Food Research 3 (2) : 171 - 176 (April 2019)
Journal homepage: http://www.myfoodresearch.com

FOOD RESEARCH

FULL PAPER

Rice yield prediction model with respect to crop healthiness and soil fertility

[1]*Putri, R.E. [2] Yahya, A., [3]Adam, N.M. and [2]Abd Aziz, S.

[1]*Department of Agricultural Engineering, Andalas University, West Sumatera 25163, Indonesia*
[2]*Department of Biological and Agricultural Engineering, Univeristi Putra Malaysia, 43400 Serdang, Selangor, Malaysia.*
[3]*Department of Mechanical Engineering, Univeristi Putra Malaysia, 43400 Serdang, Selangor, Malaysia.*

Article history:
Received: 5 July 2018
Received in revised form: 9 October 2018
Accepted: 20 October 2018
Available Online: 14 November 2018

Keywords:
Rice Yield,
Crop Healthiness,
Soil Nutrient,
Multiple Linear Regressions.

DOI:
https://doi.org/10.26656/fr.2017.3(2).117

Abstract

The main staple food for most people in the world precisely in Asia is rice also known as another name *Oryza sativa* L. This study aimed to establish a rice yield prediction model based on crop healthiness and soil fertility in Malaysia. Multiple linear regressions (MLR) model was used to develop a relationship between a dependent variable, Y (rice yield) and three predictor variables, X (crop healthiness and soil nutrient). A model was developed to predict rice yield based on crop healthiness and soil nutrient variability including all interaction variables with an overall R^2 = 0.6403. The information obtained from variability could assist farmers in making management decisions to improve cropping practices for succeeding rice crops. This study may aid in enhancing the rice yield and the profitability of Malaysia farmers.

1. Introduction

In Malaysia, rice is regarded as a huge insurance and strategically crucial agricultural industry. As of 2013, Malaysia has harvested 690,000 hectares area with the annual rice production reached 2.63 million tons (FAOSTAT, 2014). In the same year, the country's number population had reached 29.72 million which makes the total guesstimate to be 3.13 million tons. The rice production in the country itself was insufficient and to balance the rice, rice was imported from other countries such as Vietnam, Thailand, and Pakistan. The country's net rice imports are estimated to grow by 2.2% for each year. In 2011, it started from 1.08 million ton and by the year 2021, it will peak at 1.37 million ton (Wailes, 2012).

The application of precision farming practices in crop management is described as understanding the content of leaf chlorophyll and the NPK status of harvest soil in each part of the agricultural field as a mean to develop the production profits within the field and to gain a higher per unit production costs. Maintaining the inputs (i.e. seed, fertilizer, chemical, etc.) is a fundamental of precision farming and at the most economic production time and location, implementing every input exclusively is necessary. In addition, essentials such as the evaluation, exploration and control of temporal, spatial variability of plant and soil nutrient must be maintained in order to have a successful implementation of the systems (Chan, 2013).

Malaysian crop productions are well known to have a quite degree of variability that influences crop yield. The factors that influence crop yield include the contour of the area, condition and variety of soil, the level of leaf chlorophyll and more primary determinants. Recently, a research has been conducted to understand the crop growth using yield monitoring technology that seeks after the relationship between the output of crop yield and the influential crucial factors (Yanai *et al.,* 2000; Yanai *et al.,* 2001; Yanai *et al.,* 2002; Liu *et al.,* 2008). There were also several studies to understand the collision or responsiveness to stimulate the chlorophyll content and the soil nutrient on rice yield using spatial variability management (Yana *et al.,* 2000; Mzuku *et al.,* 2005; Aimrun *et al.,* 2007; Liu *et al.,* 2007; Aishah *et al.,* 2010; Gholizadeh, Amin, Soom *et al.,* 2011; Gholizadeh, Amin, Anuar *et al.,* 2011; Teoh *et al.,* 2012; Tilaki *et al.,* 2013, Putri, Adam and Aziz, 2016). As a result of comparing the farmer's fertilizer practice to locale-specific N management during the farm research, it was discovered that yields with site-specific N management increased by reducing N fertilizer rate (Balasubramanian *et al.,* 2000; Dobermann *et al.,* 2002; Peng *et al.,* 2006). Soil fertility variation is an important factor to improve the quality of rice. For that reason, many laboratories had

*Corresponding author.
Email: *rennyekaputri@ae.unand.ac.id*

eISSN: 2550-2166 /

evaluated the soil to quantify the nutrients level. A few studies had investigated on the spatial variability of the content in leaf chlorophyll and harvest soil (Nitrogen, Phosphorus, and Potassium/NPK) status that might be available within the rice field.

The main purpose of the study is to formulate a relationship to predict the rice yield based on the leaf chlorophyll content and after-harvest-soil NPK status levels. The modelling will be based on the instantaneous harvested rice from the actual rice field.

2. Material and Methods

The rice fields placed at Blok E5 Parit Timur 5 of Sungai Besar, Selangor was used as the research locations. The area was selected due to it is plain flat coastal under Integrated Agricultural Development Authority (IADA) Rice Granary located between Kuala Selangor and Sabak Bernam and it is one of the well-known main growing areas in Sungai Besar District, Malaysia. Out of the forty (40) available rice plots in this area, three rice plots were randomly selected. The size of each plot was 1.09 ha. Field observation and data collection at the selected plots were conducted on two consecutive of the rice growing seasons. Figure 1 displays the field layout of the selected rice plots.

Figure 1. Aerial view of Blok E5 Parit 5 Timur, Sungai Besar, Selangor.

A calibration instrument combined with rice harvester was used to measure the leaf chlorophyll content and the NPK status of the soil after harvesting (Putri *et al.,* 2014). During the growing period, the soil nutrient values and the SPAD values were measured. The Crop Cut Test (CCT) was performed first on the harvesting period, followed by the field harvest operation using the combined instrument. This research also involved the farmers in the experimental plots specifically on agronomic and cultivation activities as recommended by the Department of Agriculture Farm Officers at Kuala Selangor (Putri, Yahya, Adam *et al.,* 2016).

2.1 SPAD values

Leaf samples were randomly collected from grown rice plants on the 45-day after planting (DAP), 70-DAP and 95-DAP. In total, twenty sampling points were picked from a sampling grid (30 x 18 m). The chlorophyll content of the leaf samples collected at the sampling points was measured using the SPAD-502 meter in triplicates. The mean value was determined as the SPAD value at each sampling point (Putri, Yahya, Adam *et al.,* 2016).

2.2 Soil sampling analysis

In order to determine the variation of soil nutrient levels, a laboratory analysis of soil samples that had been collected on a grid pattern was conducted. In addition, the spatial variation of soil properties allowed sophisticated farming technologies to change the rate of the input application on-the-go condition. Data that were included on the specification of soil nutrient levels were nitrogen (N), phosphorous (P) and potassium (K).

A cone sampler was used to collect soil samples (15 cm depth) at the sampling points. As many as twenty sampling points were taken from each plot using an 18 x 30 m sampling grid. The positions of the grid points were opted using a handled DGPS receiver. The grid sampling was designed as a sophisticated soil sampling approach as it provided highly detailed information about the variability of nutrient. The soil samples were then analysed at Agricultural Services Sdn Bhd (353791-M) Felda Analytical Laboratory.

2.3 Rice yield model

In order to obtain the relationship of response Y (rice yield) to variable X (after-harvest-soil NPK status), and also some unknown parameters (β), Multiple Linear Regression (MLR) was used to create every model. Y was a function of X and β. SAS was used to determine the linear regression relationship. This determination was conducted using the chlorophyll content data (SPAD value at 45-DAP, 70-DAP and 95-DAP), soil nutrient data (N, P and K) and yield data. Equation (1) shows the rice yield prediction model:

$$Y = a_o + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 \qquad (1)$$

Where Y = yield (ton/ha); a_o= Intercept value; X_1 = SPAD value at 45-DAP; X_2= SPAD value at 70-DAP; X_3 = SPAD value at 95-DAP; X_4 = Total Nitrogen Content (%); X_5 = Available phosphorus (mg/kg); X_6 = Exchangeable K (cmol/kg); $a_{1,2,3,4,5,6}$ = Corresponding coefficients of X1 through X6

The SAS analysis application aimed to determine every single parameter's model. This model included R^2,

Table 1. Pearson' correlation coefficient matrix for all parameters

	IY (Y)	SPAD 45-DAP (X_1)	SPAD 70-DAP (X_2)	SPAD 95-DAP (X_3)	N (X_4)	P (X_5)	K (X_6)
Y	1						
SPAD 45-DAP (X_1)	0.46^{***}	1					
SPAD 70-DAP (X_2)	0.62^{***}	0.63^{***}	1				
SPAD 95-DAP (X_3)	0.53^{***}	0.52^{***}	0.70^{***}	1			
N (X_4)	-0.25^{***}	0.36^{***}	-0.01^{ns}	-0.06^{ns}	1		
P (X_5)	0.28^{***}	-0.04^{ns}	0.14^{ns}	0.10^{ns}	-0.07^{ns}	1	
K (X_6)	-0.07^{ns}	-0.01^{ns}	-0.15^{ns}	-0.07^{ns}	0.07^{ns}	0.11^{ns}	1

*Correlation is significant at the 0.1 level, **Correlation is significant at the 0.05 level, ***Correlation is significant at the 0.01 level, and ^{ns} Not Significant

sum of square, mean square, F value, errors, regression, P_{prob}>F and the value of $a_{1,2,3,4,5,6}$ (SAS Institute, 1996). The predicted rice yield model was evaluated based on the R^2 value for the prediction parameter. In fact, the R^2 value will show the percentage of variance in the variable Y calculated by the variable X. From R^2 value between 0.50 to 0.65, it can be summed up that 50% to 65% of the variable Y is taken into account by the variable X. An R^2 between 0.66 and 0.81 shows an estimate of quantitative predictions, whereas, values between 0.82 and 0.90 reveal good predictions. The model that has R^2 above 0.91 is categorized as a very good prediction model (Williams, 2003). The other interpretations based on R^2 values derived from Best and Kahn, (2003) in case it is up to 0.20 then they are considered negligible, 0.20-0.40 is low, 0.40-0.60 mean moderate, 0.6-0.80 substantially and from 0.80 to 1.01 is categorized high to very high.

3. Results and discussion

Multiple regression was used to develop the rice yield prediction models. The model was based on the chlorophyll content and after-harvest-soil NPK status. Pearson's correlation coefficient (r) was also used to analyze the patterns of relationships among the variables. Table 1 tabulates the correlation between the yield and all the parameters involved. According to this research, it could be concluded that there was a significant positive correlation between the yield and the SPAD values at all growth stages (45-DAP, 70-DAP, and 95-DAP), nitrogen and phosphorus except potassium.

The models of multiple regression were evaluated to discover the relationship between independent and dependent variables. In this case, the dependent variable was variable Y (rice yield) while the independent variables were X_1 to X_6 (six variables). Equation (2) shows the linear model.

$$Y= b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 \quad (2)$$

Where Y = yield (ton/ha); b_0 = Intercept value; X_1 = SPAD value at 45-DAP; X_2 = SPAD value at 70-DAP; X_3 = SPAD value at 95-DAP; X_4 = Total Nitrogen Content (%); X_5 = Available phosphorus (mg/kg); X_6 = Extractable Potassium (cmol/kg); $b_{1,2,3,4,5,6}$ = Corresponding coefficients of X_1 through X_6

The regression model of yield used six explanatory variables against the rice yield. Five models were evaluated to have their own predictive power and were explained.

3.1 Model 1

Model 1 used all the six variables to predict the yield. Table 2 presents the details of the regression model. This table shows a highly significant relationship between the chlorophyll content and after-harvest-soil NPK status to predict the yield.

Table 2. ANOVA of Model 1.

Source	DF	Sum of Squares	Mean square	F Value	P Value
Model	6	43.23	7.20	23.00	<.0001
Error	113	35.40	0.31		
Corrected Total	119	78.63			

The model 1 prediction model can be shown as:

$$Y_1 = 1.77189 + 0.11138X_1 + 0.08063X_2 + 0.01684X_3 - 8.95345X_4 + 0.06807X_5 - 0.26292X_6 \quad (3)$$

With $R^2 = 0.5498$ or adjusted $R^2 = 0.53$

3.2 Model 2

As there was no significant correlation between the yield and X_6, variable X_6 was deleted to produce Model 2. The analysis of variance for model 2 is presented in Table 3. This model showed a significant difference at α 0.05.

Table 3. ANOVA of Model 2

Source	DF	Sum of Squares	Mean square	F Value	P Value
Model	5	43.20	8.64	27.80	<.0001
Error	114	35.43	0.31		
Corrected Total	119	78.63			

Model 2 can be shown as:

$$Y = 1.62015 + 0.11047X_1 + 0.08220X_2 + 0.01672X_3 - 8.96464X_4 + 0.06714X_5 \quad (4)$$

With $R^2 = 0.5494$ or adjusted $R^2 = 0.53$

From the equation above, the R^2 value was lower than the other models, suggesting the X_6 should be included in the prediction model.

3.3 Model 3

Table 4 presents the ANOVA for model 3 when the model was incorporated with the interaction of variables. High significant correlation of variables between the SPAD values were showed via Pearson's correlation coefficient of the following: i) 45-DAP (X_1) and 70-DAP (X_2) with r = 0.6264; ii) 45-DAP (X_1) and 95-DAP (X_3) with r = 0.5156; and iii) 45-DAP (X_1) and 95-DAP (X_3) with r = 0.6984. These interaction variables were added to this model as X_1X_2, X_1X_3 and X_2X_3.

Table 4. ANOVA of Model 3

Source	DF	Sum of Squares	Mean square	F Value	P Value
Model	9	49.74	5.53	21.05	<.0001
Error	110	28.89	0.26		
Corrected Total	119	78.63			

The predicting model 3 can be shown as:

$$Y = 16.6820 + 0.2194X_1 + 0.7772X_2 - 0.1691X_3 - 7.6821X_4 + 0.0577X_5 + 0.1179X_6 + 0.0092X_1X_2 - 0.0184X_1X_3 + 0.0233X_2X_3 \quad (5)$$

With $R^2 = 0.6326$ or adjusted $R^2 = 0.60$

3.4 Model 4

This model was a continuation of the first three models with the addition of interaction of three variables, $X_1X_2X_3$. Based on Table 5, the developed model showed a significant difference in predicting the rice yield using all parameters include the interaction of variables. This model gave the best R^2 (0.6403), which defined the necessity to include the interaction of variables.

Table 5. ANOVA of Model 4

Source	DF	Sum of Squares	Mean square	F Value	P Value
Model	10	50.34	5.03	19.40	<.0001
Error	109	28.28	0.26		
Corrected Total	119	78.63			

The predicting model 4 can be shown as:

$$Y = -44.04973 + 2.01624X_1 + 0.97993X_2 + 2.39686X_3 - 7.12933X_4 + 0.05267X_5 + 0.19644X_6 - 0.04269X_1X_2 - 0.09405X_1X_3 - 0.05020X_2X_3 + 0.00216X_1X_2X_3 \quad (6)$$

With $R^2 = 0.6403$ or adjusted $R^2 = 0.61$

3.5 Model 5

Model 5 was a modification of model 4, but the number of variables was reduced. The objective of developing this model was to determine the reduction of the interaction variable would improve R^2. The ANOVA results for model 5 is presented in Table 6.

Table 6. ANOVA of Model 5

Source	DF	Sum of Squares	Mean square	F Value	P Value
Model	7	45.52	6.50	22.00	<.0001
Error	112	33.10	0.30		
Corrected Total	119	78.63			

The developed model 5 is as shown in Equation (7):

$$Y_5 = 10.34460 - 0.02355X_1 - 0.03220X_2 - 0.18173X_3 - 9.17146X_4 - 0.07138X_5 - 0.02188X_6 + 0.00015752X_1X_2X_3 \quad (7)$$

With $R^2 = 0.5790$ or adjusted $R^2 = 0.55$

In conclusion, all models were evaluated based on R^2values to have a predictable yield. The highest R^2 was obtained using model 4 which included all interaction among variables. Based on variable X of the chlorophyll content and after-harvest-soil NPK status, the analysis of variance has been accounted for and a resulted R^2 value of 0.64. It is denoted that an R^2 value between 0.5 and 0.65 contain well 50% of the variance yield (Williams, 2003). A low R^2 value might be caused by several factors. Those basis factors are the chlorophyll content, the after-harvest-soil NPK status, the management, the circumstances of the environmental and the climate (Dahal and Routray, 2011). Compared to soil fertility, crop fertility has more variability. It may have been affected by many factors such as the multifariousness of microclimate, the distribution of nutrient supply, the phase of yield growth, and the struggle of the plant against pest (Heege, 2013).

4. Conclusion

Between the crop yield and the SPAD reading, a positive correlation was found between these variables. In paddy plot association, at the age of 70-DAP, the Pearson's correlation ranged from 0.7280 to 0.8336. The correlation coefficients indicated that the yield was not significantly correlated to several soil nutrients. It appeared that the relationship between soil test nutrient level and yield may vary. Multiple linear regressions (MLR) models were developed with the main purpose to have a better description for the relationship between the dependent variable, rice yield, and the more predictive variables, such as the chlorophyll content and after-harvest-soil NPK status. A combination of the model with the interaction of variables developed a rice yield predicting model with an $R^2 = 0.6403$.

References

Aimrun, W., Amin, M.S.M., Ahmad, D., Hanafi, M.M. and Chan, C.S. (2007). Spatial variability of bulk soil electrical conductivity in a Malaysian paddy field: key to soil management. *Paddy and Water Environment*, 5(2), 113-121. https://doi.org/10.1007/s10333-007-0072-z

Aishah, A.W., Zauyah, S., Anuar, A.R. and Fauziah, C.I. (2010). Spatial variability of selected chemical characteristics of paddy soils in Sawah Sempadan, Selangor, Malaysia. *Malaysian Journal of Soil Science*, 14, 27–39.

Balasubramanian V, Morales, A.C., Cruz, R.T., Thiyagarajan T.M., Nagarajan, R., Babu M., Abdulrachman, S. and Hai, L.H. (2000). Adaptation of the chlorophyll meter (SPAD) technology for real-time N management in rice: a review. *International Rice Research Notes*, 25(1), 4-8.

Best, W.B. and Kahn, J.V. (2003). Research in Education. New Delhi: Printice-Hall of India Pvt-Ltd.

Chan, C.W. (2013). Precision Agriculture: The Way Forward in Mechanised Agriculture. Malaysia: Malaysian Agricultural Research and Development Institute (MARDI).

Dahal, H. and Routray, J.K. (2011). Identifying associations between soil and production variables using linear multiple regression models. *Journal of Agriculture and Environment*, 12, 27-37.

Dobermann, A., Witt, C., Dave, D., Abdulrahman , S., Abdulrachman, S., Gines, H.C., Nagarajan, R., Satawathananont, S., Son, T.T., Tan, P.S., Wang, G.H., Chien, N.V., Thoa, V.T.K., Phung, C.V., Stalin, P., Muthukrishnan, P., Ravi, V., Babu, M., Chatuporn, S., Sookthongsa, J., Sun, Q., Fu, R., Simbahan, G.C. and Adviento, M.A.A. (2002). Site-specific nutrient management for intensive rice crop system in Asia. *Field Crops Research*, 74(1), 37-66. https://doi.org/10.1016/S0378-4290(01)00197-6

FAO (Food and Agriculture of the United Nations) (2014). FAOSTAT. Retrieved on September 2014 from FAO website: http://www.fao.org/faostat/en/?#data/QC

Gholizadeh, A., Amin, M.S.M., Anuar, A.R., Aimrun, W. and Saberioon, M.M. (2011). Temporal variability of spad chlorophyll meter readings and its relationship to total nitrogen in leaves within a Malaysian paddy field. *Australian Journal of Basic and Applied Sciences*, 5(5), 236-245.

Gholizadeh, A., Amin, M., Soom, M., Rahim, A.A. and Wayayok, A. (2011). Using soil plant analysis development chlorophyll meter for two growth stages to assess grain yield of Malaysian Rice (*Oryza sativa*). *American Journal of Agricultural and Biological Sciences*, 6(2), 209–213. https://doi.org/10.3844/ajabssp.2011.209.213

Heege, H.J. (2013). Precision in crop farming. site specific concept and sensing methods : application and results. New York: Springer. https://doi.org/10.1007/978-94-007-6760-7

Liu, X., Zhao, K., Xu, J., Zhang, M., Si, B. and Wang, F. (2007). Spatial variability of soil organic matter and nutrients in paddy fields at various scales in southeast China. *Environmental Geology*, 53(5), 1139–1147. https://doi.org/10.1007/s00254-007-0910-8

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F. and MacDonald, L. (2005). Spatial Variability of Measured Soil Properties across Site-Specific Management Zones. *Soil Science Society of America Journal*, 69, 1572. https://doi.org/10.2136/sssaj2005.0062

Peng, S., Buresh, R.J., Huang, J.L., Yang, J.C., Zou, Y.B., Zhong, X.H., Wang, G.H. and Zhang, F.S. (2006). Strategies for overcoming low agronomic nitrogen use efficiency in irrigated rice systems in China. *Field Crops Research*, 96(1), 37–47. https://doi.org/10.1016/j.fcr.2005.05.004

Putri, R.E., Adam, N.M. and Aziz, S.A. (2016). Performance Evaluation of Yield Monitoring System for Rice Combine Harvester in Selangor, Malaysia. *International Journal on Advanced Science, Engineering and Information Technology*, 6(1), 35-39. https://doi.org/10.18517/ijaseit.6.1.612

Putri, R.E., Yahya, A., Adam, N.M. and Aziz, S.A. (2016b). Variability of Rice Yield with Respect To Crop Health. *Jurnal Teknologi*, 78(1-2), 79-85.

Putri, R.E., Yahya, A., Adam, N.M., Aziz, S.A. and Ishola, T.A. (2014). Calibration and Accuracy Determination of a Microwave Type Sensor for Measuring Grain Flow. *Advanced Materials Research*, 931-932, 1592-1596. https://doi.org/10.4028/www.scientific.net/AMR.931-932.1592

SAS Institute. (1996). SAS system for window. North Carolina, US.

Teoh, C.C., Hassan, D.A., Radzali, M.M. and Jafni, J.J. (2012). Prediction of SPAD chlorophyll meter readings using remote sensing technique. *Journal of Tropical Agriculture and Food Science*, 40(1),127–

136.

Tilaki, G.A.D., Kamarei, R. and Vafakhah, M. (2013). Determining the relationship between soil properties and spatial variability of *Nitraria schoberi* Linn. using geostatistical analysis: a case study in Meighan Playa in Iran. *Polish Journal of Ecology*, 61(1), 93-104.

Wailes, E.J. and Chavez, E.C. (2012). ASEAN and global rice situation and outlook. ADB Sustainable Development Working Paper Series. No. 22. Philippines: Asian Development Bank

Williams, P. (2004). Near infrared Technology: Getting the Best out of Light. Canada: Value Added Wheat CRC

Yana, J., Lee, C.K., Umeda, M. and Kosaki, T. (2000). Spatial variability of soil chemical properties in a paddy field. *Soil Science and Plant Nutrition*, 46(2), 473-482.

Yanai, J., Lee, C.K. and Kaho, T. (2001). Geostatistical analysis of soil chemical properties and rice yield in a paddy field and application to the analysis of yield-determining factors. *Soil Science Plant Nutrition,* 47 (2), 291-301. https://doi.org/10.1080/00380768.2001.10408393

Yanai, J., Lee, C.K. and Takashi, K. (2000). Spatial variability of chemical properties in a paddy field. *Soil Science Plant Nutrition,* 46(2), 473-482.

Yanai, J., Lee, C.K., Umeda, M. and Kosaki, T. (2002). Spatial variability of soil properties and rice yield in paddies: Is precision agriculture worth practicing? Proceedings of the 17th World Congress of Soil Science, 14-21 August 2002. Bangkok: Srimuang.

