CAFEi2012-74

ARTIFICIAL NEURAL NETWORK IN PREDICTING RICE YIELD

Samihah Mustaffha[1,2], Siti Khairunniza Bejo[1*], Wan Ishak Wan Ismail[1].

[1] Biological and Agricultural Engineering Depertment, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2] Malaysian Nuclear Agency, 43000 Kajang, Selangor, Malaysia
Email: skbejo@eng.upm.edu.my

ABSTRACT

Rice production is one of the major sectors that play an important role on the national economy. Hence, site specific nutrient management is crucial for a sustainable agriculture. Therefore, precision agriculture and information technology is really important to balance crop productivity. The application of neural network to the task of predicting crop yield is essential. The objectives of this paper were to: 1) investigate whether artificial neural network (ANN) model could predict rice yield based on soil parameters; 2) determine the most affected soil properties towards rice yield; 3) compare the effectiveness of multiple linear regression model to ANN. Models were developed using historical data collected in Block C, Sawah Sempadan, Selangor, Malaysia for two continuous seasons. Season 1 is dry season while Season 2 is wet season. External factors such as weather, farmer's practices etc. were not being considered in this study. ANN showed more accurate results than regression model. ANN model resulted in r2 of 0.71 and 0.69 for Season1 and Season 2 respectively. While in linear regression, r^2=0.12 and 0.02 for Season1 and Season 2 respectively. The results show that ANN model is more reliable than regression model in predicting rice yield. It can be conclude that ANN model is simple yet accurate.

Keywords: *Artificial neural network, Paddy, Yield prediction model, Multiple linear regression.*

INTRODUCTION

Rice productions are one of the major sectors in Malaysia and play an important role on the national economy. In Peninsular Malaysia, paddy fields are covering almost 300,500 hectares. It grown vastly in this country due to the temperature regime and rainfall distribution and make it suitable to be cultivated throughout the year. However, most farmers plant and harvest rice twice a year at almost the same period.

A lot of factors influence crop production where they directly or indirectly affected the crop performance. Most of factors that normally highlighted in researches are soil factors, such as pH, available nutrients, texture, organic matter content and soil-water relationships, weather and climatic factors including temperature, rainfall and light intensity, the crop and cultivar, postharvest handling and storage and fertilizer applications and cultural practices (Hornick, 1992). These factors have formed a complex system for agriculture since it always deals with a large set of data for each problem it faced with. Crop growth and yield models normally are based on a combination of chemical and physical properties of soil, crop and climate variable [1]. However, in this study, only soil properties are considered to be examined. External factors such as farmers' practices, rainfall distribution and weather are not taken into account.

Since rice is a staple food for the nations, site specific nutrient management is crucial for a sustainable agriculture to ensure the supply is adequate. For that reason, precision agriculture and information technology is really important to balance crop productivity. Researchers have suggested many techniques to predict crop yield. Shearer et al. [2] has listed two methods to predict crop yield which are using mathematical models and using application of artificial intelligent for prediction of crop performance. Some researcher considered correlation analysis [3] while others like to use multiple linear regressions [4-5]. However, traditional methods tends to be unreliable [6] and not very useful in explaining yield [7].

The application of neural network to the task of predicting crop yield is essential. Artificial Neural Network (ANN) structure imitates human brain's biological neural process [8].

The main goal of this study was to develop rice yield prediction model with historical data of two continuous seasons. Season 1 is dry season (harvested in April) and season 2 is wet season (harvested in November). The specific objectives of this paper are to investigate whether ANN model can predict paddy yield based on soil properties and to compare the effectiveness of multiple linear regression (MLR) model to ANN.

MATERIALS AND METHODS

Data of soil properties and rice yield production was obtained in 118 lots of Block C, Sawah Sempadan, Tanjung Karang, Malaysia. The data were from two continuous seasons, where Season 1 was harvested in April is a dry season and Season 2 which harvested in November is a wet season. For each lot, a total of 18 soil properties being examined in laboratory. The 18 properties are apparent electrical conductivity (Ec_a), pH, potassium (K), sulphur (S), natrium (N), phosphorous (P), cation exchange capacity (CEC), calcium (Ca), magnesium (Mg), carbon (C), base saturation (BS), aluminium (Al), ferum (Fe), moisture content (MC) and percentage of clay, silt and sand.

A few techniques have been applied in this study to see the relationship between soil properties and rice yield. However, in this paper, only two methods is discussed. The first method is multiple linear regression (MLR). Since the study using more than one independent variables, MLR can be one of the tool to see the relationship between soil properties (independent variables) and yield (dependent variable). MLR by stepwise method is used to construct the prediction model, where a simple equation can be made for later used.

The second analysis method is artificial neural network (ANN). Typically, a minimum of three layers which are the input layer, the hidden layer and the output layer is required to develop an ANN system (Figure 1). The number of hidden nodes are depends on specific problem of the study can easily extended to more hidden layers. The input contains nodes that correspond to input variables while the output contains nodes that correspond to output variables (Kaul et al., 2005). The input layer used to distribute the inputs to a number of hidden layers, the output of which are connected to an output layer, where the outputs of units are connected to the inputs of the next via connection weight (Marchant & Onyango, 2003). Comparison between these two methods is discussed later.

RESULTS AND DISCUSSIONS

MLR by stepwise regression was used for construction of prediction model using SAS. It showed only three significant variables for Season 1, as shown in Table 1. Hence, a MLR equation has suggested in Eq. 1.

$$Y_p = 7108.12 - 482.07\ (pH) - 1484.23\ (N) - 101.33\ (Mg) \quad (1)$$

For Season 2, only one variable showed significant relationship between parameters and rice yield, which is Ca. Fig. 1 and Fig. 2 below show the value r^2 of Season 1 and Season 2 respectively. The calculated root means square error (RMSE) for Season 1 is 792.63 and for Season 2 is 1398.55.

Table 1: Significant parameters by stepwise method in Season 1

Parameter	Constant value	Model (r^2)
Intercept	7108.12	
pH	-482.07	0.126
N	-1484.23	0.057
Mg	-101.33	0.093

Neural Tools 1.0.1 was used to analyse the relationship between soil properties and yield. All networks were set to train 70% of random cases and the remaining data was for testing. For each season, every dataset was repeated five times to test the reliability. The result of the analyses as shown in Table 2.

As can be seen, ANN shows more reliable result rather than MLR. Based on coefficient of determination value, r^2, ANN shows promising result than MLR. Unlike the MLR model, the calculated RMSE for both seasons shows lower value in ANN. The average predicted yield for MLR model is 3905.67 kg/ha and 3508.47 kg/ha for Season 1 and Season 2 respectively. Although yield value for Season1 using MLR model is higher than ANN model, Season 2 shows the opposite results. The big gap between actual yield and predicted yield in MLR model does not happened in ANN model.

e-ISBN 978-967-960-322-4

Table 2: Result of ANN analyses for Season 1 and Season 2

Season	r^2	RMSE	Actual yield	Predicted yield
Season 1	0.71	754.29	3877.4	3818.05
Season 2	0.69	794.93	4399.25	4455.23

CONCLUSIONS

In overall, as can be seen, ANN model is more trustworthy in predicting rice yield rather than MLR. Indeed acknowledge that MLR is far easier to be built, but ANN is more reliable. Further analysis using ANN should be carried out by including external factors such as rainfall distribution and fertilizer used.

ACKNOWLEDGEMENT

This work was financially supported by the Science Fund.

REFERENCES

[1] Jaafar, A.A, Abu Talib, A.R., Mokhtar, A.S., Abd.Rahim, I., Arora, P.R and Megat Ahmad, M.M.H (2005) Combined experimental and numerical investigation on a film cooling flow over a flat surface. Proceedings, National Conference on Advances in Mechanical Engineering, Kuala Lumpur.

[2] Davis, M.L. and Cornwell, D.A. (1998) *Introduction to Environmental Engineering*, 3rd Edition, McGraw-Hill, Singapore.

[3] Baumann, J.W. (1981) Application of enzymes in fruit juice technology. In G.G. Birch, N. Blakebrough, and K.J. Parker (Eds.), *Enzymes and Food Processing* (pp.129-147), Applied Science Publication, London.

[4] Mohd Saiah, H.R (2006) Heat transfer measurements on flat plate surface film cooling. Bachelor of Engineering (Aerospace) Project Report, Department of Aerospace Engineering, UPM, Malaysia.

[5] Abu Talib, A.R (2003) Detailed investigation of the low-temperature analogy of an aircraft engine standard fire-test. D.Phil Thesis, Department of Engineering Science, University of Oxford, United Kingdom.

[6] Mills, R. (1973) Self-diffusion in normal and heavy water in the range 1-45 deg C, http://nmr-relaxation.com/handbook/general/water.htm, NMR-Relaxation.com, 835 Sextant Ave., Roseville, 55113 MN, USA. (accessed: 17 Dec 2002).

[7] Gillespie, D. R. H., Wang, Z., and Ireland, P. T. (1995) Heating element, PCT/GB96/02017, United Kingdom.

[8] Norman, M. (2003) Business Development Manager, Hallcrest Ltd., 8-9 Stepnell Reach, 541 Blanford Road, Poole, Dorset BH16 5BW, United Kingdom, (Personal Communication).