Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 2183-2186

http://www.ripublication.com

Forecasting Rice Yield In The Northern Thailand With Multiple Linear Regression Model

K. Saithanu[1] J. Mekparyup[2*] and P. Wichaikul[3]

[1,2,3]Department of Mathematics, Faculty of Science, Burapha University
169 Muang, Chonburi, Thailand
[1]ksaithan@buu.ac.th, corresponding author: [2]jatupat@buu.ac.th,*
Phodjana333@gmail.com

Abstract

To forecast rice yield (Y) in the northern Thailand, multiple linear regression (MLR) model with best subsets and stepwise technique was chosen for training data. The estimated MLR equation rectified in agreement of regression assumptions containing five predictor variables (total rainfall: X_1, number of rainy days: X_2, extreme minimum temperature: X_5, mean of relative humidity: X_6 and type of rice: X_7) was exemplified as $\hat{Y}' = Y^{0.2} = -16.3 + 0.00291X_1 - 0.0566X_2 + 0.254\ X_5 + 0.343\ X_6 + 3.42\ X_7$ with 2.44738 for standard error of estimation. The MLR model was then finally evaluated from the validation data set with the root mean square error (RMSE). It revealed the performance of MLR model could potentially forecast rice yield with quite small value of RMSE (2.99692).

Keywords: Rice yield, MLR model, best subsets, stepwise regression

Mathematics Subject Classification: 62J05

INTRODUCTION

Rice is the most essential food for many Asian countries including Thailand. Rice production in Thailand represents a significant portion of the Thai economy and labor force [1]. Rice is also an important main Thai export. Its rice exports surged from around 6.2 million ton in 1995 to 8.9 million ton in 2010 and grew at about 4% per year in 2005-2010 [2]. Thailand is classified 4 regions for rice harvesting: central, north, northeast and south. The northern part contains almost one-third of the land area of Thailand. Its region has about 20% of the total rice land in the country. Some papers reported the significant factors affected to rice yield in Thailand were water

[3], exposure for photosynthesis, short or long day relating to reproduction or even temperature [4], [5]. In addition, rice production in Thailand is derived from in-season and off-season rice. In-season rice is harvested in rainy season (April-February) while off-season is able to cultivate throughout the year. The analysis tools applied in forecasting rice yield generally are a simple method like multiple linear regression (MLR) and a complicated one like neural network scheme [3], [4], [5], [6] etc. As of such these reasons, this study aimed to build the MLR model for forecasting rice yield in the northern Thailand which was included 13 provinces; Chiang Rai, Phayao, Lampang, Lamphun, Chiang Mai, Mae Hong Son, Tak, Kamphaengphet, Sukhothai, Nan, Phrae, Phitsanulok and Phetchabun.

MATERIALS AND METHODS

Office of Agricultural Economics, Ministry of Agriculture and Cooperatives [7] provided yearly data of rice yield (Y: ton) and type of rice (X_7; X_7=1 stand for in-season rice otherwise X_7=0 denoted off-season rice) in the northern Thailand measured during 2008-2013. The six subsequent meteorological predictor variables were collected by National Statistical office [8] supported from Thai Meteorological Department: total rainfall (X_1: mm.), number of rainy days (X_2), daily maximum rain (X_3: mm.), extreme maximum temperature (X_4: °c), extreme minimum temperature (X_5: °c) and mean of relative humidity (X_6: %). The procedure of data analysis for this study was following these steps.

1. Dividing whole data into 2 sets. One was the training data set utilized for training MLR model. It contained each of 54 cases equally for both types of rice. The remaining data was the validation data set employed for checking suitability of model.
2. Determining the estimated MLR equation from the training data set. The best subsets method was often initially applied to roughly obtain as few predictor variables as possible. The stepwise tool was then later used for fitting the MLR model. The estimated MLR equation was finally justified whether the obtained MLR equation was appropriate in agreement with 3 assumptions of regression analysis:

(2.1) The distribution of error term was examined for normality with the statistic of Anderson-Darling [9].

(2.2) The error term was proved for independent and constant variance with the statistic of Breusch-Pagan [10].

(2.3) The correlation among predictor variables was tested for multicollinearity problem with the Variance Inflation Factor [11]. If the value of VIF was less than 5, it can be stated there is no multicollinearity problem.

3. Diagnosing performance of the obtained MLR model from the previous step with the Root Mean Square Error (RMSE)

RESULTS

The best subsets method indicated the MLR equation contained 5 variables (X_1, X_2, X_5, X_6 and X_7) was firstly considered because it gave the smallest value of Mallows C-p. Then, stepwise technique was used for fitting the MLR model. It indicated no variable was removed from the model because each of p-value of *t* statistic for all these variables was quite small. When this MLR equation was examined for the normality assumption, it was not satisfied. Box-Cox was therefore necessarily applied for transformation of the rice yield. The adjusted of estimated MLR equation was then reviewed as

$$\hat{Y}' = Y^{0.2} = -16.3 + 0.00291X_1 - 0.0566X_2 + 0.254X_5 + 0.343X_6 + 3.42X_7$$

The justification of the appropriateness of this MLR equation showed the satisfied results in accordance with all assumptions as follows.

(1) The p-value of AD test statistic was 0.084.

(2) The Breusch-Pagan test statistic was 6.9767. It was smaller than the critical value (11.0706).

(3) The VIF of all these 5 predictor variables were less than 5.

The performance of the obtained MLR model was finally diagnosed from the validation data set. It gave small values of RMSE=2.99692 and S=2.44738.

DISCUSSION

The estimated MLR equation really obtained small value of RMSE and S. That meant, the influential five predictor variables to efficiently forecast rice yield in the northern Thailand were total rain (X_1), number of rainy days (X_2), extreme minimum temperature (X_5), mean of relative humidity (X_6) and type of rice (X_7).

ACKNOWLEDGEMENT

The authors were grateful to the Office of Agricultural Economics and National Statistical office, for kindly providing all data.

REFERENCES

[1] Country Profile: Thailand, 2007, "Library of Congress, Washington DC: Federal Research Division," Jul 2007. Retrieved January 10, 2015, from Web site: http://ricepedia.org/thailand

[3] Phuphak, S., 2008, "Risk analysis of rainfed rice production using GIS and crop growth simulation model," Thai Rice Research Journal (Thailand), 2(1), 1-2.

[4] Bangphan, S., Bangphan, P., & Boonkang, T., 2013, “Multiple Regression Analysis With Test Percentage For Good Rice,” Engineering Journal Chiang Mai University, 20(3), 1-7.

[5] Kunlerd, A., & Seresangtakul, P., 2014, “Rice Yield Prediction Model based on Artificial Neural Networks,” The Tenth National Conference on Computing and Information Technology.

[6] Hassan O.M.I., 2012, “Simulation of Barley grain yield using artificial neural networks and multiple linear regression models,” Egyption Journal of Applied Sciences, 27(1), 1-11.

[7] Office of Agricultural Economics, 2015, “Production Data of Agricultural Commodity,” Retrieved May 1, 2015, from Web site: http://www.oae.go.th/ewt_news.php?nid=13577

[8] National Statistical office, 2015, “Rain Temperature,” Retrieved May 1, 2015, from Web site: http://service.nso.go.th/nso/web/statseries/statseries27.html

[9] Romeu, J. L., 2003, “Anderson-Darling: a goodness of fit test for small samples assumptions,” Selected Topics in Assurance Related Technologies, 10(5), 1-6.

[10] Michael, H. K., John, N., Christopher, J. N., & William Li., 2005, “Applied Linear Regression Models, 5th Edition,” New York: McGraw-Hill.