Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(1) 2016 63-70

ARIMA-WNN Hybrid Model for Forecasting Wheat Yield Time-Series Data

Mrinmoy Ray[1], Anil Rai[1], Ramasubramanian V.[2] and K.N. Singh[1]

[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

[2]*ICAR-Central Institute of Fisheries Education, Mumbai*

Received 19 February 2016; Revised 18 April 2016; Accepted 19 April 2016

SUMMARY

The present study proposed a new hybrid model combining Autoregressive integrated moving average (ARIMA) and Wavelet Neural Network (WNN). ARIMA is the most widely used technique for forecasting in divergent domains for several decades. WNN is the recently developed neural networks which utilize wavelet activation function in the hidden neuron. As a case study, wheat yield of India has been considered to evaluate the forecasting performance of the proposed hybrid model. The proposed method was compared with ARIMA and existing hybrid ARIMA-ANN approach. Empirical results clearly reveal that the forecasting accuracy of the proposed method is better as compared to the existing approach.

Keywords: Forecasting, ARIMA, ANN, WNN, Hybrid model.

1. INTRODUCTION

Agriculture is one of the most prominent sectors in India as about 70% population of our country is directly dependent on agriculture and about 43% of India’s geographical area is used for agricultural activities. Since Green Revolution, India has been growing steadily in terms of agricultural productivity and growth of food grain and commercial crops in India have risen considerably over the last four decades. Modern farming practices and inclusive technologies have been implemented in many parts of rural India to foster growth. The increased agricultural production has been achieved largely throughan increase in the yield per hectare, rather than from an increased cultivated area. India is now in a position to export surplus agricultural commodities and earns a lot of foreign exchange. In such a scenario, forecasting of crop yields or any agricultural production is a formidable challenge.

Auto-Regressive Integrated Moving Average (ARIMA) models (Box *et al*. 1994) have been appreciated for crop yield or any other agricultural production forecasting. Sarika *et al.* (2011) employed, ARIMA model for modeling and forecasting India’s pigeon pea production data. Suresh *et al.* (2011) employed ARIMA model for forecasting sugarcane area, production and productivity of Tamilnadu state of India.

Recent research activity shows that combining different model enhances the accuracy of forecasting as compared to individual model. Zhang (2003) hybrid methodology is one of the

Corresponding author: Mrinmoy Ray
E-mail address: mrinmoy.ray@icar.gov.in

most popular hybrid method which combine ARIMA and ANN models. Koopman *et al.* (2007) proposed an approach which combines periodic ARFIMA and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. Che and Wang (2010) proposed a hybrid model which combine ARIMA and SVM (Support Vector Machine). Jha and Sinha (2014) combine ARIMA and TDNN (Time Delay Neural Network) for forecasting monthly wholesale price of oilseed in India. Chaâbane (2014) proposed a hybrid model combing ARFIMA and Least square SVM. Paul (2015) forecasted volatile data combining ARIMAX, GARCH and wavelet approach. Paul *et al.* (2015) combine AR and FIGARCH (Fractionally integrated GARCH) and applied it for forecasting spot price of lentil.

In this study a new hybrid model is proposed which combine Autoregressive integrated moving average (ARIMA) and Wavelet Neural Network (WNN). WNN is the recently developed neural network which utilizes wavelet activation function in the hidden neuron. A good account on Wavelet Neural Network (WNN) is given in Alexandridis and Zapranis (2013).

2. MATERIAL AND METHODS

2.1 Data Description

Yearly data on wheat yield (1951-52 to 2013-14) of all-India level was collected from Department of Agriculture and Cooperation (Agricultural Statistics at a Glance 2014) given in http:/agricoop.nic.in/agrilstatistics.htm.. Data from 1951-52 to 2003-04 were used for model construction and 2004-05 to 2013-14 were used to check the forecasting performance.

2.2 Arima Model Fitting

An ARIMA model is given by:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t$$

where

$\phi(B) = 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$ (Autoregressive parameter)

$\theta(B) = 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_p B^q$ (Moving average parameter)

ε_t = white noise or error term

D = differencing term

B = Backshift operator i.e. $B^a Y_t = Y_{t-a}$

ARIMA methodology is carried out in three stages, viz. Identification, estimation and diagnostic checking. Identification of *d* is necessary to make the non-stationary time series to stationary. A formal statistical test for the existence of stationarity, known as the test of the unit-root hypothesis or Augmented Dickey Fuller test was utilized to test the stationarity. A good account on Augmented Dickey Fuller test can be found in Makridakis *et al*. (1998). At the estimation stage, parameters are estimated for the ARIMA model tentatively chosen at the identification stage. Estimation of parameters for ARIMA model is generally done through iterative least squares method. The adequacy of the selected model is then tested at the diagnostic checking stage. At this stage, testing is done to see if the estimated model is statistically adequate i.e. whether the error terms are white noise which means error terms are uncorrelated with zero mean and constant variance. For this purpose, Ljung-Box test is applied to the original series or to the residuals after fitting a model. A good account on Ljung-Box test can be found in Box *et al*. (1994). 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.

2.3 BDS (Brock-Dechert-Scheinkman) Test for Testing of Nonlinearity

BDS test utilizes the concept of spatial correlation from chaos theory. The computational procedure is given as follows.

1. Let the considered time series is

   $$\{x_i\} = [x_1, x_2, x_3, ..., x_N]$$

2. The next step is to specify a value of *m* (embedding dimension), embed the time

series into m dimensional vectors, by taking each m successive points in the series. This transforms the series of scalars into a series of vectors with overlapping entries

$$x_1^m = (x_1, x_2, ..., x_m)$$

$$x_2^m = (x_2, x_3, ..., x_{m+1})$$

.

.

.

$$x_{N-m}^m = (x_{N-m}, x_{N-m+1}, ..., x_N)$$

3. In the third step correlation integral is computed, which measures the spatial correlation among the points, by adding the number of pairs of points (i, j), where $1 \leq i \leq N$ and $1 \leq j \leq N$, in the m-dimensional space which are "close" in the sense that the points are within a radius or tolerance ε of each other.

$$C_{\varepsilon,m} = \frac{1}{N_m(N_m - 1)} \sum_{i \neq j} I_{i,j;\varepsilon}$$

where $I_{i,j;\varepsilon} = 1$ if $\left\| x_i^m - x_j^m \right\| \leq \varepsilon$

= 0 otherwise

4. If the time series is i.i.d. then $C_{\varepsilon,m} \approx [C_{\varepsilon,1}]^m$

5. The BDS test statistics is as follows

$$BDS_{\varepsilon,m} = \frac{\sqrt{N}[C_{\varepsilon,m} - (C_{\varepsilon,1})^m]}{\sqrt{V_{\varepsilon,m}}}$$

where

$$V_{\varepsilon,m} = 4[K^m + 2\sum_{j=1}^{m-1} K^{m-j} C_\varepsilon^{\,2j} + (m-1)^2 C_\varepsilon^{\,2m} - m^2 K C_\varepsilon^{\,2m-2}]$$

$$K = K_\varepsilon = \frac{6}{N_m(N_m - 1)(N_m - 2)} \sum_{i<j<N} h_{i,j,N;\varepsilon}$$

$$h_{i,j,N;\varepsilon} = \frac{[I_{i,j;\varepsilon} I_{j,N;\varepsilon} + I_{i,N;\varepsilon} I_{N,j;\varepsilon} + I_{j,i;\varepsilon} I_{i,N;\varepsilon}]}{3}$$

The choice of m and ε depends on number of data. The null hypothesis is data are independently and identically distributed (i.i.d) against the alternative hypothesis the data are not i.i.d.; this implies that the time series is non-linearly dependent. BDS test is a two-tailed test; the null hypothesis should be rejected if the BDS test statistic is greater than or less than the critical values.

2.4 Artificial Neural Network (ANN) Model

Artificial neural networks (ANNs) model are considered as a class of generalized nonlinear model that are able to capture various nonlinear structures present in the data set. The main advantage of this model is that it does not require prior assumption of the data generating process, instead it is largely depend on characteristics of the data popularly known as data-driven approach. Single hidden layer feed forward network is the most popular for time series modeling and forecasting. This model is characterized by a network of three layers of simple processing units, and thus termed as multilayer ANNs. The first layer is input layer, the middle layer is the hidden layer and the last layer is output layer.

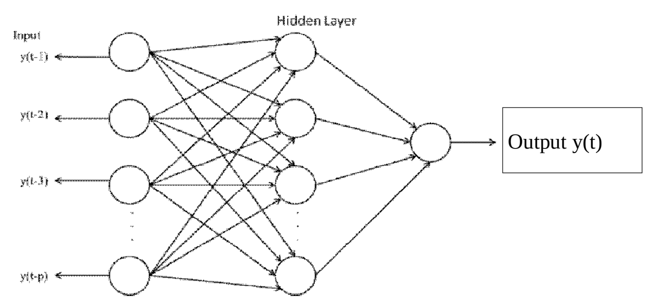

Fig. 1. Neural network architecture

The relationship between the output (y_t) and the inputs ($y_{t-1}, y_{t-2},\ldots,y_{t-p}$) can be mathematically represented as follows:

$$y_t = f\left(\sum_{j=0}^{q} \omega_j g\left(\sum_{i=0}^{p} \omega_{ij} y_{t-i}\right)\right) \qquad (1)$$

where, $\omega_j (j = 0,1,2,\ldots,q)$ and $\omega_{ij} (i = 0,1,2,\ldots,p$, $j = 0,1,2,\ldots,q)$ are the model parameters often called the connection weights, p is the number of input nodes and q is the number of hidden nodes, g and f denote the activation function at hidden and output layer respectively. Activation function defines the relationship between inputs and outputs of a network in terms of degree of the non-linearity. Most commonly used activation functions are as follows:

Activation Function	Equation
Identity	x
Sigmoid	$\frac{1}{1+e^{-x}}$
TanH	$\tanh(x) = \frac{2}{1+e^{-2x}} - 1$
ArcTan	$\tan^{-1}(x)$
Sinusoid	$\sin(x)$
Gaussian	e^{-x^2}

For time series forecasting sigmoid activation function is employed in hidden layer and identity activation function is employed in the output layer (Banakar and Azeem 2008, Jha and Sinha 2014, Zhang 2003).

When radial basis functions are employed as activation function in the hidden neuron the neural network is known as Radial Basis Function (RBF) network. These activation functions can take many forms, but they are usually found as one of three functions

RBF Activation Function	Equation
Gaussian	$e^{\left(-\frac{\|v_i - c_i\|^2}{2\sigma^2}\right)}$
Multiquadratics	$\sqrt{\|v_i - c_i\|^2 + a^2}$
Inverse Multiquadratics	$\left(\|v_i - c_i\|^2 + a^2\right)^{-\frac{1}{2}}$

where c_i is the vector representing the function center a and σ are parameters affecting the spread of radius.

When wavelets are employed as activation function in the hidden neurons it is known as the wavelet neural network. There are different types of wavelet activation functions. In this paper morlet wavelet (Banakar and Azeem 2008) was employed as activation function. Morlet function is represented as follows:

$$\psi(x) = e^{(-x^2)} \cos(5x) \tag{2}$$

This wavelet is derived from a function that is proportional to the cosine function and Gaussian probability density function. Research article Alexandridis and Zapranis (2013), Banakar and Azeem (2008) show that wavelet activation function is better than sigmoid function, hence in this article the comparison between sigmoid and wavelet activation function is done.

Thus ANN model performs a nonlinear functional mapping between the input and output which characterized by a network of three layers of simple processing units connected by acyclic links.

$$y_t = f(y_{t-1} + y_{t-2}, ..., y_{t-p}, w) + \varepsilon_t \tag{3}$$

where, w is a vector of all parameters and f is a function of network structure and connection weights. Therefore, the neural network resembles a nonlinear autoregressive model.

The selection of appropriate number of hidden nodes as well as optimum number of lagged observation p for input vector is important in ANN modeling for determination of the autocorrelation structure present in a time series. Though there are no established theories available for the selection of p and q, hence experiments are often conducted for the determination of the optimal values of p and q. For time series forecasting there are two common learning method one is levenberg-marquardt back propagation and the other is Gradient decent back propagation. Most of the literature (Alexandridis and Zaprani 2013, Banakar and Azeem 2008, Zang 2003) employed Gradient decent back propagation algorithm hence in this article gradient decent method was utilized. The objective of training is to minimize the error function that measures the misfit between the predicted value and the actual value. The error function which is widely used is mean squared error which can be written as:

$$E = \frac{1}{N}\sum_{n=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{n=1}^{N}\left\{y_t - f\left(\sum_{j=0}^{q}\omega_j g\left(\sum_{i=0}^{p}\omega_{ij} y_{t-i}\right)\right)\right\}^2 \tag{4}$$

where N is the total number of error terms. The parameters of the neural network are ω_j and ω_{ij} estimated by iteration. Initial connection weights are taken randomly from uniform distribution. In each iteration the connection weights changed by an amount $\Delta\omega_j$

$$\Delta\omega_j(t) = -\eta \frac{\partial E}{\partial \omega_j} + \delta\Delta\omega_j(t-1) \quad (5)$$

where, η is the learning rate and $\frac{\partial E}{\partial \omega_j}$ is the partial derivative of the function E with respect to the weight ω_j. δ is the momentum rate. The $\frac{\partial E}{\partial \omega_j}$ can be represented as follows

$$\frac{\partial E}{\partial w_j} = -e_j(n) \times f'(x) \times y_j(n) \quad (6)$$

where $e_j(n)$ is the residual at n^{th} iteration and $f'(x)$ is derivative of the activation function in the output layer. As in time series forecasting the activation function in the output layer is identity function hence $f'(x)$=1. $y_j(n)$ is the desired output. Now connection weights in from input to hidden nodes changed by an amount $\Delta\omega_{ij}$

$$\Delta\omega_{ij}(t) = -\eta \frac{\partial E}{\partial \omega_{ij}} + \delta\Delta\omega_{ij}(t-1) \quad (7)$$

where

$$\frac{\partial E}{\partial w_{ij}} = g'(x) \times \sum_{j=0}^{q} e_j(n) * w_j(n) \quad (8)$$

where $g'(x)$ is the activation function in the hidden layer. For sigmoid activation function

$$g'(x) = \frac{\exp(-x)}{(1+\exp(-x))^2}$$

For morlet wavelet function

$$g'(x) = -5\sin(5x)\exp(-x^2) - 2x\cos(5x)\exp(-x^2)$$

Learning rate is user defined parameter known as tuning parameter of neural network which determine how slow or fast the optimal weight is obtained. The learning rate must be set small enough to avoid divergence. The momentum term prevents the learning process from setting in a local minimum. Though there are no established theories available for the selection of learning rate and momentum, hence experiments are often conducted for the determination of the learning rate and momentum. Typically learning rate and momentum values lies between zero to one. After obtaining final weights by employing the equation 1 final output is obtained.

2.5 Proposed Hybrid Approach

The proposed approach considered time series (y_t) as a function of linear and nonlinear components. Hence

$$y_t = f(L_t, N_t) \quad (9)$$

where L_t and N_t represents the linear and nonlinear component respectively. This approach follows the Zhang's (2003) hybrid approach, accordingly the relationship between linear and nonlinear components can be written as following.

$$y_t = L_t + N_t \quad (10)$$

The main strategy of this approach is to model the linear and nonlinear components separately by different model. The methodology consists of three steps. Firstly, an ARIMA model is employed to fit the linear component. Let the prediction series provided by ARIMA model denoted as $\hat{L}_t$. In the second step, instead of predicting the linear component, the residuals denoted as e_t which are nonlinear in nature are predicted. The residuals can be obtained by subtracting the predicted value $\hat{L}_t$ from actual value of the considered time series y_t.

$$e_t = y_t - \hat{L}_t \quad (11)$$

Now the residuals are predicted employing a WNN model. Let the prediction series provided by WNN model denoted as $\hat{N}_t$. Finally, the predicted linear and nonlinear components are combined to generate aggregate prediction.

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{12}$$

The proposed approach can be graphically represented as below.

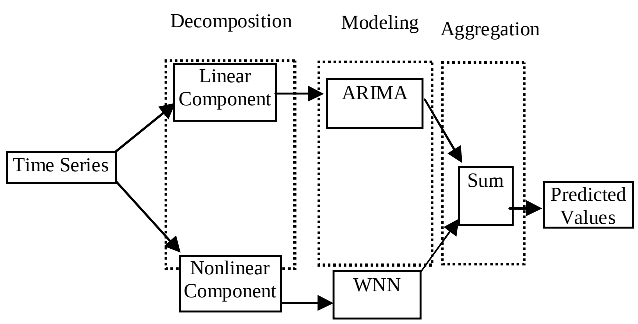

Fig. 2. Overview of the forecasting framework

2.6 Forecasting Performance

Forecasting performance of the model has been judged by computing Mean Absolute Percent Error (MAPE) and Mean square Error (MSE). The model with less MAPE and MSE is preferred for forecasting purposes. The MAPE and MSE is computed as

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| / y_t \times 100 \tag{13}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2 \tag{14}$$

where n is the total number of forecast values. y_t is the actual value at period t and $\hat{y}_t$ is the corresponding forecast value.

ARIMA model was fitted utilizing Statistical Analysis Systems (SAS), USA, Version 9.4, Module SAS-ETS. BDS test was employed using R software package tseries. A code has been developed for WNN [reader can email the first author for the code] in Matlab software.

3. RESULTS AND DISCUSSION

The time series plot reveals that there is a positive trend over time which indicates the time series non-stationary in nature. ADF test was applied for assessing non-stationary in time series. The results of the ADF test are given below in Table 1. The plot of time series is given below.

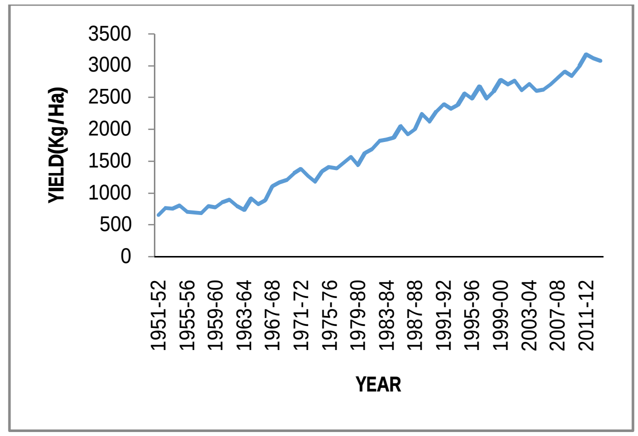

Fig. 3. Wheat yield of all-india (1951-52 to 2013-14)

Table 1. ADF test

Series	Single Mean		With Trend	
	ADF test Statistics	Probability	ADF test Statistics	Probability
Actual series	0.19	0.9696	-3.00	0.14
1st difference series	-7.59	0.0001	-7.64	<.0001

From Table 1 it can be infer that there exists non-stationary is the actual series but after first differencing the series become stationary. ARIMA (2, 1, 0) was found adequate for the considered time series which is given below.

Table 2. Parameters of ARIMA (2,1,0)

Parameter	Estimate	Std. Error	t value	Pr>|t|
μ	39.25	7.83	5.02	0.0002
ϕ_1	-0.48	0.14	-3.53	0.004
ϕ_2	-0.29	0.14	-2.03	0.0419

$$y'_t = 39.25 - 0.48 y'_{t-1} - 0.29 y'_{t-1} + \varepsilon_t$$

where $y'_t = y_t - y_{t-1}$

In the next step residuals are obtained from the fitted ARIMA model. The Brock, Dechert and Scheinkman (BDS) test Broock *et al.* (1996) was employed to test the existence of nonlinearity. The results of the test in given in Table 3 which indicate that nonlinear pattern exist in the residual data.

Table 3. BDS test

Dimension (m)	Epsilon (Ɛ)		Statistic	Probability
2	eps(1)	48.76	-5.26	<.0001
	eps(2)	97.51	-2.59	0.009687
	eps(3)	146.27	-3.51	0.0004406
	eps(4)	195.02	-6.34	<.0001
3	eps(1)	48.76	-8.95	<.0001
	eps(2)	97.51	-2.83	0.00464
	eps(3)	146.27	-3.99	<.0001
	eps(4)	195.02	-6.71	<.0001

Next the residual was fitted employing WNN as well as ANN model. The summary of the neural network model is given below in Table 4.

Table 4. Neural network summary

Parameters	ANN	WNN
Number of input (lag)	5	5
Number of hidden unit	3	3
Activation function in hidden unit	Sigmoid	Morlet (wavelet)
Number of iterations	1500	1500
Learning algorithm	Gradient decent back propagation	Gradient decent back propagation
Learning rate	0.04	0.04
Momentum	0.02	0.02

To evaluate the forecasting performance last ten observations of the considered time series was predicted employing the proposed approach. This approach was compared with the conventional ARIMA as well as Zhang hybrid approach (ARIMA-ANN). The results are given Table 5.

Table 5. Forecasted values from different model

Year	Actual	ARIMA	ARIMA -ANN	Proposed Approach
2004-05	2602	2776.30	2726.07	2683.80
2005-06	2619	2785.71	2736.80	2700.04
2006-07	2708	2832.45	2805.20	2753.53
2007-08	2802	2876.63	2883.66	2871.35
2008-09	2907	2911.36	2916.64	2928.65
2009-10	2839	2951.38	2905.13	2870.12
2010-11	2988	2991.55	2977.86	3065.83
2011-12	3177	3030.14	3055.79	3122.50
2012-13	3117	3069.44	3074.72	3153.72
2013-14	3075	3108.86	3041.15	3094.58
MAPE	-	**3.18**	**2.50**	**1.83**
MSE	-	**11686.41**	**6757.90**	**3233.70**

From Table 5 it can be infer that proposed approach perform better as compare to conventional ARIMA as well as Zhang hybrid approach (ARIMA-ANN) for the considered time series.

4. CONCLUSION

In this article a new hybrid approach is proposed which combine ARIMA and WNN. The difference between ANN and WNN is only the activation function in the hidden layer. For time series forecasting sigmoid activation function is employed in the hidden neuron. But in this study we employed morlet wavelet activation function in the hidden neuron. Wheat yield of India has been considered to evaluate the forecasting performance of the proposed hybrid model. Firstly ARIMA model was fitted for the considered time series. The residual obtained from fitted model was tested employing BDS test which reveals that nonlinearity pattern exists in the residual data. The performance of the proposed approach was compared with the conventional ARIMA model as well as the most popular Zhang hybrid approach (ARIMA-ANN). Based on the results obtained from this work it can be infer that proposed approach perform better as compare to conventional ARIMA as well as Zhang hybrid approach (ARIMA-ANN) for the considered time series. In further studies, one can improve the forecasting accuracy by applying some other hybrid models. This approach can be further evaluated by using some other data sets so that practical validity of the model can be well known**.**

ACKNOWLEDGEMENTS

We would like to acknowledge the reviewer for his constructive comments.

REFERENCES

Alexandridis, A.K. and Zapranis, A.D. (2013). Wavelet neural networks: A practical guide. *Neural Networks*, **42,** 1-27.

Banakar, A. and Azeem, M.F. (2008). Artificial wavelet network and its application to neuro-fuzzy models. *Appl. Soft. Computing*, **8**, 1463-1485.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994), *Time Series Analysis: Forecasting and Control (3rd ed.)*. Holden-Day, San Francisco.

Broock, W., Scheinkman, J.A., Dechert, W.D. and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econ. Rev.*, **15**, 197-235.

Chaâbane, N. (2014). A novel auto-regressive fractionally integrated moving average least squares support vector machine model for electricity spot prices prediction. *J. Appl. Statist.*, **41(3)**, 635-651.

Che, J. and Wang, J. (2010). Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Convers. Manage.*, **51**, 1911-1917.

Jha, G.K. and Sinha, K. (2014). Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Comput. Appl.*, **24(3).**

Koopman, S., Ooms, M. and Carnero, M. (2007). Periodic seasonal reg-ARFIMA–GARCH models for daily electricity spot prices. *J. Amer. Statist. Assoc.*, **102,** 563-571.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998). *Forecasting: Methods and Applications (3rd ed.)*. Wiley, Chichester.

Paul, R.K. (2015). ARIMAX-GARCH-WAVELET model for forecasting volatile data. *Model Assist. Statist. Appl.*, **10(3)**, 243-252.

Paul, R.K., Gurung, B., Samanta, S. and Paul, A.K. (2015). Modeling long memory in volatility for spot price of lentil with multi-step ahead out-of-sample forecast using AR-FIGARCH model. *Econ. Affairs-Quarterly Journal of Economics*, **60(3),** 457-466.

Sarika, Iquebal, M.A. and Chattopadhyay, C. (2011). Modelling and forecasting of pigeonpea (Cajanus cajan) production using autoregressive integrated moving average methodology. *Ind. J. Agric. Sci.*, **81(6),** 520-523.

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

Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**, 159-175.