ABSTRACT
This study examines the forecasting performance of Adaptive Neuro Fuzzy Inference System (ANFIS) compared in comparison to statistical autoregressive integrated moving average (ARIMA) and the artificial neural network (ANN) model in forecasting of rice yield production. To assess the effectiveness of these models, we used 9 years of time series records for rice yield data in Malaysia from 1995 to 2001. The rice yield forecasting models having various input structures are trained and tested to investigate the applicability of ANFIS and ANN methods. The results of ANFIS and ANN models for both training and testing are evaluated and the best fit forecasting model structure and method is determined according to criteria of performance evaluation. The best fit model is also trained and tested by ARIMA method and the performances of all models are compared in order to get more effective evaluation. The results demonstrate that ANFIS model is superior to the ANN and ARIMA forecasting models in term of accuracy and reliability. Thus, ANFIS can be successfully utilized for rice yield forecasting.

Categories and Subject Descriptors : I.6.5 [Model Development]:Modeling methodologies; I.2.8 [Problem Solving, Control Methods, and Search]: Heuristic methods

General Terms
Experimentation.

Keywords
ANFIS, Neural Network, ARIMA, rice yield, forecasting, time series.

1. INTRODUCTION
The 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 [21]. 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 decision processes[25]. One of the most important and widely used time series model is the ARIMA model. The popularity of the ARIMA model is due to its statistical properties as well as the well known Box-Jenkins methodology. In the literature, there have been many approaches generally used in the forecasting of time-series, such as the Box and Jenkins[2] autoregressive (AR), AR moving average (ARMA), AR integrated moving average, and autoregressive moving average with exogenous inputs methods. These approaches have employed conventional methods of time-series forecasting and modelling [1,4,11,13,16,20,23,24].

Recently, artificial neural network (ANN) are being used more frequently in the analysis of time series forecasting, pattern classification and pattern recognition capabilities [11]. ANN provides an attractive alternative tool for both forecasting researchers and has shown their nonlinear modeling capability in data time series forecasting. 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 [24]. Since the real world is highly complex, there exists some linear and nonlinear patterns in the time series simultaneously.

Another approach is the fuzzy logic method, first developed to explain the human thinking and decision system by Zadeh [22]. Several studies have been carried out using fuzzy logic in hydrology and water resources planning [7,14,15,19,22].

Recently, an adaptive neuro-fuzzy inference system (ANFIS), which consists of the ANN and fuzzy logic methods, has been used for several application, such as database management, system design and planning/forecasting of water resources [6,7,8,9,10,18].

The main purpose of this study is to investigate the applicability and capability of the ANFIS, ANN and ARIMA methods for modeling of rice yields time-series forecasting. To verify the application of this approach, the rice yields data form 27 stations in Peninsular Malaysia is chosen as the case study.
2. LINEAR

George Box and Gwilym Jenkins [2] developed ARIMA models to become popular at the beginning of the 1970s. The model has seen one of the most popular approaches to the analysis of the time series and prediction. The general ARIMA models are compound of a seasonal and non-seasonal part are represented by the following way:

\[ \Phi_p(B) \Phi_p(B^s) \sum_{j=1}^{D} \theta_j(B) \Theta_q(B^s) a_t \]

where \( \Phi(B) \) and \( \theta(B) \) are polynomials of order \( p \) and \( q \), respectively; \( \Phi(B^s) \) and \( \Theta(B^s) \) are polynomials in \( B^s \) of degrees \( P \) and \( Q \), respectively; \( p \) 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.

This model can be expressed as

\[ \text{ARIMA } (p,d,q), (P,D,Q) \]

where \( (p,d,q) \) nonseasonal part of the model and \( (P,D,Q) \) seasonal part of the model. The general non-seasonal model is known as \( \text{ARIMA } (p,d,q) \) can be written as the linear expression

\[ y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{d} \theta_j a_{t-j} + \epsilon_j \]

The Box-Jenkins methodology is basically divided in the four stages: identification, estimation, diagnostic checking and forecasting. The identification stage involved transforming the data if necessary to improve the normality and the stationary time series. The next step is choosing the suitable model by analyzing both the autocorrelation (ACF) and partial autocorrelation function (PACF) of the stationary series. Once a model is identified, the parameters of the model are estimated. It is necessary to check if the assumptions are satisfied. Diagnostic checking using the ACF and PACF of residuals was carried out, which can be referred to Brockwell & Davis [3]. The forecasting model was then used to compute the fitted values and forecasts values.

3. NEURAL NETWORK FORECASTING MODEL

Recently, ANN has been extensively studied and used in time series forecasting. Zhang presented a recent review in this area [24]. The major advantage of ANN is their ability to model complex nonlinear relationship without a priori assumptions of the nature of the relationship. The ANN with single hidden layer feedforward network is the most widely used model for modeling and forecasting. The model is characterized by a network of three layers of simple processing units connected by a cycle links. The relationship between the input observations \( (y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \) and the output value \( y_t \) has following:

\[ y_t = a_0 + \sum_{j=1}^{q} a_j f\left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i}\right) + \epsilon_t \]

where \( a_j \) \( (j = 0, 1, 2, \ldots, q) \) is a bias on the \( j \)th unit, and \( w_{ij} \) \( (i = 0, 1, 2, \ldots, p; j = 0, 1, 2, \ldots, q) \) is the connection weights between layers of the model, \( f(\ast) \) is the transfer function of the hidden layer, \( p \) is the number of input nodes and \( q \) is the number of hidden nodes.

Training a network is an essential factor for the success of the neural networks. Among the several learning algorithms available, back-propagation has been the most popular and most widely implemented learning algorithm for all neural network paradigms [26]. In this paper algorithm of back-propagation is used in the following experiment.

Actually, the ANN model in (3) performs a nonlinear functional mapping from the past observation \( (y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \) to the future value \( y_t \), i.e.,

\[ y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, w) + \epsilon_t \]

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.

4. THE ADAPTIVE NEURAL FUZZY INFERENCE SYSTEM (ANFIS)

The fuzzy logic approach is based on a linguistic uncertainty expression rather than on numerical uncertainty. Since Zadeh proposed the fuzzy logic approach to describe complicated systems, it has become popular and has been used successfully in various engineering problems [14, 7, 22, 18, 6, 8, 9]. ANFIS, consisting of the combination of ANNs and fuzzy logic, has been used by many researchers to organize the network structure itself and to adapt the parameters of the fuzzy system for many engineering problems, such as the modeling of agricultural time-series. The fuzzy inference system is a rule-based system consisting of three conceptual components. These are: (1) a rule base, containing fuzzy if-then rules, (2) a database, defining the membership function and (3) an inference system, combining the fuzzy rules and producing the system results [22]. The first phase of fuzzy logic modeling is determination of the membership functions of the input-output variables, the second phase is the construction of fuzzy rules, and the last phase is the determination of the output characteristics, output membership function and system results [9, 17]. Two methods, called the back-propagation algorithm and the hybrid-learning algorithm, provide the learning of the ANFIS and construction of the rules and are used to determine the membership function of the input-output variables. A general structure of fuzzy system is demonstrated in Figure 1.
ANFIS has been shown to be powerful in modeling numerous processes, such as rainfall-runoff modeling and real-time reservoir operation [6, 9, 8]. ANFIS uses the learning ability of the ANN to define the input-output relationship and construct the fuzzy rules by determining the input structure. The system results were obtained by the thinking and reasoning capability of the fuzzy logic. The hybrid-learning algorithm and subtractive function are used to determine the input structure. The detailed algorithm and mathematical background of the hybrid-learning algorithm can be found in [12]. There are two types of fuzzy inference system in the literature: the Sugeno-Takagi inference system and the Mamdani inference system. In this study, the Sugeno-Takagi inference system is used for modeling of agricultural time-series. The most important difference between these systems is the definition of the consequence parameter. The consequence parameter in the Sugeno inference system is a linear equation, called a 'first-order Sugeno inference system', or a constant coefficient, called a 'zero-order Sugeno inference system' [12]. It is assumed that the fuzzy inference system includes two inputs, x and y, and one output, z. For the first-order Sugeno inference system, typical two rules can be expressed as

Rule 1: IF x is $A_1$ and y is $B_1$ THEN $f_1 = p_1 x + q_1 y + r_1$

Rule 2: IF x is $A_2$ and y is $B_2$ THEN $f_2 = p_2 x + q_2 y + r_2$

Where x and y are the crisp inputs to the node i, $A_i$ and $B_i$ are are the linguistic labels such as low, medium, high, etc., which are characterized by convenient membership functions, $P_i$, $Q_i$, and $R_i$ are the consequence parameters. The structure of this fuzzy inference system is shown in Figure 1.

The ANFIS scheme consists of five layers, and the model can be briefly presented step by step as follows.

**Input nodes (layer 1).** Each node in this layer generates membership grades of the crisp inputs which belong to each of convenient fuzzy sets by using the membership functions. Each nodes’s output $O_i$ is calculated by

$$O_i = \mu_{A_i}(x) \quad \text{for } i = 1, 2;$$

$$O_i = \mu_{B_i}(y) \quad \text{for } i = 3, 4$$

where $\mu_{A_i}$ and $\mu_{B_i}$ are the membership functions for the $A_i$ and $B_i$ fuzzy sets respectively. Various membership functions, such as trapezoidal, triangular, Gaussian generalized bell membership function, etc., can be applied to determine the membership grades. If generalized bell membership function is used, the membership function is given by

$$O_i = \frac{1}{1 + \left(\frac{x - c_i}{a_i}\right)^2 b_i}$$

where $a_i$, $b_i$, and $c_i$ are the parameter set.

**Rule nodes (layer 2).** In this layer, the AND/OR operator is applied to get one output that represents the results of the antecedent for a fuzzy rule, i.e. firing strength. The outputs of the second layer, called firing strengths $w_i$, are the products of the corresponding degrees obtained from layer 1, termed w, as follows:

$$w_i = \mu_{A_i}(x) \mu_{B_i}(y) \quad i = 1, 2$$

**Average nodes (layer 3).** The main target is to compute the ratio of firing strength of each i\(^{th}\) rule to the sum of the firing strengths of all rules. The firing strength in this layer is normalized as

$$O_i = \frac{w_i}{\sum_i w_i} \quad i = 1, 2$$

**Consequent nodes (layer 4).** The contribution of the i\(^{th}\) rule towards the total output or the model output and/or the function defined is calculated by

$$O_i = \frac{w_i f_i}{\sum_i w_i} \quad i = 1, 2$$

where $w_i$ is the i\(^{th}\) node output from the previous layer as demonstrated in the third layer. $\{p_i, q_i, r_i\}$ is the parameter set in the consequence function and also the coefficients of linear combination in the Sugeno inference system.

**Output nodes (layer 5).** This layer is termed the output node, in which a single node computes the overall output by summing all incoming signals and it is the last step of the ANFIS. The output of the system is calculated as
The ANN is trained based on supervised learning. The objective is to train adaptive networks having convenient unknown functions given by the training data and to find the proper values of the input and output parameters. ANFIS applies the hybrid-learning algorithm to achieve this aim, which consists of a combination of the 'gradient descent' and the 'least-squares' methods. The gradient descent method is used to assign the non-linear input parameters, and the least-squares method is employed to identify the linear output parameters. The antecedent parameter, i.e. MF given in layer 2, is applied to construct the rules of the ANFIS model. Since the input variables within a range might be clustered into several classes, the structure of the input layer needs to be determined accurately. The 'subtractive fuzzy clustering' function, which offers an effective result using less rules, is applied to solve the problem in the ANFIS modeling [18].

5. EMPIRICAL RESULTS

5.1 Data Sets
The data were collected from Muda Agricultural Development Authority (MUDA) Kedah, Malaysia ranging from 1995 to 2001. There are 4 areas with 27 locations. There are two types of season symptom that influenced the rice yield in Malaysia. The rice yield series data is used in this study to demonstrate the effectiveness of the hybrid method. These time series come from different location and have different statistical characteristics. The rice yields data contains the yields data from 1995 to 2001, giving a total of 351 observations. Given a set of 351 observations made at uniformly spaced time intervals, the locations of rice yield are rescaled to the time axis becomes the set of integers \{1, 2, ..., 432\}. For example the first location in 1995 is written as time 1, the second location in 1995 as time 2 and so on. The time series plot is given in Figure 2.

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 1995 to 2001 producing 351 observations for training purpose and the remainder as the output sample data set with 27 observations for forecasting purpose. The performances of the each model for both the training data and forecasting data are evaluated and is 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

\[
\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t| \\
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}
\]

where \(y_t\) and \(\hat{y}_t\) are the observed and the forecasted rice yields at the time \(t\). The criterions to judge for the best model are relatively small of MAE and RMSE.

6. FITTING ARIMA MODELS TO THE DATA
The plots in Figure 2 shows the plots of the rice yield time series indicate that the time series are non-stationary in mean and variance. The natural logarithm was taken to reduce the variance, and then the first difference was applied in order to remove the trend.

The sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for the transformed series are plotted in Figure 3. The plot shows that there is seasonality in the series. We find the first-difference series becomes stationary. For ACF, there are major spikes at lags 1 and 27, while for PACF, we observe major spikes at lags 1, 2, 23 and 27.
Several models were identified and the statistical results during training are compared in the following Table 1. The criterions to judge for the best model based on MAE and MSE show that the ARIMA (0,1,1)x(1,0,1) is a relatively best model. This model has both non-seasonal and seasonal components.

Table 1  Comparisons of ARIMA models’ statistical results

<table>
<thead>
<tr>
<th>Criteria</th>
<th>ARIMA Model</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0,1,1)x(1,1) (0,1,1)x(1,0,1) (0,1,1)x(1,1) (0,1,1)x(1,0,1) (0,1,1)x(1,0,1)</td>
<td>0.006</td>
<td>0.129</td>
</tr>
</tbody>
</table>

The ACF and PACF plots of residuals for the rice yields series are shown in Figure 4. From the residual plot of the best ARIMA model, it was observed that the selected ARIMA model passed the diagnostic checks and they were all white noise.

Figure 4  ACF and PACF of residuals for rice yields

7 FITTING NEURAL NETWORK MODELS TO THE DATA

In this investigation, we only consider the situation of one-step-ahead forecasting with 27 observations. Before the training process begins, data normalization is often performed. The linear transformation formula to [0, 1] is used

\[ x_n = \frac{y_0}{y_{\text{max}}} \]

where \( x_n \) and \( y_0 \) represent the normalized and original data; and \( y_{\text{max}} \) represent the maximum values among the original data. In order to conform the neural network used in the forecast, ACF and PACF were used to determine the maximum number of input neurons used during the training. Figure 3 presents the ACF and PACF of data sets for the rice yields time series. The input variable selection for the ANN is selected from lags with high ACF and PACF. Based on these analyses, the maximum number of lags, 27, was identified suitable to use as inputs for the proposed ANN. The one only neuron in the output layer represented being modeled. All the data were normalized in the range 0 and 1. After the input and out variables were selected, the ANN architecture of 27-H-1 was explored for capturing the complex, non-linear and seasonality of rice yields data. 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. Table 2 shows the performance of ANN during training with varying the number of neurons in the hidden layer (H).

Table 2 Performance Variation of a Three-Layer ANN during training with the number of neurons in the hidden layer for ANN

<table>
<thead>
<tr>
<th>N</th>
<th>Number of neurons in the hidden layer</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.50</td>
<td>3.51</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0.12</td>
<td>1.10</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>0.12</td>
<td>1.10</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>0.12</td>
<td>1.10</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>0.12</td>
<td>1.10</td>
</tr>
</tbody>
</table>

It is observed that the performance of ANN is improved as the number of hidden neurons increases. However, too many neurons in the hidden layer may cause over-fitting problem, which results in the network can learn and memorize the data very well, but lacks the ability to generalize. If the number of neurons in hidden layer is not enough then the network may not be able to learn. So, an ANN with 63 neurons in the hidden layer seems to be appropriate.

8 FITTING ANFIS MODELS TO THE DATA

The ANFIS configurations obtain through a trial and error process. One of the most important steps in developing a satisfactory forecasting ANFIS model is the selection of the input variables among the available variables. In this study firstly, the four models having various inputs are trained and tested by ANFIS method and the performances of models for rice yields are compared and evaluated based on training and forecasting performances. The structures of the forecasting models can be expressed as

\[ y(t)_{351} = f(y(t-1)_{351}, y(t-2)_{351}, \ldots, y(t-k)_{351}) \]

where \( y_t \) represents the rice yields at time \( t \).

It worked with different numbers of membership functions: two, three, four, and five. Various membership functions, such as trapezoidal, sigmoid, Gaussian, generalized bell membership function were also considered. For output set before defuzzification process we select Linear Models, which indicates that the generated model are Type-I Takagi-Sugeno model.

The results in terms of various performance statistics from all the ANFIS models are presented in Table 3. From experiment results, using ANFIS algorithm, we find the most of ANFIS suffer the problem of slow convergence and almost all of them cannot reach the target of training especially when the number of membership more and the number of input more than 4 are used.

The final model was chosen according to the smallest value of errors. The analysis revealed that three symmetric Gaussian membership functions for four inputs has shown the lowest value of the RMSE and MAE. The network that performs best is chosen as the final model for forecasting of 27 observation rice yields.

Table 3 Different Structure of the ANFIS

<table>
<thead>
<tr>
<th>Membership function</th>
<th>Membership Number</th>
<th>Performance rating</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigmoid</td>
<td>2</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

9. COMPARISON OF PERFORMANCES

The value predicted by the adaptive neural network with fuzzy inference system (ANFIS) were compared with the data set. The forecasting accuracy was evaluated by undertaking the comparison with the ANN and ARIMA methods. The best results of ANFIS during the training are compared with the ARIMA and ANN model in order to get the best forecasting model of rice yields. The performances of the best models developed by ARIMA, ANN and ANFIS models for forecasting data sets are summarized in Table 4. As it can be seen from Table 4, ANFIS model produces minimum RMSE and MAE errors followed by ANN and ARIMA model.

Table 4 Rice yields forecast results

<table>
<thead>
<tr>
<th>Performance criterion</th>
<th>ARIMA</th>
<th>ANN</th>
<th>ANFIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE (%)</td>
<td>0.1469</td>
<td>0.1103</td>
<td>0.084</td>
</tr>
<tr>
<td>RMSE</td>
<td>20072.59</td>
<td>16782.406</td>
<td>12558.9</td>
</tr>
</tbody>
</table>

Figure 5 shows overall summary statistics forecasting for rice yields with three models by using box-plot. Figure 5 demonstrate
that the ANFIS model performance is in general, accurate and satisfactory, where all data points of errors are quite near the zero.

Figure 5 Comparison of the ARIMA, ANN and ANFIS models

10 CONCLUSION
This study investigated the applicability and capability of the ANFIS method in rice yields forecasting. The data set includes 351 rice yields data for the period 1995-2003. The data set was divided in to two subsets, i.e. training and forecast. The performances of the ANFIS models and observations were compared and evaluated based on their performance in the training and testing sets. The ARIMA and ANN models were also investigated for the same set of data and the results are reported.

Based on the performance of three models, it can be concluded that ANFIS is an effective method to forecast rice yields. The result suggests that the ANFIS method is superior to the ANN method in the modeling of time-series. The results of the ANFIS model show that the ANFIS can be applied successfully to establish time-series forecasting models, which could provide accurate forecasting and modeling of time-series.

7. ACKNOWLEDGMENTS
Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

8. REFERENCE

