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

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

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

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; Airmun 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
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.](image)

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:

\[ Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 \]  

Where \( Y \) = yield (ton/ha); \( a_0 \) = 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); \( X_6 \) = 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 \( R^2 \),
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 X₁ to X₆(six variables). Equation (2) shows the linear model.

\[ Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 \]  
(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.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>43.23</td>
<td>7.20</td>
<td>23.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>113</td>
<td>35.40</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>119</td>
<td>78.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model 1 prediction model can be shown as:

\[ Y_1 = 1.77189 + 0.11138X_1 + 0.08063X_2 + 0.01684X_3 - 8.95345X_4 + 0.06807X_5 - 0.26292X_6 \]  
(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 \( \alpha = 0.05 \).

Table 3. ANOVA of Model 2

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>43.20</td>
<td>8.64</td>
<td>27.80</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>114</td>
<td>35.43</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>119</td>
<td>78.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Model 2 can be shown as:
\[ Y = 1.62015 + 0.11047X_1 + 0.08220X_2 + 0.01672X_3 - 8.96464X_4 + 0.06714X_5 \]

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

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>9</td>
<td>49.74</td>
<td>5.53</td>
<td>21.05</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>110</td>
<td>28.89</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>78.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The predicting model 3 can be shown as:
\[ Y = 16.6820 + 0.2194X_1 + 0.7772X_2 - 0.1691X_3 - 7.6821X_4 + 0.0577X_5 + 0.1179X_6 + 0.0092X_1X_2 - 0.0184X_1X_3 + 0.0233X_2X_3 \]

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

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>10</td>
<td>50.34</td>
<td>5.03</td>
<td>19.40</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>109</td>
<td>28.28</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
<td>78.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The predicting model 4 can be shown as:
\[ Y = -44.04973 + 2.01624X_1 + 0.97993X_2 + 2.39686X_5 - 7.12933X_1 + 0.05267X_5 + 0.19644X_6 - 0.04269X_2 - 0.09405X_3 - 0.05020X_3 + 0.00216X_1X_2X_3 \]

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

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean square</th>
<th>F Value</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>45.52</td>
<td>6.50</td>
<td>22.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>112</td>
<td>33.10</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td>119</td>
<td>78.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The developed model 5 is as shown in Equation (7):
\[ Y_5 = 10.34460 - 0.02355X_1 - 0.03220X_2 - 0.18173X_3 - 9.17146X_4 - 0.07138X_5 - 0.02188X_6 + 0.00015752X_1X_3X_6 \]

With \( R^2 = 0.5790 \) or adjusted \( R^2 = 0.55 \)

In conclusion, all models were evaluated based on \( R^2 \) values 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 (Hheege, 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


