

A Comparative Study on FFNN and ARIMA Model in the Presence of Outliers

K.Senthamarai kannan
Professor
Department of Statistics
Manonmaniam Sundaranar
University

V.Deneshkumar
Research Scholar
Department of Statistics
Manonmaniam Sundaranar
University

S.Arumugam
CEO
Nandha Educational Institutions

ABSTRACT

Time series data mining (TSDM) techniques explores large amount of time series data in search of interesting relationships among variables. The TSDM methods overcome limitations including stationarity and linearity requirements of traditional time series analysis by adapting data mining concepts for analyzing time series data. The Feed Forward Neural Net is one of the most widely used neural nets. In this paper, the Feed Forward Neural Nets architecture is examined and compared with Statistical Time Series Auto regressive integrated moving average (ARIMA) model for prediction of agricultural production. The performance by ANN model and Time series model for prediction are examined using visualization technique and statistical test and the results are illustrated numerically and graphically

Keywords

Feed Forward Neural Nets, ANN, ARIMA, Outliers, Forecasting and MSE.

1. INTRODUCTION

Artificial neural networks (ANNs) have attracted increasing attentions in recent years for solving many real-world problems. Neural network computing is a key component of any data mining. Neural networks have been used in many business applications for pattern recognition, forecasting, prediction, and classification. ANN is an information processing paradigm that is inspired by the way of biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements working in unison to solve specific problems. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. As the neural network possesses segmentation and identification ability (Zhang et al., 1998) it is widely used in various aspects. For example it is helpful for credit card fraud, stock prices, interest rates and bankruptcy prediction, financial analysis, weather forecasting, judgment for medical images and fingerprint recognition system (Fish et al., (1995), Lee and Chiu, (2002)). Time series is a sequence of observations arranged in specific order of time. The main goals of time series analysis are identifying the nature of the phenomenon represented by the sequence of observations and Forecasting (predicting future values of the time series variable). Arrangement of statistical data in chronological order is known as "Time Series". Such series have a unique

important place in the field of Economic and Business statistics. An economist is interested in estimating the likely population in the coming year so that proper planning can be carried out with regard to food supply, job for the people etc. Similarly, a business man is interested in finding out his likely sales in the near future, so that the businessman could adjust his production accordingly and avoid the possibility of inadequate production to meet the demand. In this connection one usually deal with statistical data, which are collected, observed or recorded at successive intervals of time. Such data are generally referred as 'time series' data. The list of areas in which time series is observed and studied is endless. Outlier detection is important in many fields. In statistics, an outlier is a observation that is numerically far away from the rest of the data. The handling of outlying observations in a data set is one of the most important tasks in data pre-processing. For many data mining applications, finding the outliers is more interesting than finding the common patterns of the data. Data that have been incorrectly entered or that do not belong to the population from which the rest of the data came can bias estimates and give misleading results. Methods have been devised to identify outlier observations in a variety of situations. With recent advances in technology, scientists are collecting larger data sets and the analyst is getting further and further from the data or even sees every data point. So, it is important to have good methodology for dealing with outlier observations that might not be noticed in a typical data analysis.

Artificial neural networks (ANN) have been developed as generalizations of mathematical models of biological nervous systems. A first wave of interest in neural networks (also known as connectionist models or parallel distributed processing) emerged after the introduction of simplified neurons by McCulloch and Pitts (1943). A typical artificial neuron and the modelling of a multilayered neural network are illustrated in Figure 1 and Figure 2 the signal flow from inputs X_1, X_2, \dots, X_n is considered to be unidirectional, which are indicated by arrows, as is a neuron's output signal flow (O). The neuron output signal O is given by the following relationship

$$o = f(\text{net})f\left(\sum_{j=1}^n \omega_j x_j\right) \quad \dots (1)$$

where W_j is the weight vector, function $f(\text{net})$ is referred to as an activation (transfer) function.

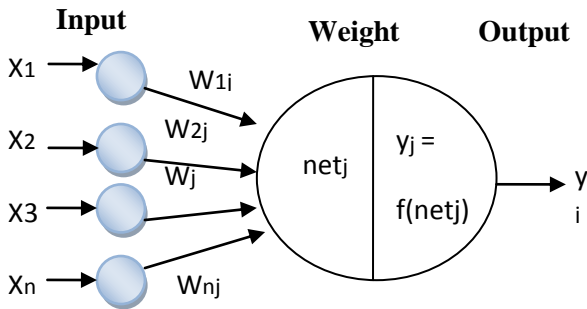


Figure1: Artificial Neuron

The variable (net) is defined as a scalar product of the weight and input vectors.

$$\text{net} = \omega^T x = \omega_1 x_1 + \dots + \omega_n x_n \quad \dots (2)$$

where T is the transpose of a matrix, and, in the simplest case, the output value O is computed as

$$o = f(\text{net}) = \begin{cases} 1 & \text{if } \omega^T x \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad \dots (3)$$

where θ is called the threshold level and this type of node is called a linear threshold unit.

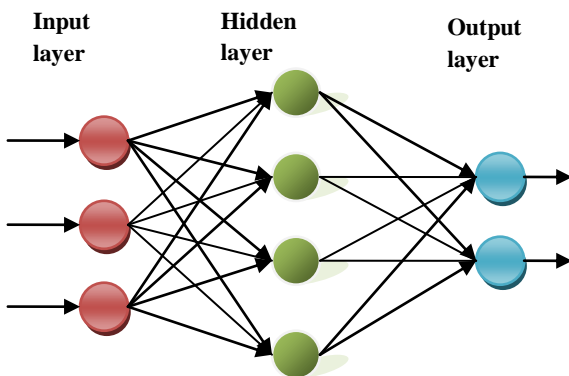


Figure 2: Feed forward multiayered perception

A neural network has three layers in its structure. First layer is input layer which is directly interact with external worlds, second layer is of hidden layer where computation is done according to function provided and the last layer is output layer from where we get output.

2. RELATED WORK

Dasgupta et.al., (1994) has compared the performance of two statistical market response models logistic regression model and discriminant analysis model to that of a back propagation neural network model. Chu and Widjaja (1994) have discussed a backpropagation based an appropriate forecasting method selection. Stern (1996) has introduced the principles

of the multilayer feedforward network which is the most commonly used neural networks in practical problems. Glorfeld (1996) has discussed a new methodology for the development of highly simplified backpropagation neural network models. Zhang et al., (1998) have discussed the back propagation neural network is a feed forward network and it is the most widely applied neural network technique in time series forecasting and business. According to the research of Vellido et al., (1999), 78% of the researcher from 1992 to 1998, used artificial neural network which has BPN for the commercial aspects. Also BPN has the advantage of a high learning accuracy and quick retrospect speeds, etc. hence BPN is used for an analytical tool in this study. Law (2000) has used the backpropagation model of ANN analysis for investigating the demand for tourists from Taiwan to Hong Kong. Hansen and Nelson (2002) has discussed extension methods of neural network for planning and budgeting. Petoskey (2003) has studied outlier detection algorithms used in data mining systems and discussed their advantages and disadvantages. Lin et al., (2004) have discussed the time series data mining for multimodal bio-signal data. They classified time series data mining into three large tasks: (1) subsequence matching (2) anomaly detection and (3) time series motif discovery. They also mentioned visualization aspects of time series data. Bodyanskiy et al.,(2006) have proposed a novel ANN architecture for prediction of time series containing an arbitrary number of quasiperiodic components with aliquant frequencies, which may vary in time. Premchand et al., (2006) has discussed two neural network models for cash forecasting for a bank branch. One is daily model – taking the parameter values for a day as input to forecast cash requirement for the next day and the other is weekly model, which takes the withdrawal affecting input patterns of a week to predict cash requirement for the next week. Cadenas and Rivera (2010) has proposed the Hybrid models consisting of Autoregressive Integrated Moving Average model and Neural Network model and developed for the wind speed forecasting in three different regions of Mexico. The Hybrid model could be a better alternative for the wind speed forecasting where linear and nonlinear tendencies be established. Tak Chung Fu (2011) has reviewed a comprehensive revision on the existing TSDM research. They are generally categorized into representation as indexing, similarity measures, segmentation, visualization and mining. More over the state-of-the-art research issues are also highlighted. The primary objective is to serve as a glossary for interested researchers to have an overall picture on the current TSDM development and identify their potential research direction to further investigation. Esling and agon (2012) has provided a survey of the several techniques applied for time-series data mining (TSDM). They presented the three core implementation components that constitute most of time-series systems, namely representation techniques, similarity measures, and indexing methods. Abhishek et al., (2012) have discussed the applicability of ANN approach by developing effective and reliable nonlinear predictive models for weather analysis also compare and evaluate the performance of the developed models using different transfer functions, hidden layers and neurons to forecast maximum, temperature for 365 days of the year. Chen et al., (2012) have discussed novel forecasting model based on empirical mode decomposition and proposed Neural Network to predict tourism demand. During the development of neural network, many sophisticated models have been proposed. They can be divided into three structural networks of learning strategies: supervised, non-supervised and associative learning. Among all the network models, the Back-propagation network (BPN)

of supervised learning is the most representative and most widely used.

3. METHODOLOGY

3.1 Forecasting procedure using artificial neural network

Step1: Input variable selection

Input variable production of sugarcane (in tones) in India is considered for the designing ANN for 60 observations from the years 1951 to 2011 time series data are initially chosen.

Step2: Data pre processing

Improperly recorded data and observation error are inevitable, hence, bad and abnormal data are identified and discarded or adjusted using a statistical method to avoid contamination of the model.

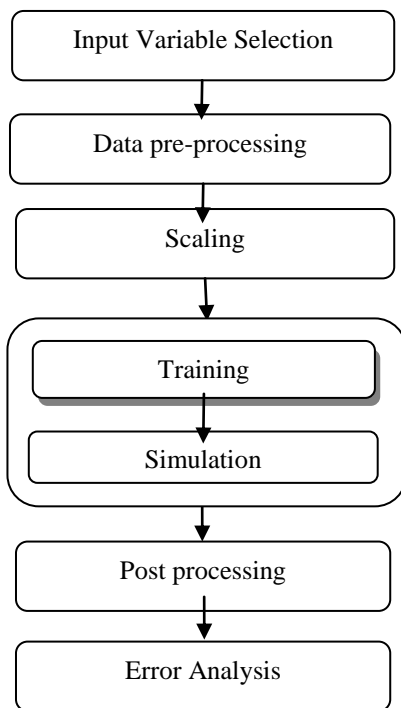


Figure 3: ANN based Forecasting

Step3: Scaling

Since the variables have very different ranges, the direct use of network data may use convergence problems. The scaling scheme all input and output variables are scaled to be in the [0,1] range, hence the input and output variables are scaled as follows:

$$x_i^{(k)'} = \frac{x_i^{(k)}}{\max(x_i^{(k)})}$$

$$o_i^{(k)'} = \frac{o_i^{(k)}}{\max(o_i^{(k)})}$$

Where k is the index of input and output vectors/patterns.

Step4: Training

Each layer's weights and biases are initialized when the neural network is setup. The network adjusts the connection strength among the internal network nodes until the proper transformation that links past inputs and outputs from the training cases is learned.

Step 5: Simulation

Using the trained neural network, the forecasting output is simulated using the input patterns.

Step6: Post processing:

The neural network output needs de scaling to generate the desired forecasted loads.if necessary, special events can be considered at this stage.

Step7: Error analysis

As characteristics of the load vary, error observations are important for the forecasting process. Hence, the following mean absolute percentage error ε and root mean square error σ are used here for after the fact error analysis

Standard statistical measures: $e_t = Y_t - F_t$

Y_t is actual observation for time period t and F_t is forecast for the same period

$$\text{Percentage error } PE_t = \left(\frac{Y_t - F_t}{Y_t} \right) * 100$$

$$\varepsilon = \frac{1}{n} \sum_{t=1}^n |PE_t| \quad \dots (4)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad \dots (5)$$

3.2 Evaluation criteria for ANN prediction

The correlation coefficient (R-value) has been widely used to evaluate the goodness-of-fit of model validation. This is obtained by performing a linear regression between the ANN-predicted values and the targets and is computed by

$$R = \frac{\sum_{i=1}^N t_i p_i}{\sqrt{\sum_{i=1}^n t_i^2} \sqrt{\sum_{i=1}^n p_i^2}} \quad \dots (6)$$

where

R- is correlation coefficient, N- is the number of samples

$$t_i = T - \bar{T}; p_i = P_i - \bar{P},$$

t_i and p_i are the target and predicted values for $i=1, \dots, N$ and \bar{T} and \bar{P} are the mean values of the target and predicted data set, respectively.

Here the R value is equal to 1 refers to a perfect correlation and the predicted values are either equal or very close to the target values, whereas there exists a case with no correlation between the predicted and the target values when R is equal to zero. Intermediate values closer to 1 indicate better agreement between target and predicted values.

3.3 Directional Measure

A typical measure here is to use a 2×2 contingency table that summarizes the number of “upward” and “downward” of the model in predicting ups and downs of $X_{T+\ell}$ in the forecasting subsample. Specifically, the contingency table is given as

Actual	Predicted		
	up	down	
up	m_{11}	m_{12}	m_{10}
down	m_{21}	m_{22}	m_{20}
	m_{01}	m_{02}	m

where

m is the total number of l -step ahead forecasts in the forecasting subsample,

m_{11} is the number of “upward” in predicting upward movements,

m_{22} is the number of “downward” in predicting downward movements and so on. Larger values in m_{11} and m_{22} indicate better forecasts

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(m_{ij} - m_{i0}m_{0j}/m)^2}{m_{i0}m_{0j}/m} \dots (7)$$

The test statistic can be used to evaluate the performance of the model. A large χ^2 signifies that the model outperforms the chance of random choice. Under some mild conditions, χ^2 has an asymptotic chi-squared distribution with 1 degree of freedom.

3.4 Statistical Time Series Model

Time Series analysis Autoregressive Integrated Moving Average (ARIMA) model was developed by Box and Jenkins (1976). It consists of three parts auto regression AR(p), moving averages MA(q) and differencing to the series (d) and forms ARIMA (p,d,q). A time series is said to be stationary if its underlying generating process is based on constant mean and constant variance with its Autocorrelation Function (ACF). The Autocorrelation refers to the way the observation in a time series are related to each other and is measured by the simple correlation between the current observation (Y_t) and observation from p periods before the current one (Y_{t-p}). The ACF is defined by,

$$r_p = \frac{\sum_{t=p+1}^n (Y_t - \bar{Y})(Y_{t-p} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \dots (8)$$

That is for a given series Y_t , autocorrelation at lag p =correlation (Y_t, Y_{t-p}) and is given by r_p . It ranges from -1 to +1. Box-Jenkins has suggested that the maximum of useful r_p are roughly $N/4$, where N is the number of periods upon which information on Y_t is available. The Partial Autocorrelation Function (PACF) is used to measure the degree of association between Y_t and Y_{t-p} , when the Y -effects at other time lags $1,2,3 \dots p-1$ are removed.

ARIMA

Autoregressive Integrated Moving average process

$$\phi_p(B)Y_t \Delta^d = \theta_q(B)\epsilon_t \dots (9)$$

where

$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ is the autoregressive process of order p .

$\theta_q(B) = 1 - \theta_1(B) - \dots - \theta_q B^q$ is the moving average process of order q . and Δ is the difference operator.

Additive Outlier (AO)

The Additive outlier which is also known as Type I outlier (Fox, 1972). An AO only affects a single observation, which is either smaller or larger in value compared to the expected values in the data. After this disturbance, the series returns to its normal path as if nothing has happened.

The effect of an additive outlier is independent of the ARIMA model and is bounded. Assuming that an AO outlier occurs at time $t=T$, the observed series can be represented as

$$Y(t) = U(t) + \omega I_T(t) \dots (9)$$

Where $I_T(t) = \begin{cases} 0 & t \neq T \\ 1 & t = T \end{cases}$ is a pulse function and ω is the deviation from the true $U(T)$ caused by the outlier. The true rule suggests that the shock caused by an AO affects the original observation at $t = T$ only with the magnitude of ω and the rest remained unaffected.

Innovational Outlier (IO)

In contrast to the AO, innovational outlier is known as Type II outlier that affects several observations. An AO affects only one residual, at the date of the outlier. The effect of the IO on an observed series consists of an initial shock that propagates in the subsequent observations with the weights of the moving average (MA) representation of the ARIMA model. Assuming that an IO outlier occurs at time $t=T$, then

$$Y(t) = \mu(t) + \frac{\theta(B)}{\Delta\phi(B)} (a(t) + \omega I_T(t)) \quad \dots (10)$$

4. RESULT AND DISCUSSION

The Production of sugarcane (in tones) in India is considered for the model building for 60 observations from the years 1951 to 2011 are used for fitting the both ANN and ARIMA model. Its aim is to find out the more accurate model for forecasting agricultural production demand.

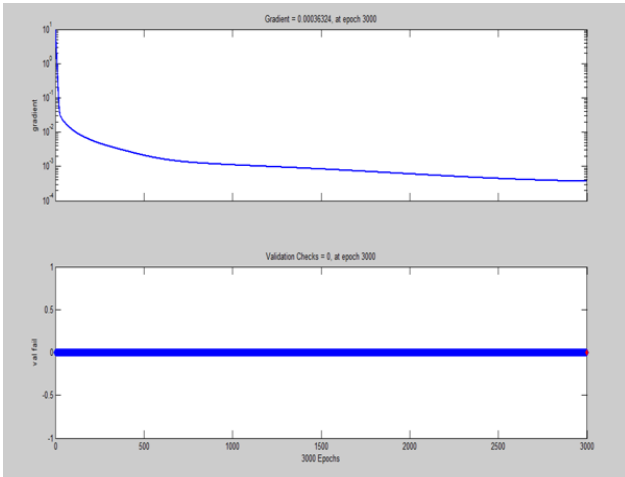


Figure 4: Training state diagram

After the initial value of premise parameter and the architecture of the predictive model were determined, the network was trained by the backpropagation algorithm. The figure 4 shows that variation in gradient coefficient with respect to number of epochs. The final value of gradient coefficient at epoch number 3000 is 0.0029628 which is near to zero. Minimum the value of gradient coefficient will be better training and testing of networks.

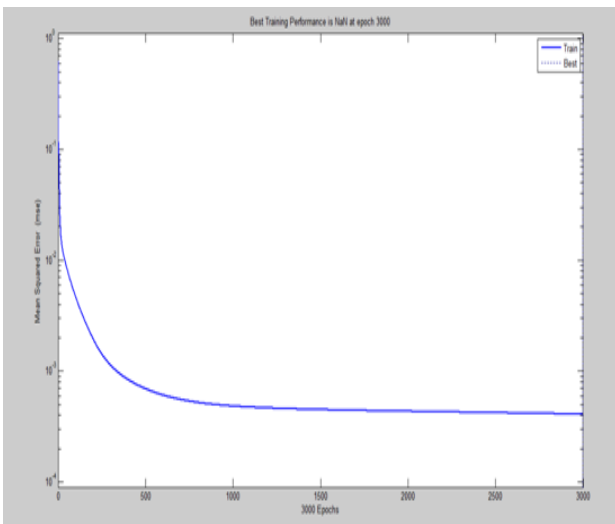


Figure 5: Mean square error vs Epochs

The figure 5 shows variation in mean square error with respect to epochs. It is clear from figure that as increasing the

number of epochs for training and testing the error rate decreasing.

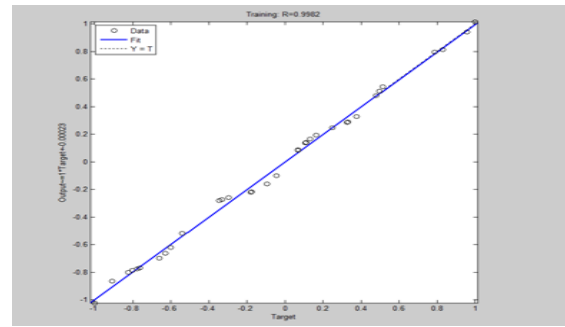


Figure 6: Train performance of training data by FFNN predictive model

The above figure shows that FFNN predictions for the training data and R value was 0.9962 from figures 4 and 5, a good agreement is achieved in that the predicted values of the network were able to follow the desired values well, which indicates that the network has a strong learning capability.

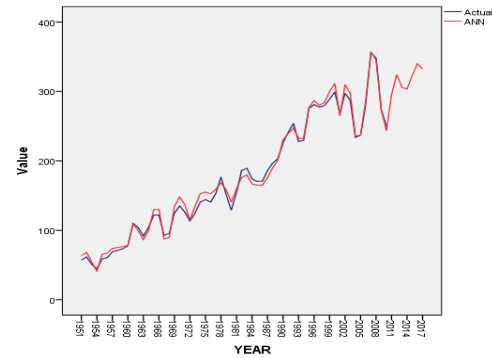


Figure 7: ANN-Actual and Forecasting Production of Sugarcane

4.1 Testing relative forecast accuracy using contingency tables.

Figure 7 is obtained by the method of obtained by feed forward neural network. From the figure 7, table 1 showing the forecasted values of 2 × 2 contingency table for the combinations of downward and upward comparisons in each 60 years of the sample period (1951 to 2011) is estimated.

Table 1: Actual vs ANN Crosstabulation

		ANN		Total
		Down	Up	
Actual	Down	20	2	22
	up	1	37	38
Total		21	39	60

From the above table shows that 39 upward differential values and 21 downward values actual observations out of 60 observations. The chi-squared statistic of the table is (Pearson Chi-Square) 47.72 with p-value 0.00 significantly above the

1% critical value and indicating that reject the null hypothesis that the forecasts and observations are independent.

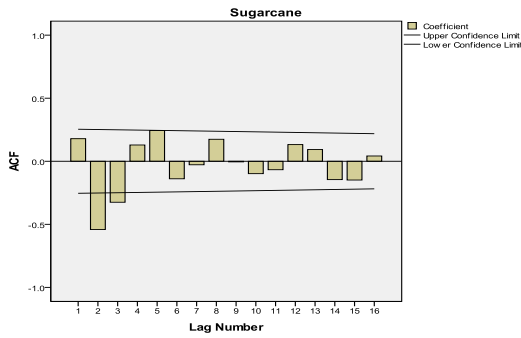


Figure: 8 ACF of Sugarcane

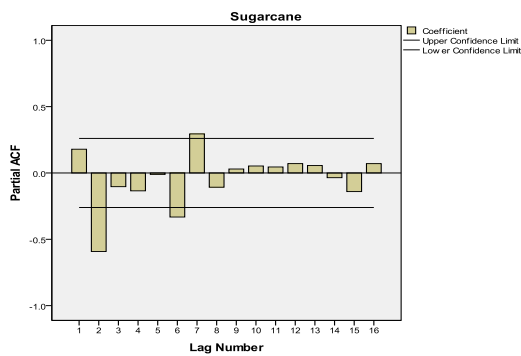


Figure 9: PACF of Sugarcane

From the ACF and PACF given in the model (2,1,1) is suitable model for this series. Hence ARIMA (2,1,1) is fitted for this time series. The co-efficient and parameters are given in the table (2). The fit of the series

$$\left[1 - (1 + \phi_1)B + (\phi_1 - \phi_2)B^2 + \phi_2 B^3\right] = e_t - \theta_1 e_{t-1}$$

Figure 8 represents SPSS output for the ACF and PACF of the original sugarcane production shows particular pattern to understand the behavior of the series it would be helpful to analyze the data series.

Table 2: Model Statistics of Sugarcane production

Model	Stationary R-squared	Normalized BIC	No.of Outliers
Sugarcane	0.628	5.694	2

The model, ARIMA (2, 1, 1) is found as the best fit for the production of sugarcane with BIC=5.694. The productions of actual sugarcane and forecast are given in the figure 10 at 95% confidence limit.



Figure 10: ARIMA- Actual and Forecasting Production of Sugarcane

Table 3: Forecasted values of Sugarcane Production

YEAR	2013	2014	2015	2016	2017
ARIMA	291.78	321.89	344.36	337.39	322.59
ANN	306.03	303.62	323.21	340.0	332.45
UCL	331.26	362.13	390.21	387.35	373.10
LCL	252.31	281.66	298.52	287.44	272.09

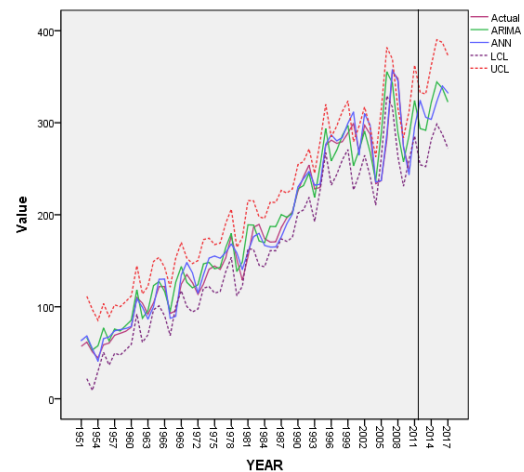


Figure 11: ARIMA and ANN- Actual and Forecasting Production of Sugarcane

Table 4: Forecast Accuracy of Agriculture production

Production	Error	ARIMA	ANN
Sugarcane	RMSE	17.261	7.04

The table 4 reveals that the RMSE is between the ARIMA and ANN models representing its forecast effect is normal. They predict the future values in an appropriate way. The RMSE of ARIMA is the highest representing and its forecast effect is the worst.

5. CONCLUSION

Agricultural industry has become an important industry globally. Agricultural production has a growing trend in the future, thus every country put more importance on the agricultural industry and the forecast demand which becomes more and more essential. We have examined possibility of FFNN and tried to find an appropriate way of applying them into agricultural production time series data. In this paper, two forecast models Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN) are used for analysis. These results show that forecasting based on ANN model gives better forecast than the ARIMA model.

6. ACKNOWLEDGMENTS

The first author thanks the University Grants Commission, New Delhi for providing research fellowship under the scheme of BSRF to carry out this work.

7. REFERENCES

- [1] Abhishek.K, Singha.M.P, Ghosh.S, Anand.A.,2012. Weather forecasting model using Artificial Neural Network, *Procedia Technology*, Vol.4, pp.311 – 318.
- [2] Ananda Rao, M. and Srinivas J., 2002. *Neural Networks Algorithms and Applications*, Narosa Publishing House Pvt.Ltd, New Delhi.
- [3] Barnett. V. and Lewis.T, 1984.Outliers in statistical data (3rd ed.), New York: John Wiley & Sons.
- [4] Bodyanskiy.Y and Popov.S. 2006. Neural network approach to forecasting of quasiperiodic financial time series, *European Journal of Operational Research*, Vol.175, pp.1357–1366.
- [5] Box, G.E.P. and Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control*, Second Edition, Holden Day.
- [6] Cadenas.E and Rivera.W.,2010.Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model, *Renewable Energy*, Vol.35, pp.2732-2738.
- [7] Chen, C.F, Lai, M.C and Yeh .C.C., 2012. Forecasting Tourism Demand Based on Emprirical Mode Decomposition and Neural Network. *Knowledge Based Systems*, Vol.26, pp.281-287.
- [8] Chu, C.H and Widjaja. D, 1994. Neural Network system for forecasting method selection, *Decision Support Systems*, Vol.12, pp13-24.
- [9] Dasgupta, C.G., Dispensa,G., and Ghose,S. 1994. Comparing the Predictive Performance of a Neural Network Model with Some Traditional Market Response Models. *International Journal of Forecasting*, Vol.10, pp.235-244.
- [10] Esling, P. and Agon, C., 2012. Time-Series data mining. *ACM Computing Surveys*, Vol. 45, No.1, Article 12, pp.1-34.
- [11] Fish, K. E., Barnes, J. H. and Aiken, M. W. 1995. Artificial Neural Networks: A New Methodology for Industrial Market Segmentation. *Industrial Marketing Management*, Vol.24, pp.431-438.
- [12] Freeman, J. A. and Skapura, D. M. 1992. *Neural Networks Algorithms, Applications, and Programming Techniques*. Addison-Wesley Publishing Company.
- [13] Glorfeld, L.W.1996. A Methodology Based for Simplification and Interpretation of Backpropagation-Based Neural Network Models, *Expert Systems with Applications*, Vol.10, pp.37-54.
- [14] Hawkins. D.M. 1980. *Identification of outliers*, Chapman and Hall.
- [15] Hansen.J.V and Nelson.R.D. 2002. Data mining of Time Series using stacked generalizers, *Neurocomputing*, Vol. 43, pp.173-184.
- [16] Last.M, Kandel.A and Bunke.H, 2004. *Data Mining in Time Series Databases*, World Scientific Publishing.
- [17] Law.R., 2000. Back Progration Larning in Improving the Accuracy of Neural Network Based Tourism Demand Forecasting, *Tourism Management*, Vol.21, pp.331-340.
- [18] Lee, T. S. and Chiu, C. C., 2002. Neural Network Forecasting of an Opening Cash Price Index. *International Journal of Systems Science*, Vol.33, No.3, pp.229-237.
- [19] Lin, J., Keogh, E., Lonardi, S., Lankford, J. P., and Nystrom, D. M. Visually., 2004. Mining and Monitoring Massive Time Series, *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, Seattle, WA, pp.460-469.
- [20] Pankratz, A., 1983. *Forecasting with univariate Box-Jenkins models: Concepts and cases*, John Wiley & sons, New York.
- [21] Petrovskiy, M.I. 2003. *Outlier Detection Algorithms in Data Mining Systems*, Programming and Computer Software, Vol.29, No.4, pp.228–23.
- [22] Premchand.K and Wekta.W, 2006. Cash Forecasting: An Applications of Artificial Neural Networks in Finance, *International Journal of Computer science and applications*, Vol.3, No.1, pp.61-77.
- [23] Stern.H.S. 1996. Neural Network in Applied statistics, *Technometrics*, Vol.3, No.3, pp.205-216.
- [24] Tak Chung Fu, 2011. A review on time series data mining, *Engineering Applications of artificial Intelligence*, Vol.24, pp.164-181.
- [25] Vellido, A., Lisboa, P. J. G. and Vaughan, J., 1999. Neural Networks in Business: A Survey of Applications (1992-1998), *Expert Systems with Applications*, Vol.17, pp.51-70.
- [26] Zhang. G., Patuwo. B. E. and Hu. M. Y., 1998. Forecasting with Artificial Neural Networks: The State of the Art, *International Journal of Forecasting*, Vol.14, pp.35-62.
- [27] Fox, A.J. 1972. Outliers in Time series, *Journal of the Royal Statistical Society*, Vol.17, No.55, pp.559-567.