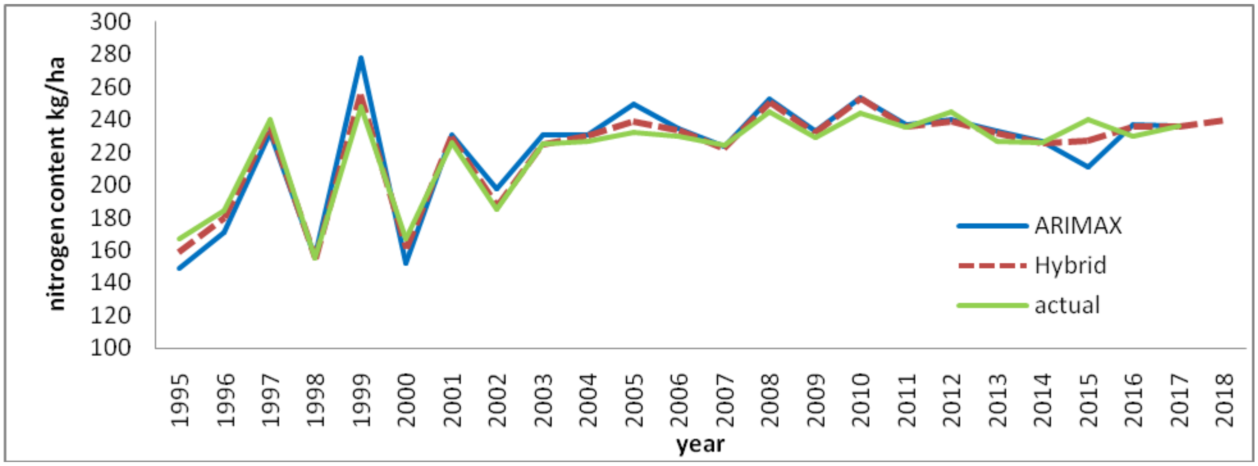


Fig.7 Time series plot for the 50N+100%PK+GM series



different p and q ARMA & ARMAX model fitted and appropriate model was chosen corresponding to minimum value of selection criteria i.e. lower RMSE along with higher R^2 for different treatment of nutrient combination.

A multilayer perception neural network was fitted to the data with the help of ANN out of sample forecast with trail version of STATISTICA software. Sigmoid and linear functions were used as activation function in hidden and output layers respectively. The Levenberg-Marquardt (LM) back propagation algorithm was used for ANN model building and based on repetitive experimentation. More than 80 percent of the observations in data set are used for model training and remaining data are used for and testing or validation. Different numbers of neural network models are tried before arriving at the final structure of the model.

As discussed hybrid methodology in above methodology section, first step is to test the nonlinearity of the residuals by ARIMAX residuals of nitrogen content on different treatment (control, 100%NPK, 100% NPK+Zn, 100%NPK+ FYM, 50%N+ 100% PK + GM) which was nonlinear in nature. Further a statistical test *i.e.* (Dechert-Scheinkman test) BDS test was applied to test the nonlinearity of the residuals, the all residual series has been found to be nonlinear then, it has been modelled and predicted using nonlinear models. The nonlinear models namely ANN is used for modeling. For the hybrid model the ARIMAX residuals predicted from ANN was combined with forecast obtained from ARIMAX model.

For nitrogen content time series of Raipur district of Chhattisgarh, the above mentioned model has been fitted for modeling and forecasting performance has been assessed in terms of their prediction ability measured by model errors.

Among the linear models the ARIMAX model performed better as compare to ARIMA model (Table 2). Comparison of model performance has been done on the basis of lowest RMSE. The reason for better performance of ARIMAX model may be due to consideration of exogenous variables in ARIMA time series model.

The performance of machine intelligence techniques like ANN and Hybrid of linear and nonlinear model is better as compared to linear time series models (Table 2).

The variations in nitrogen content data for all treatments are large (Table 1) this could be the reason that nonlinear machine learning techniques can capture the heterogeneous trend in the data set and performed well as compare to ARIMA and ARIMAX.

The actual and predicted values has been depicted in the figure 3,4,5,6 and 7 for control, 100% NPK, 100%NPK+Zn, 100%NPK+FYM and 50%N+100%PK+GM treatments respectively.

In conclusion, among the linear models the ARIMAX model performed better as compare to ARIMA model. The reason for better performance of ARIMAX model may be due to consideration of exogenous variables in ARIMA time series model. The performance of machine intelligence techniques like ANN and Hybrid of linear and nonlinear model is better as compared to linear time series models. The variations in nitrogen content data for all treatments are large (Table 1) this could be the reason that nonlinear machine learning techniques can capture the heterogeneous trend in the data set and performed well as compare to ARIMA and ARIMAX.

Further the highest forecasted value by hybrid model for nitrogen content for the year 2018, was found to be 270.39 kg/ha by using 100%NPK+FYM treatment combination. On the basis of forecasted value we can say that the combination of recommended dose with farm yard manure might be useful as the best combination for establishing higher nitrogen in the soil.

References

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Fransisco.

E. Ogasawara, D. Oliveira, (2013). ARIMA model for forecasting of fertilizer consumption in Brazil.,*Inter. Journal of Agricultural & emvironmetal information system*,4(2): 23-36.

Prabakaran, K., C. Sivapragasam, C. Jeevapriya and Narmatha, A. (2013). Forecasting Cultivated Areas And Production Of Wheat In India Using ARIMA Model, *Golden Research Thoughts*, Volume-3, Issue-3: 2-7.

Kumari Prity, Mishra, G. C., Pant, A.K., Shukla, G. and Kujur, S.N. (2014). Autoregressive Integrated Moving Average (ARIMA) approach for prediction of rice (*Oryza sativa l.*) yield in India. *The Bioscan*, 9(3): 1063-1066.

Fauset, L. (1994). Fundamental of Neural Network: Architectures, Algorithm, and Applications, *New Jersey: Prentice Hall Inc.*

Mitra, D. and Paul, R.K.,(2017). Hybrid time-series models for forecasting agricultural commodity prices, *Model Assisted Statistics and Applications*, vol. 12: 255–264

Nolan, B. T., Ruddy, B. C., Hill, K. J. and Helsel, (1997). D. R., *Environ. Sci. Technol.*, 31, 2229–2236.

Mishra, P., PK Sahu, JPS Uday, (2014). ARIMA Modeling Technique in Analyzing and Forecasting, *Fertilizer Statistics in India*,vol.7,issue 13:170-176.

Sanjeev, Verma, Urmil, (2015).Time-series modeling and forecasting of sugarcane yield in Haryana, *M. Sc. Dissertations*, Chaudhary Charan Singh Haryana Agricultural University.

How to cite this article:

Neelam Chouksey, G.C. Mishra, S.K. Singh, Munneswar Singh and Rajesh Chouksey. 2018. Forecasting of Nitrogen Content in the Soil by Hybrid Time Series Model. *Int.J.Curr.Microbiol.App.Sci.* 7(07): 1634-1642. doi: https://doi.org/10.20546/ijcmas.2018.707.191