Indonesian Journal of Electrical Engineering and Computer Science
Vol. 3, No. 3, September 2016, pp. 626 ~ 633
DOI: 10.11591/ijeecs.v3.i2.pp626-633

Oil Palm Yield Forecasting Based on Weather Variables Using Artificial Neural Network

Nadia Dwi Kartika, I Wayan Astika, Edi Santosa
Postgraduate student MIT of IPB postgraduate school
Department of Mechanical and Biosystem Engineering Bogor Agricultural University
Department of Agronomy and Horticulture Bogor Agricultural University
Corresponding author, e-mail: nadia.dwi@bppt.go.id, wayanastika@ipb.ac.id, edisang@gmail.com

Abstract

Forecasting of oil palm yield has become a main factor in the management of oil palm industries for proper planning and decision making in order to avoid monthly high cost in harvesting. Predicting future value of oil palm yield with minimum error becomes an important issue recently. A lot of factors determine the productivity of oil palm and weather variables play an important role that affect plant growth and development that may reduce yield significantly. This research used secondary data of yield and weather variables available in company administration. It proposed feed forward neural network with back propagation learning algorithm to build a monthly yield forecasting model. The optimization procedure of ANN architecture obtained the best using 60 neurons in input layer, five hidden layers and one neuron in the output layer. Training data were from January 2005 to June 2008 while testing data were from July 2008 to December 2009. ANN architecture using five hidden layers gave the best accuracy with MAE 0.5346 and MSE 0.4707 while the lowest accuracy occurred by using two hidden layers with MAE 1.5843and MSE 4.087.

Keywords*: oil palm, yield forecasting, weather variables, feed forward neural network*

1. Introduction

Oil palm (Elaeis guineensis Jacq.) is one of the most rapidly expanding crops in tropical countries especially Indonesia that has been cultivated on more than 16 million hectares [1]. This sector gives a big contribution to the economic in rural areas by creating jobs and increasing income of farmers and rural people. This study explores the possibility of improving accuracy of the model by inputting some variables. Generally, the productivity of oil palm in a plantation is caused by a lot of factors such as the origin place of the plant [2], genotype [3], light intensity that influences the manufacture of plant food, length of stem, flowering and leaf color [4], rainfall [5], water requirement [6], population of oil palm trees [7], the age of oil palm trees [8], fertilizer application [9], land suitability [10], and management of pest and diseases [6]. Oil palm trees need humid equatorial conditions to thrive. Seasonal droughts at higher tropical latitudes greatly reduce yields and water-stressed oil palm trees will produce fewer female flowers and abort unripe fruits [11]. Oil palm is a perennial crop, it generally begins to produce fruits 30 months after being planted in the field. However, the yield of an oil palm tree is relatively low at this stage. As the oil palm continues to mature, its yield increases and reaches its peak production in from the age of 9-18, and gradually decline thereafter. The typical commercial lifespan of an oil palm tree is approximately 25 years and after 25 – 30 years, the trees will become too tall to harvest and get replaced [11]. For optimal growth and production, the trees requires stable climatic conditions, in particular with respect to light and moisture supply and any deviation from these conditions will enhance a yield decrease [12].

Weather is one of the key components that control crop yield. It affects plant growth and development, particularly at critical crop growth stages that may reduce yield significantly [13]. It is well known that one of the main factors causing yield to change from year to year is the climate variability as no two growing seasons experience exactly the same weather [14]. In some cases, it has been stated that as much as 80% of the variability of agricultural production is due to the variability in weather conditions [15, 16]. The occurrence of abnormal weather during the growing season or during critical development stages may hinder growth processes,

Received April 2, 2016; Revised July 25, 2016; Accepted August 10, 2016

resulting in yield reduction. That is why weather variables become the main factor especially during dry season because it can make oil palm trees stressful and affect the yield, so monitoring weather parameters such as rainfall, temperature, solar radiation, relative humidity, and wind speed is useful in forecasting oil palm yield [17]. A research study has been conducted regarding oil palm yield, agro-ecological factors such as relative humidity, sunshine hours, rainfall, rainfall days and water deficit, age of oil palm tree, and fertilizer application affected the yield and determined the accuracy of oil palm yield forecasting [18]. The study was carried out at a private plantation that is well established in Angsana District, South Kalimantan.

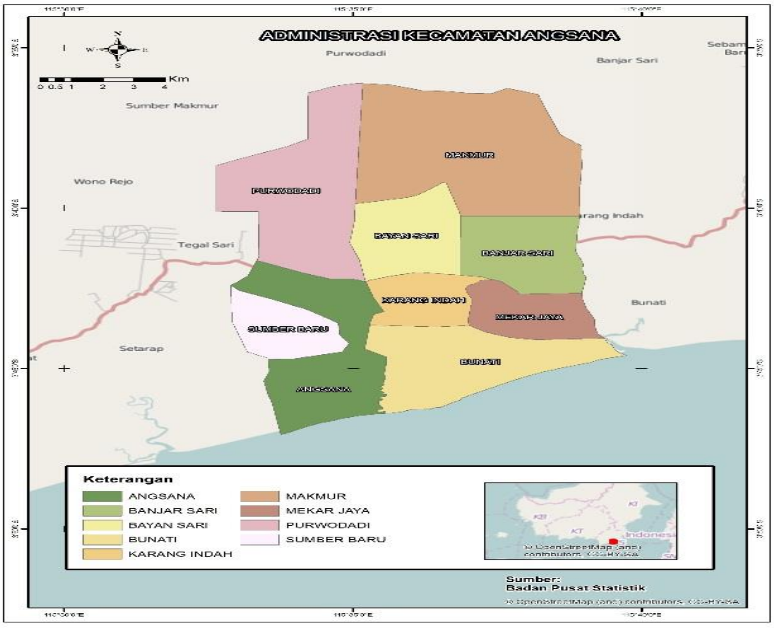

Figure 1. Map of Angsana District

Oil palm plantation aims to increase the yield efficiency so that the price of oil products become more competitive. Yield forecasting is an important part in increasing the yield efficiency. It affects the operations, scheduling of production in a plantation where the total labors, fertilizer, equipment supply and transportation for harvesting need to be estimated accurately. The accuracy of current forecasting misses up until 50% from the targets of production and to build a new forecasting equation needs time. Therefore, to obtain an accurate yield forecasting, it needs to have a further exploration in modeling of oil palm where the main issue in yield forecasting is to predict the future value with the minimum error.

Crop yield forecasting plays an important role in farming planning and management, domestic food supply, international food trade, ecosystem sustainability, and so on [19, 20]. Yield forecasting in oil palm plantations is an important element to the organization to plan or adopt the necessary policies in order to make a better development and decision making in setting targets of production, operational and financial planning firm. Few studies have been done to predict crop yield by using weather parameters [21]. A research was conducted in crop yield forecasting model based on weather variables by considering weather data on maximum and minimum temperature, morning relative humidity and rainfall (MAXT, MINT, RHI, and RF) for each agro-climatic zone in India [22]. Rainfall data were used o forecast oil palm yield at Kumai plantation in Central Kalimantan, and the forecasting missed up 20 – 50% [23]. The fluctuation of dry and wet season makes the monthly yield become fluctuating and hard to predict [10]. As a result, the choice of the model of forecasting is one important factor that would affect the accuracy of forecasting.

Artificial intelligence-based approach is important for time series forecasting. Due to the popularization, there is a large interest in Artificial Neural Network (ANN) technique. One of major application area of ANN models is forecasting [24]. It is difficult to ignore the use of ANN models to predict the reliable forecast. ANN has received a great deal of attention because it can treat complicated problems especially if the data are imprecise. ANNs provide an attractive alternative tool for both forecasting researches and practitioners. ANN models find relationships by observing a large number of input and output examples to develop a formula that can be used for prediction [25]. The number of hidden nodes determines the number of connections between inputs and outputs and may vary depending on the specific problem under study. If there are too many nodes used then the ANN model may become over-trained and cause it to memorize the training data and result in poor predictions [26]. ANNs are well suited for problems whose solutions require knowledge that are difficult to specify although there are enough data of observations.

The scope that limits this research is that the age of trees are at their peak production, the plantation is implementing good agricultural practices where the plantation is well irrigated, fully mechanized and well equipped with all modern agricultural technologies and it is governed with a high level management practices with the regular use of high yielding varieties, recommended doses of fertilizers, proper plant protection measures including soil water management.

2. Research Method

The ANN is a computational tool based on the biological nervous system. ANN finds relationships by observing a large number of input and output examples to develop a formula that can be used for predictions. These layers are connected by links which consist of weights. The weights are adjustable parameters of the network and are determined from a dataset through the process of training [27]. A three layered feed forward back propagation neural network is used in this study. It consists of one input layer that represents the considered variables, one or more hidden layers that contain some nodes which process every element of the network and output layer as the result of the network. The input signal propagates through the network in a forward direction, layer by layer then the hidden units extract useful information from inputs and use them to predict the output. These nodes are known as neurons which are fundamental elements of a network and interconnected in a definite rule to form architecture through network weights.

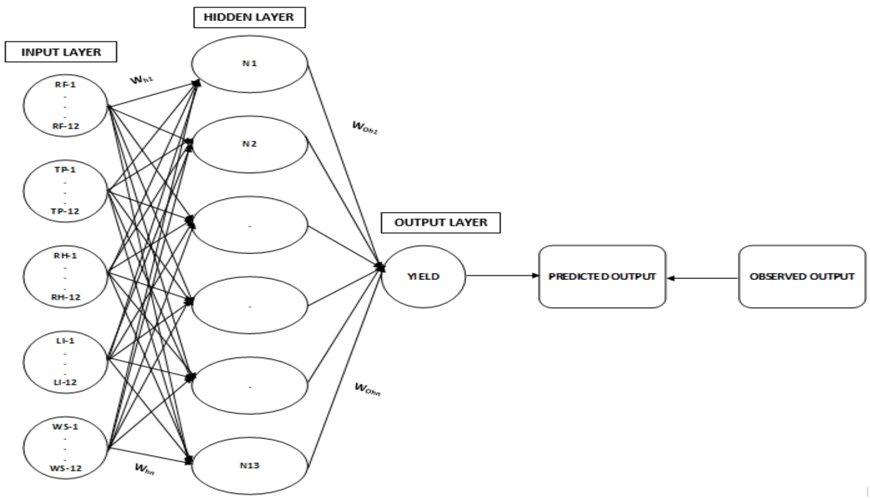

Figure 2. Feed Forward Neural Network with Back Propagation Learning Algorithm

ANN architecture has a set of requirements that must be satisfied. When modeling oil palm yield, the date are reorganized to fit the ANN requirement of the parameters involved. Secondary data were used in this research where most of the data of yield and weather variables are in tabular form and obtained from the plantation. Five weather parameters were included in this study, namely rainfall, temperature, relative humidity, light intensity, and wind speed. The time-series weather data and yield data of six years from 2004–2009 were obtained from undergraduate thesis where the location of the plantation was in Tanah Bumbu, South Kalimantan. The assumption from an expert said that the current oil palm yield was influenced by the weather variables some months on previous year before [28] and according to Risza [29] the productivity of oil palm plantations is also depending on the age composition of the plant in a plantation. Based on that assumption the prediction of oil palm yield started at t-12. Gaining a large dataset of oil palm yield was the limitation in this research because production data are confidential being shared to the public.

This study proposed feed forward neural networks with back propagation learning algorithm to describe how the neural network processed and recalled patterns. It is a supervised training algorithm where the network must be provided with both sample inputs and anticipated output that will be compared against the actual output for given input. The ANN model using of this study can be seen in Figure 2.

The raw data used were checked, validated and partitioned them into training, validation and testing set of data. The validation is essential in ANN because it will indicate the presence of faulty data. By definitions, training sets are used to update the weights in a network, validation sets are used to decide the architecture of the network and the testing sets are used to examine the final performance of the network. The oil palm yield prediction model also could be validated by using yield real data obtained from the plantation. The most important thing in building the ANN model is to make sure that the training set contains enough data and the distribution is suitable enough to demonstrate every property of the network. The back propagation learning phase for a pattern consists of a forward phase followed by a backward phase. In the forward phase, the hidden layer weight matrix (W) is multiplied by the input vector (X) to calculate the hidden layer output. Using the anticipated outputs the back propagation training algorithm will take a calculated error and adjust the weights that minimize the error between target values and network outputs.

$$\boldsymbol{y_{h,j} = f\left(\sum_{i=1}^{N_i} w_{h,ji} x_i - \theta\right)}$$

Where $w_{h,ji}$ is the weight connecting input unit i to unit j in the hidden neuron layer. The θ is an offset termed bias that is trained like an ordinary weight. If the network gives the wrong answer the weights are corrected so that error is lessened and as a result, future responses of the network are more likely to be corrected. The Error measure (E_p) for a training pattern p, the function is defined as:

$$\boldsymbol{E_p = \frac{1}{2}\sum_{k=1}^{N_o}\left(d_{p,k} - y_{p,o,k}\right)^2}$$

3. Result and Analysis

In this study the oil palm yield has been predicted 12 months earlier. The data have not been applied to any research related to forecasting using intelligence method. Each variable (rainfall, temperature, relative humidity, light intensity, and wind speed) deviations of the data were used as independent or input variables and yield data as dependent or output variable in development of ANN model.

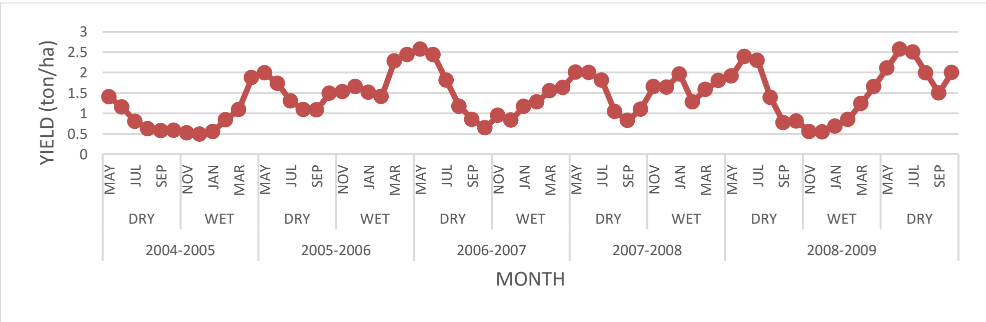

Figure 3. Oil Palm Yield Data in 2004-2009

Figure 3 showed how the oil palm yield at Angsana Estate over last six years had been fluctuating. The actual production in 2004-2009 had not reached the standard production of Indonesian Oil Palm Research Institute (IOPRI) for land suitability classes S3 where the productivity was supposed to be 13-18 ton/ha/year. The fluctuation of monthly yield can be caused by climatic conditions which also tend to fluctuate and it affects the spread of production and becomes difficult to predict the monthly yield [10].

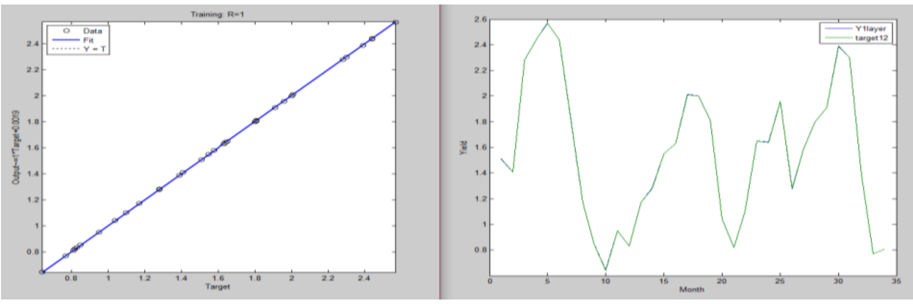

Figure 4. Result of Training Process Using One Hidden Layer

The feed forward neural networks were built by using three layers at first, one input layer, one hidden layer with 61 nodes and one output layer with one node and epoch maximum was set at 500 with 0.01 of MSE. The performance of the training process with one hidden layer is shown in figure 4 where $R^2 = 1$ and RMSE = 0.0014.

3.1. Training Process

Back propagation is the best-known learning algorithm for training process. During the training process of neural network with one until three hidden layers, the algorithm progressed fastly that only required 2-4 iterations to stop the process. On the other hand, the training process took longer by adding four and five hidden layers with 61 nodes for each layer. The initial network configuration on training process is random and on each iteration, the weights from training process were submitted to the network and target then the actual outputs compared to the error calculated. This error was used to adjust the weights and then the process repeated until it stopped after reaching an acceptable level. In this study the network configuration model set iteration to 500, the resilient back propagation learning algorithm set to learning parameter of 0.01. All variables were standardized by normalization of means and standard deviation. Training process was developed and governed by minimalizing the mean square error (MSE) between actual yield and predicted yield.

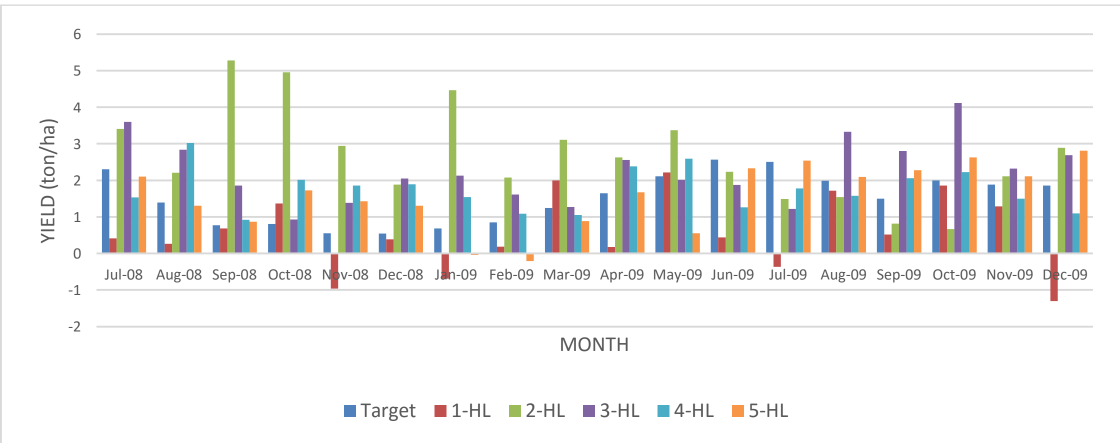

Figure 5. Result of Testing Process after Being Validated

3.2. Testing Process

The data sets were splitted in two parts, 70 % of the values were used for training and 30% for testing purpose. There were 1080 data with 60 parameters as input variables from July 2008 to December 2009 used in this testing process (Figure 5).

After getting a high accuracy from the training process, the model would be applied to the testing process. Apparently, this model with one hidden layer was supposed to be good enough to be applied into the testing process but the result after being validated showed otherwise, so the training process needed to be continued by adjusting hidden layer into the network until the model gave a low accuracy after being validated by current yield data (Figure 5).

Table 1. Validation Result of Error Value By Adjusting Hidden Layer

No. of Hidden Layer	MAE	MSE	RMSE
1	1.102303	2.039054	1.427955
2	1.584289	4.086959	2.021623
3	0.974465	1.245605	1.116067
4	0.737069	0.740118	0.860301
5	0.534546	0.4707	0.686076

Table 1 showed the validation result of error value for each model with different hidden layer and the ANN model with five hidden layer gave the lowest error 0.4707 of MSE value and ANN model with 2 hidden layer gave the highest error or 4.086959 of MSE value. It took almost 1 hour for running ANN model with five hidden layer and the training process with more hidden layer could not continue due to the performance of the mobile computer. From the result, it can be concluded that the number of hidden layer affects the accuracy of the ANN model because the error value is getting lower.

Low accuracy of forecasting can occur because a lot of factors. For an instance, from climatic factor, the amount of rainfall. Oil palm tree is commonly grown in tropical areas where rain is abundant throughout the entire year. Production can be negatively effected in areas where rain does not meet the water demands. Oil palm is a fast growing crop with high productivity and biomass production. The spread of uneven rainfall can increase the opportunity of dry month (rainfall < 60mm/month) in the wet season that causes water deficit which is very influential in the production of oil palm because it can affect the flowering. The lack of water during dry season and poor water management can reduce 8 – 10% of normal production in the first year after water deficit and 3-4% in the second year after water deficit [6]. The climate changes also causes longer periods of time without rain. Although oil palm is growing naturally in tropical climate productivity will be effected. The other factor that gives low accuracy is regarding the yield and weather variables data, especially the weather data. If the recording data relied upon are inaccurate and or size of the data are not large enough to build an ANN model, the impact of inaccurate data will affect the output of the model where the yield prediction will always miss the real target in a large range of value.

4. Conclusion

Oil palm yield forecasting model aims to increase the yield efficiency so the price of oil products become more competitive. It affects the operations, scheduling of production in a plantation where the total labors, fertilizer, equipment supply and transportation for harvesting need to be estimated accurately. Building a monthly yield forecasting model based on weather variables needs a large dataset which is accurate both yield data and weather variables data has become an important factor in the management of oil plam industries for proper planning and decision making in order to avoid monthly high cost in harvesting.

The application of a feed forward neural network with back propagation learning algorithm to the yield data and weather variables data practically to forecasting the yield of oil palm because current oil palm yield was influenced by the weather variables some months on previous year before. The ANN model gives a high accuracy when it comes to forecasting. The validation result of ANN model with five hidden layer gave the lowest error 0.4707 of MSE value and ANN model with 2 hidden layer gave the highest error or 4.086959 of MSE value. It took almost 1 hour for running ANN model with five hidden layer and the training process with more hidden layer could not continue due to the performance of the mobile computer. The number of hidden layer affects the accuracy of the ANN model because the error value will get lower. The low accuracy of forecasting model of oil palm can occur because of a lot of factors such as the amount or the uneven spread of rainfall where production can be negatively effected in areas where rain does not meet the water demands and the recording data relied upon are inaccurate and or size of the data are not large enough to build an ANN model, the impact of inaccurate data will affect the output of the model where the yield prediction will always miss the real target in a large range of value.

References

[1] Food and Agriculture Organization. FAOSTAT Database. Food and Agriculture Organization of the United Nations. Rome, Italy. 2013.
[2] Hetharie H, GA Wattimena, M Thenawidjaya, H Aswidinnoor, N Toruan-Mathius, G Ginting. Karakterisasi morfologi bunga dan buah abnormal kelapa sawit (*Elaeis guineensis* Jacq.) hasil kultur jaringan. *Buletin Agronomi*. 2007; 3: 50-57.
[3] Palupi RE, Y Dedywiryanto. Kajian karakter toleransi terhadap cekaman kekeringan pada empat genotipe bibit kelapa sawit (*Elaeis guineensis* Jacq.). *Bul. Agron.* 2008; 34: 24-32.
[4] Van Kraalingen DWG, CJ Breure, CJT Spitters. Simulation of oil palm growth and yield. *Agric. Forest Meteorol.* 1989; 46: 227-244.
[5] Siregar HH, NH Darian, TC Hidayat, W Darmosakoro, IY Harahap. Hujan sebagai Faktor Penting untuk Perkebunan Kelapa Sawit. Medan: Pusat Penelitian Kelapa Sawit. 2006
[6] Mangoensoekarjo S, H Semangun. Manajemen Agribisnis Kelapa Sawit. Yogyakarta: Gajah Mada University Press. 2005
[7] Corley RHV, CK Mok. Effect of nitrogen, phosporus, potassium and magnesium on growth of the oil palm*. Exp. Agric.* 1972; 8: 347-353.
[8] Lubis AU. Kelapa Sawit (*Elaeis guineensis* Jacq.) di Indonesia. Medan: Pusat Penelitian Marihat Pematang Siantar. 2008.
[9] W Darmosarkoro, ES Sutarta, Winarna. Lahan dan Pemupukan Kelapa Sawit. Medan: Pusat Penelitian Kelapa Sawit. 2007: 65-78
[10] Harahap IY, Winarna, ES Sutarta. Produktivitas tanaman kelapa sawit tinjauan dari aspek tanah dan iklim. 2007: 246-268.
[11] Basiron Y. Palm oil production through sustainable plantations. *European Journal of Lipid Science and Technology.* 2007; 109: 289-295.
[12] Verheye W. Growth and Production of Oil Palm. In: Verheye W. *Editor.* Land Use, Land Cover and Soil Sciences. Encyclopedia of Life Support Systems (EOLSS). Oxford: UNESCO-EOLSS Publishers. 2010.
[13] Satake T, S Yoshida. High temperature-induced sterility in Indica rices at flowering. *Japanese Journal of Crop Science*. 1978; 47: 6-17.
[14] Lobell DB, Burkeb MB. On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology.* 2010; 150: 43-52.
[15] Petr J Weather, Yield. Amsterdam: Elsevier. 1991.
[16] Fageria NK. Maximizing Crop Yields. New York: Marcel Dekker. 1992.
[17] Awan A Majid, Mohd Noor Md Sap. *A framework for predicing oil palm yield from climate data*. Proceedings, Postgraduate Annual Research Seminar. University Technology Malaysia, Malaysia. 2006: 360-366.
[18] Basiron Y. Palm oil production through sustainable plantations. *European Journal of Lipid Science and Technology.* 2007; 109: 289-295.
[19] Jame YW, HW Cutforth. Crop growth models for decision support systems. *Canadian Journal of Plant Science*. 1996; 76(1): 9-19.
[20] Tao F, M Yokozawa, Z Zhang, Y Xu, Y Hayashi. Remote sensing of crop production in China by production efficiency models: models comparisons, estimates and uncertainties. *Ecological Modelling.* 2005; 183(4): 385-396.
[21] Huda AKS, BP Ghildyal, Tomar VS, Jain RC. Contribution of climate variables in predicting rice yield. *Agric. Met.* 1975; 15: 71-86.
[22] Laxmi, Ratna Raj, Amrender Kumar. Weather based forecasting model for crops yield using neural network approach. *Statistics and Applications.* 2011; 9: 59-69.
[23] Prasetyo EI. Analisis hubungan curah hujan dan produksi kelapa sawit dengan model fungsi transfer. Undergraduate Thesis. Bogor: Institut Pertanian Bogor; 2009.
[24] Sharda R. Neural networks for the MS/OR analyst: An application bibliography. *Interfaces.* 1994; 24(2): 116-130.

[25] Pachepsky YA, Timlin D, Varallyay G. Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Science Society of America Journal.* 1996; 60: 727-733.
[26] Lawrence J. Introduction to Neural Networks. Nevada City: Scientific Software Press. 1994.
[27] Zhang GP, Patuwo BE, Hu MY. A simulation study of artificial neural networks for nonlinear time-series forecasting, computers and operations research. *International Journal of Forecasting*. 2001; 28: 381-396.
[28] Sunarko. Petunjuk Praktis Budidaya dan Pengolahan Kelapa Sawit. Jakarta: Agromedia Pustaka. 2007: 70.
[29] Risza S. Kelapa Sawit: Upaya Peningkatan Produktivitas. Yogyakarta: Kanisius. 2009: 189.