

NEURAL NETWORKS METHOD IN PREDICTING OIL PALM FFB YIELDS FOR THE PENINSULAR STATES OF MALAYSIA

YOUSIF Y HILAL*; YAHYA, A**; ISMAIL, W I W** and ASHA'ARI, Z H†

ABSTRACT

Reliable and accurate predictions in oil palm production can provide the basis for management decisions of budgeting, storage, distribution, and marketing. Artificial Neural Network (ANN) and Non-linear Autoregressive Exogenous Neural Network (NARX) models were developed based on 19 440 data set of 15 inputs variables, namely, percentage of mature area and percentage of immature area, rainfall, rainy days, humidity, radiation, temperature, surface wind speed, evaporation and cloud cover, ozone (O₃), carbon monoxide (CO), nitrogen dioxide (NO₂), sulphur dioxide (SO₂), and particulate matter of less than 10 microns in size (PM₁₀) for predicting oil palm fresh fruit bunch (FFB). The results were validated with an independent validation dataset. Results showed that NARX models performed more accurately with multiple coefficients of determination (R²) reached 97% and mean square errors (MSE) between 0.0104%-0.0665%, besides being an easy-to-use tool. Generally, NARX models proved to give more accurate predictions than the predictions of common ANN and Multi Linear Regression (MLR) models. Finally, 15-10-4-1 is chosen as the architecture of NARX for the states of Kedah, Kelantan, Perak, Pahang, Selangor, and Terengganu. The 15-7-4-1 is the best architecture of NARX for the state of Melaka and Pulau Pinang, while 15-13-4-1 architecture is for the state of Johor. This study showed that all of these architectures gave high accuracy with acceptable MSE values.

Keywords: climate factors, forecasting, linear regression, NARX, oil palm yield predictions.

Received: 22 May 2020; **Accepted:** 8 September 2020; **Published online:** 25 November 2020.

INTRODUCTION

Environmental change is widely believed to bring the world into a scarcity of food. It is one of the significant factors affecting development and oil palm production. Climate change has some effects

on oil palm growth and yield. The dominant factors for climate parameters are rainfall, temperature, sunshine periods per day, water deficit and many others (Corley and Tinker, 2016; Saadon *et al.*, 2014). In other words, environmental factors can significantly influence oil palm growth and geographic distribution.

Vegetation has been affected by the amount of rainfall and period (number of rainy days) which are the most vital factors in agriculture. Rainfall in the period of pollinator's mating or several hours after pollination may cause pollination process failure (Oettli *et al.*, 2018; Corley and Tinker, 2016). Fresh fruit bunch (FFB) yield reductions are critically influenced by water deficit. For example, there is an exponential decline down to <10 t of fruit bunches per hectare per year at water deficits of >500 mm (Paterson

* Department of Agricultural Machines and Equipment, College of Agriculture and Forestry, University of Mosul, Iraq.
E-mail: yousifyakoub@yahoo.com

** Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

† Department of Environmental Sciences, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

et al., 2017). Temperature or intensity of heat energy plays an active role in agricultural processes and air temperatures which govern chemical processes and physical states within the palm. Diffusion rates of liquids, the solubility of different substances, and gases in oil palm are dependent on temperature. Corley and Tinker (2016) found a 30% decrease in oil palm yields for every 2°C increase in temperature above the optimal range of growing oil palm. Humidity influences the water requirement of plants. Mean minimum relative humidity for oil palm lies above 45%, with the best relative humidity lying between 70% and 95 % (Do Amaral Teles *et al.*, 2016).

The objective of modelling applications is to create a high-resolution predictive model, based on the readily available response variables in the database. The number of available predictions has significant effects on the model formulation (Kapadia *et al.*, 2017; Gaber *et al.*, 2020).

Many of the studies to understand crop production potential under climate change in Malaysia have focused in using statistical approaches. Studies were often based on single climate scenarios, without relating to differences and overlapping variables. Several analyses have also been conducted to study climate change impacts at regional levels in Malaysia, using statistical models. For example, models by Soon and Hong (2001), Foster (2003), Verdooren (2003), Ahmad Tarmizi *et al.* (2004) and Shanmuganathan and Narayanan (2012) had indicated the effects of weather on oil palm yield and prediction results showed a lower value of R^2 that ranged from 0.15-0.36.

Multi Linear Regression (MLR) is very commonly used in the oil palm industry (Rahman and Robson, 2016; Antle *et al.*, 2017). MLR models in oil palm production, such as models by Keong and Keng (2012), have shown a much reliable performance, the previous developed with a much better value of R^2 of 0.68. Traditional methods are unable to capture the full variability of the sub-seasonal weather and are limited in their ability to project changes in the future. Furthermore, conventional statistical models have limited illustrative capacity, and do not apply to the development of adaptation measures to climate change.

To date, an appropriate framework for modelling an integrated system in palm oil production has not yet been developed. However, the components to create such a system are currently available. Operations Simulator (OPSIM) and Agricultural Production Systems Simulator (APSIM) oil palm model have been developed and tested in the main production regions (Huth *et al.*, 2014; Garrett *et al.*, 2016). Fan *et al.* (2015) conclude that most models have not yet aimed at the full picture of water, air pollution, and energy exchanges between land and atmosphere and remain to be coupled with a limited number of climatic elements.

The challenge in modeling oil palm yield is that it does not follow a linear behaviour. The crop growth and yield production normally have a non-linear and sigmoidal type of relationship (Hilal *et al.*, 2018). In modeling a non-linear curve, the problem becomes more complex with their numerous independent variables. The major hindrance in modeling the behavior of yield is the challenge of extracting the constants of the mathematical models (Quaranta *et al.*, 2020).

Non-linear methods can be used to address many issues as powerful predictive tools. Neural Networks (NN) are one method of modeling non-linear (accommodating multivariate) and non-parametric data with a model-free estimation (Harahap and Lubis, 2018; Waheeb and Ghazali, 2019). Exploratory research on artificial intelligence has revealed that little has been done on oil palm yield (Khamis *et al.*, 2006; Zuhaimy and Azme, 2011). NN are a diverse set of techniques concocted to meet the diverse needs of forecast modeling. Unlike traditional regression-based methods, NN can reliably model non-linear relationships (Majumdar *et al.*, 2016; Da Silva *et al.*, 2017). The concept of the NN is based on the human brain, which comprises billions of interconnected neurons through their synapses. NN have three significant advantages: parallelism, the ability to learn, and the ability to generalise. The parallelism is ideal for complicated tasks such as pattern recognition, which would be much more challenging to accomplish serially. Another significant advantage of NN is their ability to learn by training. Training consists of supplying the NN with many training samples, each of which includes a set of inputs and the desired sets of outputs (Asta, 2015). Due to their efficient capacity to learn and generalise (Khairunniza-Bejo *et al.*, 2014), NN have garnered widespread attention among researchers in various fields, to generate and analyse experimental data. NN are a significant modeling instrument that has terrific potential in many areas of studies, especially in the area of oil palm production where modeling is very useful. Therefore, the main objectives of the research are to develop and propose the preferred Artificial Neural Network (ANN) and Non-linear Autoregressive Exogenous Neural Network (NARX) architectures for predicting oil palm fresh fruit bunch (FFB) yield within multiple areas in Peninsular Malaysia. It also develops MLR models to predict FFB, and compares with the final NN models.

MATERIAL AND METHODS

Description of the Study Area

Peninsular Malaysia is one of the major oil palm production regions in Malaysia. Its area is

130 598 km² and extends from latitude 1° 20' N to 6° 40' N and from longitude 99° 35' E to 104° 20' E. The latitude and longitude for Peninsular Malaysia include the nine states, as described in *Table 1*.

Input and output data for FFB models. The models were developed to predict FFB oil palm yield based on prior values of the environment (climatic change, air pollution) and types of oil palm areas (mature and immature). Monthly data were collected using 15 variables as inputs and one variable for the output. Data were obtained from the Malaysian Palm Oil Board (MPOB), Meteorological Department, and Department of Statistics, Prime Minister's Department, Malaysia as described in *Table 2*.

The whole data for nine states were 19 440 numbers. The total data for each state was 2160 numbers. The data were partitioned into two data sets consisting of the development set and the

simulation set. The simulation set was not used during model development that was held back in the evaluation of the final model. The available data, from January 2005 to December 2015 were used in the development set, while the data, from January to March 2020 were applied in the simulation set.

Development of ANN Models for FFB Yield

The best ANN model simulations that use hypothetical data were conducted to forecast the FFB yield by the effect of input data that affects the productivity of oil palm plantations. The ANN is programmed by MATLAB's Neural Network Toolbox. The architecture of ANN consists of three layers; the input layer, a hidden layer, and an output layer. There are three kinds of neurons, input, hidden and output neurons. Each of the following needs to be examined to understand how a neuron produces an output from its arbitrary inputs.

TABLE 1. DESCRIPTIONS OF THE STUDY SITES

Fresh fruit bunch area			
District	State	Location	Oil palm planted areas (ha)
Kubang Pasu	Kedah	6.3726° N, 100.4370° E	2 373
Kota Bahru	Kelantan	6° 09' 49" N, 102° 18' 02" E	24
Kota Tinggi	Johor	1.7294° N, 103.8992° E	149 684
Melaka Tengah	Melaka	2° 16' N, 102° 15' E	1 997
Seberang Prai Selatan	Pulau Pinang	5.1958° N, 100.4921° E	6 895
Hilir Perak	Perak	4.0209° N, 101.0230° E	107 630
Kuantan	Pahang	3.7634° N, 103.2202° E	81 592
Kuala Langat	Selangor	2.8038° N, 101.4951° E	33 114
Kuala Terengganu	Terengganu	5° 23' N, 103° 6' E	3 679

TABLE 2. INPUT AND OUTPUT VARIABLES

Input data	Source
Types of the oil palm area	Malaysian Palm Oil Board (MPOB)
Percentage of mature (%)	
Percentage of immature (%)	
Climate	Malaysian Meteorological Department
Rainfall (mm)	
Rain day	
Humidity (%)	
Radiation (Mjm ²)	
Temperature (°C)	
Surface wind speed (m s ⁻¹)	
Evaporation (mm)	
Cloud cover (oktas)	
Air pollution (ppm)	
Ozone (O ₃)	
Carbon monoxide (CO)	
Nitrogen dioxide (NO ₂)	
Sulphur dioxide (SO ₂)	
Particulate matter of less than 10 microns in size (PM ₁₀)	Department of Environment, Ministry of Natural Resources and Environment, Malaysia
Output	MPOB
FFB yield t ha ⁻¹	

Input neurons. The neuron can have an arbitrary number of inputs. Typically, a neuron has more than one input. In this study, we have used variables, 15 input variables to predict FFB yield. In general, ANN randomly selects a sample of 70% for training, a sample of 15% for cross-validation and 15% used for a test (Harahap and Lubis, 2018).

Hidden neurons and output neurons. The hidden neurons do an intermediate computation before directing the input to output neurons. The treatments in the hidden neurons vary according to the Neural Network structure. The number of neurons in the hidden layer is determined through experimentation and the discretion of the network designer. It usually has been decided by trial and error. Actually, 5, 7, 10 and 13 hidden neurons (nodes) are suggested for this network topology. There are output neurons that are connected to input fields. Each output neuron computes one dependent variable.

Training function. The training process was employed to find optimal weight matrices and bias vectors that minimise a predetermined error function. The proposed ANN model used in this study is the supervised learning, and Levenberg-Marquardt backpropagation (LM algorithm), which is the most widely-utilised optimisation algorithm. It is commonly used because it outperforms simple gradient descent and other conjugate gradient methods.

Transfer function. This study used the continuous non-linear function which publishes values by entering them in accordance with the scale of schedule in the emerging amount. The non-linear,

curved S-shape function is called the sigmoid function, which is the most popular transfer function for a non-linear relationship. It is mathematically well behaved, differentiable and strictly increasing function. It publishes values by entering them by the scale of schedule in the emerging amount (Kyurkchiev and Markov, 2015).

Development of NARX Models for FFB Yield

The NARX is a recurrent dynamic network, with feedback connections including many layers of the net. The series-parallel NARX dynamic network will be used as a basis for the models in this study with a structure as shown in Figure 1. The NARX model is commonly used in time-series modelling. The NARX consists of the following parts:

Part 1. Layers and neurons of the NARX. The structure of a neuron in NARX is similar to ANN. Therefore, the architecture of the NARX consists of three layers which are: input, hidden and output layers, with some variations that include, Tapped Delay Line (TDL), of which the delay line is applied to feed the network with the past values of inputs. In this study, the delay values are four for each NARX model architecture (Guzman *et al.*, 2017). The number of neurons in the input layer will be equal to the number of inputs provided to the networks. Using additional layers makes the model more complicated and increases the time required for training and simulation. The default number of hidden neurons at a hidden layer is set to 10; in this work 5, 7, 10 and 13 hidden neurons are suggested for NARX topology. Each output neuron computes an architecture of NARX for FFB yield models.

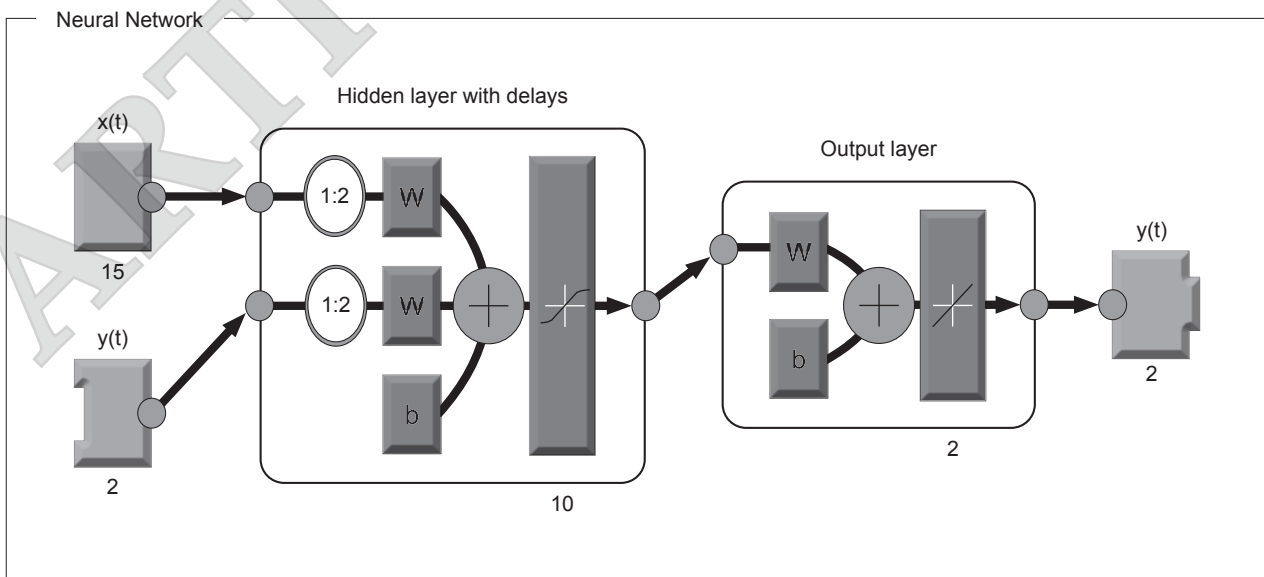


Figure 1. Snapshot of series-parallel Non-linear Autoregressive with Exogenous Neural Network (NARX) dynamic network.

Part 2. Training and transfer functions. Training is mapping the objectives within the neuron range and training the network using an algorithm that is suited for such an implementation. The MATLAB® Neural Network Toolbox provides a detailed survey of algorithms suitable for different applications. LM is the best algorithm for NARX networks approximation and actual output (default by MATLAB®). This study will use the sigmoid neurons (default by MATLAB®), which have been extensively used in most of the Neural Network applications. The datasets were distributed into three partitions as training, validation and testing partitions. The constructions of 70%:15%:15% were set up for implementation (Harahap and Lubis, 2018).

Development of MLR for FFB Yield

The data was analysed using the Design-Expert Version 9.0.1 software (Stat-Ease Inc., Minneapolis, USA). The objective is to build MLR models to predict the oil palm FFB yield. Briefly, steps to build a multiple regression model are as follows:

- Step 1. Define parameters and read the data in the linear models. This step is designed for importing data that already exists. Fifteen input variables and one output variable were selected for the most significant variables in FFB models.
- Step 2. Transformation of the response. Design-Expert provides extensive diagnostic capabilities to check if the statistical assumptions underlying the data analysis are met.
- Step 3. Model selection. The most appropriate mathematical models were chosen for predicting the interactive effects of input parameters on oil palm production. Evaluate a model after it has been run; there are three choices; the p-value, Adjusted R-Squared and the analysis of variance (ANOVA). The MSE results from performing a linear method in the ANOVA. In this work, the following performance measurement functions were employed:

$$R\text{-Squared} = 1 - [SS_{\text{Pure Error}} / (SS_{\text{Model}} + SS_{\text{Pure Error}})] \quad (1)$$

$$\text{Adj.R-Squared} = 1 - (SS_{\text{Pure Error}} / DF_{\text{Pure Error}}) / [(SS_{\text{Model}} + SS_{\text{Pure Error}}) / (DF_{\text{Model}} + DF_{\text{Pure Error}})] \quad (2)$$

where:

$SS_{\text{Pure Error}}$ - pure error sum of squares.

SS_{Model} - the model sum of squares.

$DF_{\text{Pure Error}}$ - pure error degrees of freedom.

DF_{Model} - model degrees of freedom.

Forecasting Simulation Models

Simulation modeling is necessary to measure the robustness and efficiency of any developed forecasting algorithm. It allows verifying the robustness of the different algorithms, to consider the component combination on the model to obtain the test scenario. The performance of the model can be evaluated by the validity of its estimation, its ability to reproduce the actual yield in simulation and its stability. Hence, the estimated model in this study will be validated and evaluated on the basis of its forecasting power by using MSE. We want to regularly forecast monthly in 2020 ahead and produce a new forecast each month.

The simulation and prediction functions were developed and compiled into an application with a functional user interface in the MATLAB network programmer. The following performance measure functions were employed (Ranjit and Sinha, 2016):

1-Mean square error (MSE) for evaluating the accuracy of the proposed ANN, NARX and linear models.

$$MSE = \frac{1}{N} \sum_{i=1}^N (T_i - Y_i)^2 \quad (3)$$

where:

T_i - the actual yield.

Y_i - the forecasted yield.

N - the number of data points for $i=1, 2, \dots, N$.

RESULTS AND DISCUSSION

Development of Oil Palm FFB Yield Models

For the best training, the initial connectivity weights of neurons and biases are selected as following the connection weights and biases between the input layer, the hidden layer and the output layer are selected randomly as small numbers.

Tables 3 and 4 present the performance of the ANN, NARX and MLR models at nine states. The outcome of the ANN and NARX models contains 15 inputs and one output (FFB). It is seen from Table 3 the values of R^2 are very high for ANN and NARX models, almost all of the R^2 values were between 0.842%-0.974% during the network training. On the other hand, the R^2 values of MLR models significantly decrease in most states. Moreover, the best performance of the network model is shown by its sensitivities to the difference of the number of hidden neurons.

Table 4 shows that the effect of a different number of hidden neurons on MSE in nine states. MSE is the lowest that provided the best model performances. It can be seen the values of MSE are very low for most ANN and NARX models, where

most MSE values were between 0.0104%-0.0665%. There was a significant increase in MSE values for MLR models in most states. For example, MSE values calculated for the MLR models for Pahang, Selangor and Terengganu, was 0.7410%, 0.1272% and 0.4182%, respectively.

Based on the evaluations, there are a critical number of hidden units existing for minimising the error rate for each state. If there are too many hidden units, a network can simply memorise the correct response to each pattern in its training set instead of learning a general solution. The number of hidden neurons to be used in a layer is rather arbitrary and it has usually been decided through trial and error. In this study, the variation of R² and MSE for all phases against several iterations for a different number of hidden neurons were 5, 7, 10 and 13. On the other hand, utilising more than 13 neurons in the network makes the computation process more

complicated and expensive. Currently, there is no reliable method exists despite several efforts to arrive at an approximate formula. It also takes a long time to train the network. Therefore, trial and error are still employed to derive some empirical rules to find the optimal network sizes (Gaurang *et al.*, 2011).

With several trials and errors, the variation of the learning rate and the best performance of the model is obtained with seven hidden neurons in Kedah, Johor and Perak. There were 10 hidden neurons in Kelantan, Melaka, Pulau Pinang, Selangor and Terengganu, while 13 hidden neurons for Pahang for the best ANN architecture in nine states. The best number of hidden neurons for NARX models is seven hidden neurons of Melaka and Pulau Pinang, and 10 hidden neurons in Kedah, Kelantan, Perak, Pahang, Selangor, and Terengganu. The best number of hidden neurons for Johor is 13 neurons as shown in Table 5.

TABLE 3. EFFECT OF HIDDEN NEURON NUMBERS ON COEFFICIENT OF DETERMINATION (R²) OF MODELS IN NINE STATES

Models	ANN R ²				NARX R ²				MLR R ²
	5	7	10	13	5	7	10	13	
No. of hidden neurons	5	7	10	13	5	7	10	13	-
Kedah	0.897	0.946	0.895	0.888	0.884	0.896	0.953	0.903	0.249
Kelantan	0.908	0.909	0.960	0.902	0.893	0.911	0.960	0.914	0.407
Johor	0.868	0.960	0.869	0.889	0.903	0.895	0.915	0.974	0.352
Melaka	0.887	0.893	0.954	0.875	0.910	0.962	0.891	0.910	0.191
Pulau Pinang	0.909	0.906	0.950	0.902	0.892	0.958	0.891	0.890	0.540
Perak	0.902	0.962	0.886	0.906	0.904	0.906	0.968	0.908	0.327
Pahang	0.891	0.906	0.905	0.961	0.917	0.914	0.964	0.918	0.336
Selangor	0.883	0.905	0.962	0.888	0.883	0.888	0.969	0.895	0.310
Terengganu	0.878	0.876	0.942	0.842	0.876	0.877	0.949	0.871	0.434

Note: ANN – Artificial Neural Network; NARX – Non-linear Autoregressive Network with exogenous inputs; MLR – Multi Linear Regression.

TABLE 4. EFFECT OF HIDDEN NEURON NUMBERS ON MSE OF MODELS IN NINE STATES

Models	ANN MSE				NARX MSE				MLR MSE
	5	7	10	13	5	7	10	13	
No. of hidden neurons	5	7	10	13	5	7	10	13	-
Kedah	0.050	0.036	0.051	0.047	0.036	0.059	0.027	0.066	0.090
Kelantan	0.015	0.011	0.011	0.017	0.014	0.015	0.010	0.012	0.055
Johor	0.023	0.017	0.019	0.020	0.021	0.022	0.018	0.014	0.149
Melaka	0.041	0.047	0.031	0.037	0.033	0.020	0.033	0.026	0.088
Pulau Pinang	0.020	0.016	0.020	0.019	0.029	0.015	0.029	0.021	0.083
Perak	0.016	0.014	0.012	0.013	0.019	0.016	0.012	0.016	0.045
Pahang	0.020	0.018	0.018	0.016	0.018	0.021	0.018	0.019	0.741
Selangor	0.022	0.020	0.014	0.019	0.018	0.013	0.011	0.017	0.127
Terengganu	0.025	0.023	0.022	0.026	0.018	0.025	0.011	0.022	0.418

Note: ANN – Artificial Neural Network; NARX – Non-linear Autoregressive Network with exogenous inputs; MLR – Multi Linear Regression; MSE - mean square error.

TABLE 5. ARCHITECTURE OF BEST NEURAL NETWORKS MODELS

States	Architecture of best ANN models	Architecture of best NARX models
Kedah	15-7-1	15-10-4-1
Kelantan	15-10-1	15-10-4-1
Johor	15-7-1	15-13-4-1
Melaka	15-10-1	15-7-4-1
Pulau Pinang	15-10-1	15-7-4-1
Perak	15-7-1	15-10-4-1
Pahang	15-13-1	15-10-4-1
Selangor	15-10-1	15-10-4-1
Terengganu	15-10-1	15-10-4-1

Note: ANN - Artificial Neural Network; NARX - Non-linear Autoregressive Network with exogenous inputs.

The total results demonstrated that the forecasting ability of ANN and NARX NN is sensitive to the number of hidden neurons. These results concurred with Arbain and Wibowo (2012). Due to results which contradict with Taskaya-Temizel and Casey (2005), they then concluded that the forecasting ability of NN is not sensitive to the number of hidden nodes.

Scatter plots of predicted output oil palm FFB yield against actual values are shown in *Figures 2 and 3* for the testing data sets. It can be seen that actual and predicted FFB are visual representation which simply suggests deviation in the direction of a line between the actual and predicted for all training with an entire data set for the NARX model. Almost a linear trend can be observed between the actual and predicted FFB, which justifies the efficacy of the model, except for a clear deviation in Pahang as presented. For the ANN model, visual representation strongly suggests deviation in the direction of a line between the actual and predicted FFB for the entire data set because the MSE values varied between models, and the NARX models outperformed ANN models in nine states.

Simulation and Selecting Model

The accuracy of the developed models should always be evaluated with a set of data, not used in the training process (Kline, 2000). During the last section of the FFB model evaluation, the model's response is compared by testing the data. The model verification focused on checking the residuals of the model to determine if any systematic pattern could be removed to improve the chosen model. As shown in *Tables 6 and 7*, a comparison between the MLR with ANN and NARX forecasts, with the actual values based on the MSE within the studied states. The predicted values were very close to the measured values for NARX models. The outputs of

the network were found to be closer to the actual values of the FFB yield amount than the linear model. Moreover, the average accuracy for the forecasts by the linear model was very low and largely differed with the actual values.

The finding indicated that the average MSE values were between 15%-75% for the FFB models, which indicated the good performance of the multivariate forecasting model based on the statistical results. This result is in agreement with the theorem, which states that a neural network with a single hidden layer and a sufficiently large number of neurons can well predict any arbitrary continuous function (Khamis *et al.*, 2006; Zuhaimy and Azme, 2011).

The best models were used for forecasting the NARX models are superior to that of the other models. The average of MSE values for the forecasts by NARX models had a record lowest value of between 0.0019% and 0.006%. The NARX model achieved high accuracy in the overall forecast during 2020. It indicated that the results fit the available data sufficiently to satisfy the forecasting. Consequently, the NARX model was proven to be an appropriate model for forecasting the data being studied. NARX model can be used as an alternative model to predict the FFB yield of oil palm.

Test the Significance of Independent Variables

The capacity of the oil palm production was explored from the perspective of the NARX models. The analysis was conducted by changing independent variable indicators for different possible changes in various regions. Results showed that the predictive power of the analysis of changes in some of the factors. The independent variable power can be determined with the distinction between the highest and lowest values for each scenario.

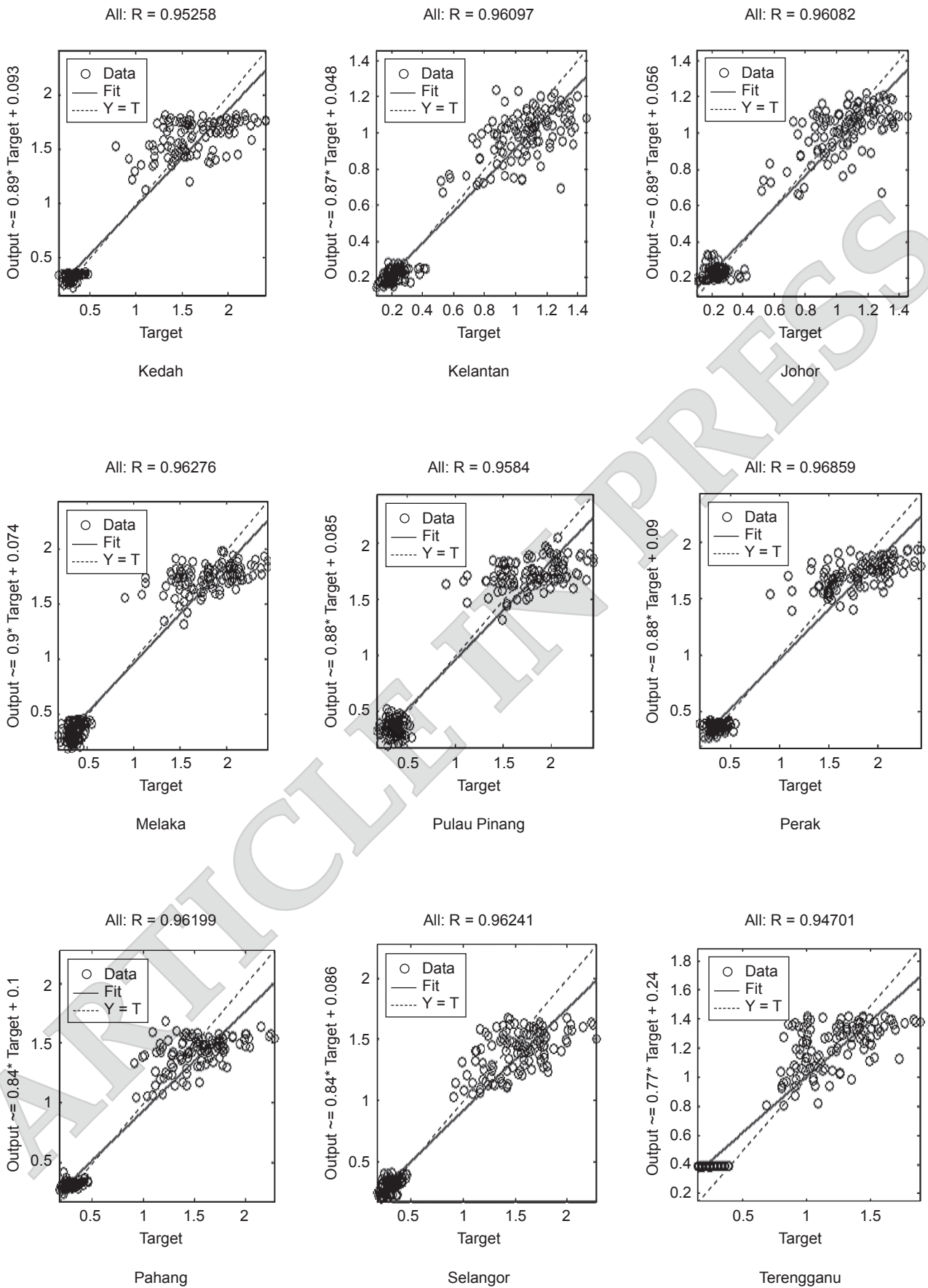


Figure 2. The scatter plots of Artificial Neural Network (ANN) models predicted vs. actual values.

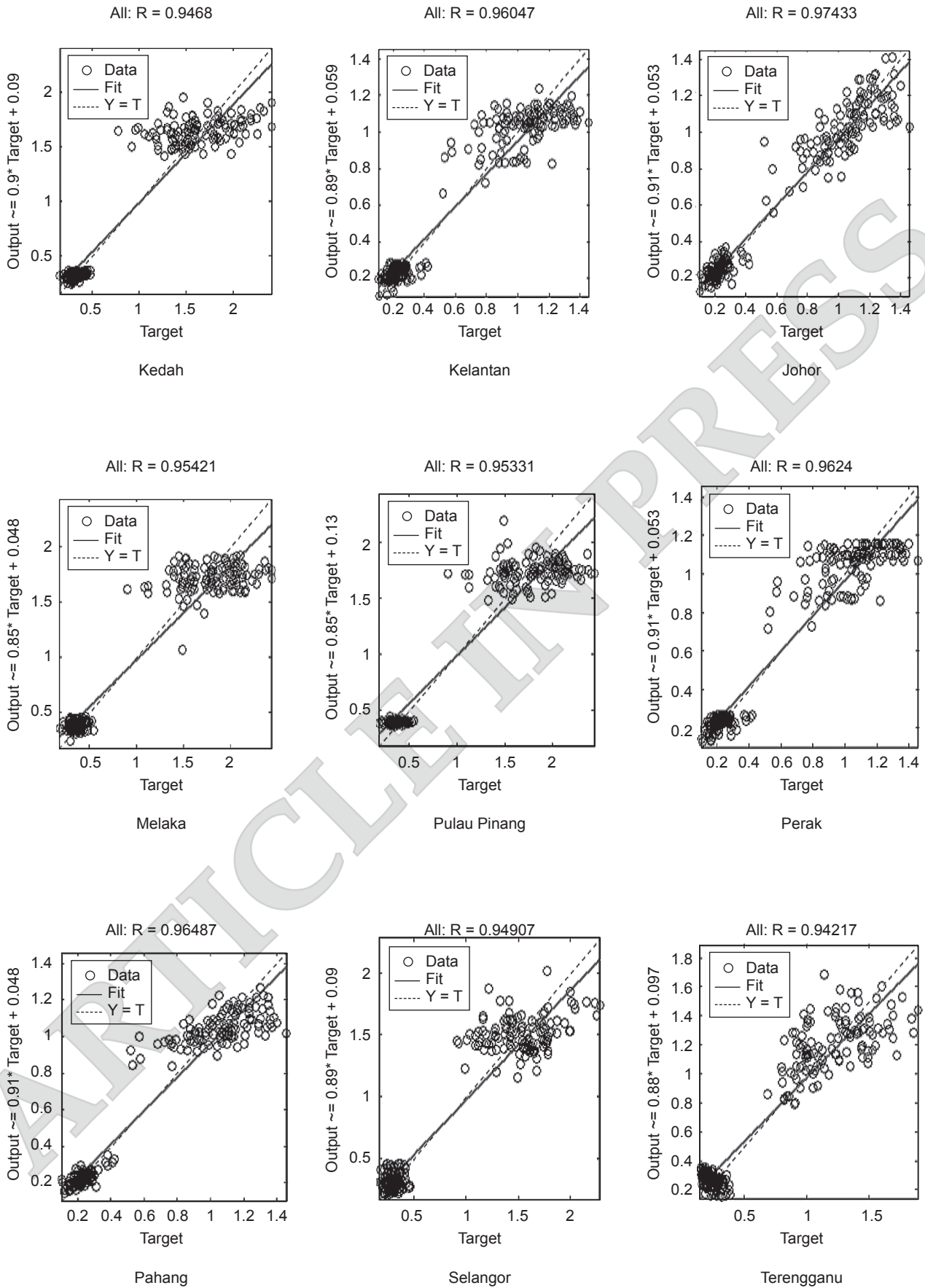


Figure 3. The scatter plots of Non-linear Autoregressive Exogenous Neural Network (NARX) models predicted vs. actual values.

TABLE 6. FORECAST MODELS WITH THE ACTUAL YIELD IN 2020 FOR THE EAST COAST OF PENINSULAR MALAYSIA

States	Forecasting by	Date of 2020			Forecast performance
		t			MSE
		Jan	Feb	Mar	
Kelantan	FFB yield	0.49	0.55	0.79	-
	ANN	0.46	0.49	0.74	0.0023
	NARX	0.47	0.52	0.75	0.0011
	MLR	0.12	0.13	0.19	0.2244
Terengganu	FFB yield	0.80	0.82	0.97	-
	ANN	0.74	0.76	0.90	0.0040
	NARX	0.76	0.78	0.92	0.0019
	MLR	0.39	0.40	0.47	0.1980
Pahang	FFB yield	0.79	0.92	1.16	-
	ANN	0.7426	0.8648	1.0904	0.0033
	NARX	0.7505	0.8740	1.1020	0.0023
	MLR	0.1975	0.230	0.290	0.5280

Note: ANN - Artificial Neural Network; NARX - Non-linear Autoregressive Network with exogenous inputs; MLR - Multi Linear Regression; MSE - mean square error. FFB - fresh fruit bunch.

TABLE 7. FORECAST MODELS WITH THE ACTUAL YIELD IN 2020 FOR THE WEST COAST OF PENINSULAR MALAYSIA

States	Forecasting by	Date of 2020			Forecast performance
		t			MSE
		Jan	Feb	Mar	
Kedah	FFB yield	1.07	1.55	1.63	-
	ANN	1.0058	1.457	1.5322	0.0074
	NARX	1.0165	1.4725	1.5485	0.0024
	MLR	0.675	0.875	0.475	0.6485
Johor	FFB yield	1.18	1.36	1.42	-
	ANN	1.1092	1.2784	1.3348	0.0063
	NARX	1.121	1.292	1.349	0.0043
	MLR	0.595	0.4	0.55	0.6735
Melaka	FFB yield	0.89	1.33	1.50	-
	ANN	0.8366	1.2502	1.41	0.0057
	NARX	0.8455	1.2635	1.425	0.0040
	MLR	0.522	0.533	0.437	0.6335
Pulau Pinang	FFB yield	0.62	1.20	1.34	-
	ANN	0.5828	1.128	1.2596	0.0043
	NARX	0.589	1.14	1.273	0.0030
	MLR	0.155	0.43	0.5335	0.4865
Perak	FFB Yield	1.35	1.61	1.68	-
	ANN	1.269	1.5134	1.5792	0.0086
	NARX	1.2825	1.5295	1.596	0.0060
	MLR	0.5375	0.7025	0.842	0.7286
Selangor	FFB Yield	1.17	1.52	1.72	-
	ANN	1.0998	1.4288	1.6168	0.0079
	NARX	1.1115	1.444	1.634	0.0055
	MLR	0.5925	0.638	0.643	0.7571

Note: ANN - Artificial Neural Network; NARX - Non-linear Autoregressive Network with exogenous inputs; MLR - Multi Linear Regression; MSE - mean square error. FFB - fresh fruit bunch.

TABLE 8. EFFECT OF INDEPENDENT VARIABLES IMPORTANCE IN FRESH FRUIT BUNCH (FFB) YIELD FOR THE PENINSULAR STATES OF MALAYSIA

Experiment variables	Peninsular states of Malaysia								
	Kedah	Kelantan	Johor	Melaka	Pulau Pinang	Perak	Pahang	Selangor	Terengganu
% of mature	0.019	0.080	0.08	0.044	0.031	0.056	0.02	0.033	0.039
% of immature	0.071	0.032	0.06	0.057	0.086	0.081	0.047	0.029	0.035
Rainfall	0.063	0.078	0.05	0.038	0.038	0.027	0.037	0.131	0.051
Rain day	0.178	0.065	0.03	0.080	0.047	0.041	0.116	0.048	0.043
Humidity	0.020	0.077	0.04	0.051	0.090	0.053	0.034	0.073	0.098
Radiation	0.066	0.051	0.09	0.122	0.080	0.083	0.096	0.092	0.079
Temperature	0.062	0.057	0.06	0.082	0.059	0.084	0.085	0.085	0.064
Wind speed	0.042	0.127	0.04	0.039	0.139	0.114	0.051	0.078	0.177
Evaporation	0.040	0.043	0.03	0.036	0.066	0.071	0.079	0.055	0.030
Cloud cover	0.031	0.070	0.14	0.094	0.07	0.124	0.155	0.044	0.074
Ozone (O ₃)	0.127	0.105	0.07	0.048	0.063	0.014	0.122	0.035	0.048
Carbon monoxide (CO)	0.039	0.034	0.10	0.088	0.069	0.036	0.04	0.044	0.090
Sulphur dioxide (SO ₂)	0.084	0.093	0.06	0.029	0.055	0.061	0.041	0.095	0.040
Nitrogen dioxide (NO ₂)	0.137	0.033	0.03	0.136	0.054	0.115	0.045	0.030	0.064
Particulate matter of less than 10 microns in size (PM ₁₀)	0.021	0.054	0.120	0.055	0.053	0.039	0.032	0.127	0.067

As shown in Table 8, the findings of this research, using 11 years of climate change and air pollution, have significantly affected the oil palm production. Surface wind speed was recorded at an impact ratio of up to 12.7%, 13.9% and 17.7%, which highly correlates to the productivity of oil palm plantations in Kelantan, Pulau Pinang and Terengganu, respectively. Additionally, the cloud cover is considered as the most essential variable recorded with an impact ratio of up to 14%, 12.4% and 15.5% on FFB yields in Johor, Perak and Pahang while the rainy days, total rainfall and average NO₂ in the air recorded the most significant impact up to 17.8%, 13.1% and 13.6% of FFB yields in Kedah, Selangor and Melaka, respectively. The effect of corresponding variables on production and the possible oil palm productivity which benefitted from direct sunshine was suggested as most of South-east Asia has low incidence of cloud cover, which probably caused higher oil palm yields compared with those of West Africa (Hilal *et al.*, 2018). According to the data collected, wind speed, cloud cover, rainfall, and the number of rainy days correlated positively with the productivity of oil palm plantations in the Peninsular states of Malaysia. Corley and Tinker (2016) concluded that the pattern of rainfall and number of rainy days were markedly visible, between July and October. August was recorded as the month with the highest days of rainfall, with an average of 25.6 days. However, the month with the highest total rainfall was in September, which recorded 500 mm. The

number of rainy days was most influential during the harvest season.

Air pollution in Peninsular Malaysia can be highly variable from year to year due to biomass burning. The meteorological data for each station in 2006, 2009, 2013, 2014 and 2015, showed monthly averages of ozone (O₃), carbon monoxide (CO), nitrogen dioxide (NO₂), sulphur dioxide (SO₂) and particulate matter of less 10 microns in size (PM₁₀) values higher than normal. Higher values appeared to occur in June to October, which roughly correspond to the drier season and an increase in biomass burning.

CONCLUSION

The current study developed and applied a Neural Network model to predict oil palm FFB yield. The power of a Neural Network was used to produce a fast and efficient solution promptly. The results of this study indicated that the NARX models can be successfully used to predict the variables based on large variables and allowing a much more comprehensive search of the solution. Through the model analysis of the variables, one sees the effect of climate change, air pollution and quality of land for oil palm areas on several elements simultaneously on land productivity. NARX models are easy to apply on a wide range to obtain optimisation prediction models with high precision oil palm yield models.

The results of this study are only based on nine states in Malaysia. Future works will focus in expanding the data, increasing inputs and samples to further explore other areas through similar studies.

REFERENCES

- Ahmad Tarmizi, M; Zin, Z; Mohd, T and Ariffin, D (2004). Oil palm fertilizer program: A proposal for higher yield. Presented in MPOB and RISDA Action Plan Meetings, Prime City, Kluang. August 2004.
- Antle, J M; Jones, J W and Rosenzweig, C E (2017). Next generation agricultural system data, models and knowledge products: Introduction. *Agricultural Systems*, 155(7): 186-190.
- Arbain, S H and Wibowo, A (2012). Neural networks based nonlinear time series regression for water level forecasting of Dungun River. *J. Computer Science*, 8(9): 1506-1513.
- Asta, S (2015). *Machine Learning for Improving Heuristic Optimization*. Doctoral dissertation. University of Nottingham, United Kingdom. 173 pp.
- Corley, R H V and Tinker, P B (2016). Reference list and index of citations. *The Oil Palm*. Fifth edition. Wiley. p. 537-626.
- Da Silva, I N; Spatti, D H; Flauzino, R A; Liboni, L H B and Dos Reis Alves, S F (2017). *Artificial Neural Networks*. Cham: Springer International Publishing. 305 pp.
- Do Amaral Teles, D A; Braga, M F; Antoniassi, R; Junqueira, N T V; Peixoto, J R and Malaquias, J V (2016). Yield analysis of oil palm cultivated under irrigation in the Brazilian Savanna. *J. Amer. Oil Chem. Soc.*, 93(2): 193-199.
- Fan, Y; Roupsard, O; Bernoux, M; Le Maire, G; Panferov, O; Kotowska, M M and Knohl, A (2015). A sub-canopy structure for simulating oil palm in the Community Land Model (CLM-Palm): Phenology, allocation and yield. *Geoscientific Model Development*, 8(11): 3785-3800.
- Foster, H L (2003). Assessment of oil palm fertilizer requirements. *Oil Palm: Management for Large and Sustainable Yields* (Fairhurst, T and Hardter, R eds.). Potash and Phosphate Institute (PPI), Potash and Phosphate Institute Canada (PPIC) and Int. Potash Inst. (IPI), Singapore. p. 231-257.
- Gaber, J (2020). *Qualitative Analysis for Planning and Policy: Beyond the Numbers*. Routledge, Taylor and Francis. 165 pp.
- Garrett, R D; Carlson, K M; Rueda, X and Noojipady, P (2016). Assessing the potential additionally of certification by the Round Table on Responsible Soybeans and the Roundtable on Sustainable Palm Oil. *Environmental Research Letters*, 11(4): 1-21.
- Gaurang, P; Amit, G; Kosta, Y and Devyani, P (2011). Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *Int. J. Computer Theory and Engineering*, 3(2): 332-337.
- Guzman, S M; Paz, J O and Tagert, M L M (2017). The use of Narx neural networks to forecast daily groundwater levels. *Water Resources Management*, 31(5): 1591-1603.
- Harahap, I and Lubis, M (2018) Application of an Artificial Neural Network (ANN) model for predicting oil palm fresh fruit bunch (FFB) yield based on rainfall and the yield previously. *J. Pen. Kelapa Sawit*, 26(2): 59-70.
- Hilal, Y Y; Ishak, W; Yahya, A and Asha'ari, Z H (2018). Development of genetic algorithm for optimization of yield models in oil palm production. *Chilean J. Agricultural Research*, 78(2): 228-237.
- Huth, N I; Banabas, M; Nelson, P N and Webb, M (2014). Development of an oil palm cropping systems model: Lessons learned and future directions. *Environmental Modelling and Software*, 62: 411-419.
- Kapadia, A S; Chan, W and Moyé, L A (2017). *Mathematical Statistics with Applications*. USA: CRC Press. 896 pp.
- Keong, Y K and Keng, W M (2012). Statistical modelling of weather-based yield forecasting for young mature oil palm. *APCBEE Procedia*, 4: 58-65.
- Khairunniza-Bejo, S; Mustaffha, S and Ismail, W I W (2014). Application of artificial neural network in predicting crop yield: A Review. *J. Food Science and Engineering*, 4(1): 1-9.
- Khamis, A; Ismail, Z; Haron, K and Mohammed, A T (2006). Neural network model for oil palm yield modelling. *J. Applied Sciences*, 6(2): 391-399.
- Kline, P (2000). *The Handbook of Psychological Testing*. Psychology Press. 744 pp.

- Kyurkchiev, N and Markov, S (2015). *Sigmoid Functions: Some Approximation and Modelling Aspects. Some Moduli in Programming Environment Mathematica*. LAP Lambert Acad. Publ., Saarbrucken. 110 pp.
- Majumdar, J; Mal, A and Gupta, S (2016). Heuristic model to improve feature selection based on machine learning in data mining. *Sixth International Conference on Cloud System and Big Data Engineering (Confluence)*. p. 73-77.
- Oettli, P, Behera, S K and Yamagata, T (2018). Climate based predictability of oil palm tree yield in Malaysia. *Scientific Reports*, 8(1): 1-13.
- Paterson, R R M; Kumar, L; Shabani, F and Lima, N (2017). World climate suitability projections to 2050 and 2100 for growing oil palm. *The J. Agricultural Science*, 155(5): 689-702.
- Quaranta, G; Lacarbonara, W and Masri, S F (2020). A review on computational intelligence for identification of nonlinear dynamical systems. *Nonlinear Dynamics*. p. 1-53.
- Rahman, M M and Robson, A J (2016). A novel approach for sugar cane yield prediction using Landsat time series imagery: A case study on Bundaberg region. *Advances in Remote Sensing*, 5(2): 93-102.
- Ranjit, K P and Sinha, K (2016). Forecasting crop yield: A comparative assessment of ARIMAX and NARX model. *RASHI*, 1(2): 72-87.
- Saadon, S; Uemura, Y and Mansor, N (2014). Torrefaction in the presence of oxygen and carbon dioxide: The effect on yield of oil palm kernel shell. *Procedia Chemistry*, 9: 194-201.
- Shanmuganathan, S and Narayanan, A (2012). Modelling the climate change effects on Malaysia's oil palm yield. 2012 IEEE Symposium on E-Learning, E-Management and E-Services (IS3e). 21-24 October 2012. Kuala Lumpur, Malaysia. 6 pp.
- Soon, B B F and Hong, H W (2001). Oil palm responses to N, P, K and Mg fertilizers on two major soil types in Sabah. *Proc. of the Int. Palm Oil Congress - Agriculture*. p. 318-334.
- Taskaya-Temizel, T and Casey, M C (2005). A comparative study of autoregressive neural network hybrids. *Neural Networks*, 18(5): 781-789.
- Verdooren, R (2003). Design and analysis fertilizer experiments. *Oil Palm: Management for Large and Sustainable Yields* (Fairhurst, T and Hardter, R eds.). Singapore: Potash & Phosphate Institute/Potash & Phosphate Institute of Canada (PPI/PPIC) and International Potash Institute (IPI). 382 pp.
- Waheeb, W and Ghazali, R (2019). A novel error-output recurrent neural network model for time series forecasting. *Neural Computing and Applications*. p. 1-27.
- Zuhaimy, I and Azme, K (2011). Neural network in modeling Malaysian oil palm yield. *Amer. J. Applied Sciences*, 8(8): 796-803.