

Research Article

An Artificial Neural Network with Stepwise Method for Modeling and Simulation of Oil Palm Productivity Based on Various Parameters in Sarawak

¹Yousif Y. Hilal, ¹W. Wan Ishak, ¹Azmi Yahya and ²Zulfa H. Asha'ari

¹Department of Biological and Agricultural Engineering, Faculty of Engineering,

²Department of Environmental Sciences, Faculty of Environmental Studies,
University Putra Malaysia, Malaysia

Abstract: Aim of study to optimize the oil palm yield amount by studying parameters of land quality and climate, determines which of them is distinctly effective on oil palm yield amount, develops ANN model and simulation of Oil Palm production by using MATLAB software and Design Expert software, conducted an experiment to determine the effect of the number of neurons and the number of hidden layers in the network ANN is used. Across the optimization procedures obtained the best ANN architecture is 8 neurons in input layer -5 neurons in the hidden layer and -2 neuron in the output layer to obtain the best model of oil palm productivity prediction with a value of R 0.989 and MSE: 0.013, training Error 1.1%, testing error 1.9% and validation error 1.19%. The results of simulation and Independent Variable Importance show that the average accuracy percentage simulation is 0.9867% and MSE 0.0513%. The climatic changes that influenced the simulation are very high, where the relative humidity recorded on the proportion of impact of up to 100%, while the recorded rainy days, which is ranked second in influence was almost 90% and the effect of temperature was up to 70%. The influence of several climatic changes that decrease the quantity of rainfall, Rain days, Temperature rise, Evaporation and increasing Humidity, reduces the productivity of oil palm plantations for 2.35 tons/ha/year. This research concludes that ANN can be used to predict the production of palm oil based on the quality of land and local climate with very good results.

Keywords: Artificial neural network, climate, FFB, land quality, oil palm, stepwise method

INTRODUCTION

Forecast of production for a plantation is required since it began evaluation the suitability of land to acquire land economic value of a particular land usage or query periodically in production estimates. Land productivity data is as needed as will in land suitability and land usage planning to reduce the risk of investment failure. Growth and crop production in certain regions, are heavily dependent on interactions between climatic parameters, soil, plants and management. In other words, production plants with specific management system is a function of quality, maturity stage of the plant and the surrounding climate (Barcelos *et al.*, 2015).

The oil palm is the most efficient oilseed crop in the world. One hectare of oil palm plantation is able to produce up to ten times more oil than other leading oilseed crops. The most efficient producers may achieve yields as high as eight tons of oil per hectare. Among the 10 major oil seeds, oil palm is accounted for 5.5% of global land use for cultivation; but

produced 32.0% of global oils and fats output (Schwarze *et al.*, 2015). In 2015, production efficiency is about 3.9 ton/ha/year. Production efficiency is now considered the lowest compared to the theoretical efficiency of production (Chemura, 2012). In theory, the production efficiency can reach 18.5 ton/ha/year (Lee, 2011). Production efficiency is a measurement of utilizing land usage and, this reflects the significant reduction in the ratio of production rate.

A lot of factors influence oil palm tree production where they directly or indirectly affect the oil palm performance. Few studies have focused on this differentiation and if so just did for one parameter while oil palm productions may be effected by many factors negatively or positively as any crop yield (Corley and Tinker, 2016; Tao *et al.*, 2016). Ignoring or considering any factor cannot done randomly; because it's effect on the prediction accuracy of the yield amount is compared with actual yield. The productivity of oil palm plantations is strongly influenced by the quality of kinds of oil palm areas and climate, which are: for instance: Immature, high yield mature and

Corresponding Author: Yousif Y. Hilal, Department of Biological and Agricultural Engineering, Faculty of Engineering, University Putra Malaysia, Malaysia

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

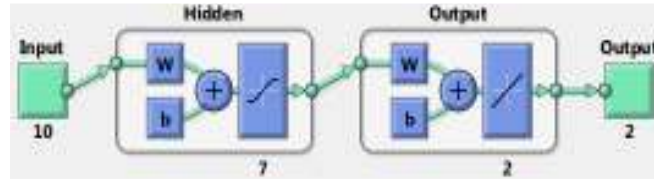


Fig. 1: Block diagram of ANN

deteriorated mature oil palm area, rainfall mm/year, average temperature, water deficit mm/year, humidity, radiation sun and many others ones (Muhammad *et al.*, 2015). Crop production as a function of the quality of the oil palm areas and the climate can be predicted using various methods. Artificial neural network ANN is one of the recognized superiority prediction methods, especially for the prediction that involves many parameters that work simultaneously to form functional relationships that are not linear (Ficken, 2015).

Artificial Neural Network is a computational structure that was developed based on the process of biological neural systems in the brain tissue. ANN is a translation function of the human brain (*biological neurons*) in the form of mathematical functions that run in parallel calculation process (Rad *et al.*, 2015). Kant and Sangwan (2015) stated that the ANN is flexible to the input data and produce a consistent response. The network consisting of several layers (multilayer) may indicate capabilities that are perfect for solving various problems. ANN learning can solve parallel computation for complex tasks, such as prediction and modeling; classification and pattern recognition; clustering; and optimization (Neto *et al.*, 2015).

Multilayer feed-forward back-propagation Artificial Neural Network consists of three layers which are called input layer, hidden layer and output layer. The input layer has n nodes, hidden layer and output node has the h layer has m nodes, as in Fig. 1 ANN method is expected to provide a better answer to predict tree crop production as a function of the characteristic parameter/land quality. Non-linear nature of which is the strength of other artificial neural networks can overcome the shortcomings of conventional methods that are cumbersome and unpopular when entering the non-linear model, even though the ANN wasn't very accurate especially when the inputs set is large (Han and Kim, 2016).

In statistics, stepwise is a tool that is used to select the most significant inputs when several parameters are being analyzed in regression. Parameters are added and removed one at a time and the model selects the top few based on the alpha value chosen, step wise regression includes regression models in which the choice of predictive variables is carried out by an automatic procedure (Wagner and Shimshak, 2007; Khamis *et al.*, 2016). Vaidyanathan and Volos (2016)

stated that stepwise methods have the same ideas as best subset selection but they look at a more restrictive set of models. Between backward and forward stepwise selection, there is just one fundamental difference, which is whether you are starting with a model with no predictors forward and with all the predictors backward. Stepwise selection is a method that allows moves in either direction, dropping or adding variables at the various steps. Backward stepwise selection involves starting off in a backward approach and then potentially adding back variables if they later appear to be significant (Mousavi *et al.*, 2015).

The process is one of alternation between choosing the least significant variable to drop and then re-considering all dropped variables except the most recently dropped for re-introduction into the model. This means that two separate significance levels must be chosen for deletion and addition to the model. The second significance must be more stringent than the first. The variables that show little significance to the model can be removed allowing for a more concise and less complex model. This allows an easier updating, since only the significant parameters are needed to be collected. Overall, Stepwise regression is helpful when a large number of inputs are given to a model and the user wants to see the significance of the most important ones (Sandry, 2012).

The purpose of this research is to develop a predictive model based on land quality, land productivity and climate using ANN models with Stepwise selected inputs, so it can overcome the shortcomings of the Back-propagation algorithm in a finer way. Stepwise as optimization tools that are meant to reduce the parameters number and select optimal set that has a higher effective role over other ones, in order to get high accurate result with lowest error. The obtained ANN model is used to simulate the effect of land quality and climate change simultaneously on the productivity of oil palm plantation in Sarawak.

METHODOLOGY

Description of the study area: Sarawak province is one of the major oil palm production regions in Malaysia. It is the largest state in Malaysia covering an area of nearly 125,000 km² in the Borneo land mass and stretches over 750 km of the northeast Borneo. It makes up 37.5% of land area in the country.

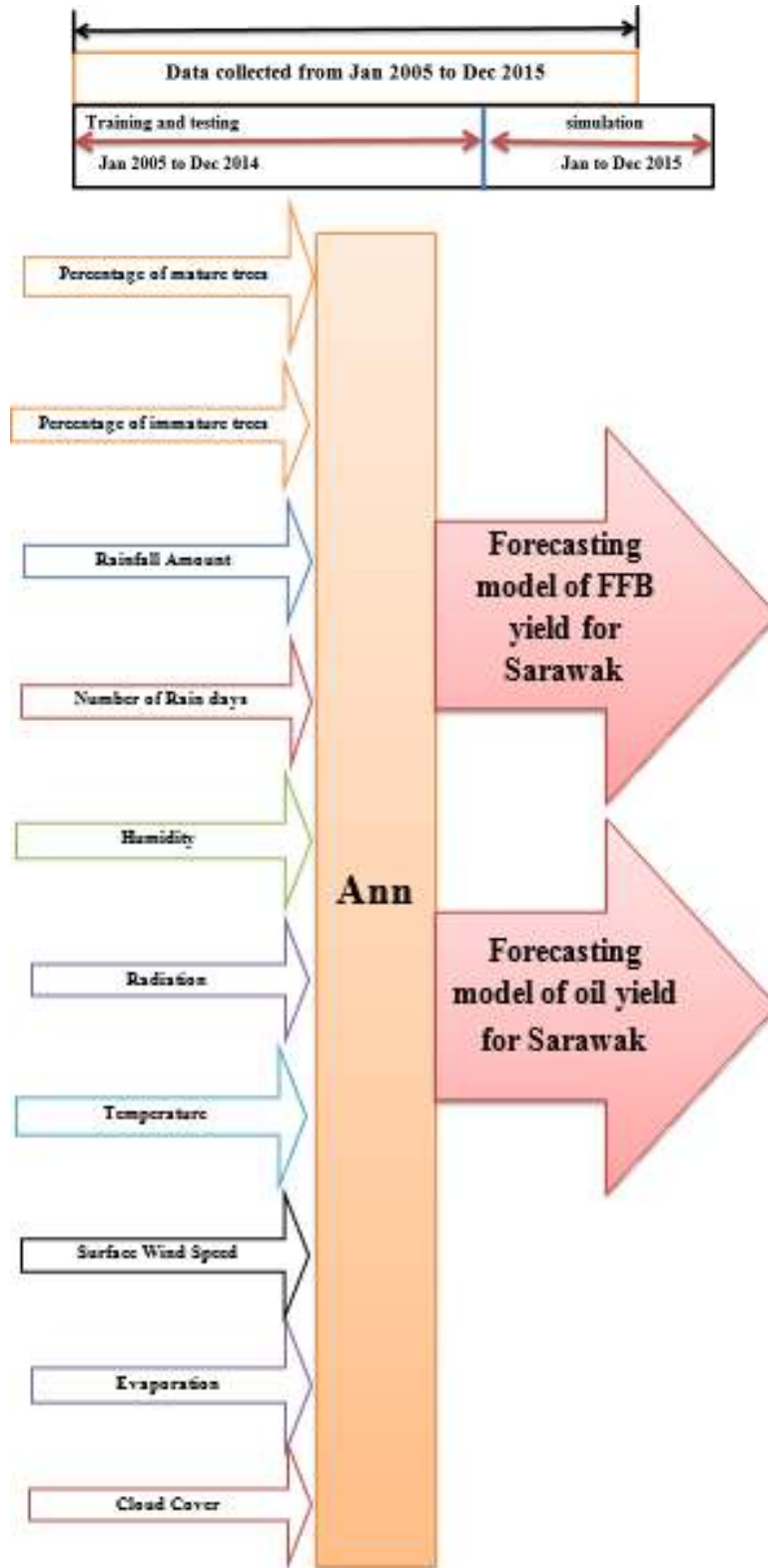


Fig. 2: Schematic of input and output vectors of the ANN

Description of data used: The materials used in this study are a couple of production data, quality of land

and climate. Secondary data in this study and the dataset were collected from the

Malaysian Meteorological Dept., Dept. of Statistics Malaysia Prime Minister's and Malaysian Palm Oil Board (MPOB).

The data that are taken from each of the tests done on the input parameters used were: percentage of mature and immature area %, rainfall amount mm, number of rain days day, the relative humidity %, daily global radiation Mjm^{-2} the average temperature $^{\circ}C$, surface wind speed m/sec, mean daily Evaporation mm and cloud cover oktas. The data of the output: Fresh Fruit Bunch (FFB) Yield ton/hectare and average oil yield ton/hectare, the data were used throughout the 10 years, ranging from Jan -2005 to Dec-2015 of the years of harvest in the province Sarawak. These 10 input factors are used in this study to forecast the yield amount For 10 years. The output (target) is the Fresh Fruit Bunch and average oil yield for nine month ahead (Fig. 2).

Neural network model: The research was conducted in three phases:

- Build program ANN
- To learn and test models of neural networks to obtain the optimal model
- Simulation model of ANN was elected to the changes in climate quality

ANN programs by MATLAB by Neural Network Toolbox (Haykin, 2009). Step learning (training) is a supervised learning process of an ANN for value-weighted w the best. The method used for training is Back propagation algorithm. The weight of the network is modified in ways that minimize the sum of squared errors calculated against all vertices output. In step training input data layer is the quality of the land and climate parameters, while as the output target is the productivity of plantations. Step model test is a method to test the weight that has been obtained at the time of training. Testing was conducted to see the consistency of the best models which were acquired during the training using different input data. Best ANN model simulations using hypothetical data input is done to determine the effect of climate change in the quality of some of the elements simultaneously on the productivity of oil palm plantations.

Designing the ANN model: In order to generate an output vector that is as close as possible to the target vector from an ANN, a training process was employed to find optimal weight matrices and bias vectors that minimize a predetermined error function. The proposed ANN model used in this study is the supervised and the three training algorithms were used are:-Levenberg-Marquardt LM, Resilient back-propagation RP and Gradient descent with momentum and adaptive learning rate back-propagation GDX which are

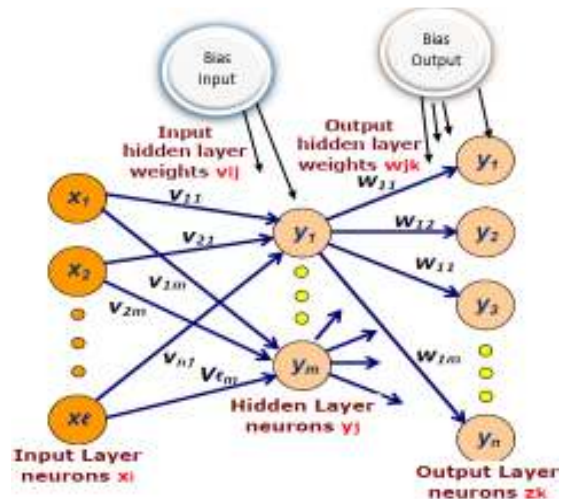


Fig. 3: Represents the mathematical form of the ANN

employed to evaluate their suitability. Multiple layers of neurons with non-linear transfer functions allow the network to learn nonlinear and linear relationships between input and output parameters. The linear output layer allows the network to take any values even outside the range -1 to +1; while if the last layer of a multilayer network has sigmoid neurons, the outputs of the network will be only in a limited range. This research used two transfer function Log-sigmoid and tangent sigmoid Package which randomly selected a sample of 70% for training, a sample of 15% for cross validation and 15% were used for test. 5 to 13 hidden nodes is suggested for this network topology.

Neural network is a network of digital Facts that is related to the specified links weighed and processed mediated mathematical model. Facts alone does not undertake any effect; but if this data is united with Weights in addition to the value of bias, it leads into defining the task of the neural network and Fig. 3 illustrates the mathematical form of artificial neural network (Graupe, 2013). Artificial Multilayer networks consist of the following parts:

- The hidden layer does an intermediate computation before directing the input to output layer. The input layer neurons are linked to the hidden layer neurons; the weights and bias on these links are referred to as input-hidden layer weights and bias. Using the following equation (Du and Swamy, 2013):

$$v = \sum_{i=1}^m (w_{iinput} * xi) + \beta_{iinput} \quad (1)$$

- Part NET, which represents the amount inclusive of the values of the variables involved, is multiplied by weights in addition to the value of bias. The hidden layer neurons and the

corresponding weights and bias are referred to as output-hidden layer weights and bias:

$$net = \beta_{i_{output}} + \sum_{i=1}^n (v_i * w_{i_{output}}) \quad (2)$$

$$y_{FFB} = f(net_{FFB})$$

$$y_{oil\ yield} = f(net_{oil\ yield})$$

where,

i = Input neurons

m = Neurons in the first hidden layers

n = Neurons in the second hidden layers

w_i = The weights representing the strength of the connection between the node

β_i = The bias associated with node

Regarding the final part of neural network, the transfer functions are Continuous non-linear function which called activation function (activation) and it publishes Value by entering them in accordance with the scale of schedule in the emerging value. The function approximation regression is sigmoidal function, Log-sigmoid and Tan-sigmoid (Kyurkchiev and Markov, 2015). The most popular transfer function for a nonlinear relationship is the sigmoidal function (Schmidhuber, 2015). The general Log-sigmoid form of this function is indicated below:

$$y_{FFB} = \frac{1}{1+e^{-net_{FFB}}} \quad (3)$$

$$y_{oil\ yield} = \frac{1}{1+e^{-net_{oil\ yield}}} \quad (4)$$

The general Tan-sigmoid form of this function is indicated below:

$$y_{FFB} = \frac{2}{(1 + \exp(-2 * net_{FFB}))} - 1 \quad (5)$$

$$y_{oil\ yield} = \frac{2}{(1 + \exp(-2 * net_{oil}))} - 1 \quad (6)$$

In briefly, writing back-propagation neural network algorithm into a computer programming language is as follows:

- Input pair of data input, output targets and training parameters
- Normalization of data input and output targets
- Providing initial value weighted random
- Repeat training
- Repeat a couple of data
- The calculation of the value of assets
- Calculation error
- Gradient calculation error
- Until all data pairs counted

- Calculation of total gradient error
- Correction (adjustment) weighting
- Until criteria for dismissal of reached training.

STEPWISE IMPROVED PROCESS OPTIMIZATION FOR NEURAL NETWORK MODEL

Stepwise Backwards approach is achieved by design expert program. We begin by describing the basic stepwise procedure using the backwards approach for the modeling of the quality of the land and climate parameters variables. The backwards approach starts by considering all possible input and output variables in the quality of the land, climate parameters model and production.

The process is one of alternation between choosing the least significant variable to drop and then re-considering all dropped variables except the most recently dropped for re-introduction into the model. This means that two separate significant levels must be chosen for deletion from the model and for adding to the model. The second significance must be more stringent than the first. Briefly writing stepwise into a computer programming language is as follows (Wagner and Shimshak, 2007):

- Start
- Run a single variables analysis that includes the full set of input variables and output variables and record the efficiency scores for each models.
- Dropping one input variable at a time in each run. For each analysis: record the efficiency scores for each model for all runs. Calculate the average difference in efficiency over the set of differences.
- Choose the single input to be dropped by selecting the variable with the minimum average difference in efficiency scores from above. At least one input variable must be kept in the analysis. If the model has only one input variable remaining, then this one variable cannot be dropped and another variable must be considered based on the selection procedures above.
- For the variables selected to be dropped? Is based on the efficiency scores of the models for the remaining input variables.
- Stop.

After we determine the factors affected by used stepwise method, these factors obtained were inserted into neural network which was according to the steps that were described previously (Fig. 4).

Performance of the network was evaluated. Finally suitable and acceptable results were obtained. In this study, the following performance measuring functions were employed: mean square error MSE and mean

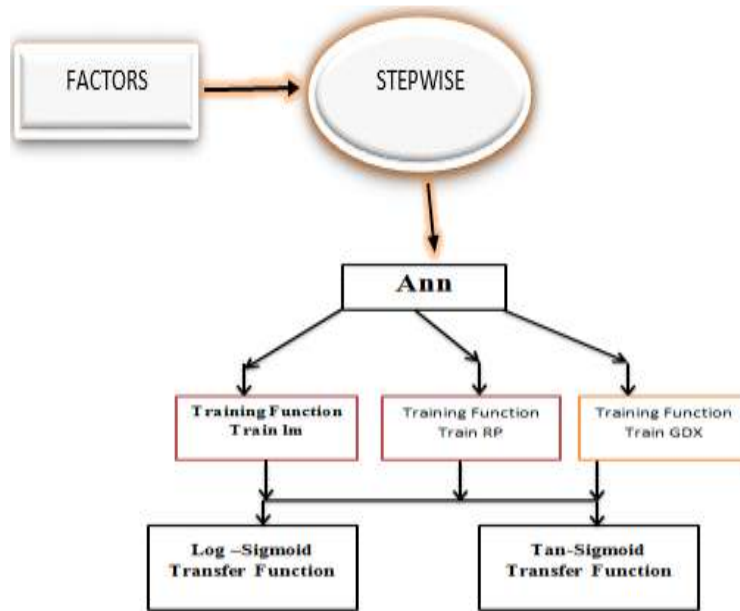


Fig. 4: Block diagram of Stepwise and ANN

absolute percentage error MAPE % to evaluate the accuracy of the proposed ANN model. These functions are given in the following equations:

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \quad (7)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|(X_i - Y_i)|}{X_i} \times 100 \quad (8)$$

$$\text{Average Accuracy Percentage (AA\%)} = 100\% - \text{MAPE} \quad (9)$$

where,

- X_i : The actual yield
- Y_i : The forecasted yield
- N : The data points
- I : 1, 2, ..., N

RESULTS AND DISCUSSION

The prediction models tested on the data are the NN model using the supervised the three training algorithms were used Levenberg-Marquardt LM, Resilient back-propagation RP, Gradient descent with momentum and adaptive learning rate back-propagation GDX and stepwise-NN models.

Results of neural network: The feed-forward PB neural network usually has one or more hidden layers, which enable the network to model non-linear and complex functions (Graupe, 2013). It is crucial to emphasize that determining the number of neurons in the hidden layers is very important because it affects the training time and dissemination of property neural

networks. It may impose a higher value of neurons in the hidden layer of the network for keeping rather than circular the patterns that have been witnessed during the training, while the decline in the value of neurons in the hidden layer will waste a lot of time training to find the optimal representation (Kant and Sangwan, 2015). Nevertheless, there is no general rule that is used to choose the number of neurons in the hidden layer (Rad *et al.*, 2015). Thus, it depends on the complexity of the system that is being modeled. The most common and widely used approach to find the optimal number of neurons in the hidden layer is by trial and error. Therefore, the approach has used trial and error to determine the optimal neurons in the hidden layer of the network in this study. There are many types of learning algorithms in the literature, which can be used to train the network. For that, it is difficult to know which algorithm is more efficient for a certain problem. The algorithm that is used to train ANN in this study is Levenberg-Marquardt back-propagation LM, Resilient back-propagation RP, Gradient descent with momentum and adaptive learning rate back-propagation GDX. These algorithms are very well suited to the training of the neural network. For studying the effect of the training function transfer function and number of hidden layers that was chosen best tests results are shown in Table 1. The results show that the best performance of the network model is obtained when using the training function LM compared with both other training functions and it can be seen that neural network is almost insensitive performance to the difference of the number of hidden layers. Using functional function Log-Sigmoid in the faculties of two layers of moderation hidden layer outputs had produced

Table 1: The values of the best ANN models

Transfer function :log sig						
Coef. Correlation (r) % & MSE						
Training function	NO. of hidden layers	Training	Test	Validation	All	MSE
Train LM	5	99.27	94.52	97.93	98.29	0.015
	7	98.10	98.40	98.65	98.20	0.009
	10	98.18	97.50	98.13	98.04	0.010
	12	98.66	96.21	97.82	98.13	0.012
Train RP	9	97.33	96.19	97.77	97.20	0.012
	13	95.55	94.86	98.31	95.87	0.010
Train GDX	8	96.56	94.84	97.80	96.31	0.013
	9	96.71	95.21	97.36	96.45	0.014
	10	96.43	96.36	96.65	96.34	0.019
Transfer function: tan sig						
Coef. Correlation (r) % & MSE						
Training function	NO. of hidden layers	Training	Test	Validation	All	MSE
Train LM	5	97.57	95.27	96.85	97.05	0.017
	7	98.95	98.32	98.91	98.58	0.010
	10	98.72	91.80	97.25	97.40	0.018
	12	97.13	98.28	97.24	97.13	0.015
Train RP	9	97.77	95.36	97.71	97.36	0.015
	13	96.57	95.54	96.99	96.20	0.019
Train GDX	8	96.21	95.95	95.21	95.82	0.028
	9	97.01	96.82	99.022	97.18	0.015
	10	96.25	94.75	97.95	96.14	0.013

Table 2: The effect of ANN architecture on the performance of models

ANN architecture	Training error %	Testing error %	Validation error %	Data Division %		
				Training %	Validation %	Testing %
10-5-2*	0.0123	0.0200	0.0120	70	15	15
10-7-2*	0.0180	0.020	0.0170			
10-10-2*	0.0170	0.020	0.0140			
10-12-2*	0.0180	0.018	0.0170			
10-7-2**	0.0189	0.023	0.0188			

*: log sig; **: tan sig

the best performances in the mean squared error MSE and Coefficient Correlation r compared with Tan-sigmoid and were determined as the best mean value at LM Training function as 0.985 for training set, 0.981 for validation set and 0.966 for testing set. The MSE and Coefficient Correlation r for all data sets were also calculated as 0.011 and 0.981. These results are similar to previous work (Neto *et al.*, 2015).

Through Table 2 it can be observed that the selection of the data division of the default to MATLAB program is 70% for a set of training, the 15% for a set of validation and 15% for a set of testing. The best choice of the ANN architecture depends on less training error, testing error and validation error. The results obtained from the training stage models have the best ANN architecture 10-5-2 input layer that consists of ten neurons, five hidden layer neurons and output layer two neurons. Reliability models are indicated by the value of training error 1.23%, testing error 2% and validation error 1.2%.

Figure 5 shows scatter plots of the ANN model predicted versus actual values using LM algorithm for

training, validation and testing and all data sets. According to the predicted model fully fitted to the actual values for training, validation and testing and all data sets. In the same investigation Lamba and Dhaka (2014) showed why Neural Network Model is important from other models for nonlinear data behavior system like crop yield prediction.

Simulations and independent variable importance:

The output of the network was found to be close to the actual values of the yield amount. The optimal structure for the model with minimum prediction error, correlation coefficient r between the actual and predicted values. Forecasting of the average monthly FFB and oil yield of nine month ahead starting at January to September 2015, has been carried out in the province Sarawak. The input data of previous per month which are percentage of mature and immature area, rainfall amount, number of rain days, the relative humidity, daily global radiation, the average temperature, surface wind speed, mean daily Evaporation and cloud cover all these data presents to

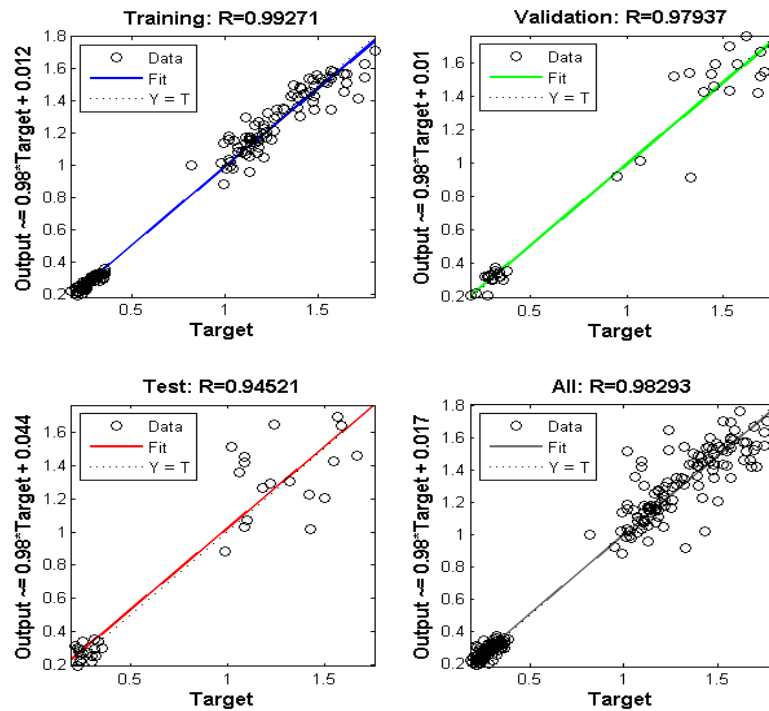


Fig. 5: The scatter plots of ANN model predicted versus actual values

Table 3: Forecasting for nine month by using ANN model

Date	Target FFB	Forecasted	Error	Target oil yield	Forecasted	Error
JAN-2015	0.98	1.038655	-0.05865	0.19	0.219405	-0.02941
FEB- 2015	0.89	1.208545	-0.31854	0.17	0.244061	-0.07406
MAR-2015	1.08	1.702779	-0.62278	0.22	0.330657	-0.11066
APR-2015	1.24	1.240114	-0.00011	0.25	0.264017	-0.01402
MAY-2015	1.33	1.001983	0.328017	0.27	0.232634	0.037366
JUN-2015	1.40	1.287955	0.112045	0.28	0.258476	0.021524
JUL-2015	1.51	1.460616	0.049384	0.30	0.289029	0.010971
AUG-2015	1.75	1.618887	0.131113	0.34	0.319535	0.020465
SEP-2015	1.70	1.704152	-0.00415	0.35	0.342933	0.007067
MSE		0.0702			0.00235	
MAEP		1.5196			1.5261	
AA		98.4804%			98.4739%	

the MLP neural network model proposed. The results obtained for the forecasted FFB and oil yield for nine month ahead are summarized in Table 3.

In mathematical models that involve a lot of input sensitivity test, it is considered an essential element for building the model and quality assurance. It can use sensitivity analysis to simplify the forms and to verify the robustness of the model predictions. Also, it is used to find out the factors to find the relative importance of the factors affecting the often contribution to change output productivity making use of allergy testing. Notice in Table 4, we find that the variable 5, which is the relative humidity is considered the most important variable, as its importance relative 0.226, Based on the data collected, it can be analyzed in a single relationship between the parameter with the productivity of oil palm plantations as follows humidity correlated negatively on

the productivity of oil palm plantations such as when the relative humidity was 78.92%, the Average FFB Yield 1.51 ton/ha/month and the average oil yield 0.30 ton/ha/month, while production fell to 0.98, 0.19 ton/ha/month when humidity recorded 85.59%. Followed variables 4 and 7 are the Rain days and Temperature as if their are importance relative is 0.204 and 0.158 respectively. Since the dry months can reduce the production of palm oil, as for instance as a result of a rain days reduced from 28 to 21 days, it will reduce the yield of 3-4% in the 2005 year and 8-10% in the 2015 year.

The effect of drought stress occurs not only in the vegetative stage but also on the generative phase. Temperatures were recorded in oil palm plantations have averages between 24-28°C and optimal 25-27°C, annual temperature variations should not be too high,

Table 4: Independent variable importance

Q	Independent variable	Importance	Normalized importance (%)
1	MATURE	0.018	8.0
2	IMMATU	0.028	12.5
3	Rainfall	0.082	36.3
4	Rain days	0.204	90.2
5	Humidity	0.226	100.0
6	Radiation	0.025	11.0
7	Temperature	0.158	70.0
8	Wind speed	0.035	15.7
9	Evaporation	0.110	48.9
10	Cloud cover	0.114	50.7

Table 5: The values of the best ANN with stepwise models

		Transfer function :log sig				
		Coefficient correlation (r) % & Mean squared error				
Training function	NO. of hidden layers	Training	Test	Validation	All	MSE
Train LM	5	99.29	99.18	97.93	98.99	0.013
	7	98.19	98.59	98.65	98.29	0.009
	9	97.40	96.59	98.08	97.32	0.010
	10	98.20	97.59	98.23	98.39	0.011
	12	98.77	96.29	97.85	98.23	0.011
Train RP	9	97.37	96.23	97.79	97.43	0.011
	12	98.2	94.49	96.83	97.35	0.015
	13	96.54	95.8	98.37	96.8	0.010
Train GDX	8	96.67	9.86	97.89	96.67	0.012
	9	96.79	95.27	97.37	96.67	0.014
	10	96.63	96.39	96.75	96.4	0.017
	12	95.93	95.37	96.04	95.83	0.028
		Transfer function: tan sig				
		Coefficient correlation (r) % & Mean squared error				
Training function	NO. of hidden layers	Training	Test	Validation	All	MSE
Train LM	5	97.67	95.34	96.89	97.55	0.015
	7	98.98	98.88	98.99	98.78	0.010
	9	97.35	97.36	96.25	97.16	0.021
	10	98.79	91.89	97.34	97.49	0.016
	12	97.15	98.36	97.45	97.33	0.014
Train RP	9	97.78	95.4	97.78	97.63	0.014
	12	97.22	95.55	96.87	96.67	0.015
	13	96.56	95.59	97.1	96.78	0.017
Train GDX	8	96.39	95.96	95.25	95.85	0.026
	9	97.23	96.88	99.23	97.29	0.014
	10	96.85	94.77	97.94	96.18	0.013
	12	97.31	93.74	97.18	96.71	0.016

for example three degree greater the variation of the a decline production to 10%. The results agreed with Corley and Tinker (2016), but the difference in the proportion of the effect of relative humidity and the percentage of rainfall, that mentions rainfall for optimal palm oil between 2000-3000 mm/year, while according to the Malaysian Meteorological Dept., a decline 5-11% in the 2005 years, 11-20% in 2012, 4-18% in 2014. These climate elements simultaneously cause a decrease in production output by the Average FFB Yield of 2.15 tones/ha/yr. We find that the variables 9 and 10 which are the Evaporation and Cloud cover are considered the most important 0.11 and 0.11, while less values variables recorded for Mature, Immature area and wind speed. In the study carried out by Shanmuganathan *et al.* (2014) looked at the possible

climate change effects on Malaysia’s oil palm yield using 36 monthly average temperatures as lag variables along with yield data at the regional scale.

Results of stepwise with neural network model: The input number has been changed by using the method of stepwise by raising a number of factors that have a limited impact, such as the percentage mature and immature area. After, the repetition of the experiment for the introduction of eight factors and the usage of hidden layers 5 to 13, the best results are shown in Table 5. The results showed the best performance of the network model is obtained when using the training function LM compared with both other training functions. The best five models were selected on the basis of the highest accuracy, the network using the

Table 6: The effect of ANN architecture with stepwise method on the performance of models

ANN architecture	Training error %	Testing error %	Validation error %	Data division %		
				Training %	Validation %	Testing %
8-5-2*	0.0110	0.0190	0.0119	70	15	15
8-7-2*	0.0160	0.02	0.0167			
8-10-2*	0.0168	0.021	0.0137			
8-12-2*	0.0180	0.019	0.0178			
8-7-2**	0.0189	0.022	0.0182			

*log sig **tan sig

Table 7: Forecasting for nine month by using ANN model with stepwise

Date	Target FFB	Forecasted	Error	Target oil yield	Forecasted	Error
JAN-2015	0.98	1.017612	-0.03761	0.19	0.20640	-0.0164
FEB- 2015	0.89	1.198230	-0.30823	0.17	0.23242	-0.06242
MAR-2015	1.08	1.612700	-0.53270	0.22	0.31121	-0.09121
APR-2015	1.24	1.181236	0.05876	0.25	0.26001	-0.01001
MAY-2015	1.33	1.091260	0.23874	0.27	0.24331	0.02669
JUN-2015	1.40	1.299900	0.10010	0.28	0.26847	0.01153
JUL-2015	1.51	1.492520	0.01748	0.30	0.29081	0.00919
AUG-2015	1.75	1.646662	0.10333	0.34	0.319977	0.020023
SEP-2015	1.70	1.682430	0.01757	0.35	0.34942	0.00058
MSE		0.0513			0.00154	
MAEP		1.3229			1.16293	
AA		98.677			98.837	

training function LM and all the five models are constructed using the same input dataset. After the models have been constructed and validated. The ANN architectures are shown in the Table 6. The results obtained from the training stage models have the best ANN architecture 8-5-2. Reliability models are indicated by the value of training error 1.10%, testing error 1.90% and validation error 1.2%. Forecasts for the next nine month from the new models are given in Table 7. The prediction using ANN network model with stepwise method has 98.677% for FFB and 98.837% for oil yield prediction bounds. Forecasts are determined by providing new dataset that was not used during the training session to the constructed models and comparing their outputs with the expected output in terms of percentage accuracy. This model has outperformed in comparison with the ordinary one by allowing prediction accuracy of the yield convergence with actual yield.

CONCLUSION

Two prediction models, Artificial neural networks and ANN with Stepwise are able to predict the productivity of oil palm plantations based on land quality and climate simultaneously with either one. Through the simulation model of the ANN, one see the effect of climate change on several elements simultaneously on land productivity.

REFERENCES

Barcelos, E., S.D.A. Rios, R.N.V. Cunha, R. Lopes, S.Y. Motoike, E. Babiychuk, A. Skiryecz and S. Kushnir, 2015. Oil palm natural diversity and the potential for yield improvement. *Front Plant Sci.*, 6: 190.

Chemura, A., 2012. Determining oil palm age from high resolution satellite imagery. M.Sc. Thesis, International Institute for Geo-Information Science and Earth Observation (ITC), the Netherlands.

Corley, R.H.V. and P.B.H. Tinker, 2016. *The Oil Palm*. 5th Edn., John Wiley and Sons, Hoboken, New Jersey, USA.

Du, K.L. and M.N.S. Swamy, 2013. *Neural Networks and Statistical Learning*. Springer-Verlag, London.

Ficken, F.A., 2015. *The Simplex Method of Linear Programming*. Dover Edn., Dover Publications Inc., Mineola, New York.

Graupe, D., 2013. *Principles of Artificial Neural Networks*. 3rd Edn., World Scientific, New Jersey, USA.

Han, J.S. and G.J. Kim, 2016. A method of unsupervised machine learning based on self-organizing map for BCI. *Cluster Comput.*, 19(2): 979-985.

Haykin, S.S., 2009. *Neural Networks and Learning Machines*. Pearson Education, Upper Saddle River, NJ, USA, Vol. 3.

Kant, G. and K.S. Sangwan, 2015. Predictive modelling for energy consumption in machining using artificial neural network. *Proc. CIRP*, 37: 205-210.

Khamis, N.N., R.M. Satava, S.A. Alnassar and D.E. Kern, 2016. A stepwise model for simulation-based curriculum development for clinical skills, a modification of the six-step approach. *Surg. Endosc.*, 30(1): 279-287.

Kyurkchiev, N. and S. Markov, 2015. Sigmoid functions: Some approximation and modelling aspects. *Biomath. Commun.*, 2(2): 673-687.

- Lamba, V. and V.S. Dhaka, 2014. Wheat yield prediction using artificial neural network and crop prediction techniques (A survey). *Int. J. Res. Appl. Sci. Eng. Technol.*, 2(9): 330-341.
- Lee, Y.N., 2011. Increase Malaysia palm oil production efficiency. M.Sc. Thesis, System Dynamics Group, Faculty of Social Sciences, University of Bergen, the Norway.
- Mousavi, S.M.A., S.R. Shadzadeh and M.A. Riahi, 2015. NMR log prediction from seismic attributes: Using multiple linear regression and neural network methods. *Energ. Source. Part A*, 37(7): 781-789.
- Muhammad, L.N., A.B. Shahrman, B.M.T. Shamsul, B.M. Deros, A.S. Rambely, Y.G. Ng, M.R. Zuradzman, K. Wan, H. Desa, E.M. Cheng and M. Afendi, 2015. Conceptual design for oil palm evacuation to improve ergonomics issues of collecting fresh fruit bunch. *Appl. Mech. Mater.*, 786(6): 281-286.
- Neto, O.C.D.R., A.D.S. Teixeira, A.P.D.S. Braga, C.C.D. Santos and R.A.D.O. Leão, 2015. Application of artificial neural networks as an alternative to volumetric water balance in drip irrigation management in watermelon crop. *Eng. Agric.*, 35(2): 266-279.
- Rad, M.R.N., S. Koohkan, H.R. Fanaei and M.R.P. Rad, 2015. Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (*Cucumis melo* L). *Sci. Hortic-Amsterdam*, 181: 108-112.
- Sandry, T.D., 2012. Introductory statistics with R. *Technometrics*, 45(3): 274-275.
- Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Networks*, 61: 85-117.
- Schwarze, S., M. Euler, M. Gatto, J. Hein, E. Hettig, A.M. Holtkamp, L. Izhar, Y. Kunz, J. Lay, J. Merten, S. Moser, O. Mußhoff, F. Otten, M. Qaim, E. Soetarto, S. Steinebach, K. Trapp, M. Vorlauffer and H. Faust, 2015. Rubber vs. Oil Palm: An Analysis of Factors Influencing Smallholders' Crop Choice in Jambi, Indonesia. *EFForTS Discussion Paper Series No. 11, Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia)*, Collaborative Research Centre 990, University of Goettinge.
- Shanmuganathan, S., A. Narayanan, M. Mohamed, R. Ibrahim and H. Khalid, 2014. A Hybrid Approach to Modelling the Climate Change Effects on Malaysia's Oil Palm Yield at the Regional Scale. In: Herawan, T. *et al.* (Eds.), *Recent Advances on Soft Computing and Data Mining*. Springer International Publishing, Switzerland, 287: 335-345.
- Tao, H.H., E.M. Slade, K.J. Willis, J.P. Caliman and J.L. Snaddon, 2016. Effects of soil management practices on soil fauna feeding activity in an Indonesian oil palm plantation. *Agr. Ecosyst. Environ.*, 218: 133-140.
- Vaidyanathan, S. and C. Volos, 2016. *Advances and Applications in Nonlinear Control Systems. Studies in Computational Intelligence*, Springer International Publishing, Switzerland, 635: 683.
- Wagner, J.M. and D.G. Shimshak, 2007. Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *Eur. J. Oper. Res.*, 180(1): 57-67.