



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Forecasting of the Rice Yields Time Series Forecasting using Artificial Neural Network and Statistical Model

¹A. Shabri, ²R. Samsudin and ¹Z. Ismail

¹Department of Mathematic, Faculty of Science,

²Department of Software Engineering, Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia, 81310, Skudai, Malaysia

Abstract: Accurate forecasting of the rice yields is very important for the organization to make a better planning and decision making. In this study, a hybrid methodology that combines the individual forecasts based on artificial neural network (CANN) approach for modeling rice yields was investigated. The CANN has several advantages compared with conventional Artificial Neural Network (ANN) model, the statistical the autoregressive integrated moving average (ARIMA) and exponential smoothing (EXPS) model in order to get more effective evaluation. To assess the effectiveness of these models, we used 38 years of time series records for rice yield data in Malaysia from 1971 to 2008. Results show that the CANN model appears to perform reasonably well and hence can be applied to real-life prediction and modeling problems.

Key words: ARIMA, exponent smoothing, artificial neural network, combining forecasting, rice yields

INTRODUCTION

Almost 90% of rice is produced and consumed in Asia and 96% in developing countries. In Malaysia, the Third Agriculture Policy (1998-2010) was established to meet at least 70% of Malaysia's demand a 5% increase over the targeted 65%. The remaining 30% comes from imported rice mainly from Thailand, Vietnam and China. Raising level of national rice self-sufficiency has become a strategic issue in the agricultural ministry of Malaysia. The ability to forecast the future enables the farm managers to take the most appropriate decision in anticipation of that future.

The accuracy of time series forecasting is fundamental to many decisions processes. One of the most popular and commonly used models for the forecasting research and practice is the autoregressive integrated moving average (ARIMA) model. The popularity of the ARIMA model is due to its statistical properties as well as the well known Box-Jenkins methodology (Zhang, 2003). In the ARIMA models, the desired forecasting is generally expressed as a linear combination. But real world time series are often full of nonlinearity and irregularity. Consequently, the forecasting result predicted by the ARIMA model is generally not very accurate.

The exponential smoothing (EXPS) model has been found to be one of the effective forecasting methods and has been widely used in time series forecasting (Hyndman *et al.*, 2002; De Gooijer and Hyndman, 2006; Co and Boosarawongse, 2007). This method is most effective when the components describing the time series may be changing slowly over time (Cho, 2003). However, the exponential smoothing method is only a class of linear model and thus it can only capture linear feature of data time series. Obviously, the approximation of linear models to complex real-world problems is not always sufficient. Hence, it is necessary to consider other nonlinear methods to complement the exponential smoothing model (Lai *et al.*, 2006).

In recent years, an Artificial Neural Network (ANN) has been increasingly used in the analysis of time series forecasting because of its excellent and powerful learning capability of modeling nonlinearity (Zhang *et al.*, 1998; Maier and Dandy, 2000; Tareghian and Kashefipour, 2007). The ANNs are flexible computing frameworks for modeling a broad range of nonlinear problems. One significant advantage of the ANN models over other classes of nonlinear model is that ANNs are universal which can approximate a large class of functions with a high degree of accuracy. The ARIMA models and ANN are often compared with mixed conclusions in terms of

superiority in forecasting performance. Survey of the literature shows that both ARIMA and ANN models have performed well in different cases (Zhang *et al.*, 1998; Hill *et al.*, 1996). Co and Boosarawongse (2007) compared the performance of ANN with EXPS and ARIMA models in forecasting rice exports from Thailand.

Several empirical studies have already suggested that by combining several different models, forecasting accuracy can often be improved over the individual model. Zhang *et al.* (1998) proposed a hybrid approach to time series forecasting using ARIMA and ANN models. Valenzuela *et al.* (2008) proposed a hybridization of intelligent techniques such as ANNs, fuzzy systems and evolutionary algorithms. Ince and Trafalis (2006) proposed a two stage forecasting model which combining parametric techniques such as ARIMA, vector autoregressive (VAR) and co-integration techniques and nonparametric techniques such as support vector regression and ANN. There are many combination methods have been proposed to improve the forecast accuracy such as a simple average method, weighted average method (Zou *et al.*, 2007), a neural network method (Sfetsos and Siriopoulos, 2004) and the self-organizing data mining algorithms (He and Xu, 2005). Combining forecasting based on artificial neural network (CANN) has recently attracted the interest of researchers (Cho and Yoon, 1997; Ruhaidah *et al.*, 2008). Experimental results have shown that this kind of combination can increase the fitting accuracy and the reliability of forecasting. Every model has some degree of uncertainty, including structure uncertainty and parameter uncertainty. However, by combining individual models in a proper way, the fully information from each model can be used and the overall performance of forecasts can be improved.

The main purpose of this study is to investigate the applicability and capability of the combining forecasting based on artificial neural network compared with the individual forecast ARIMA, EXPS and ANN methods for modeling of rice yields time-series forecasting. To verify the application of this approach, the rice yields data from MADA in Peninsular Malaysia are chosen as the case study.

MATERIALS AND METHODS

The autoregressive integrated moving average model: Box and Jenkins developed ARIMA models to become popular at the beginning of the 1970s. The models are a class of linear models that were designed for linear time series (Box *et al.*, 1994). ARIMA models that are compound of a seasonal and non-seasonal part are represented by the following way:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)a_t$$

where, ϕ_i ($i = 1, 2, \dots, p$), θ_j ($j = 1, 2, \dots, p$), Φ_l ($l = 1, 2, \dots, p$) and Θ_k ($k = 1, 2, \dots, Q$) are model parameters. p order of nonseasonal auto regression; d number of regular differencing; q order of the nonseasonal moving average; P order of seasonal auto regression; D number of seasonal differencing; Q order of seasonal moving average; and s length of season. The general non-seasonal model is known as ARIMA (p, d, q) can be written as the linear expression

$$y_t^{BJ} = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + \epsilon_j \quad E = \frac{1}{2} \sum_k (t_{kj} - o_{kj})^2 \quad (1)$$

The exponential smoothing model: In the application of the exponential smoothing (EXPS) model, there are three types of models that are widely used in different time series. Simple exponential smoothing is used when the time series has no trend, double exponential smoothing for handling a time series that displays a slowly changing linear trend and the Winters' method which is an exponential smoothing approach to predicting seasonal data. In this study, the double exponential smoothing is suitable for this time series forecasting for the specified time period. The model employs a level component and a trend component at each period. It uses two weights, or smoothing parameters to update the components at each period. The double exponential smoothing equations are

$$\begin{aligned} y_t^{EXPS} &= L_{t-1} + T_{t-1}, \quad t = 1, 2, \dots, n \\ L_t &= \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad 0 < \alpha \leq 1 \\ T_t &= \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \end{aligned}$$

where, L_t is the level at time t , α is the weight for the level, T_t is the trend at time t , γ the weight for the trend, Y_t is the data value at time t and \hat{Y}_t is the fitted value.

The artificial neural network forecasting model: An Artificial Neural Network (ANN) is a computer system that simulates the learning process of human brain. The greatest advantage of a neural network is its ability to model complex nonlinear in the data series. The basic architecture consists of three types of neuron layers: input, hidden and output layers. The ANN model performs a nonlinear functional mapping from the input observations ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) to the output value (y_t), i.e.,

$$zy_t = a_0 + \sum_{j=1}^q a_j f(w_{0j} + \sum_{i=1}^p w_{ij} y_{t-i}) + \epsilon_t \quad (2)$$

where, a_j ($j = 0, 1, 2, \dots, q$) is a bias on the j th unit and w_{ij} ($i = 0, 1, 2, \dots, p; j = 0, 1, 2, \dots, q$) is the connection

weights between layers of the model, $f(\cdot)$ is the transfer function of the hidden layer, p is the number of input nodes and q is the number of hidden nodes (Lai *et al.*, 2006). The activity function utilized for the neurons of the hidden layer was the logistic sigmoid function that is described by:

$$f(x) = \frac{1}{1 + e^{-x}}$$

This function belongs to the class of sigmoid functions and has advantages characteristics such as being continuous, differentiable at all points and monotonically increasing.

Training a network is an essential factor for the success of the neural networks. Among the several learning algorithms available, back-propagation (BP) has been the most popular and most widely implemented learning algorithm for all neural network paradigms (Zou *et al.*, 2007; Baroutian *et al.*, 2008). The procedure of the BP is repeated by adjusting the weights of the connection in the network so as to minimize the error. The gradient descent method is utilized to calculate the weight of the network and adjust the weight of interconnection to minimize the sum-squared error, SSE, of the network which is given by:

$$SSE = \frac{1}{2} \sum_k (y_k - \hat{y}_k)^2 \tag{3}$$

where, y_k and \hat{y}_k are the true and predicted output vector of the k th output node.

Actually, the ANN model in Eq. 2 performs a nonlinear functional mapping from the past observation ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) to the future value (y_t), i.e.,

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w) + \epsilon_t \tag{4}$$

where, w is a vector of all parameters and f is a function determined by the network structure and connection weights.

Thus, in some senses, the ANN model is equivalent to a nonlinear autoregressive (NAR) model. A major advantage of neural networks is their ability to provide flexible nonlinear mapping between inputs and outputs. They can capture the nonlinear characteristics of time series well.

RESULTS AND DISCUSSION

In order to validate the combining forecasting methods for rice yields modeling, the data were collected from Muda Agricultural Development Authority (MADA)

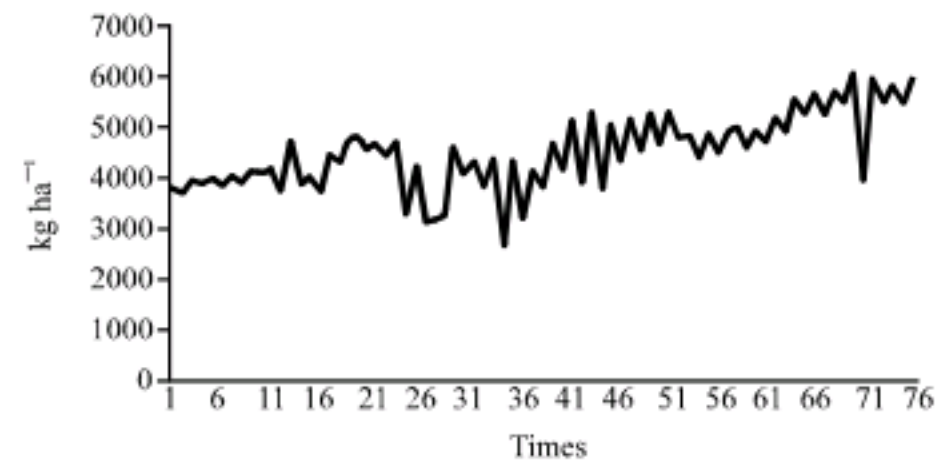


Fig. 1: Rice yields series (1971-2008)

Kedah, Malaysia. The rice yields data contains the yields data from 1995 to 2001, giving a total of 78 observations. The time series plot is given in Fig. 1.

To assess the forecasting performance of different models, each data set is divided into two samples. The first series was used for training the network (modeling the time series) and the remaining were used for testing the performance of the trained network (forecasting). We take the data from 1971 to 2001 producing 63 observations for training purpose and the remainder as the output sample data set with 15 observations for forecasting purpose.

The performances of the each model for both the training data and forecasting data are evaluated and selected according to the Mean Absolute Error (MAE) and root-mean-square error (RMSE), which are widely used for evaluating results of time series forecasting. The MAE and RMSE are defined as:

$$MAE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}$$

where, y_t and \hat{y}_t are the observed and the forecasted rice yields at the time t , N equals the number of observations. The criterions to judge for the best model are relatively small of MAE and RMSE in the modeling and forecasting.

Fitting ARIMA models to the data: The plots in Fig. 2 shows the plots of the rice yield time series indicate that the time series are non-stationary in mean and variance. The first difference was applied in order to remove the trend.

The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for the first difference series are plotted in Fig. 2a and b. The plot shows that there is seasonality in the series. We find the first-difference series becomes stationary. For ACF, the

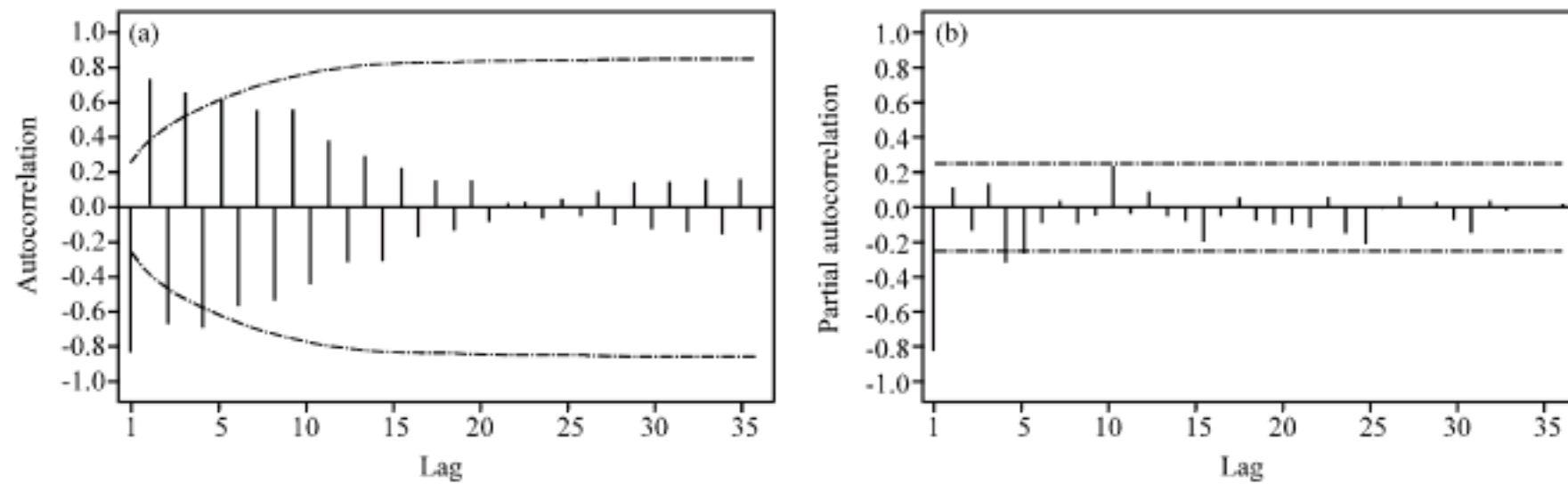


Fig. 2: (a) ACF and (b) PACF for differenced series of natural logarithms

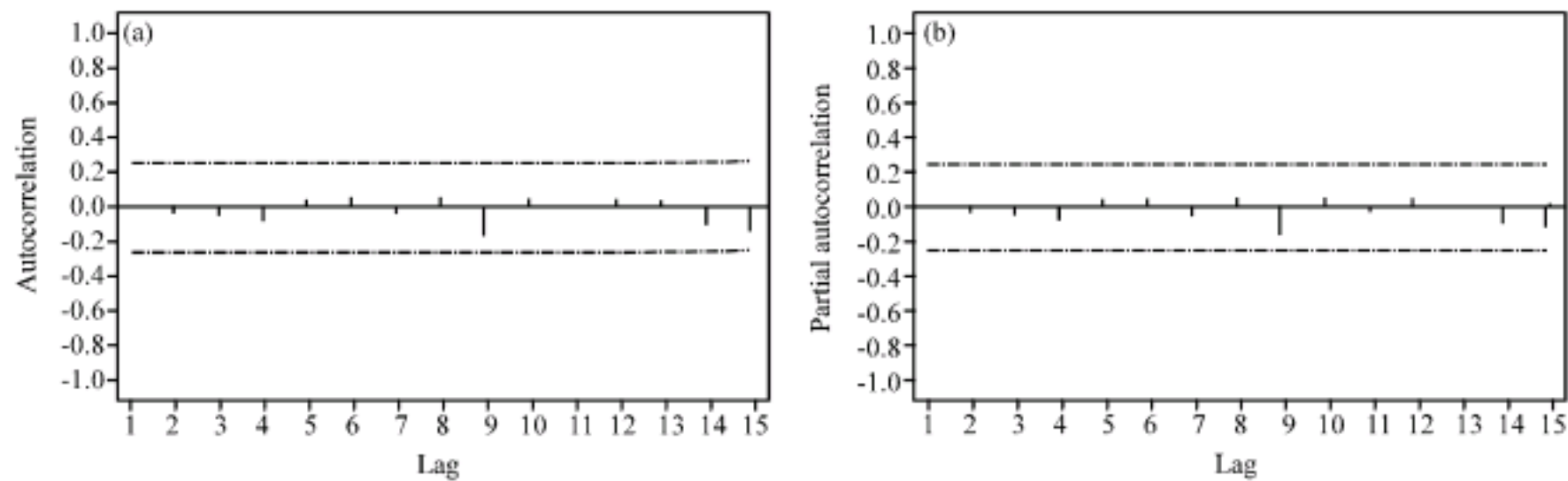


Fig. 3: (a) ACF and (b) PACF of residuals for ARIMA (1,1,1)x(1,0,1)₅

Table 1: Comparison of ARIMA models' statistical results

ARIMA model	Training	
	RMSE	MAE
(1,1,0)x(1,0,0) ₅	372.431	0.0676
(1,1,0)x(1,0,1) ₅	360.963	0.0664
(1,1,1)x(1,0,0) ₅	372.085	0.0677
(1,1,1)x(1,0,1) ₅	360.958	0.0664

series tends to oscillate rapidly, while for PACF, we observe major spikes at lags 1 and 5.

Several models were identified and the statistical results during training are compared in the Table 1.

The criteria to judge for the best model based on MAE and MSE show that the ARIMA (1,1,1)x(1,0,1)₅ is a relatively best model. This model has both non-seasonal and seasonal components.

The ACF and PACF plots of residuals for the rice yields series are shown in Fig. 3a and b. From the residual plot of the best ARIMA model, it was observed that the selected ARIMA (1,1,1)x(1,0,1)₅ model passed the diagnostic checks and they were all white noise.

Fitting exponential smoothing models to the data: The computer software, MINITAB, which is a powerful statistical package was used to analysis the double exponential smoothing models. In order to find the optimal values for the parameter of α and γ , the values that

provides the smallest Sum of Square Errors (SSE) by grid search is selected. The grid search chooses a combination for the two parameters by employing a method of trial and error. The parameters that produce the smallest SSE is recorded and used to formulate models for the series. In this study, the optimal parameters of the model are $\alpha = 0.3893$ and $\gamma = 0.0493$ and Table 3 illustrate the results.

Fitting neural network models to the data: One of the key tasks in time series forecasting is the selection of the input variables and the number of neurons in the hidden layer. For the ANN models, there is no systematic approach which can be followed. The universal approximation theorem shows that a neural network with a single hidden layer with a sufficiently large number of neurons can in principle relate any given set of inputs to a set of outputs to an arbitrary degree of accuracy. As a result, the ANN designed in this study are equipped with one single hidden layer. The determination of the number of neurons in the hidden layer is more art than science.

Determining the size of the network (the number of neurons) has important consequences for its performance. Too small a network may not reach an acceptable level of accuracy. Too many neurons may result in an inability for the network to generalize (it may rote learn the training patterns).

Table 2: Comparison of neural network model

No. of neurons			Training	
Input	Hidden	Output	RMSE	MAE
2	2	1	402.592	0.0741
2	4	1	398.529	0.074
3	3	1	395.419	0.072
3	6	1	370.818	0.0684
5	5	1	380.125	0.0722
5	10	1	324.715	0.0602

Table 3: Comparing forecasting performance

Model	Training		Forecasting	
	RMSE	MAE	RMSE	MAE
ARIMA	360.958	0.0664	781.462	0.1363
EXPS	334.851	0.0627	736.449	0.1226
ANN	324.715	0.0602	740.977	0.122
CANN	257.741	0.0489	689.152	0.1135

The minimal errors are highlighted in bold

Since, the number of inputs varies depending on the input determination method used, it is not possible to use the same network architecture for each model. In this study the number inputs (I), 2, 3 and 5 were used for all data set. To help avoid the over fitting problem, some researchers have provided empirical rules to restrict the number of hidden nodes. To select an appropriate architecture, the guidelines using 2I proposed by Maier and Dandy (2000) and I proposed by Tang and Fishwick (1993).

The network was trained for 5000 epochs using the back-propagation algorithm with a learning rate of 0.001 and a momentum coefficient of 0.9. The networks that yielded the best results for the forecasting set were selected as the best ANN for the corresponding series. The experiment is repeated 10 times and afterwards the average of RMSE and MAE are computed. Table 2 shows overall summary statistics for each ANN model in training model selection results for the rice yields.

In this study, a trial and error method is performed to optimize the number of neurons in the hidden layer. Table 2 shows the performance of ANN varying with the number of neurons in the hidden layer. The best neural networks structure found consists of the layers 5-10-1.

Combination model: In this study the outputs y_t^{BJ} , y_t^{EXP} and y_t^{ANN} of the above three individual forecasting models are used as the input nodes. For the artificial neural network combination (CANN) forecasting model, the model has the form:

$$y_t^{CANN} = f(y_t^{BJ}, y_t^{EXP}, y_t^{ANN}) + \epsilon_t \text{ (No explicit form)}$$

Three inputs with two values (3 and 6) for the number of neurons in the single hidden layer were tried in order to investigate the effects of model complexity on forecasting performance.

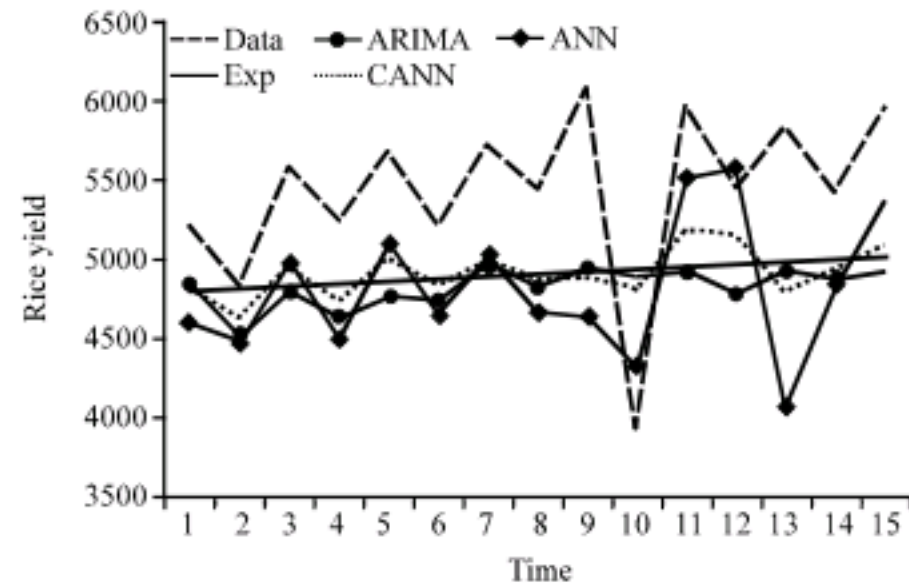


Fig. 4: The forecasted and actual values of the rice yields data

Table 3 reports the training and forecast results MAE and RMSE for each of the individual forecasts and a combined forecast for the rice yields data. A comparison between the actual value and the forecast value for the all models is given in Fig. 3.

Based on the MAE and RMSE, it is clear that the combined forecasts based on ANN (CANN) are more accurate than any of the individual forecasts. Figure 4 graphically supports this result. For individual forecasts, it was found that ANN outperforms EXPS and ARIMA.

It has been shown in recent statistical studies that combining forecasts may produce more accurate forecasts than individual ones. However, the literature on combining forecasts has almost exclusively focused on linear combining forecasts and most of them are designed to combine the similar method. The issue and methods of nonlinear combining forecasts have not yet been fully explored and not wide, even through forecast improvements may be possible using nonlinear combination techniques.

The outcome of the comparison proved that nonlinear methods are superior combining forecasts and neural network is an effective method in solving non-linear problems. The result suggests that the ANN method can be used as alternative to conventional linear combing methods to achieve greater forecasting accuracy.

CONCLUSIONS

This study examines the forecasting performance of combining forecasting based on artificial neural network (CANN) model compared with others statistical model (ARIMA and EXPS) and ANN model in forecasting the rice yields in Peninsular Malaysia. The case study on rice yields training and forecasting demonstrated that the fitting accuracy of CANN is superior to those of the conventional ARIMA, EXPS and ANN. Generally, the

CANN model is robust in the modeling of complex nonlinear structures, but its superiority may be subtle in simple structure models. The results show that the CANN model can be applied successfully to establish time-series forecasting models, which could provide accurate forecasting and modeling of time-series. The performance of the ANN largely depends on the network architecture. For the individual forecasts, results show the ANN models forecasts are considerably more accurate than the traditional ARIMA and EXPS models which used as a benchmark.

ACKNOWLEDGMENTS

The research is supported by Minister of Science, Technology and Innovation (MOSTI), Malaysia with Grant Vote No. 79346. The authors would like to thank lecturers and friends for many helpful ideas and discussions. We also wish to thank Muda Agricultural Development Authority (MADA), Ministry of Agricultural and Agro Based Industry, Malaysia for providing necessary data in this study.

REFERENCES

- Baroutian, S., M.K. Aroua, A.A. Raman and N.M. Sulaiman, 2008. Prediction of palm oil-based methyl ester biodiesel density using artificial neural networks. *J. Applied Sci. Data*, 8: 1938-1943.
- Box, G.E.P., G.M. Jenkins and G.C. Reinsel, 1994. *Time Series Analysis: Forecasting and Control*. 3rd Edn., Prentice Hall, New Jersey, ISBN-13: 978-0130607744.
- Cho, S. and S. Yoon, 1997. Reliable roll force prediction in cold mill using multiple neural networks. *IEEE Trans. Neural Networks*, 8: 874-882.
- Cho, V., 2003. A comparison of three different approaches to tourist arrival forecasting. *Tourism Manage.*, 24: 323-330.
- Co, H.C. and R. Boosarawongse, 2007. Forecasting Thailand's rice export: Statistical techniques vs artificial neural networks. *Comput. Ind. Eng.*, 53: 610-627.
- De-Gooijer, J.G. and R.J. Hyndman, 2006. 25 years of time series forecasting. *Int. J. Forecasting*, 22: 443-473.
- He, C. and X. Xu, 2005. Combination forecasts using self-organizing algorithms. *J. Forecasting*, 24: 269-278.
- Hill, T., M. O'Connor and W. Remus, 1996. Neural network models for time series forecasts. *Manage. Sci.*, 42: 1082-1092.
- Hyndman, R.J., A.B. Snyder and S. Grose, 2002. A state space framework for automatic forecasting using exponential smoothing methods. *Int. J. Forecasting*, 18: 439-454.
- Ince, H. and T.B. Trafalis, 2006. A hybrid for exchange rate prediction. *Decision Support Syst.*, 42: 1054-1062.
- Lai, K.K., L. Yu, S. Wang and W. Huang, 2006. Hybridizing Exponential Smoothing and Neural Network for Financial Time Series Prediction. In: *Computational Science-ICCS 2006*, Alexandrov, V.N. *et al.* (Eds.). LNCS 3994, Springer Verlag, Berlin, Heidelberg, ISBN: 978-3-540-34385-1, pp: 493-500.
- Maier, H.R. and G.C. Dandy, 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and applications. *Environ. Model. Software*, 15: 101-124.
- Ruhaidah, S., S. Puteh and S. Ani, 2008. A comparison of neural network, arima model and multiple regression analysis in modeling rice yields. *Int. J. Soft Comput. Appl.*, 3: 113-127.
- Sfetsos, A. and C. Siriopoulos, 2004. Combinatorial time series forecasting based on clustering algorithms and neural networks. *Neural Comput. Applic.*, 13: 56-64.
- Tang, Z. and P.A. Fishwick, 1993. Feed forward neural nets as models for time series forecasting. *ORSA J. Comput.*, 5: 374-385.
- Tareghian, R. and S.M. Kashfipour, 2007. Application of fuzzy systems and artificial neural networks for flood forecasting. *J. Applied Sci.*, 7: 3451-3459.
- Valenzuela, O., I. Rojas, F. Rojas, F. Pomares and L.J. Herrera *et al.*, 2008. Hybridization of intelligent techniques and arima models for time series prediction. *Fuzzy Set Syst.*, 159: 821-845.
- Zhang, G., B.E. Patuwo and M.Y. Hu, 1998. Forecasting with artificial neural networks: The state of the art. *Neurocomputing*, 14: 35-62.
- Zhang, G.P., 2003. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50: 159-175.
- Zou, H.F., G.P. Xia, F.T. Yang and H.Y. Wang, 2007. An investigation and comparison of artificial neural network and time series models for chinese food grain price forecasting. *Neurocomputing*, 70: 2913-2923.