RASHI 1 (1) :77-85 (2016)

FORECASTING CROP YIELD : A COMPARITVE ASSESSMENT OF ARIMAX AND NARX MODEL

Ranjit Kumar Paul and Kanchan Sinha

ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012

ABSTRACT

Weather variability within and between seasons is uncontrollable source of variability in yields. The extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season. Therefore, when forecasting is carried out for dynamic behaviour of crop yield, it should be able to take advantage not only of historical data of crop yield, but also of the impact of various driving forces from the external environment. In the present investigation, an attempt has been made to forecast wheat yield at Kanpur district of Uttar Pradesh by considering most important weather variable i.e. maximum temperature at Critical Root Initiation (CRI) stage of wheat crop which comes around 21 days after sowing of the crop. Both parametric (ARIMA model) and nonparametric approach (NARX model) have been employed. It is observed that NARX model outperformed the ARIMAX model as far as modelling and forecasting is concerned. Besides Mean absolute prediction error (MAPE), Relative MAPE (RMAPE) and Root mean square errors (RMSE), Diebold-Mariano test has also been employed to compare the predictive accuracy of two competing models.

Keywords **:** ARIMAX, Forecasting, NARX, Max temperature

1. Introduction

A reliable forecast of crop yield forms a basis for its policy decision in regards to marketing of agricultural commodities. Crop yield forecasting plays an important role in farming planning and management, domestic food supply, international food trade, ecosystem sustainability, and so on (Guo and Xue, 2014). Under the changing scenario, forecasting of various aspects relating to agriculture are becoming essential. But in-spite of its strong need, the current status is far from satisfaction. Weather variability both within and between seasons is uncontrollable source of variability in yields. Weather variables affect the crop differently during different stages of development. Thus extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season. There is need to develop statistically sound objective forecasts of crop yield based on weather variables so that reliable forecasts can be obtained. If wheat yield and wheat quality response to weather conditions could be predicted early and accurately, the information could be widely used. The information could be particularly important to farmers optimizing late season agronomic and marketing decisions and to grain elevators and millers for purchasing decisions. In the statistical modeling framework, a lot of simplifications is embodied in sacrifice for the feasibility and efficiency. The most investigated statistical crop-yield-weather models are multiple regression models (Alexandrova and Hoogenboom, 2000; Prasada *et al.,* 2006; Yu, 2011). However, considering the inherent and irreparable disadvantages of the multiple regression model, a more scientific methodology to incorporate weather data into crop yield models, is still under exploration, and remains of great importance to government, and private sector insurers, and reinsurers. A widely used approach to crop yield prediction is to rely on numerical models that emulate the main processes of crop growth and development. These models are typically developed and tested using experimental trials. But the main disadvantage of these models is that they require extensive input data on cultivar, management, and soil conditions that are unavailable in many parts of the world. More significantly, even in the presence of such data these models can be very difficult to calibrate because of a large numbers of uncertain parameters.

Email: ranjitstat@gmail.com

Kumar *et al.* (2001) studied the effect of different weather variables on wheat yield and found that maximum temperature was negatively correlated with yield of late sown wheat in Tarai region. Lobell *et al.* (2006, 2010) developed weather based yield forecast model for 12 California crops. The authors combined weather and yield data in a linear regression model to test how well yield anomalies could be predicted before harvest based on monthly weather measurements. But the authors did not take care of the time-series behaviour of the data. Maximum temperature plays a very important role in wheat crop development and wheat yield (Pathak and Wassmann, 2009). Asseng *et al.* (2011) reported that the effect of temperature on wheat production is underestimated. The authors observed that variations in average growing-season temperatures of ±2 °C in the main wheat growing regions of Australia can cause reductions in grain production of up to 50%. Kaur *et al.* (2012) reported that the effect of maximum temperature on wheat yield is more negative than that of the minimum temperature.

For the purpose of crop yield forecasting, the autoregressive integrated moving average (ARIMA) model has been widely used in past. But this model cannot incorporate exogenous variable. Hence, Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) model is preferred over ARIMA in order to forecast the crop yield more accurately. Paul *et al.* (2013) developedfive models at five important stages of wheat growth by including the most important weather variables for forecasting the pre-harvest wheat yield of the Kanpur district of Uttar Pradesh. They showed that, as wheat crop grows towards maturity; pre-harvest forecasts get closer to actual values. Paul *et al.* (2014, 2015) have developed some advanced models for forecasting the volatile crop yield. In recent years, artificial neural networks (ANN) has been developed as a powerful nonlinear model for modeling and forecasting of crop yield. The ANN modeling form depends on the available data with little a priori rationalization between variables and about the models (Zhang *et al.,* 1998; Cheng and Titterington, 1994). Laxmi and Kumar (2011) applied ANN for forecasting yield of rice, wheat and sugarcane in different zones in India using ANN approach.

More recently, the architectural approach proposed to deal with chaotic time series is one based upon Nonlinear Autoregressive models with eXogenous input (NARX model), which are therefore called NARX recurrent neural networks (Haykin, 1999; Lin *et al.,* 1996; Gao and Er, 2005). This is a powerful class of models which has been demonstrated that they are well suited for modeling nonlinear systems and specially time series. One principal application of NARX dynamic neural networks is in control systems. In the NARX networks learning is more effective in than in other neural network and these networks converge much faster and generalize better than other networks (Lin *et al.,* 1996; Gao and Er, 2005). Guo and Xue (2014) discussed the latest research outcomes from using both the spatial and temporal neural network models in crop forecasting. In the present investigation an attempt has been made to apply both ARIMAX model and NARX model for forecasting of wheat yield in Kanpur district of Uttar Pradesh by including important weather variable.

2. MATERIALS AND METHODS

The ARIMAX model (Bierens, 1987) is a generalization of the ARIMA model, which is capable of incorporating an external input variable (X). Given a (k+1) time-series process $\{(y_t, x_t)\}$ where y_t and k-components of x_t are real valued random variables, the ARIMAX model assumes the form

$$\left(1-\sum_{s=1}^{p}\alpha_s L^s\right)\Delta y_t = \mu + \times\sum_{s=1}^{q}\beta_s' L^s x_t + \left(1+\sum_{s=1}^{r}\gamma_s L^s\right)e_t$$

where L is the usual lag operator, *i.e.* $L^s y_t = y_{t-s}$

$\Delta y_t = y_t - y_{t-1}$ $\mu \in R, \alpha_s \in R, \beta_s \in R^k$ and $\gamma_s \in R$ are the unknown parameters and e_t,s are the errors, and p, q and r are natural numbers specified in advance.

The first step in building an ARIMAX model consists of identifying a suitable ARIMA model for the endogenous variable. The ARIMAX model concept requires testing of stationarity of exogenous variable before modelling. The transformed variable is added to the ARIMA model in the second step, in which the lag length r is also estimated. Nonlinear least squares estimation procedure is employed to estimate the parameters of ARIMAX model (Bierens, 1987). Fortunately, the ARIMAX model can be fitted to data by using a software package, like SAS, MATLAB, EViews and R. In the present investigation, SAS, Version 9.3 is used for data analysis.

NARX Model :

Neural networks are considered as a class of generalized non-linear, nonparametric, data driven statistical methods. A general neural network consists of an input layer that accept external information, one or more hidden layer that provide non-linearity to the model and an output layer that provides the target value. Each layer consists of one or more nodes. All the layers are connected through acyclic arc. Each input node in the input layer is associated with its corresponding weight. To compute the output, its activation function is applied to the weighted sum of the inputs. The activation function is either the identity function or sigmoidal function.

In time series analysis an extension of feed forward neural network *i.e.,* time delay neural network (TDNN) is more commonly used. In time delay neural network, the inputs are the d previous lagged observations $(y_t, y_{t-1}, \ldots, y_{t-d})$ of a time series and output is the predicted future observation $\hat{y}_{t+1}$ Time delay neural network can be trained with Levenberg Marquardt back propagation algorithm and the weights are accumulated across all samples and updated once per epoch. Unlike feed forward neural network, recurrent neural networks contain cycle. The cycles are formed by time delay connections which carry values between successive activations. One of the most promising recurrent neural network for time series analysis is nonlinear autoregressive eXogenous (NARX) model. Nonlinear autoregressive with external input is a modified nonlinear autoregressive model by including another relevant time series as extra input to the forecasting model. The model can be written as:

$$y_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-d+1}, y_t, y_{t-1}, \dots, y_{t-d+1})$$

which may be written in vector form as

$$y_{t+1} = f(\boldsymbol{x_t}, \boldsymbol{y_t})$$

Where, $\boldsymbol{x_t}$ is the external input to the forecasting model with the same number of time delays as $\boldsymbol{y_t}$. Here, two known time series are used as independent inputs to the hidden layer according to the same number of delay. The nonlinear mapping f(.) is generally unknown and can be approximated through the architecture of NARX dynamic neural network model.

The NARX neural network can be expressed as

$$y_t = \sum_i c_i \psi\left(\sum_{j=1}^{d}\left(a_{ji}x_{t-j} + b_{ji}y_{t-j}\right)\right)$$

Where, ψ is the activation function in the hidden layer ; a_{ji} and b_{ji} are the input to hidden layer weights at the hidden neuron *j;* and $\boldsymbol{C_i}$ is the hidden to output layer weight, d is number of input nodes (tapped delays)

Learning Algorithm of NARX model

The error surface of dynamic neural network such as NARX is like to be confined in local minima. A dynamic back propagation neural network is required to compute the gradients, which is more computationally rigorous than static BP. In this context, Levenberg-Marquardt back propagation algorithm has been utilized to minimize the error as well as weights of NARX model. This combines with the advantages of the simple gradient descent and Newton's method algorithm, and has rapid convergence and robust performance. The schematic representation of NARX model is presented in figure 1.

3. An Illustration

Dataset

As an illustration, annual wheat yield data of Kanpur district of Uttar Pradesh during 1972 to 2013 comprising 42 data points are obtained from Directorate of

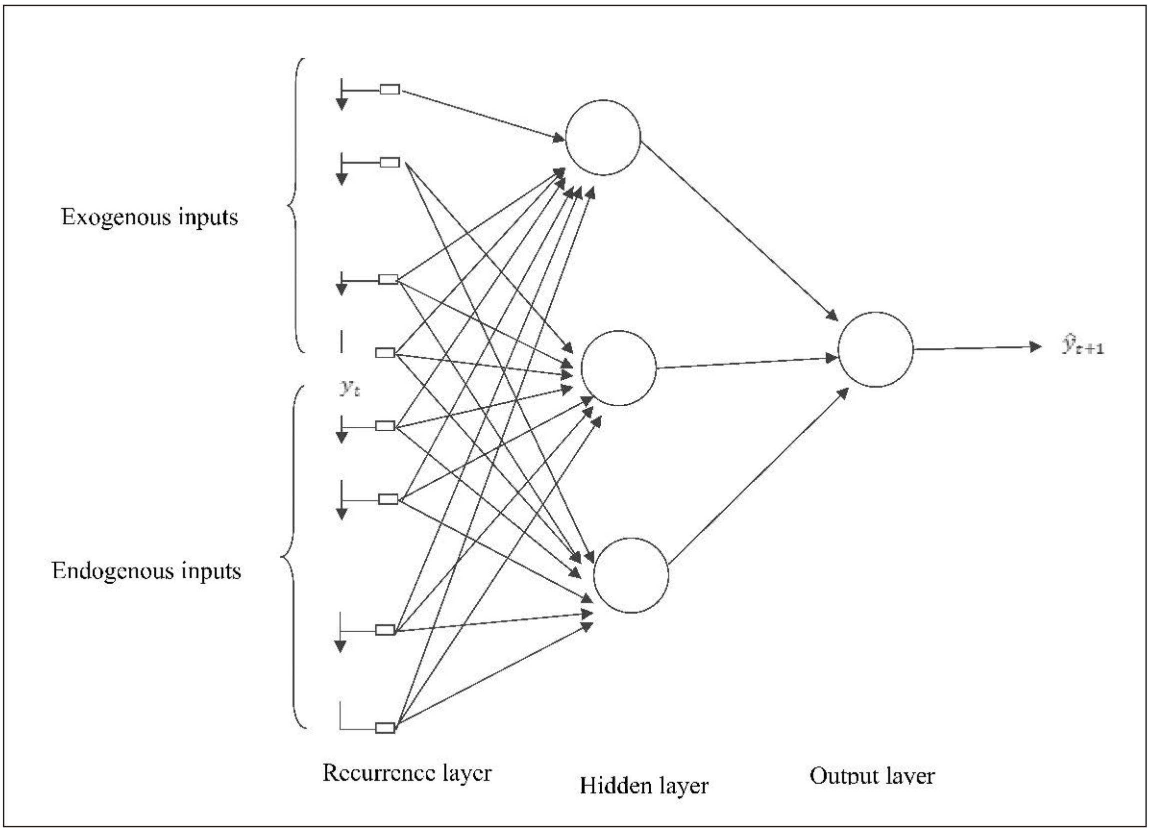

Fig 1: NARX neural network

Economics and Statistics, Ministry of Agriculture, Government of India. The first 36 observations, *i.e.* the data from 1972 to 2007 are used for model building and the remaining 6 data points, *i.e.* the data from 2008 to 2013 are used for validating the model. Along with the yield data, the daily data on maximum temperature for the same time period has been used in the present study. The daily data is first converted to weekly data. The time plot of both the yield and temperature series are given in figure 1 and 2 respectively. A perusal of figures 2 and 3 depicts that yield data is non stationary whereas the temperature series is stationary. But to confirm this claim, statistical tests have been used which is described below. In consonance with the results of Pathak and Wassmann (2009), Paul (2015), exploratory data analysis for present data showed that the correlation coefficients between wheat yield and weekly maximum temperature at Critical Root Initiation (CRI) stage is statistically significant at 5% level of significance.

Fitting of ARIMAX model

As described in section 2.1, ARIMAX model was fitted to the data. First of all both variables *i.e.* yield and maximum temperature were tested for stationarity by using Augmented Dickey Fuller (ADF) test as well as by Phillips Perron (PP) test. The value of test statistics for ADF and PP were found to be -3.36 and -4.93 respectively for maximum temperature series; whereas for yield series, the values are -0.804 and -1.028 respectively. The critical value for ADF and PP test at 5 % level is -2.93. Clearly, the maximum temperature series is stationary at 5% level of significance but yield is found to be nonstationary. On taking first differencing of yield, it becomes stationary. On the basis of minimum AIC and BIC values, best ARIMAX model found for the data set under consideration is:

$$\Delta y_t = 17.36 - 0.23\,\Delta y_{t-1} - 0.61\,x_t - 0.78\,\varepsilon_{t-1} + \varepsilon_t$$
$$(2.84)\quad (0.17)\qquad (0.10)\quad (0.09)$$

Where y_t and x_t denotes the yield of wheat and maximum temperature at CRI stage in the year t. The values in the parenthesis denote standard error of the corresponding parameter estimate. The fitted ARIMAX model along with data points is exhibited in Figure 4.

Fitting of NARX model

Among the 42 data points, we have taken 75% for training, 15% for validation and 15% for testing. The model has been examined at various delays with different number of hidden nodes as shown in table 1. In

Fig. 2: Yield (Q/h) of Wheat at Kanpur of Uttar Pradesh

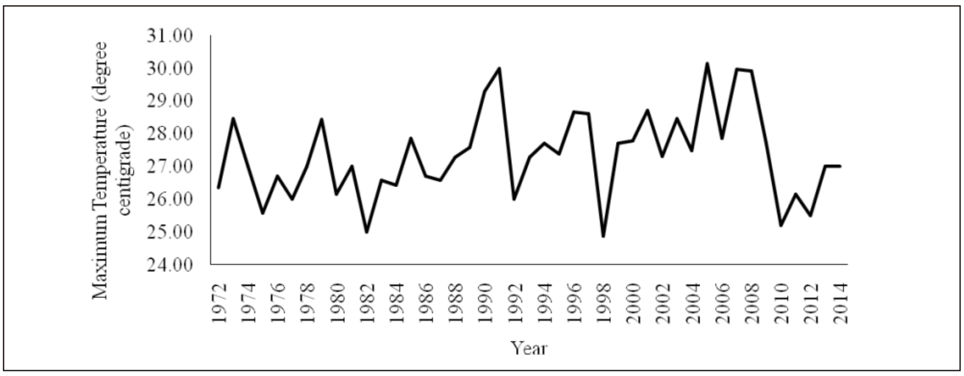

Fig. 2: Yield (Q/h) of Wheat at Kanpur of Uttar Pradesh

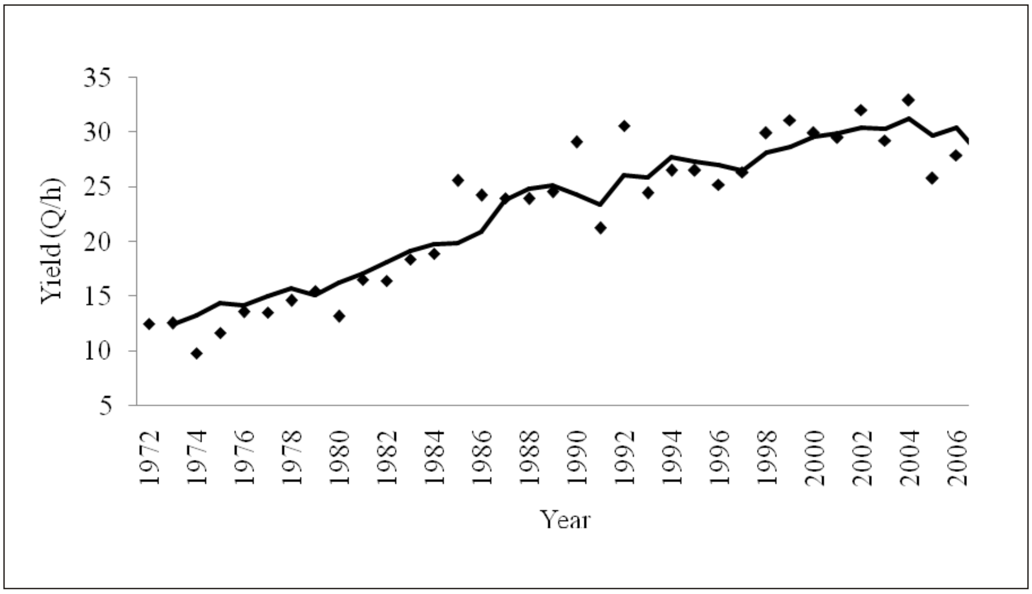

Fig. 4: Fitted ARIMAX model along with data points

the table, we see that out of a total of 20 NARX neural network structures, a neural network model with 2 hidden nodes and 2 delays performs better than other competing models in this study. The present investigation also shows that increasing number of hidden node and delays produce inferior neural networks output. The ovservedvs predicted wheat yield obtained from NARX model for the study period has been plotted in figure 5.

4. Performance evaluation of the fitted models

One-step ahead forecasts of wheat yield based on ARIMAX and NARX models have been considered. Before computing the forecast of wheat yield, the

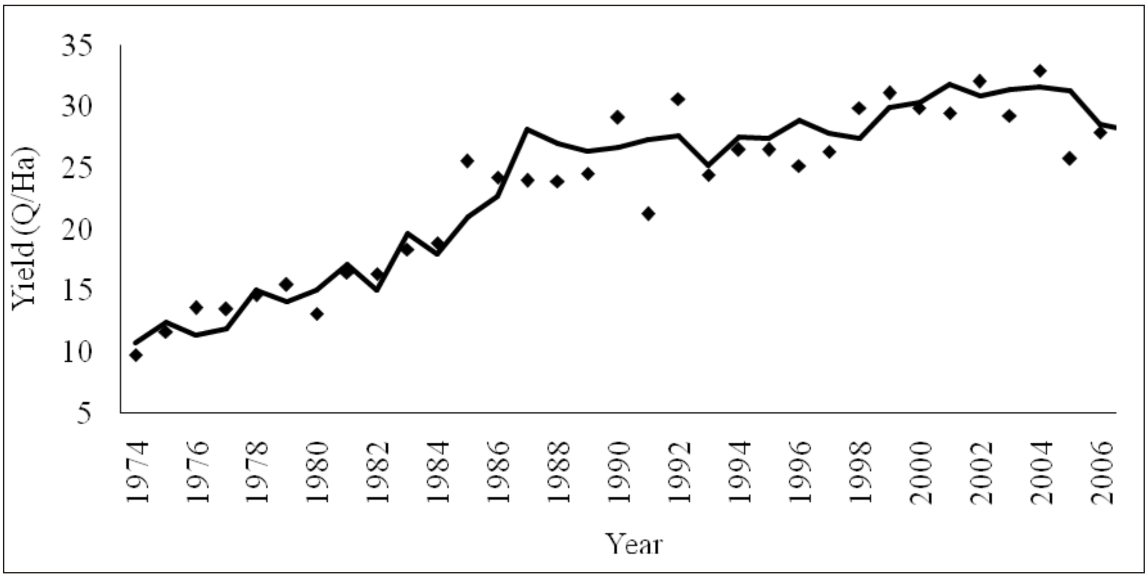

Fig. 5: Fitted NARX model along with data points

Table 1: The performance of the NARX model

No. of Lag.	Hidden node	Training MSE	Validation MSE	Testing MSE
1	1	5.985	32.716	14.914
2	1	10.577	8.461	3.528
3	1	5.104	2.405	19.534
4	1	8.427	15.547	8.929
1	2	6.401	20.098	9.220
2	2	5.958	3.758	5.257
3	2	5.517	10.841	10.458
4	2	7.391	1.806	15.125
1	3	7.336	6.288	31.316
2	3	5.798	7.305	9.075
3	3	19.957	14.350	7.267
4	3	2.186	16.568	23.113
1	4	5.164	3.434	14.478
2	4	31.187	8.954	3.795
3	4	4.539	11.536	17.142
4	4	14.923	3.731	26.149
1	5	6.855	3.985	3.342
2	5	2.026	15.189	20.122
3	5	3.796	12.230	12.278
4	5	15.046	11.551	44.515

forecast of exogenous variable *i.e.* maximum temperature at CRI stage has been computed by using AR(1) model. Once we get the forecast values of maximum temperature, the forecast for wheat yield can be computed by ARIMAX model as follows:

Where $\hat{y}_t$ and $\hat{x}_{tAR}$ denote the prediction of yield and maximum temperature at CRI stage in the year t by ARIMA methodology and wavelet methodology respectively. The forecast of wheat yield for the years 2008-13 in respect of above fitted models are reported in table 3

$$\hat{y}_t = 17.36 + 0.77\, y_{t-1} + 0.23\, y_{t-2} - 0.61\,\hat{x}_{tAR} - 0.78\,\varepsilon_{t-1}$$

Diagnostic Checking

The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen model.This has been done through

Table 2: Forecast of wheat yield (quintals/hectare)

Year	Actual yield	Forecasting by	
		Arimax	NARX
2008	26.66	25.60	29.44
2009	29.40	26.66	30.12
2010	31.50	27.49	28.95
2011	34.80	29.96	31.97
2012	34.64	31.82	33.04
2013	35.58	33.46	34.45

examining the autocorrelations and partial autocorrelations of the residuals of various lags. For this purpose, autocorrelations of the residuals were computed and reported in table 3. Interestingly, it was found that none of these autocorrelations was significantly different from zero at any reasonable level for NARX model but this is not true for fitted ARIMAX model. This proved that the selected NARX model was an appropriate model for forecasting the data under study.

The Mean absolute prediction error (MAPE), Root mean square error (RMSE) and Relative mean absolute prediction error (RMAPE) values for fitted ARIMAX and NARX models are computed and reported in table 4.

$$MAPE = 1/6\sum_{i=1}^{6}\left|y_{t+i} - \hat{y}_{t+i}\right|$$

$$RMSE = 1/6\sum_{i=1}^{6}\left\{\left(y_{t+i} - \hat{y}_{t+i}\right)^2 / y_{t+i}\right\}$$

$$RMAPE = 1/6\sum_{i=1}^{6}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right| / y_{t+i}\right\}\times 100$$

Table 3: ACF and PACF of residuals for NARX and ARIMAX models

Lag	NARX model				ARIMAX model			
	ACF	PACF	Q-Stat	Probability	ACF	PACF	Q-Stat	Probability
1	-0.076	-0.076	0.248	0.618	0.323	0.323	4.489	0.034
2	-0.073	-0.079	0.485	0.785	0.381	0.309	10.910	0.004
3	-0.037	-0.050	0.548	0.908	0.457	0.336	20.405	<0.001
4	0.005	-0.008	0.549	0.969	0.311	0.082	24.932	<0.001
5	-0.040	-0.048	0.627	0.987	0.288	0.015	28.918	<0.001
6	-0.096	-0.108	1.085	0.982	0.374	0.139	35.832	<0.001
7	-0.039	-0.066	1.163	0.992	0.072	-0.249	36.094	<0.001
8	0.147	0.120	2.296	0.971	0.198	-0.017	38.156	<0.001
9	0.107	0.118	2.914	0.968	0.158	-0.027	39.511	<0.001
10	-0.074	-0.043	3.223	0.976	0.046	-0.039	39.631	<0.001
11	-0.272	-0.284	7.517	0.756	0.086	0.008	40.064	<0.001
12	0.039	-0.033	7.607	0.815	0.166	0.150	41.727	<0.001
13	0.052	0.035	7.776	0.858	-0.087	-0.128	42.197	<0.001
14	0.179	0.248	9.855	0.773	0.136	0.101	43.394	<0.001
15	-0.141	-0.090	11.199	0.738	-0.029	-0.127	43.450	<0.001
16	-0.061	-0.172	11.460	0.780	-0.177	-0.250	45.637	<0.001
17	0.078	-0.065	11.899	0.806	0.073	0.134	46.026	<0.001
18	0.022	0.057	11.937	0.850	-0.065	-0.043	46.347	<0.001
19	-0.107	0.073	12.856	0.846	-0.158	0.017	48.336	<0.001
20	-0.146	-0.108	14.643	0.796	-0.043	-0.055	48.492	<0.001

A perusal of table 4 indicates that all the three statistics *i.e.* MAPE, RMSE and RMAPE are lower in NARXmodel as compared to ARIMAX model. The lower values of all the three statistics reflect the superiority of NARX model over ARIMAX model for forecasting purposes.

To this end, Diebold-Mariano test (Diebold and Mariano, 2005)was conducted to test the null hypothesis that the two forecasts have the same accuracy. The alternative hypothesis is that the two forecasts have different levels of accuracy. The test statistics for comparison of NARX and ARIMAXmodels are found to be 2.75, p-values less than 0.01. It clearly indicates the superiority of NARX model over ARIMAX model.

5. Conclusion

Wheat yield data of Kanpur district of Uttar Pradesh has been analysed considering the most important weather variable i.e. maximum temperature at CRI stage

Table 4: Comparison of forecast performance of different models

Year	Forecasting Model	
	Arimax	**NARX**
MAPE	2.93	1.94
RMSE	3.17	2.10
RMAPE (%)	9.00	6.15

of wheat crop. Two approaches namely: ARIMAX and NARX have been applied. The residuals from fitting of these two models have been examined. It is seen that residuals coming out from NARX model are white noise process whereas for ARIMAX model it is not true. Forecast accuracy has been compared in terms of MAPE, RMSE and RMAPE. It is observed that NARX model has better forecast accuracy than that of ARIMAX model. The Diebalod-Mariano test was also applied to compare the prediction power of two competing models. The null hypothesis of equal predictive power is significantly rejected. The selected model has demonstrated a good performance in terms of explained variability and predicting power.The findings of the present study provided direct support for the potential use of accurate forecasts in decision making.

REFERENCES

Alexandrova, V. A. and Hoogenboom,G. (2000). The Impact of Climate Variability and Change on Crop Yield in Bulgaria. *Agri. Forest Meteorology,* **104:**315-327.

Asseng S, Foster I and Turner N C. 2011.The impact of temperature variability on wheat yields. *Global Change Biology* **17**: 997-1012.

Bierens H J. 1987. ARMAX model specification testing, with an application to unemployment in *the Netherlands. Journal Econometrics* **35**: 161-90.

Cheng, B. and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science.* **9:** 2-54.

Dickey, D. and W. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *American Statistical Association,* **74:** 427-31.

Diebold, F. X. and Mariano, R. S. (2005). Comparing predictive accuracy, Journal of Business and Economic Statistics **13:** 253-263.

Gao, Y. and Er, M. J. (2005). NARMAX time series model prediction: feed forward and recurrentfuzzy neural network approaches, Fuzzy Sets and Systems, Vol. 150, 331-50

Guo, W. W. and Xue, H. (2014). Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models. *Mathematical Problems in Engineering.* http://dx.doi.org/10.1155/2014/857865.

Kaur H, Jalota S K, Kanwar R and Vashisht B B. 2012. Climate change impacts on yield, evapotranspiration and nitrogen uptake in irrigated maize (Zea mays)–wheat (Triticumaestivum) cropping system: A simulation analysis. *Indian Agric. Sciences.* **82**: 213-29.

Kumar S, Mishra H S, Sharma A K and Kumar S. 2001. Effect of weather variables on the yield of early, timely and late sown wheat in the Tarai region. *Agric. Physics* **1**: 58-62.

Lin, T., Horne, B., Peter, T. C. and Giles, L. (1996). Learning long-term dependencies inNARX recurrent neural networks, IEEE Transactions on *Neural Networks,* Vol. 7, 1329-1351.

Lobell D B and Burkeb M B. 2010. On the use of statistical models to predict crop yield responses to climate change. *Agric. Forest Meteorology* **150**: 1443–52.

Lobell D B, Field C B, Cahill K N and Bonfils C. 2006. Impacts of future climate change on California perennial crop yields: Model projections with climate and crop uncertainties. *Agric. Forest Meteorology* **141**: 208–18.

Pathak H and Wassmann R. 2009. Quantitative evaluation of climatic variability and risks for wheat yield in India. *Climate Change* **93**: 157-75.

Paul, R. K., Prajneshu and Ghosh, H. (2013). Statistical modelling for forecasting of wheat yield based on weather variables. *Indian J. Agric. Sciences.* **83(2)**: 180-183.

Paul, R. K., Ghosh, H. and prajneshu. (2014). Development of out-of-sample forecast formulae for ARIMAX-GARCH model and their application, *J. Indian Society of Agric. Statistics.* **68 (2014)**: 85-92.

Paul, R. K. (2015). ARIMAX-GARCH-WAVELET Model for forecasting volatile data. *Model Assisted Statistics and Application.* **10(3):** 243-52.

Prasada, A. K., Chai, L., Singha, R. P. and Kafatos, M. (2006). Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters. *Int. J. Applied Earth Observation and Geoinformation.* **8**:26-33.

Simon Haykin, *Neural Networks,* Second Edition, Pearson Education, 1999

Yu, T. (2011). Three Essays on Weather and Crop Yield. PhD thesis, Iowa State University.

Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *Int. J. Forecasting* **14**: 35-62.