International Journal of Ecology and Environmental Sciences
www.ecologyjournal.in
Online ISSN: 2664-7133; Print ISSN: 2664-7125; Impact Factor: RJIF 5.28
Received: 18-08-2020; Accepted: 03-09-2020; ; Published: 19-09-2020
Volume 2; Issue 3; 2020; Page No. 274-277

Hybrid time series model for forecasting of nitrogen uptake in rice

Neelam Chouksey[1*], GC Mishra[2], Rajesh Chouksey[3]
[1] RMD College of Agriculture & Res. Sta. Ambikapur, Chhattisgarh, India
[2] Banaras Hindu University, Varanasi, Uttar Pradesh, India
[3] Krishi Vigyan Kendra, Surguja, Chhattisgarh, India

Abstract
There are available several linear time-series forecasting models in the literature. In which important and most comman technique for analysis of univariate time-series is 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. However, 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 uptake for rice crop data for all treatments are large. This could be the reason that Hybrid of linear and nonlinear model found heterogeneous trend in the data set and performed well as compare to ARIMA and ARIMAX. Further the comparision of the forecasted values by hybrid model of different treatments (control, 100% NPK, 100% NPK+ Zn, 100% NPK+ FYM, 50% N+100% PK+ GM) result has been concluded that 100 % NPK+FYM treatment gave highest forecasted value in comparison to other treatments.

Keywords: ARIMA, ARIMAX, hybrid model, nitrogen uptake

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, B. T., Ruddy, B. C., 1997). In our country, Declining land fertility and crop productivity with negative effects is the main problem against the food security and sustainable agriculture. Fertilization is also one of the most important measures that can help to enhance agricultural production, so application of optimum quantity of mineral nutrients to crop and batter combination of fertilizer are also important factors to achieving higher crop production. Accurate forecasting of nitrogen uptake for best fertilizer dose and selection of best variety might be helpful for enhancing the productivity of crops. In the literature various linear time-series forecasting models are available. In which, Box Jenkins' Autoregressive integrated moving average (ARIMA) methodology is a important and most widely used model for analysis of time-series data. Popularity of ARIMA model is may be due to some important statistical properties. ARIMA is a very flexible model for pure autoregressive (AR), pure moving average (MA), and combination of AR & MA. Many researchers have been used ARIMA model to forecast. Mishra *et al.,* (2011) [5] used ARIMA for the estimation of fertilizer for coming years. 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. Sanjeev *et al*. (2015) [10] used ARIMA and ARIMAX models for pre-harvest sugarcane yield estimation and weather parameters was used as input series in ARIMAX models. After copmparision, ARIMAX models perform superior over ARIMA models for yield of sugarcane forecasts. However, 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) [7] was adopted hybrid techniques for modeling and forecasting to wholesale potato price in Agra of India and found that ARIMA & ANN hybrid model was superior to the other combinations of model. Khashei & Bijari (2011) [4] identified a new technique with combination of ARIMA and ANN models. The proposed model was effective in capturing both linear and nonlinear trends. Results revealed that combination of models perform better than individual models.
Keeping with the above point of view, the present study carried out to find better combination of nutrients for nitrogen uptake by rice crop and developed the best prediction model for forecasting of nitrogen uptake by rice.

Material and Methods
In the present study, yearly data from 1993-2017 on nitrogen uptake by rice crop (kg/ha), nitrogen content in the soil (kg/ha) and organic carbon (kg/ha) was collected over the time from IGKVV Raipur, Chhattisgarh. In the study, time series analysis technique has been used by the help of past behavior of a time based data to make inferences about its future behavior. Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) model have been applied for different treatments viz. control, 100% NPK, 100% NPK + Zn, 100% NPK + FYM and 50%N+100% PK +GM. There nitrogen content in

the soil and organic carbon used as explanatory variables in the ARIMAX model for nitrogen uptake by rice. In the nonlinear techniques, ANN and Hybrid (ARIMAX +ANN) models have been used on all treatments for nitrogen uptake.

Autoregressive Integrated Moving Average (Arima) Model

It is linear time series model. By achieving greater flexibility in model fit the actual time series with both autoregressive and moving average processes. Suppose to the values equally spaced time epochs are *t*, *t -1*, *t -2*, by y_t, $y_{t\square 1}$, $y_{t\square 2}$, y $=\varphi_1 \psi_{\tau-1} + \varphi_2 \psi_{\tau-2} + \varphi_3 \psi_{\tau-3} + \ldots + \varphi_\pi \psi_{\tau-\pi} + \varepsilon_1 - \theta_1 \varepsilon_{\tau-1} - \theta_2 \varepsilon_{\tau-2} - \theta_3 \varepsilon_{\tau-3} + \ldots\ldots - \theta\theta \varepsilon_{\tau-\theta} \ldots\ldots$ (1)

Where, ϕ is autoregressive coefficient, θ is moving average coefficient and ε_t is white noise. This is written by ARMA (p,q) model. ARIMA model is used for nonstationary series because it is a combination of autoregressive (AR), integration (I) and moving average (MA). An ARIMA model is usually stated as ARIMA (p, d, q). Box and Jenkins (1976) [1] proposed a practical four-stage procedure for finding a good model. The ARIMA model is carried out by 4 steps, identification, estimation and diagnostic checking and forecasting.

Autoregressive Integrated Moving Average with Explanatory Variable (Arimax) – Transfer Function Model

"There is not possible to express the multivariate time series changing rules using the ARIMA model with only one time series. The reason is that the only ARIMA model is imperfect. Therefore, it is necessary to create a model with multivariate ARIMAX model." (Fan *et al*. 2009; Chadsuthi *et al*. 2012).

Let there is two stationary time series Y_t and X_t, Then the model can be written as:

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

"Where Y_t is the output series (nitrogen uptake), X_t is the input series (nitrogen content and organic carbon), C is constant term and N_t is the stochastic disturbance. V (B) X_t is the transfer function, which allows *X* to influence *Y* via a distributed lag and *B* is backshift operator."

Then, we have ARIMAX model formula with maximum lag (p)

$Y = X + \varpi_0 \Xi_\tau + \varpi_1 \Xi_{\tau-1} + \varpi_2 \Xi_{\tau-2} + \varpi_3 \Xi_{\tau-3} + \ldots\ldots + \varphi_\pi \psi_{\tau-\pi} + (_\theta \psi_{\tau-\theta} + N_\tau$ (3)

Where ϕ autoregressive coefficient with p order and θ is is moving average coefficient with q order. ARIMAX is similar as construction of ARIMA model (identification, estimation and diagnostic checking).

Artificial neural network (ANN)

"A neural network is characterized by (1) It is make 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). In ANN, multilayer perception is a simple structure and a fast training process. To model time series data (TSD) using ANN, y_t is expressed as a non-linear function f of y_{t-1}, y_{t-A}, where A is the lag till which the TSD points are involved in prediction. The model equation is

$$Y_t = g(y_{t-1}, y_{t-2}, y_{t-A}) + v_t \quad (4)$$

Where v_t is the noise or error term. The Transfer Function (TF) of the hidden layer can be a linear, sigmoid, tan-sigmoid or log-sigmoid in nature.

The sigmoidal (logistic) function

$$f(y) = \frac{1}{1+e^{-y}} \quad (5)$$

Hybrid Model (ARIMAX & ANN)

Zhang (2003) developed a hybrid time series approach is a combination of linear and nonlinear components. We can write hybrid model as

$$y_t = L_t + N_t \quad (6)$$

Where L_t is linear and N_t is nonlinear component. Both the two component are to be estimated for time series. For developing hybrid model, steps are as follows:

1. ARIMAX model is be fitted in the time series data.
2. In the next step obtained residual from the ARIMAX model suppose e_t denotes residual of linear time series model.

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

Where $\widehat{L_t}$ = forecast value of linear time series

y_t = observed value time series

3. Linearity of residual is to be checked by BDS (Brock- Dechert-Scheinkman) test.
4. Once the residuals are confirmed to the nonlinear nature, then it is used as a nonlinear series in the model ANN.
5. At the last stage combine the linear and non-linear component.

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

It is easily understand by graphical representation of hybrid model techniques. (Fig.1)

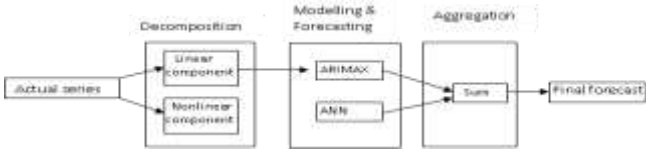

Fig 1: graphical representation of hybrid model (ARIMAX + ANN).

Accuracy of Models

Models are comparing according to the lowest valest of root mean square error (RMSE), highest value of coefficient of determination (R^2).

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

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

Results and Discussion

Time series plot (fig.2) shows that treatment NPK + FYM, performs consistently higher nitrogen uptake and increases at maximum time than others. Control and 100% N Treatment gave the decreasing trend. There was much variation exists in nitrogen uptake for different treatments and the pattern have been not systematic for all time. Fig.2 indicated that the data under consideration is heterogeneous. Hence, stationarity has been checked and found that all most series has been non-stationary in nature.

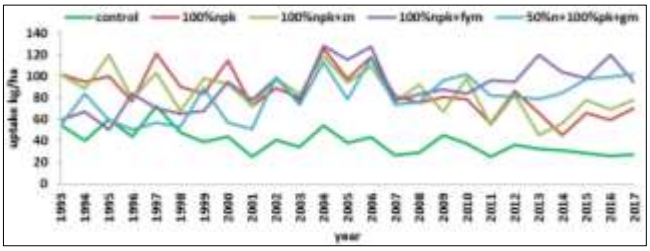

Fig 2: the Time series plot of nitrogen uptake for different treatments.

After the evaluation of trend of every series, our next task was to select model to forecasting the data for the coming years. On the basis of linear time series, ARIMA was adopted for the forecasting of nitrogen uptake. Sometimes addition of the other exogenous variables increases the prediction accuracy of ARIMA model (ARIMAX). For this aspect, there applied different p and q order ARIMAX model for five nutrient combinations of nitrogen uptake, which has been further developed by including nitrogen content and organic carbon as input (exogenous) variables. Different p and q ARMA & ARMAX model fitted and best model was chosen according to lowest value of RMSE along with highest value of R^2 for different treatment of nutrient combination.

ANN fitted to the time series data with the help of multilayer perception neural network. Sigmoid and linear functions used as activation function in hidden layers. The Levenberg Marquardt back propagation algorithm has been used for ANN model building. More than 80 percent of the observations in data set have been taken for model training and remaining data for testing and validation.

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

Table 1: Comparison of forecasting performance of all models for nitrogen uptake time series

Treatment	Model	R-squared	RMSE	Forecasted value
Control	ARIMA	0.309	7.48	29.41
	ARIMAX	0.602	6.28	29.09
	ARIMAX -ANN	0.977	5.26	36.90
100%NPK	ARIMA	0.805	11.93	73.94
	ARIMAX	0.855	11.15	69.47
	ARIMAX - ANN	0.995	7.48	76.60
NPK+Zn	ARIMA	0.603	12.71	68.51
	ARIMAX	0.746	11.39	77.19
	ARIMAX - ANN	0.969	4.6	78.80
NPK+FYM	ARIMA	0.623	16.35	99.00
	ARIMAX	0.660	12.21	98.79
	ARIMAX - ANN	0.9422	9.8	103.80
50%N+100%PK+GM	ARIMA	0.583	14.74	88.61
	ARIMAX	0.712	13.12	84.37
	ARIMAX - ANN	0.980	5.70	89.29

- 5 tones /ha farm yard manure (FYM) and 25 kg/ha green manures (GM) has been applied in the experimental field with combined recommended doses.

The above-mentioned model fitted to the nitrogen uptake time series for modeling and forecasting and performance 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 1). Comparisons of models have to be done based on lowest RMSE and higher R square. The reason for better performance of ARIMAX model may be attributed to consideration of exogenous variables in ARIMA time series model. The performance of Hybrid of linear and nonlinear model is better as Compared to linear time series models. The variation in nitrogen uptake data is 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.

Finally in other words among all models studied, ARIMAX-ANN model has been performed superior. Hybrid methodology considers both linearity and nonlinearity of the data set; hence, the performance of ARIMAX-ANN model was superior as compare to all single models for modeling and forecasting of nitrogen uptake in Raipur district of Chhattisgarh. Actual and predicted values by ARIMAX and Hybrid models has been depicted in fig 3 to fig. 7

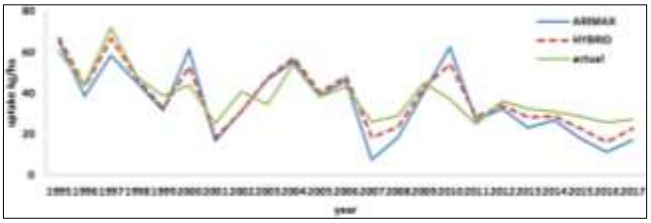

Fig 3: Time series plot for the control series

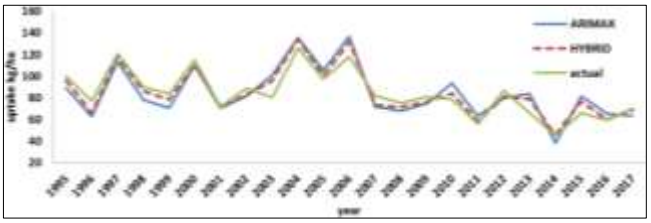

Fig 4: Time series plot for the 100%NPK series

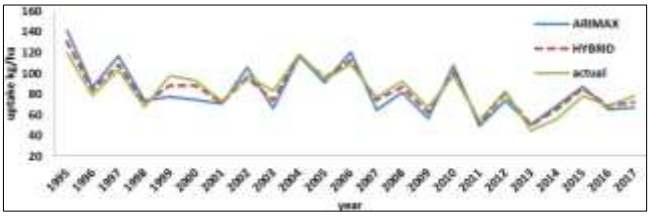

Fig 5: Time series plot for the 100%NPK+Zn series

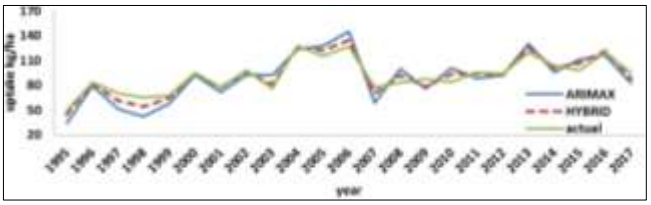

Fig 6: Time series plot for the 100%NPK+FYM series

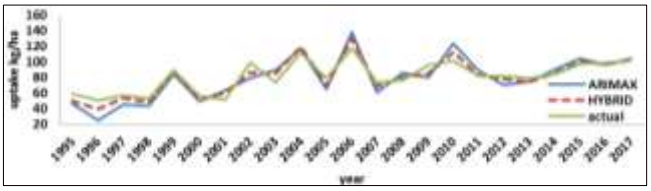

Fig 7: Time series plot for the 50N+100%PK+GM series

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. However, the performance of hybrid of linear and nonlinear model has outperformed as compared to linear time series models alone (ARIMA & ARIMAX). After comparision, the forecasted values of different treatments (control, 100% NPK, 100% NPK+ Zn, 100% NPK+ FYM, 50% N+100% PK+GM) result as been to concluded that 100 % NPK+FYM treatment gave highest forecasted value of nitrogen uptake by rice in comparison to other treatments. It means the combination of recomanded doses of fertilizer with farm yard manure (FYM) gave the highest nitrogen uptake. Result may be useful for diciding and selecting the best combinatin of treatment for imprving the yield of rice crop.

References

1. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control. Holden-Day, San Fransisco, 1976.
2. Ogasawara E, Oliveira D. ARIMA model for forecasting of fertilizer consumption in Brazil. Inter. Journal of Agricultural & environmental information system. 2013; 4(2):23-36.
3. Prabakaran K, Sivapragasam C, Jeevapriya C, Narmatha A. Forecasting Cultivated Areas And Production Of Wheat In India Using ARIMA Model, Golden Research Thoughts, ISSN 2231-5063. 2013; 3(3):2-7.
4. Khashei M, Bijari M. A new hybrid methodology for nonlinear time series forecasting. Modelling and Simulation in Engineering, Article ID 379121, 2011, 5.
5. Kumari Prity, Mishra GC, Pant AK, Shukla G, Kujur SN. Autoregressive Integrated Moving Average (ARIMA) approach for prediction of rice (Oryza sativa l.) yield in India. The Bioscan. 2014; 9(3):1063-1066.
6. Fauset L. Fundamental of Neural Network: Architectures, Algorithm, and Applications, New Jersey: Prentice Hall Inc, 1994.
7. Mitra D, Paul RK. Hybrid time-series models for forecasting agricultural commodity prices, Model Assisted Statistics and Applications, 2017; 12:255-264.
8. Nolan BT, Ruddy BC, Hill KJ, Helsel. DR., Environ. Sci. Technol, 1997; 31:2229-2236.
9. Mishra P, Sahu PK, Uday JPS. ARIMA Modeling Technique in Analyzing and Forecasting, Fertilizer Statistics in India. 2014; 7(13):170-176.
10. Sanjeev Verma, Urmil. Time-series modeling and forecasting of sugarcane yield in Haryana, M. Sc. Dissertations, Chaudhary Charan Singh Haryana Agricultural University, 2015.