RESEARCH

CHILEANJAR
CHILEAN JOURNAL of AGRICULTURAL RESEARCH

Development of genetic algorithm for optimization of yield models in oil palm production

Yousif Y. Hilal[1, 3*], Wan Ishak[1], Azmi Yahya[1], and Zulfa H. Asha'ari[2]

[1]Universiti Putra Malaysia, Faculty of Engineering, Department of Biological and Agricultural Engineering, 43400 UPM Serdang, Malaysia.
[2]Universiti Putra Malaysia, Faculty of Environmental Studies, Department of Environmental Sciences, 43400 UPM Serdang, Malaysia.
[3]University of Mosul, College of Agriculture and Forestry, Iraq. *Corresponding author (yousifyakoub@yahoo.com).

Received: 6 December 2017, Accepted: 13 May 2018; doi:10.4067/S0718-58392018000200228

ABSTRACT

For many years the Malaysian oil palm (*Elaeis guineensis* Jacq.) industry has been facing the challenge of the reduced rate of palm oil yield caused by the gap in the oil palm production and high land usage. In the oil palm industry, modelling and selecting variables play a crucial role in apprehending different issues, i.e. decision making. Nonetheless, the advance in computer technology has created a new opportunity for the study of modelling as selecting variables intended to choose the "best" subset of predictors. Owing to this great interest in the predictions, the study aims to develop a genetic algorithm (GA) to identify the relevant variables and search for the best combinations for modelling to examine the potential of oil palm production in Sarawak and Sabah, Borneo, Malaysia, under a given set of assumptions. Eleven years of high climatic change and air pollution are utilized to secure findings where the primary variable, i.e. the evaporation and surface wind speed, were recorded on the proportion of effect reached up to 100% in Sarawak and Sabah, respectively. Moreover, models were built on the basis of variables that have been selected by the GA. Across the optimization, procedures obtained the best Two Factor Interaction (2FI) models to achieve the best model of oil palm productivity prediction with a value of R^2 of 0.948, mean squared error of 0.022, and the model P-value of < 0.0001 in Sabah. This research concludes that the GA method is a user-friendly variable selection tool with excellent results because it can choose variables correctly.

Key words: Air pollution, climatic change, *Elaeis guineensis*, Sabah, Sarawak, selection variables and sensitivity test.

INTRODUCTION

Over the last 10 years, the Malaysian oil palm (*Elaeis guineensis* Jacq.) industry has encountered many challenges. One of these challenges is the gap in the production of palm oil, which leads to decrease production rate due to the increase in land usage. This issue is the main reason to reduce the average oil extraction rate and raises the number of workers. Labor shortage, in fact, is the most severe constraint, and presently the industry is highly dependent on foreign workers, which in turn increases production cost (Hoffmann et al., 2017). In Malaysia, current planting materials are believed to be capable of achieving a productivity of 40 tons of fresh fruit bunches (FFB) per hectare per year, yielding 6-7 tons of oil. But in reality, average yields are only 50%-60% of this potential (Barcelos et al., 2015; Garrett et al., 2016; Corley and Tinker, 2016).

As a primary concern, the productivity gap is the large gap between the actual production of palm oil per hectare and the crop's genetic potential. It has been widened with time as plant breeders have continued to improve the inherent

productivity of oil palm, but the yields realized have remained static or even declined. The FFB yield for 2014 was lower by 2.1% to arrive at 18.63 t ha^{-1} from 19.02 t ha^{-1} achieved in 2013. Sarawak recorded declines FFB yield by 0.6% to register at 16.13 t ha^{-1}. In 2016, the FFB yield was lower by 14% to 15.91 t ha^{-1} from 18.48 t ha^{-1} achieved in 2015. Sabah registered a decline of 6.3% to 19.99 t ha^{-1} from 21.34 t ha^{-1} achieved in the previous year (MPOB, 2015-2016; USDA Foreign Agricultural Service, 2017).

The palm oil yield in 2014, 2015, and 2016 experienced a decline of 0.3%, 1.9%, and 17% to 3.84, 3.78, and 3.21 t ha^{-1} as compared to record in 2013. The decline in palm oil yield was attributed to declining in FFB yield in last years (MPOB, 2015-2016). The oil palm yield has different amount products in multiple areas of Malaysia distributed between high yield, medium and worst, which have significantly affected the efficiency of production. In 2014, production efficiency was about 3.7 t ha^{-1} yr^{-1} (Oil World, 2014). Additionally, land availability for further expansion is limited, particularly in Sarawak and Sabah, where the cost of land is also substantially higher.

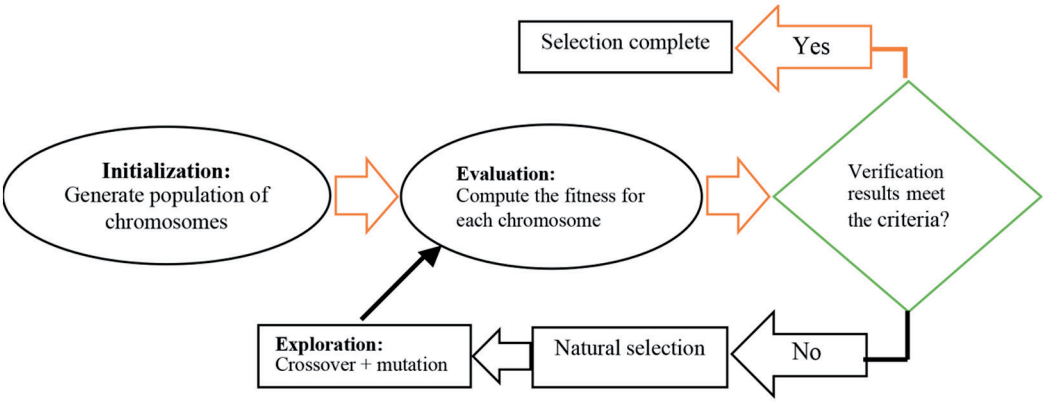
The oil palm industry is a very competitive sector in Malaysia as it contributes to the total revenue of the country (FAOSTAT, 2015). Currently, oil palm is one of the most profitable tropical crops because of its biodiesel production. The United Nations report in 2014 estimated that under the ideal management of high-yield breeding programs, different varieties of oil palm are capable of producing more than 20 t FFB ha^{-1} yr^{-1}. It translates to over 5 t oil ha^{-1} yr^{-1}. Ten percent of the dry biomass of the crop comprises of the oils, while 90% is composed of cellulosic material and fiber, which can be used as second-generation materials for the production of biofuel (Barcelos et al., 2015).

According to Awalludin et al. (2015), as the worldwide demand and consumption of oil palm products escalate, there is a pressing need to study and establish the factors lead to maximum production regarding cultivation. In fact, it is a very complicated as it deals with a broad set of data coming from a number of factors. A lot of factors influence oil palm production performance directly or indirectly. To identify these factors, one needs to know both the number and location of the underlying factors. The common approach is to determine which factors involve testing for a set of effective markers serially by examining a single efficient marker each time using traditional methods such as t-test or ANOVA. Traditionally, for logistic regression models, the variable selection issue has been addressed by stepwise forward, backward, and composite variable selection methods (Rodriguez et al., 2016). Although being well understood and relatively easy to compute, these methods consider the addition or removal of one variable at the time, conditional on the variables already selected (Szymczak et al., 2016). According to Ficken (2015), since linear relationships or linear correlation deals with one parameter at a time. It will be difficult to establish, which parameters have a high efficiency when parameters were together. This sequential approach restricts the examined number of models severely. Another approach is to explore all possible models. Given u variables to choose from, the number of potential models is of the order of 2 to the power of u which renders this exhaustive approach infeasible with other than small numbers of variables.

Responding to the quick and rapid climate changes, studies on GA and their various applications in science, agriculture and engineering have been developed considerably in recent years. The GA has been applauded all over the world as a very reliable tool in selecting data and the optimizing solution (Kelley et al., 2015). As the world moves from tedious traditional methods of modelling, which are time-consuming, less accurate and costly, to a more precise model GA that proved its efficiency in many types of research to optimize the crop yield. It is very critical to appreciate the application of its knowledge in various fields such as the selection model in oil palm. Utilizing multiple parameters, one can determine their effectiveness on oil palm yield. GAs are an optimization method to be constructed on the concepts of natural selection and heredity science (Trejos et al., 2016). The variables are epitomized as genes on a chromosome. The most widely used chromosome is a string consisting of 1's and 0's. In natural selection, 1 denotes the corresponding variable is selected; while 0 denote it is not selected (Zhu and Azar, 2015). The fitness of a chromosome is specified by computing the response function score. The main idea of the GA is to create a population of individuals first, and then, the population is evolved utilizing the principles of variation, selection and inheritance. The GA procedure consists of four main steps as outlined in Figure 1.

This research aims to optimize the oil palm yield by examining all the parameters and determining which one is effective on oil palm yield amount. Thus, the study proposes to use the Genetic algorithm as a robust macro search capability that can find a globally optimal solution with the most significant probability to overcome the shortcomings of multiple Mathematical models. As an optimization tool able to reduce the parameter number and select an optimal set, GA has a higher active role over other ones with the highest accurate result and the lowest error.

Figure 1. Main flowchart of the genetic algorithm.



MATERIAL AND METHODS

Description of data used

Monthly palm oil yield data for the two states Sarawak and Sabah, Borneo, Malaysia, for a period of 11 yr (2005-2015) were obtained from the Malaysian Palm Oil Board (MPOB). The data were taken from each of the tests done on the input parameters used. They were: Quality of land kinds of oil palm areas which included the percentage of mature and immature area. Climate included the rainfall (mm), number of rain days (d), RH (%), daily global radiation (Mj m^{-2}), average temperature (°C), surface wind speed (m s^{-1}), mean daily evaporation (mm) and cloud cover (oktas). While the air pollutants included in Malaysian's Air Pollutant Index (API) calculation are ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and particulate matter of less than10 μ in size (PM_{10}) for a period of 11 yr (2005-2015). Data were obtained from Meteorological Department and Department of Statistics Malaysia Prime Minister's Department.

Steps of proposed method

The Genetic Algorithm/Correlation Analysis (GA/CA) program presented in this study was developed into a method of selecting the quality of land kinds of oil palm areas, climate, and air pollutant factors that affect oil palm production predictions. The program uses the GA as a basic optimization method in searching for the best solution for oil palm production. The program was explicitly constructed for selecting the quality of land kinds of oil palm areas, climate, and air pollutant factors since these factors are of substantial interest in predicting oil palm production as a guide for the optimization. The proposed method uses correlation analysis as the means of calculating the fitness values.

Since we are considering the optimality of the selected subset concerning two objectives, our fitness function comprises two objective functions. In the filter phase, the first objective is to maximize the inter-correlation (i.e. the higher the correlations between input variables are included the quality of land kinds of oil palm areas, climate and air pollutant factors with the output variables the yield palm oil amount) is given by (Kantardzic, 2011):

$$R(S,y) = \frac{1}{/S/} \sum_{k=1}^{/S/} /CORR2(x,y)/ \quad [1]$$

R (S, y) is the overall correlation between the selected input variables *S* and the corresponding output variables *y*, and minimize intra-correlation (i.e., the lower mutual inter-correlations among the input variables the quality of land kinds of oil palm areas, Climate and air pollutant factors, the higher the correlation between the input variables and the output variable).

R(S) is the overall correlation between the selected variables subset (i.e., intra-correlation) is defined as (Freitas, 2013):

$$R(S,y) = \frac{1}{C(S,i)} \sum_{K=1}^{/S/} \sum_{L=K+1}^{/S/} /Corr2(X_K,X_L)/ \quad [2]$$

where *C(/S/,i)* is the number of i^{th} combinations from the selected variables subset *S*. Thus, the fitness function with respect to the first objective.

The second objective of the Fitness Function, as an important part of the program, correlation analysis is used as the method for evaluation of the chromosomes. The results of the correlation analysis are fed back to the GA program to guide the search. The first step in this process is called modeling selection. A variable is a characteristic peak; here a set of variables is represented as a chromosome in the genetic algorithm. Starting with a group of chromosomes created randomly by the GA program, the GA/CA program tries to find the chromosomes that are best capable of prediction the oil palm production. Proposed GA based algorithm for selection of an optimal subset of variables is diagrammatically represented in Figure 2.

Sensitivity test

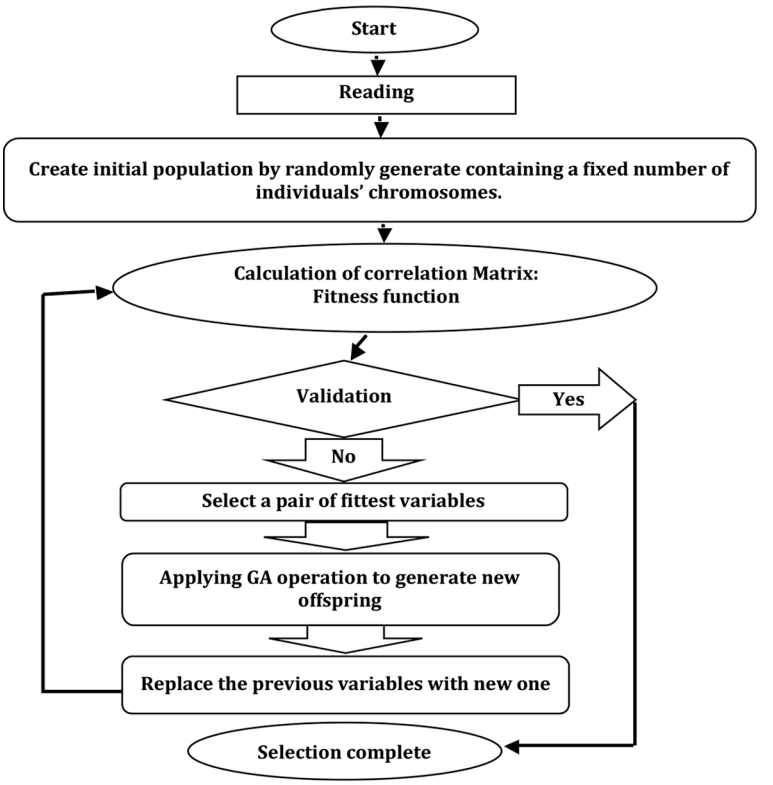
A sensitivity analysis is a technique utilized to determine how different values of an independent variable affect a certain dependent variable under a given set of assumptions. This method is used within specific boundaries that depend on one or more input variables (Svatovnová et al., 2015). Pianosi et al. (2016) stated that the 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 be used for sensitivity analysis to simplify the forms and to verify the robustness of the model predictions and to find out the factors for finding the relative importance of the factors affecting the often contribute to change output productivity making use of allergy testing. In this study, a sensitivity analysis was carried out to determine the effectiveness of a variable using the suggested model in this work. In the analysis, performance evaluations of the different possible interaction of variables were investigated. Therefore, performances of the selection group's variables were investigated by the optimal GA model using the correlation.

Modelling and estimation

The objective is to build multiple Mathematical models to predict the palm oil yield. The palm oil yield (y) is modelled as a function of the quality of land kinds of oil palm areas, climate, and air pollutant:

$$y = \beta_0 + \beta_1 X_1 + \ldots \ldots + \beta_K X_K + \varepsilon \quad [3]$$

Figure 2. Flowchart of the Genetic Algorithm/Correlation Analysis (GA/CA).



While the response variable (y) is the palm oil amount. Independent variables (X) is to determine the independent variables added to the model, which selected by GA method. K is the number of independent variables, and ε is the identically and independently distributed errors. Briefly, steps to build multiple regression models are as follows: (1) Data were analyzed using the Design-Expert Version 9.0.1 software (Stat-Ease Inc., Minneapolis, Minnesota, USA); (2) the most appropriate polynomial Mathematical models were chosen for predicting the interactive effects of parameters of quality of land, climate and air pollutants on oil palm production; and (3) the fitness of each developed model was evaluated by ANOVA.

RESULTS AND DISCUSSION

Applying GA operations and constructing optimal subset of variables

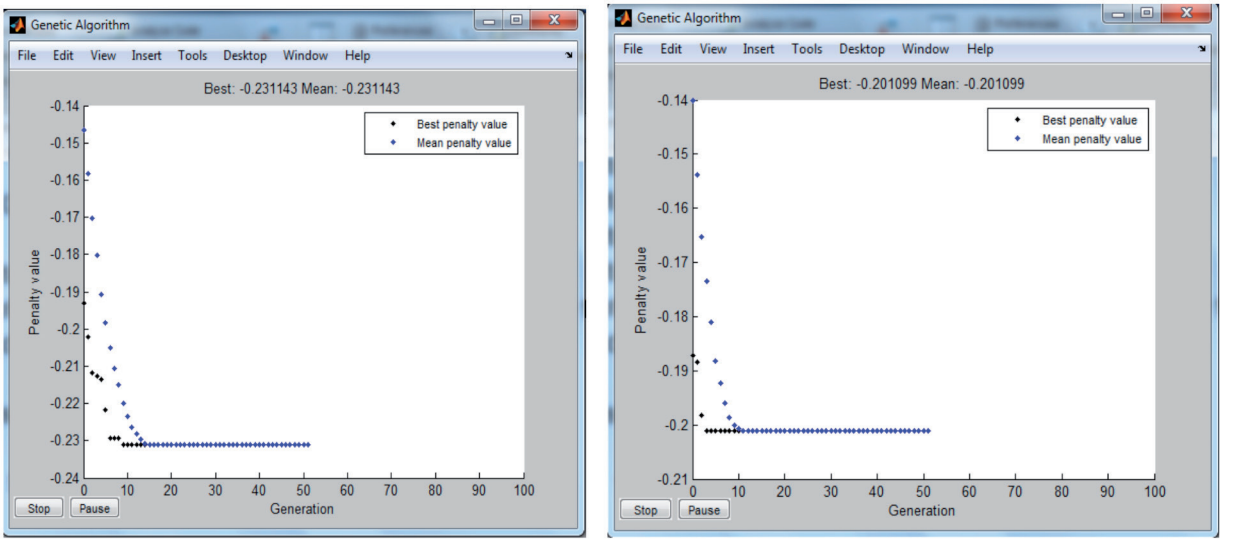
The proposed method is a filter algorithm that ranks variables subsets according to a correlation based optimizing evaluation function. The selection of the evaluation function is toward subsets that contain variables that are highly correlated with the palm oil yield and uncorrelated with each other. Irrelevant variables should be ignored because they will have low correlation with the palm oil yield. Redundant variables should be screened out as they will be highly correlated with one or more of the remaining variables.

The algorithm selects and recombines genes from the individuals in the initial population. The algorithm generates the best individual that it can use these genes at generation, where the best fitness plot becomes level (Purohit et al., 2013). The points at the bottom of the plot refer to the best fitness values, while the points above them indicate the averages of the fitness values in each generation. The scheme also displays the best and mean values in the current generation numerically at the top.

Typically, the best fitness value (fval) improves rapidly in the early generations, when the individuals are farther from the optimum. The best fitness value improves more slowly in later generations, whose populations are closer to the optimal point (Hanan et al., 2016).

The results are displayed in Figure 3 to see how the GA performs with the best and mean values of the population in every generation in Sarawak and Sabah. We can see GA converges quickly to the solution until optimization terminated: average change in the penalty fitness value less than options. The best value found to fval was -0.2312 and -0.2010 in Sarawak and Sabah, respectively. The function was automatically called with the best point found by GA. The best solution to variables selection associated with fval that the smallest one, it is very close to zero. As a result, the best fitness plot is level and the algorithm stalls at generation number 50.

Figure 3. The best fitness value of genetic algorithm performance in Sarawak and Sabah.



Sarawak **Sabah**

Table 1. The Model Selection for Borneo (Sarawak and Sabah).

Independent variable	Sarawak	Sabah
Percentage of mature trees	0	1
Percentage of immature trees	0	1
Rainfall	1	1
Number of rain days	0	0
Humidity	1	0
Radiation	1	1
Temperature	0	0
Surface wind speed	0	1
Evaporation	1	0
Cloud cover	0	1
Average O_3 in the air	0	0
Average CO in the air	0	0
Average SO_2 in the air	0	0
Average NO_2 in the air	1	0
Average PM_{10} in the air	1	0

*0 indicate the variable is not to be selected.
*1 indicate the variable is to be selected.
*PM_{10}: Particulate matter < 10 μ in size.

Table 1 clearly shows that the GA is successful in finding the selected variables. It finds a model that includes a number of extra spurious variables and the variables selected within each of the subject areas were combined. For example, the proposed method selects the six variables, rainfall amount, humidity, radiation, mean evaporation, average NO_2 and average PM_{10} in the air, which is separated from the rest of the variables. It is selected according to the highest affected relation with palm oil yield in Sarawak, while method selects the six variables in Sabah included the percentage of mature and immature trees, rainfall amount, global radiation, surface wind speed, and cloud cover which have achieved the highest correlation with the palm oil yield.

Sensitivity test and model evaluation

Sensitivity analysis facilitates us to assess the important independent variable. We explore how strong the palm oil yield from the selection models perspective. The analysis was done by changing independent variable indicators for different possible changes in various regions. The results show how sensitive is the analysis to changes in some of the factors. Sensitivity can be determined with the distinction between the highest and lowest value for each scenario. The p-value (Prob > F) approach was used for the testing. If the p-value is very small, it rejects the null hypothesis, H0 (there is no factor effect) and thus concludes that at least one of the all process parameters has a non-zero regression coefficient in the developed model.

Regression coefficient (R^2) is a measure of the amount of variation around the mean explained by the model, which also known as a degree of fit measurement that is beneficial for measuring the proportion of total variability explained by the model. The best R^2 value for a good model fitting is somewhat that is closer to 1.0 with not less than 0.8 (Jusoh et al., 2013). A value of 1.0 represents the ideal case in which the chosen model can explain 100% of the variation in the observed values. Predicted- and adjusted-R^2 should be within 0.20 of each other. Otherwise, there may be a problem with either the data or the model.

As shown in Tables 2 and 3, we find that in Sarawak that the variable evaporation is considered the most important variable, as was its importance relative 0.479. Based on the data collected could be analyzed in a single relationship between the parameter with the productivity of palm oil plantations. It followed by variables average NO_2 in the air, radiation, humidity, average PM_{10} in the air and rainfall as where they are importance relative 0.157, 0.127, 0.112, 0.093, and 0.032, respectively. The results agreed with some results by Corley and Tinker (2016), but the difference in the proportion of the effect of RH and the percentage of rainfall.

For oil yield in Sarawak, it had been chosen 2FI model. 2FI models of oil palm yield showed extremely low of the p-value. The Model P-value of < 0.0001 implies the model is significant. Values of "Prob > F" less than 0.05 indicate model terms are significant. In this case, A, B, C, D, E, F, AC, AE, BF, and CD are significant model terms are shown respectively

Table 2. Model selection and estimation in a mathematical model for Sarawak.

Model selection	Variable code	Independent variable importance			
Rainfall	A	0.032			
Humidity	B	0.112			
Radiation	C	0.127			
Evaporation	D	0.479			
Average of NO_2 in the air	E	0.157			
Average of PM_{10} in the air	F	0.093			
Mathematical model		R^2	Adj. R^2	P-value	MSE
Oil yield = 25.008 - 0.0062 * A - 0.368 * B + 0.593 * C + 10.967 * D + 449.538 * E - 1.125 * F + 0.00042 * AC - 0.086 * AE + 0.013 * BF - 0.747 * CD		0.920	0.904	< 0.0001	0.081

Table 3. The P-value range and significance levels of the independent variables (process factors) for Sarawak.

Response	Intercept	A	B	C	D	E
Oil yield	3.243	-0.138***	0.151***	-0.240***	-0.101***	0.076***
P value		0.0003***	< 0.0001***	< 0.0001***	0.0010***	0.0054***
		F	AC	AE	BF	CD
Oil yield		0.097***	0.1748***	-0.045**	0.2111***	-0.077**
P value		0.0007***	< 0.0001***	0.0345**	< 0.0001***	0.0244**

**0.01 ≤ P < 0.05.
***P < 0.01.

in Table 3. The best performances for the model when the mean squared error (MSE) is lowest. The MSE was calculated for a model that was 0.081. The R^2 values of available models are 0.92; it can be clearly seen in Figure 4, which it shows scatter plots of the 2FI model predicted versus actual. According to the predicted model fully fitted to the actual values. The model was considered for representing the fitted response values since the difference between the adjusted and predicted R^2 that was not exceeding 0.2, i.e. 0.016. 2FI were chosen for model fitting since 2FI model is a higher degree of the polynomial model as compared to linear models.

From Tables 4 and 5, in Sabah, the variable surface wind speed is considered the most important variable, as was its importance relative 0.412. Then it is followed by variables radiation, rainfall, percentage of mature trees, cloud cover, and percentage of immature trees as they were importance relative 0.141, 0.083, 0.073 and 0.06, respectively. Based on collected data it can be analyzed in a single relationship between the parameter with palm oil production as surface

Figure 4. The scatter plot of the predicted *vs*. actual for model in Sarawak.

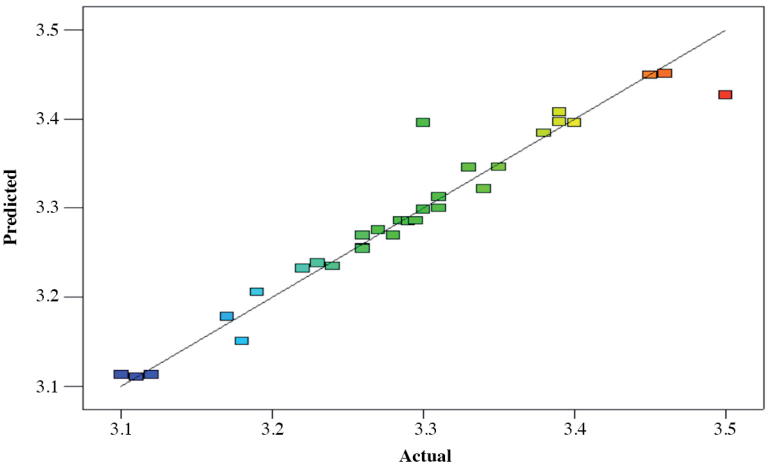


Table 4. Model selection and estimation in a mathematical model for Sabah.

Model selection	Variable code	Independent variable importance			
Percentage of mature trees	A	0.083			
Percentage of immature trees	B	0.060			
Rainfall	C	0.141			
Radiation	D	0.231			
Surface wind speed	E	0.412			
Cloud cover	F	0.073			
Mathematical model		R^2	Adj. R^2	P-value	MSE
Oil yield = - 2600.146 - 0.527 * A - 10.54 * B + 0.028 * C +17.346 * D +1229.562 * E +339.782 * F + 0.112 * AB - 0.0205 CD +0.178 * CE - 5.752 * DE -163.768 * EF		0.948	0.936	< 0.0001	0.022

Table 5. The P-value range and significance levels of the independent variables (process factors) for Sabah.

Response	Intercept	A	B	C	D	E	
Oil yield	5.416	0.916***	-0.561**	0.142***	-0.007*	0.329***	
P value		0.0035***	0.0565**	0.0074***	0.6604*	< 0.0001***	
		F	AB	CD	CE	DE	EF
Oil yield		-0.154***	0.4000***	-1.977***	2.303***	-1.170***	-1.675***
P value		0.0074***	0.0001***	< 0.0001***	< 0.0001***	< 0.0001***	< 0.0001***

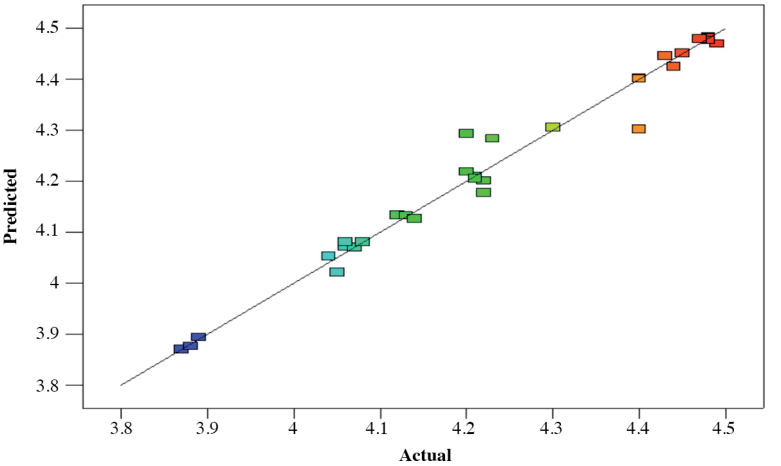
*P ≥ 0.10.
**0.01 ≤ P < 0.05.
***P < 0.01.

wind speed and drought stress that have high effect evident on the productivity of palm oil yield. The results were more inclusive to climatic factors of the many studies that have focused on one variable, such as in the study carried out by Shanmuganathan et al. (2014), who 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.

For oil yield, 2FI were chosen for model fitting in Sabah. 2FI models showed extremely low of the p-value. The Model P-value of < 0.0001 implies the model is significant. Values of "Prob > F" less than 0.05 indicate model terms are significant. In this case, A, B, C, E, F, AB, CD, CE, DE and EF are significant model terms. The values greater than 0.1 indicate the model terms are not significant are shown respectively in Table 5.

The MSE was calculated for a model that is lowest. It recorded 0.022. The R^2 values of available models are 0.948; it can be clearly seen in Figure 5, which it shows scatter plots of the 2FI model predicted *vs*. actual. According to the

Figure 5. The scatter plot of the predicted *vs*. actual for the model in Sabah.



predicted model fully fitted to the actual values. The model was considered for representing the fitted response values since the difference between the adjusted and predicted R^2 that was not exceeding 0.2, i.e. 0.012. 2FI was chosen for model fitting since 2FI model is a higher degree of the polynomial model as compared to linear models. Models were established in this article more efficient than the model by (Keong and Keng, 2012), which was established with monthly oil palm yield as the dependent variable employing agro-meteorological variables, the study showed that the model displayed performance as multiple coefficients of determination (R^2) reached 68%.

CONCLUSIONS

This study developed select variables in a mathematical model for palm oil yield using an optimization technique genetic algorithm (GA). The power of GAs uses to produce fast and efficient solutions incorrect time. The results of this study indicated that GA could be successfully used to select the variables from large variables and to allow a much more comprehensive search of the solution space. Through the sensitivity analysis for the variables chosen, one notices the effect of climate change, air pollution and quality of land of oil palm areas simultaneously on several elements on land productivity. The GA with correlation analysis are easy to apply to a wide range to obtain optimization selection models with high precision mathematical models.

ACKNOWLEDGEMENTS

Authors would like to thank Universiti Putra Malaysia UPM for financial support under Research Grant Putra IPS, the Malaysian Palm Oil Board (MPOB), Meteorological Department and Department of Statistics Malaysian Prime Minister's Department for supplying the data used in this research.

REFERENCES

Awalludin, M.F., Sulaiman, O., Hashim, R., and Nadhari, W.N.A.W. 2015. An overview of the oil palm industry in Malaysia and its waste utilization through thermochemical conversion, specifically via liquefaction. Renewable and Sustainable Energy Reviews 50:1469-1484. doi:10.1016/j.rser.2015.05.085.

Barcelos, E., Rios, S. de A., Cunha, R.N., Lopes, R., Motoike, S.Y., Babiychuk, E. et al. 2015. Oil palm natural diversity and the potential for yield improvement. Frontiers in Plant Science 6:190. doi:10.3389/fpls.2015.00190.

Corley, R.H.V., and Tinker, P.B. 2016. The oil palm. 5th ed. Wiley-Blackwell, Tunbridge Wells, UK.

FAOSTAT. 2015. Database. Available at http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Oilcrops/Documents/Food_outlook_oilseeds/Oilcrops_October_2015.pdf (accessed 12 October, 2016).

Ficken, F.A. 2015. The simplex method of linear programming. Courier Dover Publications, New York, USA.

Freitas, A.A. 2013. Data mining and knowledge discovery with evolutionary algorithms. 2nd ed. Springer Science & Business Media, New York, USA.

Garrett, R.D., Carlson, K.M., Rueda, X., and Noojipady, P. 2016. Corrigendum: Assessing the potential additionality of certification by the Round table on Responsible Soybeans and the Roundtable on Sustainable Palm Oil (2016 Environ. Res. Lett. 11 045003). Environmental Research Letters 11(7). doi:10.1088/1748-9326/11/7/079502.

Hanan, L., Qiushi, L., and Shaobin, L. 2016. An integrated optimization design method based on surrogate modeling applied to diverging duct design. International Journal of Turbo & Jet-Engines 33:395-405. doi:10.1515/tjj-2015-0042.

Hoffmann, M.P., Donough, C.R., Cook, S.E., Fisher, M.J., Lim, C.H., Lim, Y.L., et al. 2017. Yield gap analysis in oil palm: Framework development and application in commercial operations in Southeast Asia. Agricultural Systems 151:12-19.doi: 10.1016/j.agsy.2016.11.005.

Jusoh, J.M., Rashid, N.A., and Omar, Z. 2013. Effect of sterilization process on deterioration of bleachability index (DOBI) of crude palm oil (CPO) extracted from different degree of oil palm ripeness. International Journal of Bioscience, Biochemistry and Bioinformatics 3:322-327. doi:10.7763/IJBBB. 2013.V3.223.

Kantardzic, M. 2011. Data mining: concepts, models, methods, and algorithms. 2nd ed. John Wiley & Sons, New Jersey, USA.

Kelley, R.P., Rolison, L.M., Raetz, D., and Jordan, K.A. 2015. Uncertainty analysis of delayed neutron fissile material assay using a genetic algorithm. Annals of Nuclear Energy 80:460-466. doi:10.1016/j.anucene.2015.02.037.

Keong, Y.K., and Keng, W.M. 2012. Statistical modeling of weather-based yield forecasting for young mature oil palm. Asia-Pacific Chemical, Biological & Environmental Engineering Society Procedia 4:58-65. doi:10.1016/j.apcbee.2012.11.011.

MPOB. 2015-2016. Review of the Malaysian oil palm industry. Malaysian Palm Oil Board (MPOB), Kelana Jaya Selangor, Malaysia. Available at http://bepi.mpob.gov.my/index.php/en/ (accessed 12 October 2016).

Oil World. 2014. Oil world statistics. ISTA Mielke GmBh, Hamburg, Germany.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., et al. 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling and Software 79:214-232. doi:10.1016/j.envsoft.2016.02.008.

Purohit, G.N., Sherry, A.M., and Saraswat, M. 2013. Optimization of function by using a new MATLAB based genetic algorithm procedure. International Journal of Computer Applications 61.1-5. doi:10.5120/10001-4212.

Rodriguez, M.C., Dupont-Courtade, L., and Oueslati, W. 2016. Air pollution and urban structure linkages: Evidence from European cities. Renewable and Sustainable Energy Reviews 53:1-9. doi:10.1016/j.rser.2015.07.190.

Shanmuganathan, S., Narayanan, A., Mohamed, M., Ibrahim, R., and K. Haron. 2014. A hybrid approach to modelling the climate change effects on Malaysia's oil palm yield at the regional scale. In Herawan T., Ghazali R., Deris M. (eds.) Recent advances on soft computing and data mining. Advances in Intelligent Systems and Computing 287:335-345. doi:10.1007/978-3-319-07692-8_32.

Svatovnová, T., Herák, D., and Kabutey, A. 2015. Financial profitability and sensitivity analysis of palm oil plantation in Indonesia. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis 63:1365-1373. doi:10.11118/actaun201563041365.

Szymczak, S., Holzinger, E., Dasgupta, A., Malley, J.D., Molloy, A.M., Mills, J.L. et al. 2016. r2VIM: A new variable selection method for random forests in genome-wide association studies. BioData Mining 9:7. doi:10.1186/s13040-016-0087-3.

Trejos, J., Villalobos-Arias, M.A., and Espinoza, J.L. 2016. Variable selection in multiple linear regression using a genetic algorithm. p. 133-159. In Vasant, P. (ed.) Handbook of research on modern optimization algorithms and applications in engineering and economics. IGI Global, Hershey, Pennsylvania, USA.

USDA Foreign Agricultural Service. 2017. Oil seeds: World markets and trade. United States Department of Agriculture (USDA), Washington D.C., USA. Available at https://www.fas.usda.gov/data/oilseeds-world-markets-and-trade (accessed 1 September 2017).

Zhu, Q., and Azar, A.T. 2015. Complex system modelling and control through intelligent soft computations. Springer, London, UK.