Sugarcane yield Estimation of yamunanagar and panipat district of haryana using NARX model

Abstract

Uncontrollable factors that affect yield variability include seasonal and intraseasonal weather variations. The degree to which meteorological factors affect agricultural productivity depends not only on their quantity but also on how they are distributed throughout the crop season. Therefore, while estimating the dynamic behaviour of agricultural production, it should be able to take advantage from both historical data on crop yield and the impact of different external environmental driving forces. In this study, Yamunanagar and Panipat districts of Haryana's sugarcane yield have been attempted to be estimated using NARX models. In addition to percent relative deviation (RD (%)), mean absolute percentage error (MAPE) and root mean squared error (RMSE) have also been used to examine the estimation performance of NARX models.

Introduction

Statistical models are used to connect meteorological variables recorded during the growing season to yield at harvest in the yield forecasting technique currently used for the principal field crops. These models are built utilising information gathered over the course of past years. It is expected that any universal factors that affected yield throughout the years used to develop the model will also apply to the predicted year. It has been acknowledged that the existing technique may give yield projections that are significantly inaccurate due to weather variations from year to year. This is especially evident when the weather during the projected year differs noticeably from that which was experienced during the years used to build the yield model. Although the use of weather data to forecast and estimate crop yields has been investigated. With a total area of 44.2 thousand square kilometres, Haryana makes up 1.37 percent of India's overall geographic area. Net sown area in Haryana is 81.7% of the total area. Haryana produced 773 thousand tonnes of sugarcane overall in 2019–20, yielding 8028 kg per acre. The scientific name for sugarcane is *Sacchrum officinerum*, which is a long perennial grass of the genus Sacchrum. Both tropical and subtropical regions are good for growing sugarcane. Brazil, India, China, Thailand, and Pakistan are the five major nations that produce the most sugarcane, with a combined area of more than 10 lakh hectares. While the sugarcane crop has only received a small amount of research, India has recently adopted a range of yield forecasting methodologies.The following are some noteworthy evaluations of earlier research on sugarcane modeling and related works: Agrawal et al. (1980) developed models using meteorological variables, for predicting rice yield. Using 25 years' worth of yield data as well as the weekly weather variables maximum temperature, relative humidity, total rainfall, and number of rainy days, two models were created to predict rice production in Raipur district, Madhya Pradesh. Khan and Chatterjee (1987) investigated the impact of weather on predicting the rice harvest in West Bengal in the autumn Through the use of linear regression. Data on rice yield from 1958 to 1982 were used as the dependent variable, and data on weekly rainfall from the 18th to 35th standard week were used as the independent variables. McCulloch and Pitts (1943) developed the first computing machines that were designed to imitate the biological nervous system's structure and conduct logic operations that were used to convey information from one neuron to another. Kaul et al. (2005) developed a feed-forward back-propagating ANN structure in order to forecast corn and soy bean yields under normal climatic conditions. It was claimed that adjusting ANN parameters like learning rate and number of hidden nodes has an impact on agricultural yield estimates' accuracy. The range for ideal learning rates is 0.77 to 0.90. Ji et al. (2007) evaluated the ANN rice grain yield models have R-square and RMSE of 0.67 and 891 compared to 0.52 and 1977 for linear regression, respectively. Even though it took more effort to create numerous linear regression models, ANN models were shown to be more accurate in forecasting rice yields under normal climatic circumstances in Fujian. Diaconescu (2008) proposed "Nonlinear Autoregressive Model Process with Exogenous Input" (NARX) neural networks for various chaotic or fractal time series, taking into account the amount of neurons, the training procedures, and the dimensions of his embedded memory. Pisoni et al. (2009) forecasted air pollution time series using nonlinear autoregressive with exogenous variable (NARX). Obe and Shangodoyin (2010) conducted research on the use of an ANN-based model for sugarcane production forecasting. Mean Squared Error (MSE), Normalized Mean Squared Error (NMSE), correlation coefficient, and Minimum Description Length were used to evaluate the effectiveness of ANN models (MDL). Suresh and Priya (2011) discussed the "ARIMA model" for estimating the production and productivity of the Tamil Nadu sugarcane region using data gathered between 1950 and 2007. Dahikar et al. (2015) demonstrated a neural network application for forecasting agricultural yields in Maharashtra, for cotton, jawar, bajra, soyabean, corn, wheat, rice, groundnut, and sugarcane. Crop productivity, pH, nitrogen, phosphate, potassium, depth, temperature, and rainfall were taken into account as input factors. Paul and Sinha (2016) predicted wheat yield in Kanpur district of Uttar Pradesh by taking into account the most significant meteorological variable, i.e., maximum temperature at Critical Root Initiation (CRI) stage of wheat crop, which occurs about 21 days after sowing of the crop. Song et al. (2021) examined the combination of the NARX model with the MFD as exogenous variable is an effective attempt to predict and describe the long-term traffic state at the macroscopic level. Verma et al. (2021) constructed model employing regression techniques for the spring and autumn seasons revealed a strong correlation between predicted and measured values of sugarcane yield.

Materials and method

The Statistical Abstracts of Haryana were used to compile a time series of sugarcane yield data that spans 50 years, from 1972–1973 to 2020–2021 Data for the time period 49 years (1972-2014) have been used for model structure, and the time period 6 years (2015-2020) have been used for model validation in order to create the best model for forecasting the series for the upcoming year. The weather data (1972-73 to 2020-21) were taken from IMD, Pune.

Non linear autoregressive with exogenous input (NARX) model

The architectural approach for sugarcane yield estimation in this research is based on "Nonlinear Autoregressive models with exogenous input (NARX model)", also known as NARX recurrent neural networks. It has been shown that this powerful class of models is well suited for modelling nonlinear systems, particularly time series. Control systems are one of the main areas where NARX dynamic neural networks are used. Moreover, a class is computationally comparable to Turing machines. The gradient descent is better in NARX networks than in other neural networks, and these networks converge significantly faster and generalise better than other networks. These are two significant characteristics of NARX networks with gradient-descending learning gradient method. NARX is actually a nonlinear generalisation of the Autoregressive Exogenous (ARX), a common tool in linear black-box system identification. Numerous nonlinear dynamic systems can be modelled using NARX models. They have been used for a variety of purposes, including time-series modelling. An example of this type of network is the NARX. It features feedback links that surround various network layers. It is interesting to use the memory ability of the NARX neural network by using the previous values of predicted or actual time series in order to gain the full performances of the NARX neural network for nonlinear time series prediction. The NARX model is a nonlinear autoregressive model that has been modified by adding an additional relevant time series as an additional input to the forecasting model. The model can be written as:

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

which can be written in vector form as

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

Where x_t is the external input with the same amount of delays as y_t to the forecasting model.

The NARX model can be represented as

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

Where Φ 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 c_j is the hidden to output layer weight, d is number of input nodes.

Results and Discussion

The sugarcane yield of the chosen Haryana districts was predicted using NARX models. The data was split into three sets that were mutually exclusive: training (80%), testing (10%), and validity (10%). Matlab was used to fit the NARX models. Entering the training and target data, one or two neurons in the hidden layer, and the amount of delays d=5, all while training the model Lavenberg-Marquardt (trainlm) as the training method, and MSE and multiple correlation coefficients as performance metrics (R). The training data set includes the selection of several external variables using stepwise linear regression.

The NARX model was tested using the mean square error (MSE) for various hidden node counts and exogenous variables, as shown in tables 1 and 2, for the districts of Yamunanagar and Panipat, respectively.

Based on the least value of MSE for the Yamunanagar and Panipat districts, respectively, Tables 1 and 2 demonstrate that the NARX models with one hidden neuron and arf13 as an exogenous variable and the NARX models with two hidden neuron and tmx13 as an exogenous varible outperform other models.

Table 1: The performance of the NARX models for Yamunanagar district

Exogenous Variables	Hidden node	MSE		
		Training	Validation	Testing
Tmn3	1	65.53	55.05	39.43
	2	55.23	45.03	56.23
Tmn13	1	45.33	39.95	68.56
	2	67.87	67.43	118.66
Tmn15	1	32.35	30.38	27.68
	2	45.47	47.80	69.56
Tmn6	1	38.89	107.62	97.05
	2	37.25	33.02	87.73
Tmn20	1	52.92	39.87	72.52
	2	105.32	33.79	127.09
Tmx6	1	44.71	66.28	58.10
	2	34.21	40.26	98.91
Tmx7	1	47.59	44.55	43.49
	2	67.07	74.95	62.79
Tmx9	1	75.32	36.79	57.09
	2	54.71	69.28	59.10
Rf13	1	**30.74**	**29.29**	**31.75**

	2	88.82	37.93	180.68
Rf14	1	57.83	29.65	57.45
	2	33.53	83.24	100.67
Rf15	1	101.94	51.96	34.13
	2	43.46	54.12	47.40
Rf15	1	85.32	39.79	87.09
	2	49.71	68.28	68.10

Table 2: The performance of the NARX models for Panipat district

Exogenous Variables	**Hidden node**	**MSE**		
		Training	**Validation**	**Testing**
Tmn4	1	20.62	31.83	35.26
	2	58.39	90.73	81.14
Tmn6	1	27.65	82.55	24.56
	2	35.91	29.41	42.59
Tmn12	1	20.14	30.51	77.50
	2	58.55	39.10	58.90
Tmn23	1	29.88	31.21	56.48
	2	25.23	19.36	36.20
Tmx6	1	186.04	47.79	141.58
	2	28.15	13.38	132.55
Tmx7	1	38.76	67.44	47.32
	2	32.96	73.64	91.85
Tmx13	1	86.04	47.75	41.58

	2	**19.15**	**12.38**	**22.65**
Tmx17	1	58.76	69.44	77.32
	2	72.96	83.64	95.85
Rf2	1	20.96	39.12	34.49
	2	48.71	39.66	44.18
Rf15	1	24.64	17.40	33.81
	2	21.53	51.63	32.35
Rf12	1	25.96	39.56	54.49
	2	88.71	39.99	64.18
Rf19	1	64.64	47.45	53.81
	2	29.53	55.63	37.35

Table 3: NARX model with one hidden neuron and arf13for Yamunanagar district

Sr. No.	Data set	Target values	R
1	**Traning**	38	**0.92**
2	**Validation**	5	**0.95**
3	**Testing**	5	**0.90**

Figure 1: Regression plots of NARX model with one hidden neuron and arf13 for Yamunanagar district

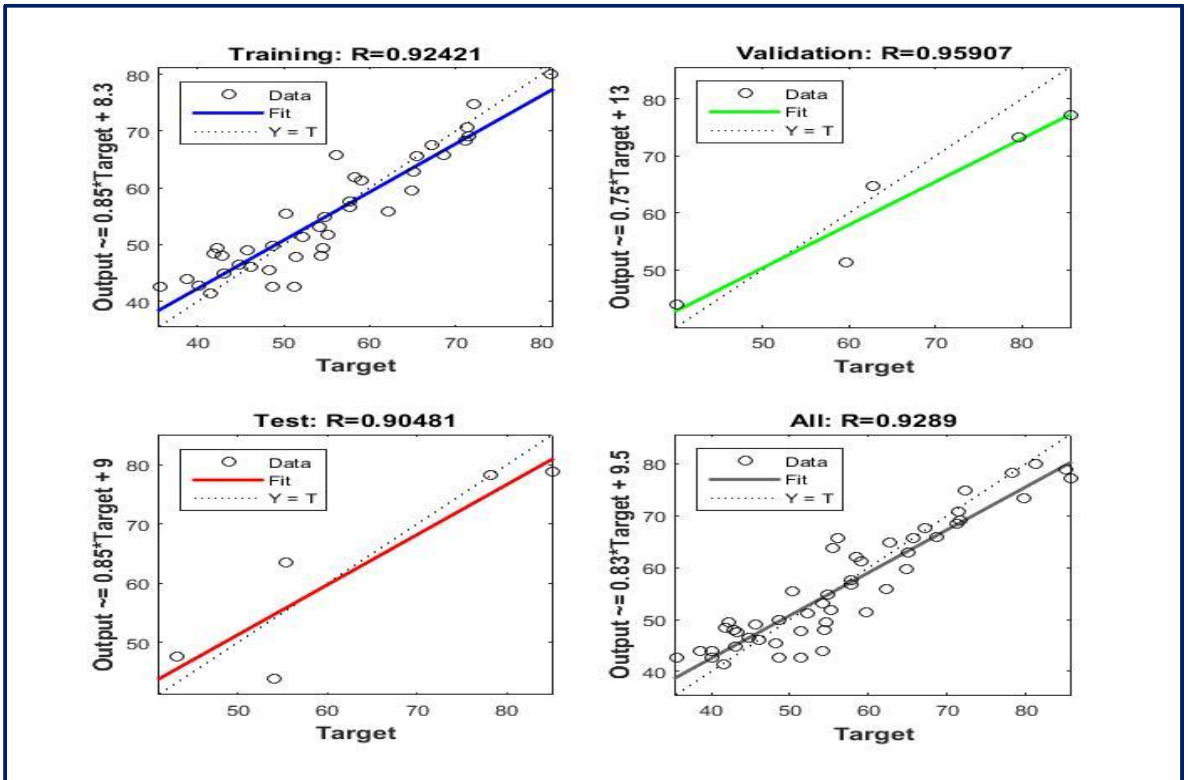
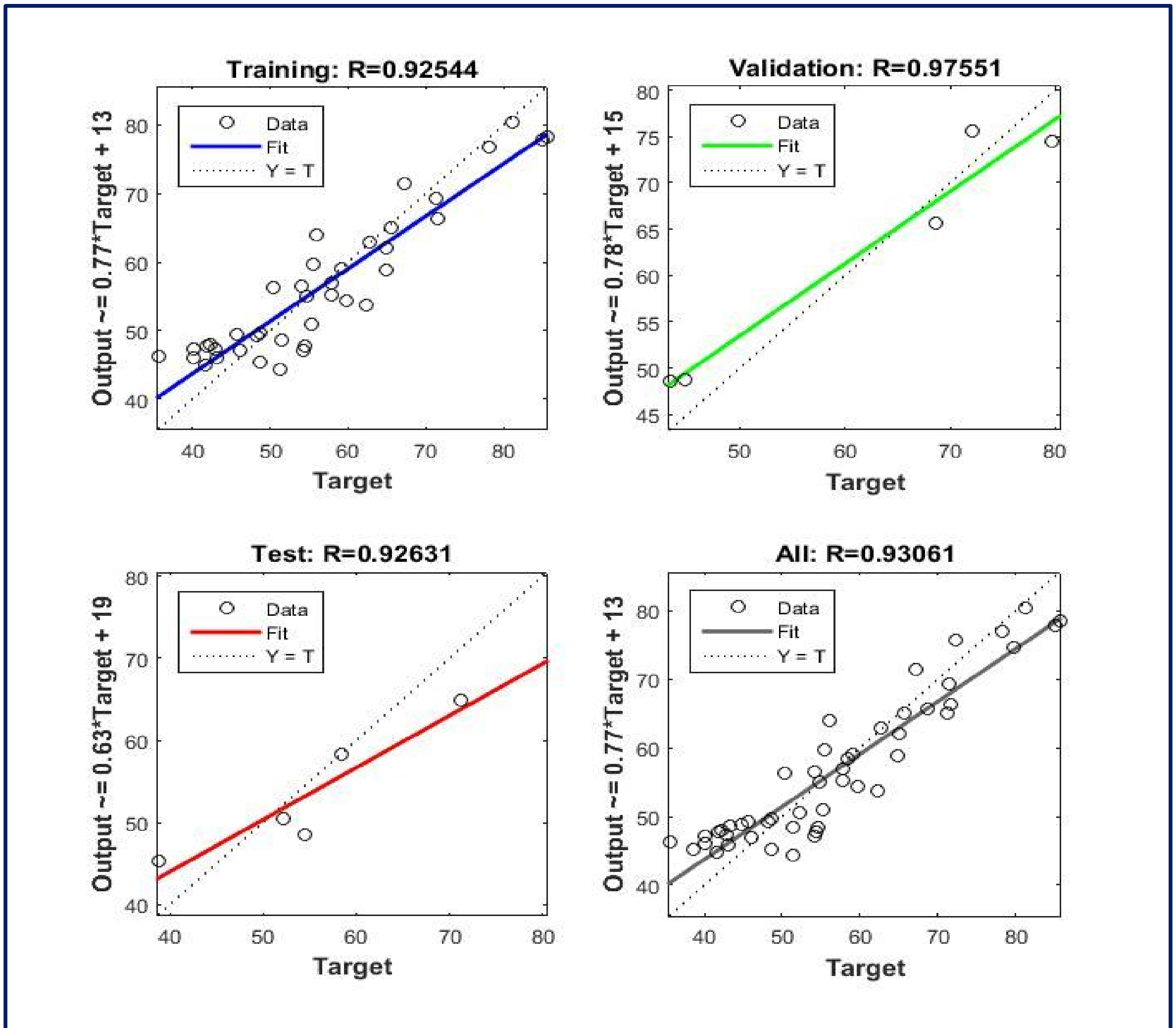


Table 4: NARX model with two hidden neuron and tmx13for Panipat district

Sr. No.	Data set	Target values	R
1	**Traning**	38	**0.92**
2	**Validation**	5	**0.97**
3	**Testing**	5	**0.92**

Figure 2: Regression plots of NARX model with two hidden neuron and tmx13 for Panipat district



Figures 1 and 2 display regression graphs for the Yamunanagar and Panipat districts, respectively. These plots show a relationship between model outputs and the corresponding targets, with R values denoting the goodness of fit for all the data sets. Tables 3 and 4 provide a summary of the findings for all districts that were chosen. The main goal of model verification is to check the residuals of the chosen model for any systematic patterns that might still be removed to improve the model. The NARX models with one hidden neuron and arf13, and the NARX models with two hidden neuron and tmx13, respectively, for Yamunanagar and Panipat districts, reveal no residual autocorrelations, or the absence of any systematic patterns in residuals. Observed and forecast value of corresponding model was shown in figure 5 and 6 for corresponding district.

Figure 3: Plot of residuals autocorrelation from selected NARX model for Yamunanagar district

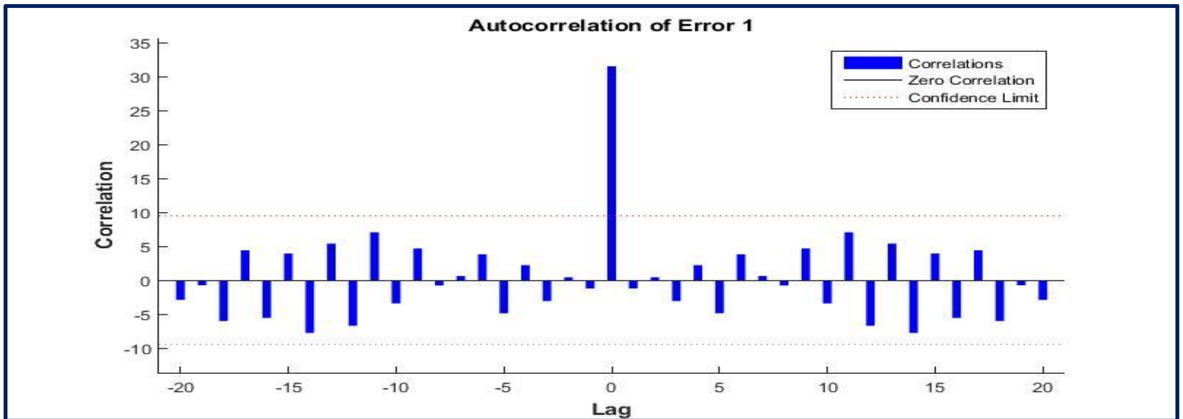


Figure 4: Plot of residuals autocorrelation from selected NARX model for Paniapat district

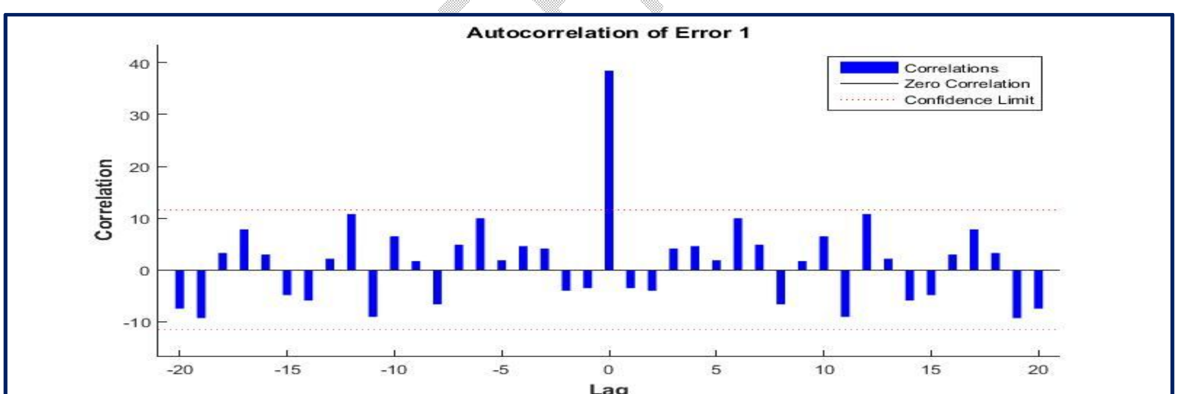


Figure 5: Plot of observed and predicted sugarcane yield by NARX model for Yamunanagar district

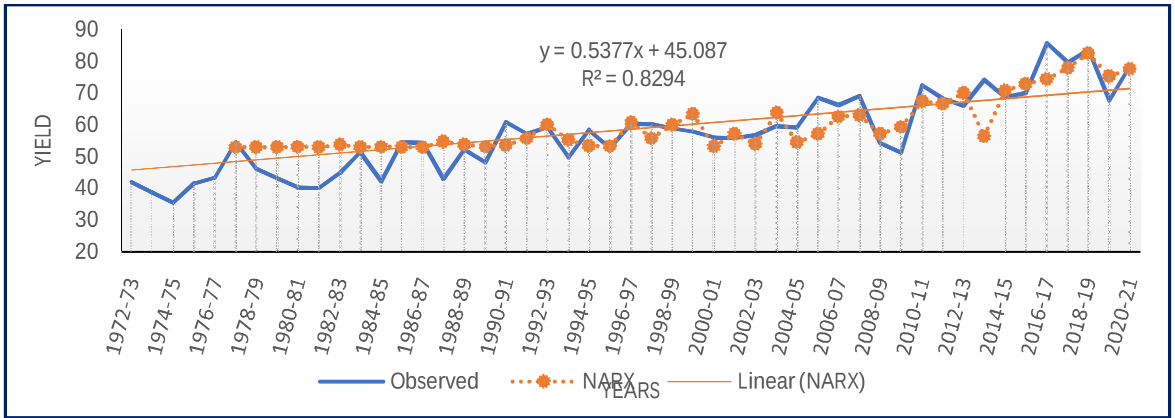


Figure 6: Plot of observed and predicted sugarcane yield by NARX model for Panipat district

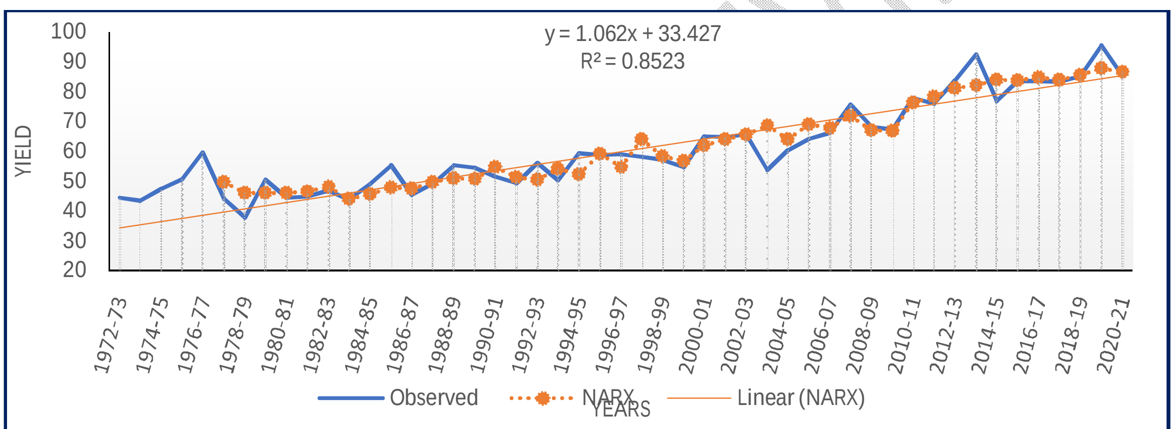


Table 5: Estimated and observed value of sugarcane yieldand associated percentage deviations (RD%) = 100×(observed yield-est. yield)/ observed yield)

Year	Yamunanagar			Panipat		
	Observed	Forecast	RD	Observed	Forecast	RD
2015-16	69.9	72.74	-4.06	83.35	83.74	-0.46
2016-17	85.57	74.25	13.23	83.55	84.79	-1.49
2017-18	79.66	77.88	2.23	83.3	84.03	-0.87
2018-19	83.53	82.36	1.40	85.1	85.49	-0.46
2019-20	67.69	75.18	-11.07	95.46	87.73	8.10

2020-21	78.40	77.52	1.12	85.12	86.56	-1.69
RMSE	6.29			3.58		
MAPE	5.52			2.18		

The period of sugarcane yield from 2015–16 to 2020–21 was utilised as the validation set for comparison purposes. To examine the effectiveness of NARX models, three statistical measures were used: percent relative deviation (RD (%)), root mean squared error (RMSE), and mean absolute percentage error (MAPE). Table 5 shows the observed, forecast, and percent relative deviation from the models that were chosen as the best fits for the validity set of sugarcane yield for Yamunanagar and Panipat based on the lowest values of RMSE and MAPE.

References

Agrawal, R., Jain, R. C., Jha, M. P. And Singh, D. (1980) Forecasting of rice yield using climatic variables. *Ind. J. Agri. Sci.*, **50(9):**680-684.

Dahikar, S. S., Rode, S. V. And Deshmukh, P. (2015) An Artificial Neural Network approach for agricultural crop yield prediction based on various parameters. *International Journal of Advanced Research in Electronics and Communication Engineering*, **4(1):**94-98.

Diaconescu, E. (2008) The use of NARX Neural Networks to predict Chaotic Time Series. *Wseas transactions on computer research*, 182-191.

Ji, B., Sun, Y., Yang, S. And Wan, J. (2007) Artificial neural networks for rice yield prediction in mountainous regions. *J. Agric. Sci.*, **145:** 249-261.

Kaul, M., Hill, R. L. And Walthall, C. (2005) Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems*, **85:** 1–18.

Khan, S. A. And Chatterjee, B. N. (1987) Crop weather analysis for forecasting the yield of autumn rice in West Bengal. *Ind. J. Agri. Sci.*, **57(11):** 791-794.

Mcculloch, W. S., and Pitts, W. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, **5:**115-133.

Obe, O. O. And Shangodoyin, D. K. (2010) Artificial Neural Network Based Model for Forecasting Sugarcane Production, Journal *of Computer Science* **(4):**439-445.

Paul, R. K. And Sinha, K. (2016) Forecasting crop yield : a comparative assessment of ARIMAX and NARX model. *RASHI*, **1 (**1**):** 77-85.

Pisoni, E., Farina, M., Carnevale, C., and Piroddi, L. (2009) Forecasting peak air pollution levels using NARX models. *Engineering Applications of Artificial Intelligence*, **22**: 593–602.

Song , Z. , Sun, F. , Zhang, R., Du, Y., and Li, C. (2021) Prediction of Road Network Traffic State Using the NARX Neural Network. Journal of Advanced Transportation, 1-17.

Suresh, K. K. and Priya, S. R. K. (2011) Forecasting sugarcane yield of Tamil Nadu using ARIMA models. *Sugar Tech*, **13**(1):23-26.

Verma, A.K., Garg, P.K., Prasad, K.S.H., Dadhwal, V.K., Dubey, S.K. and Kumar, A. (2021) Sugarcane Yield Forecasting Model Based on Weather Parameters, *Sugar Tech*, 23(1):158–166 https://doi.org/10.1007/s12355-020-00900-4.