Artificial Neural Networks for Predicting the Rice Yield in Phimai District of Thailand

Saisunee Jabjone
Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, Thailand
Email: a1102923@hotmail.com

Chatchai Jiamrum
Suranaree University of Technology, Nakhon Ratchasima, Thailand
Email: cjiamrum@gmail.com

***Abstract*—This study aimed to find the model for predicting rice yield in Phimai district, Thailand. A classic multilayer feed-forward neural network with back-propagation algorithm was used throughout this experiment. Data from 2002 to 2007 were used as the training data for predict the rice yield between 2008 and 2012. The input data from six meteorological factors; rainfall, water distribution, evapotranspiration, temperature, humidity and wind speed were used. Evapotranspiration (ET) was found by using Pennman-Montieth equation. The result showed that ANN (8, 19, and 17) provided the lowest value of RMSE (10.57) and MAPE (2.3). The rice yield predicting of ANN (8, 19, 17) and actual data have linear relationship (R^2=0.99). This predicting model, therefore, was precision and appropriate to predict the rice yield.**

***Index Terms*—rice yield, artificial neural networks, evapotranspiration, feed forward back propagation network, predicting**

I. Introduction

Rice is the most important staple crop for Thailand. Over the last 30 years, Thailand has been the largest exporter of rice in the world. In the average of the last three years (2010–2012), Thailand rice exports were around 40 – 45 percent of paddy rice production [1]. Thai farmers cultivate rice twice a year: in a rainy season, which is known as major rice, and in summer, known as second rice. Total farmland for major rice cultivation in 2010–2012 is around 10.82–11.23 million hectares in average. Major rice accounts for 9.18–9.24 million hectares or 84 percent of total cultivation area. Second rice, which is around 1.58–2.059 million hectares, composes 16 percent of total cultivation area. Farmer's cultivation relies mainly on rainfall for up to 75 percent of total cultivated area [2]. Only twenty five percent of total cultivated area is irrigated. In 2010, the low rice yield caused by a drought in the cultivation period and flooding in the harvesting period. Yield dropped from 2.54 to 2.53 tons per hectare. For the Second rice cultivation is cultivated from November 1st to April 30th. This kind of cultivation usually occurs in areas with irrigation, such as canal and waterway. This helps farmers to cultivate two to three times a year with considerately less risk from drought compared to major rice cultivation.

Phimai district locates in Nakhon Ratchasima province, northeastern of Thailand. It is about 300 kilometer from Bangkok. This district is subdivided into 12 subdistricts (tambon), which are further subdivided into 208 villages (mooban). The number of population is 129,849 which majority of population are farmers. The paddy rice has about 3.8 million rais. The rice yield for major rice cultivation and second rice cultivation were 285kilograms/rais and 640kilogram/rais respectively. In Phimai, therefore, second rice cultivation has more rice yield than major rice cultivation because of the irrigation. In addition, rice-growing environments are based on their hydrological characteristics and include irrigated and rainfall. Rice systems need water for three main purposes: i) evapotranspiration; ii) seepage and percolation; and iii) specific water management such as land preparation and drainage prior to tillering [3]. Thus the water plays a prominent role in rice production. The water management is needed to use water wisely and maximize rice yield. There is need to develop statistically sound objective predicts of rice yield based on water variables so that reliable predicts can be obtained.

Artificial neural network (ANN) is one of the most accurate and widely used predicting models that have enjoyed fruitful applications in predicting social, economic, engineering and business. It is data-driven self-adaptive methods. It learns from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to describe [4]. It is very useful for many practical problems. One of the major developments in neural networks over the last decade is the model combining or ensemble modeling [5]. ANN can identify and learn correlated patterns between input data sets and corresponding target values through training. After training, ANN can be used to predict the outcome of new independent input data and have great capacity in predictive modeling, i.e. all the characters describing the unknown situation can be presented to the trained ANN,

Manuscript received July 16, 2013; revised August 29, 2013

doi: 10.12720/ijoee.1.3.177-181

and then prediction of agricultural system may be feasible [6]. This study, therefore, aims to develop an artificial neural network (ANN) based model for rice yield predicting of Phimai, Nakhon ratchasima which will be trained and tested using data for the last ten years (January, 2002 to July, 2012).

II. Materials and Methods

Artificial Neural Network is to simulate some fundamental characteristics of human brain or natural neural network, and it is a nonlinear dynamic system which is composed of interconnecting artificial neuons [7]. Back propogation algorithm is the most typical and widely-used network model in all neural network models. Including input layer, output layer and several hidden layers, BP neural network is a kind of multilayer feed-forward network [8]. It consists of three layer; input layer, hidden layer and output layer. Every layer includes several nodes; upper and lower layer are connected in network. After training samples are input to the network, the activation valued of neurons in the same layer. After training samples are input to the network, the activation values of neurons transfer form hidden layers to output layer; if output layer does not reach anticipative effect, error signals from the output layer modify connection weight and threshold value in the middle layers, then go back to input layer. This process occurs repeatedly until the errors reach the minimum (Fig. 1).

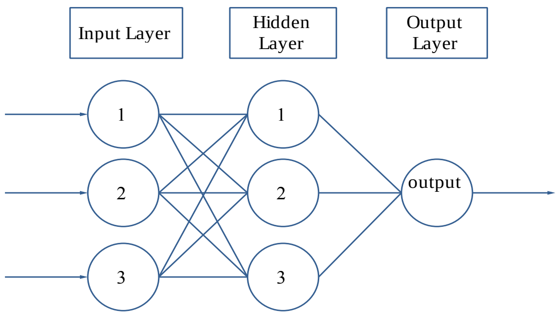

Figure 1. Network structure of BP algorithm

In this study, model was developed through water balance indices based regression approach and artificial neural networks approach using back propogation algorithm. The performances of the developed models with different learning algorithm were compared by RMSE and MAPE. To establish the model, it consists of three steps; A) preprocessing, B) establish of predicting model and C) application development.

A. Preprocessing

Model for predicting rice yield based on rainfall, evapotranspiration, temperature, humidity, water distribution and wind speed variables utilized data for past ten years in Phimai district. Evapotranspiration (ET) is the sum of evaporation and plant transpiration from the Earth's land surface to atmosphere. Evaporation accounts for the movement of water to the air from sources such as the soil, canopy interception, and water bodies [9] [10]. Transpiration accounts for the movement of water within a plant and the subsequent loss of water as vapor through stomata in its leaves. Evapotranspiration is an important part of the water cycle. FAO represent the 4 methods for determine the ET0; Pan Evaporation, Blaney-Criddle Method, Indivate Value of ET0 and Pennman-Montieth equation. In this study, we used Pennman-Montieth equation as show in equation (1).

$$ET0=\frac{0.408\Delta(R_n-G)+\gamma\frac{900}{T+273}U_2(e_a-e_d)}{\Delta+\gamma(I+0.34U_2)} \tag{1}$$

$ET0$ =water evaporation and transpiration (mm.)

R_n = net radiation for each day (Mj/m2)

Δ =evaporation slope (Kpa/C)

T =Temperature (C)

γ =phychrometric constant (KPa/C)

U_2 =wind speed (M/s)

(e_a-e_d) = vapor pressure difference

Firstly, $ET0$ is calculated by using ten years of data collection such as the data of temperature, humidity, evaporation and wind speed. These data provided by Thailand meteorological department (TMD) and Royal irrigation department (RID). To determine the water balance equation, therefore, the other input parameters such as water distribution, rainfall and ET0 are used as show in equation (2).

$$WB=WD+RF-ET0 \tag{2}$$

WB = water balance (mm.)

WD = (water distribution/Area) x1000 (mm.).

RF = rainfall (mm.)

$ET0$ =plant evaporation and transpiration (mm.)

B. Establish of Predicting Model

1) Determination of input and output variables

The input has six meteorological factors; rainfall, evapotranspiration, temperature, humidity, water distribution and wind speed. The feed forward back propagate networks was used to establish predicting model. The input data is in form of vectors of input variables pattern. Corresponding to each element of input vectors is an input node of networks layer.

2) Selection of training samples

According to the determinate input and output variables, the input data include meteorological data and agricultural water use data. The training dataset used the six years dataset of six meteorological factors (i.e. 2002-2007). The test dataset used five years dataset of rice yield in dry season (i.e. 2002-2007).

3) Determination on the number of hidden layers and their neurons

Determining the number of neurons in hidden layer was one important step to establish BP neural network [10]. If the number of neurons was too small, the network might be not trained enough; on the other hand, if the number of neurons was too large, that number of connection weights increased might make the network over parameterized [11][12]. The hidden layer of less neuron might be selected and trained at first [13]. Then if the result could not converge after being trained many times, the number of neurons should be added [14]. RMSE and MAPE were used as a factor reflecting feature of neural network to test some special structures of the network. Calculating formulas are as follow equation (3) and (4).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Xi - Fi)^2}{n}} \tag{3}$$

$$MAPE = \frac{|(Xi - Fi)|}{Xi} x100 \tag{4}$$

TABLE I. ROOT MEAN SQUARE (RMSE) ERROR AND MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) AFTER TRAINING MODEL.

ATF	HL.1	HL.2	RMSE	MAPE
Log-Log	7	5	10.59	2.31
Log-Tan	17	19	10.57	2.3
Tan-Log	5	9	12.95	2.82
Tan-Tan	10	2	12.67	2.77

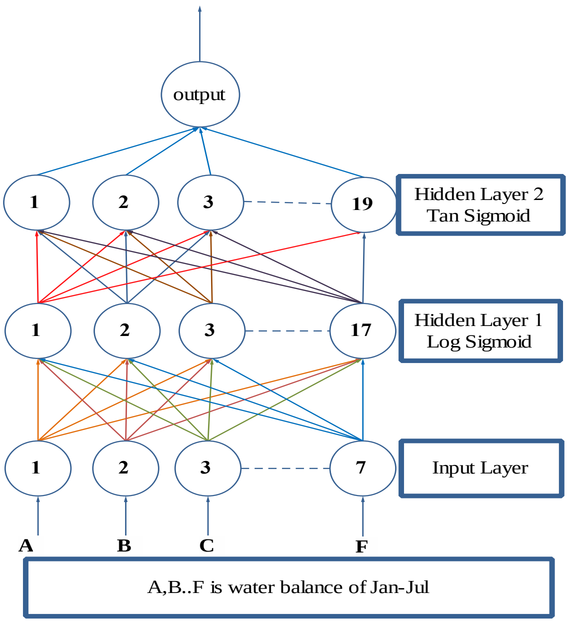

Figure 2. Predicting model

From the Table I, we choose the activation function (ATF) of Logistic (Log)-Hyperbolic (Tan) with hidden layer (HL) in sigmoid 17 and 19 which the RMSE and MAPE were less than other models. The result showed that ATF of Log-Tan (8, 19, and 7) was suitable for testing the efficiency. The matrix of data for the 2008-2012 periods was used to train the ANN and to determine the suitable parameters: number of hidden layer nodes (HN) and number of iterations. In order to test the classification quality of the model, the data matrix was randomly decomposed into two sets. The first set was used to evaluate the quality of their assignment. During the training of neural network, it has been observed that if tan extra input bias node is added in the input layer, the error decrease during training session [14]. The bias node, which as a unity input, is set to improve the learning speed in the training process. Thus with this extra unity input variables, the pruning algorithm is used. Then, the input nodes would be removed [15]. In the validation process, if the ANN model generates errors more than 5%, the model is rejected [16]. Then the hidden layer 1 or 2 is adjusted and the model is reconstructed. The process showed in Fig. 2.

C. *Application*

When the test set (i.e. 2008-2012) finds errors more than 10%, the input layer is adjusted. This process can be considered by correlation input and output by using equation (5). The input, then, is removed when the value of correlation close to zero. The process showed in Fig. 3.

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

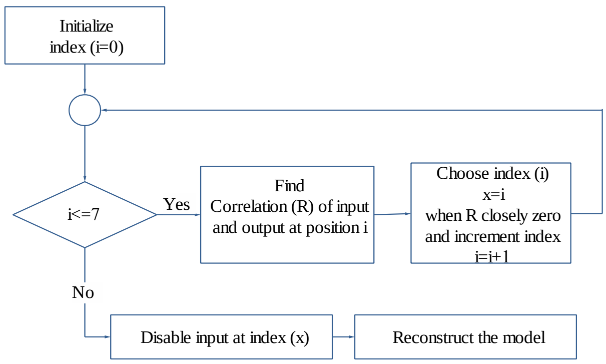

Figure 3. The reconstruct process used correlation equation

III. RESULTS

Model was developed for rice yield for Phimai district, Thailand. The water balance data during 2008 to 2012 is used to predict the rice yield. It was input to feed forward back propagate algorithm. After training data, it was adjusted interconnection node until minimum threshold value of each layer. Then the model was saved. Finally, rice yield in 2008-2012 year was used for testing the efficiency of model between the training dataset and predicting dataset. The RMSE and MAPE were compared. The ATF of Log-Tan (8, 19, and 7) returned the error lower than other topologies. The results are shown in Table II.

TABLE II. THE EFFICIENCY OF MODEL BETWEEN TRAINING DATASET AND PREDICTING DATASET.

ATF	Input	HL 1	HL 2	Training		Predicting	
				RMSE	MAPE	RMSE	MAPE
Log-Log	8	7	5	10.59	0.46	92.85	17.69
Log-Tan	8	19	17	10.57	1.54	9.94	1.46
Tan-Log	8	5	9	12.95	21.99	119.80	19.18
Tan-Tan	8	10	2	12.67	2.39	138.87	14.58

The result shows that ANN (8, 19, 17) is more accurately (RMSE<=9.94 and MAPE <=1.46) than other topology. Therefore, the rice yield predicting is closely to the actual data (Fig. 4). Therefore, the ANN (8, 17, and 19) is suitable for predicting the rice yield.

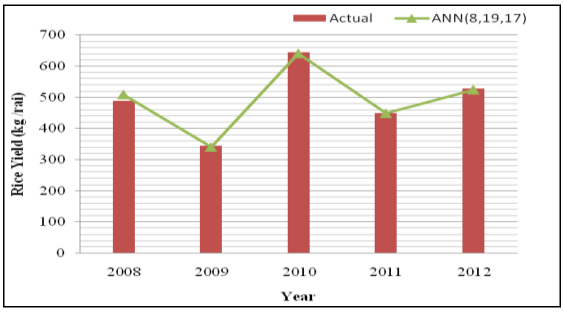

Figure 4. Comparison of ANN and the actual data.

From the Fig. 5, the graph showed the cross validation between ANN (8, 19, and 17) and actual data has linear relationship. The prediction value was nearly the actual data. The correlation (R2) was 0.99. Therefore, this predicting rice yield model is precision and appropriate to predict the rice yield.

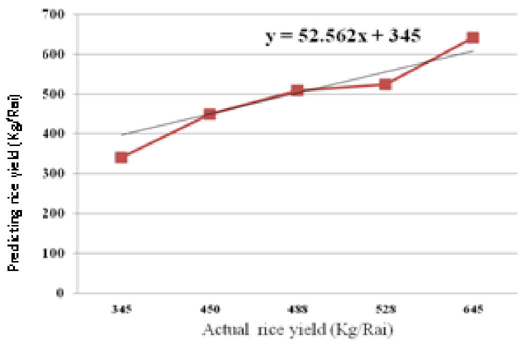

Figure 5. Cross validation between ANN and the actual data

IV. CONCLUSION AND DISCUSSION

In this study, the artificial neural networks were designed and applied to predict rice yield by using six meteorological factors; rainfall, evapotranspiration, temperature, humidity, water distribution and wind speed. The evapotranspiration (ET0) was found by using Pennman-Montieth equation. The monthly dataset during 2002 to 2007 of Phimai district was used to train the model and used to predict the rice yield in 2008 to 2012. The result showed that the BP neural computing technique could be employed successfully in ANN (8, 19, 17) modeling. The precision of predicting was closer to the actual data. The empirical rice yield predicting model ANN (8, 19, 17) produced consistently higher R^2 (0.99) and lower RMSE value (9.94) than linear regression-based yield models. The R^2 value for validated regressions was lower than those of the non-validated regression, indicating that testing regression equations with independent data is critical for the evaluation of regression-based rice yield predicting model. Empirical results with ANN (8, 19, and 17) indicate that the proposed model can be an effective way in order to yield more accurate model than others. Thus, it can be sued as an appropriate alternative for predicting the rice yield. However, the limitation of the study is the dataset come from only one source. Therefore, ANNs is unable to predict for each sub district (Tambol). In the future work, more research should investigate in each subdistrict (Tambol) and vary variables, for example field characteristic and cultivation method. The next step of our study is to extend predictions to the whole area of the rice yield in Thailand. In this study, we interested only the ability of ANN to predict rice yield by using small set of data and variables which it is easy to collect. However, before extending the model, some new analyses are needed to improve the predictions, and also to find a method of identifying the most relevant variables for modeling the prediction (logistic regression), as in ANN usually all the variables contribute to the models.

ACKNOWLEDGMENT

This research was made possible by the funding of Nakhon Ratchasima Rajabhat University and Research and Development Institute. I would like to thank Mrs. Parichart who provided raw data and contacted to district governor. Thank you to officer of Ministry of Agriculture and Cooperatives, Thai Meteorological Department (TMD), Land Development Department and Royal Irrigation Department (RID) who supported the data, gave the advice and tested the system.

REFERENCES

[1] Bureau of Agricultural Economics Research, Office of Agricultural Economics, "Rice production and marketing," *Office of Agricultural Economics Journal*, vol. 52, pp. 53-60, May 2012.

[2] Office of Agricultural Economics, "Crop cutting yield per rai second rice (Dry season), Year 2010. (Officers manual)," *Office of Agricultural Manual report*, Bangkok, 2011.

[3] A. Sayamwara and Na Ranong, W., *Compilation of Rice Knowledge*, Bangkok: Thailand Development Research Institute, 2010.

[4] G. Zhange, *et al*, "Predicting with artificial neural network," *International Journal of Predicting*, vol. 14, pp. 35-62, Jan 1998.

[5] E. Patuwo, M. Y. Hu, and M. S. Hung, "Two-group classification using neural networks," *Decision Science*, vol. 24, no. 4, pp. 825–845. July 1993.

[6] B. JI, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," *Journal of Agricultural Science*, vol. 145, pp. 249-261, 2007.

[7] Z. Tang, C. Almeida, and P. A. Fishwick, "Time series predicting using neural networks vs Box-Jenkins methodology," *Journal of Simulation*, vol. 57, no. 5, pp. 303–310. 1991.

[8] J. L. Monteith and M. H. Unsworth, *Principles of Environmental Physics.* 2nd Ed. New York: Edward Arnold, 1990, pp. 53-54.

[9] V. O. Sadras and P. A. Calvino, "Quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat," *Agronomy Journal,* vol. 93, pp. 557-583, 2011.

[10] J. K. Basak, "Weather based predicting model for crops yield using neural network approach," *Satistics and Application*, vol. 9, pp. 55-69, 2011.

[11] C. T. Lin and C. S. G. Lee, *Neural Fuzzy Systems*, New Jersey: Prentice Hall, 1995, pp. 56-78.

[12] X. Liu, X. Chen, W. Wu, and G. Peng, "A neural network for predicting moisture content of grain drying process using genetic algorithm," *Food Control*, vol. 18, pp. 928–933, 2007.

[13] A. Topuz, "Predicting moisture content of agricultural products using artificial neural networks," *J Eng Software*, vol. 41, pp. 454-470, 2010.

[14] Z. Wang, J. Sun, X. Liao, F. Chen, G. Zhao, J. Wu, and X.Hu, "Mathematical modeling on hot air drying of thin layer apple pomace," *Food Res Int*., vol. 40, pp. 39–46, 2007.

[15] Q. S. Zhang, X. Yang, G. S. Mittal, and S. Yi, "Prediction of performance indices and optimal parameters of rough rice drying using neural networks," *Biosyst Eng.*, vol. 83, pp. 281–290, 2002.

[16] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of* Neural *Computation,* New Jersey: Addison-Wesley, 1991, pp.55-78.

Saisunee Jabjonee She was born in March, 1972. She graduated Doctoral degree from the University of Technology, Sydney, Australia, major in computing. She is an Assistant Professor at the department of Informatics, Nakhon Ratchasima Rajabhat University, Thailand. Her research interest is in the area of data mining, data warehouse, ontology, data management, database, bioinformatics, and education technology.

Chatchai Jiamrum He was born in May, 1974. He graduated Master of Information Technology, major in software engineering, from Suranaree University of Technology. He works as a system analyst at Suranaree University of Technology, Thailand. His research interest is in the area of data warehouse, data mining, data management, ontology, simulation, computer network, modeling and artificial intelligence.