EnvironmentAsia
The international journal published by the Thai Society of Higher Education Institutes on Environment

Available online at www.tshe.org/EA
EnvironmentAsia 5(1) (2012) 56-62

Estimation of Crop Coefficient of Corn (Kc_{corn}) under Climate Change Scenarios Using Data Mining Technique

Kampanad Bhaktikul, Rommanee Anujit and Jongdee To-im

Faculty of Environment and Resource Studies, Mahidol University, Nakhonpathom 73170, Thailand

Abstract

The main objectives of this study are to determine the crop coefficient of corn (Kc_{corn}) using data mining technique under climate change scenarios, and to develop the guidelines for future water management based on climate change scenarios. Variables including date, maximum temperature, minimum temperature, precipitation, humidity, wind speed, and solar radiation from seven meteorological stations during 1991 to 2000 were used. Cross-Industry Standard Process for Data Mining (CRISP-DM) was applied for data collection and analyses. The procedures compose of investigation of input data, model set up using Artificial Neural Networks (ANNs), model evaluation, and finally estimation of the Kc_{corn}. Three climate change scenarios of carbon dioxide (CO_2) concentration level: 360 ppm, 540 ppm, and 720 ppm were set. The results indicated that the best number of node of input layer - hidden layer - output layer was 7-13-1. The correlation coefficient of model was 0.99. The predicted Kc_{corn} revealed that evapotranspiration (ET_{corn}) pattern will be changed significantly upon CO_2 concentration level. From the model predictions, ET_{corn} will be decreased 3.34% when CO_2 increased from 360 ppm to 540 ppm. For the double CO_2 concentration from 360 ppm to 720 ppm, ET_{corn} will be increased 16.13%. The future water management guidelines to cope with the climate change are suggested.

Keywords: crop coefficient; crop evapotranspiration; climate change; artificial neural network; data mining

1. Introduction

Carbon dioxide (CO_2) is one of significant greenhouse gas which induces temperature rising (The Institution of Professional Engineers New Zealand (IPENZ), 2001; Chinwanno, 2007; United Nations Environment Programme (UNEP), 2009). The climate change situation affects living things including human life, plants, and animals on the earth and dramatically increased for five decades (1952-2009). The unusual natural disasters such as drought, flood, tornado, hurricane, volcanic eruption, earthquake, and landslide were observed (UNEP, 2009). It is expected to spread throughout the world including Thailand (Chinwanno, 2007).

Thailand is an agricultural reliant country. Most of crop seasons depend on the climate, nature, and weather. Changes of climate could affect income of the country and agricultural products. In addition, the climate change is a cause of water supply situation. To predict the water demand and prepare water supply, the prediction of evapotranspiration (ET) is needed for better water management (IPENZ, 2001; Mohan and Arumugam, 1994). The most well known technique for ET estimation is based on the crop coefficient (Kc) approach (Allen *et al*., 1998). The reference crop evapotranspiration (ET_0) is calculated by using standard meteorological variable and Kc. The Kc is necessary for estimating the relationship between atmosphere, crop physiology, and agricultural practices. The Kc is needed to determine the water consumption of selected crop, and to estimate water usage for irrigation and agriculture. In addition, Kc is also used in the criteria and factor for water resources planning and management (Akinbile and Sangodoyin, 2009).

Kc values are derived from the relationship between the ET and ET_0. ET is measured by a lysimeter, while ET_0 is calculated by lysimeter experiment which takes long period of time and high cost (Irrigation Water Management Research Group, 2010; Vudhivanich and Udakarn, 2001). Data mining technique is a higher accuracy and fast in calculation. It can store a huge amount of data and high processing (Berry and Linoff, 1997). Therefore, Data Mining was chosen in this study for store the weather data from 1991-2000, and to predict the Kc under the different scenarios.

In light of the above, this study was developed to determine the Kc of corn, which indicated as an economic crop of Thailand (Field crops research Institute, 2001), using data mining technique, to estimate the changing of Kc under climate change scenarios, and to develop the guidelines for water management in the future based on climate change scenarios.

2. Materials and Methods

Methodological framework of this study was adopted and developed from CRoss Industry Standard Process for Data Mining (CRISP-DM) methodology (Chapman *et al*., 1999) which consists of six phases including objective understanding phases, data understanding phases, data preparation phases, modeling phases, evaluation phases, and deployment phases. The details of each phase are as following.

2.1. Objective Understanding Phase

The objectives and requirements of the study were reviewed, clarified, and converted into data mining as knowledge-base.

2.2. Data Understanding Phase

Three types of data including the calculated Kc values using Penman-Monteith method from Royal Irrigation Department, the measured daily climate data from Thai Meteorological Department, and the predicted daily climate scenarios data using Conformal Cubic Atmospheric Model (CCAM) from Southeast Asia START Regional Center (SEA START RC), were used for learning of data sources and characteristics.

The calculated Kc values were converted to daily Kc using graph plotting. The values were plotted into weekly data graph, split into daily data, and defined as a point in graph of each day. The measured daily climate data during 1991-2000 including date, maximum temperature (^oC), minimum temperature (^oC), precipitation (mm), humidity (%), wind speed (knot), and solar radiation ($Mjm^{-2}day^{-1}$) were collected, learned, and set as data mining variable for modeling phase.

The predicted daily climate scenarios data were collected, learned, and set as three climate scenarios. They were three atmospheric CO_2 concentration condition (i) 360 ppm (represented conditions during 1980-1989), (ii) 540 ppm (represented condition during 2050-2059), and (iii) 720 ppm (represented condition during 2090-2099).

2.3. Data Preparation Phase

The procedure of data preparation phase composed of data integration process, data cleaning process, data transformation process, and variable selection process.

The combination of multiple records to be new tables in relational format was conducted in data integration process. Four tables were considered and changed for more complete, accurate, and valuable in data cleaning process. The blank and missing data were deleted from the tables. The wind speed unit was changed from knot to meter per second by multiplying with 0.514.

In data transformation process, the data were changed the format into normalized form which compatibles with ANNs platform, while meaning of the data was not changed. Normalization is a reduction of redundancy, reduction of incorrect data, and setting data into normalized form which values between 0 – 1 (Swingler, 1996). Waikato Environment for Knowledge Analysis (WEKA) program version 3.6.4 was applied for data transformation.

Statistical Package for the Social Sciences (SPSS) program version 16.0 was applied for variable selection using linear regression.

2.4. Modeling Phase

Model selection, data classification, model creation, and model verification were conducted as modeling phase. Artificial Neural Network (ANNs) was chosen to be a model for Kc prediction because of its high accuracy and complex calculation (Rumelhart, 1986). Structure of ANNs model has three layers: input layer, hidden layers, and output layer (Fig. 1). The optimal number of nodes for each layer will be modeled for Kc_{corn}.

The manipulated data were categorized into three sets of data: training set (80%), testing set (10%), and validation set (10%) based on CRISP-DM (Chapman *et al*., 1999). All of records were randomized and categorized into each set of data.

The model creation phase started with finding the most optimal architecture of ANNs using WEKA. The model with lowest Root Mean Square Error (RMSE) will be set to be the optimal architecture of ANNs.

The criteria for finding the optimal number of nodes for hidden layer were set (Blum, 1992; Swingler, 1996; Berry and Linoff, 1997; Allen, 1998). The initial setting were also set as learning rate = 0.3, momentum = 0.2, and learning time = 500. The model which has lowest RMSE value is selected.

Training set data and testing set data were used for finding the optimal learning time using WEKA. The initial setting were set as number of nodes for hidden layer from previous section, learning rate = 0.3, and momentum = 0.2. Trial and error approach of learning time was 500 to 10,000 times. The comparison of RMSE from training set and testing set data was conducted. The lowest RMSE difference between RNSE from training set and testing set was the criteria for learning time optimization.

Training set data were used for finding the optimal

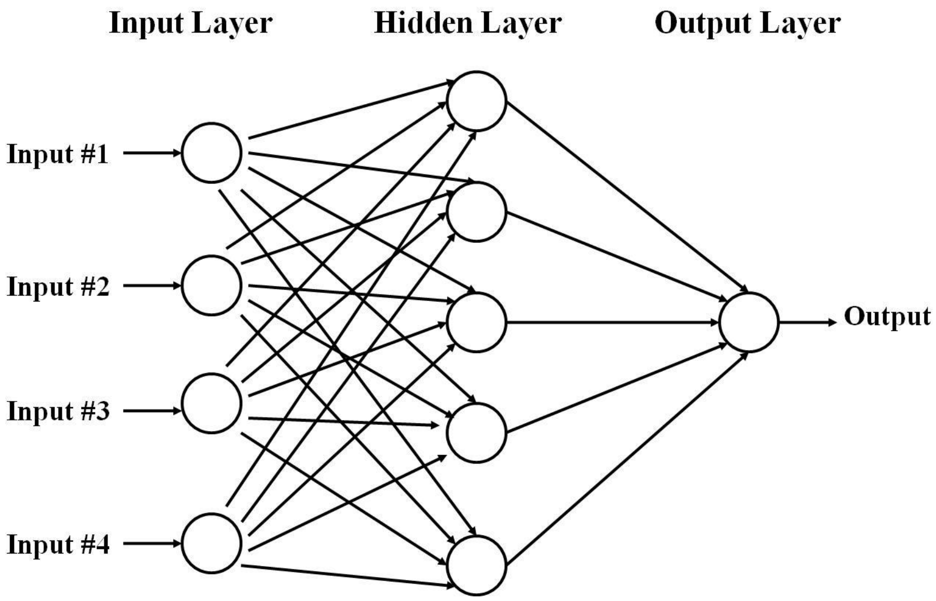

Figure 1. ANNs structure (Rumelhart, 1986)

learning rate and momentum using WEKA. The initial settings were number of nodes for hidden layer and learning time from previous section. Trial and error approach was employed. The learning rates were set as 0.1, 0.2, and 0.3. The momentums were set as 0.1, 0.2, 0.3, 0.5, and 0.7. Two models which have the lowest RMSE were selected for model verification.

Validation set data was used for verifying the model using WEKA. The initial settings were number of nodes for hidden layer = 13, learning time = 3,000, learning rate = 0.1, and momentum = 0.7. The model which has lowest RMSE value will be selected to be a model.

2.5. Evaluation Phase

Correlation and linear regression were employed for evaluation phase. Kc values from ANNs Model, Penman-Monteith method, and linear regression were used for evaluation phase. Higher correlation coefficient (R) indicated more effective model.

2.6. Deployment Phase

All of manipulated data imported into WEKA for Kc prediction. Three climate scenarios were set. The atmospheric CO_2 concentration conditions were set 360 ppm, 540 ppm and 720 ppm as Scenario 1, 2, and 3, respectively. The initial setting for Kc prediction was set using the optimal architecture of ANNs from model creation phase. The results of each scenario were analyzed and compared for the development of guidelines for water resources management coping with the changes of climate.

3. Results and Discussion

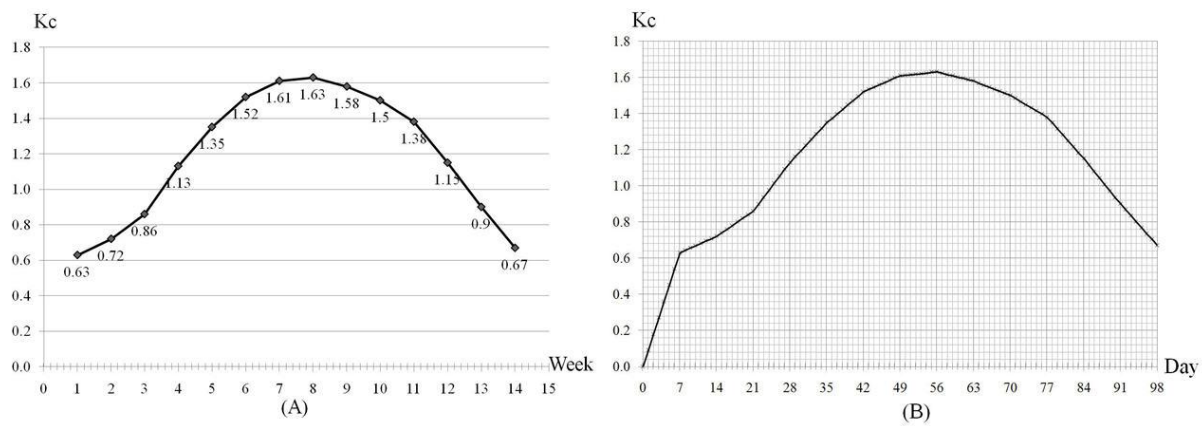

Figure 2. Graph of (A) weekly Kc_{corn}, (B) daily Kc_{corn}

Table 1. Daily Kc_{corn} values derived from graph plotting

Day	Kc	Day	Kc	Day	Kc	Day	Kc	Day	Kc	Day	Kc	Day	Kc
1	0.09	15	0.74	29	1.16	43	1.53	57	1.62	71	1.48	85	1.11
2	0.18	16	0.76	30	1.19	44	1.55	58	1.62	72	1.47	86	1.08
3	0.27	17	0.78	31	1.22	45	1.56	59	1.61	73	1.45	87	1.04
4	0.36	18	0.80	32	1.26	46	1.57	60	1.60	74	1.43	88	1.01
5	0.45	19	0.82	33	1.29	47	1.58	61	1.59	75	1.41	89	0.97
6	0.54	20	0.84	34	1.32	48	1.60	62	1.59	76	1.40	90	0.94
7	0.63	21	0.86	35	1.35	49	1.61	63	1.58	77	1.38	91	0.90
8	0.64	22	0.90	36	1.37	50	1.61	64	1.57	78	1.35	92	0.87
9	0.66	23	0.94	37	1.40	51	1.62	65	1.56	79	1.31	93	0.83
10	0.67	24	0.98	38	1.42	52	1.62	66	1.55	80	1.28	94	0.80
11	0.68	25	1.01	39	1.45	53	1.62	67	1.53	81	1.25	95	0.77
12	0.69	26	1.05	40	1.47	54	1.62	68	1.52	82	1.22	96	0.74
13	0.71	27	1.09	41	1.50	55	1.63	69	1.51	83	1.18	97	0.70
14	0.72	28	1.13	42	1.52	56	1.63	70	1.50	84	1.15	98	0.67

3.1. The results of preparation phase

The calculated Kc values were converted to daily Kc using graph plotting. Kc_{corn} values were converted from weekly to daily data using graph plotting (Fig. 2).

Table 1 show the daily Kc_{corn} data which were defined as a point in graph. Four tables from data integration were cleaned using deletion of records with missing data and unit changing. The cleaned tables have less number of records. The total records after cleaned were 6,817 from 6,860. In addition, wind speed unit was changed to meter per second (m/s).

All of data were transformed into normalized form, while the meaning of data was not changed. The results of variable selection using linear regression indicated significantly relation with Kc. All of relationships indicated low correlation coefficient (*R<0.5*). Then, multiple regressions approach was conducted for proper decision. However, the overall correlation coefficient revealed the low relationship among seven variables and Kc (*R=0.480*). Total of 6,817 records were randomized and categorized into three sets of data. There were 5,454 records of training set, 682 records of testing set, and 681 records of validation set.

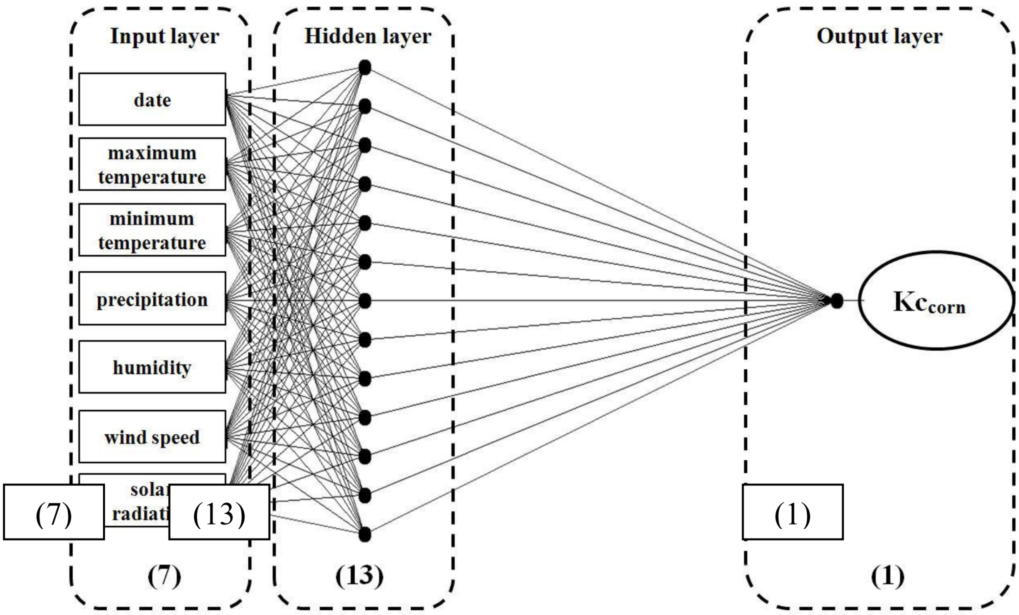

Figure 3. The structure of ANNs model for Corn (7 – 13 – 1)

Table 2. Models for Deployment Phase

Model	Number of nodes for hidden layer	Learning time	Learning rate	Momentum	RMSE
1	13	3,000	0.1	0.7	*0.0080**
2	13	3,000	0.2	0.5	0.0087

3.2. The results of modeling phase

The results of model creation indicated the optimal number of nodes for hidden layer was 13. Therefore, the structure of ANNs model of input layer – hidden layer – output layer of corn was 7-13-1 (Fig. 3).

The optimal learning time from the findings was 3,000 with the lowest RMSE (RMSE difference = 0.0002). The results of optimal learning rate: momentum of corn were 0.1 : 0.7 and 0.2 : 0.5. The results of model verification using validation set data are shown in Table 2.

After the RMSE consideration Model 1 with number of nodes for hidden layer = 13, learning time = 3,000, learning rate = 0.1, and momentum = 0.7 was selected to be the model for corn.

The correlation coefficient (*r*) comparison among Kc from Penman-Monteith method, Kc from linear regression, and Kc from ANNs was conducted (Fig. 4). The average difference of Kc from Penman-Monteith method (Kc_{Penman}) and Kc from ANNs (Kc_{ANNs}) was 0.0070, while the average difference of Kc_{Penman} and Kc from linear regression (Kc_{linear}) was 0.2954. The results indicated that the average difference of Kc_{Penman} and Kc_{ANNs} was lower than the average difference of Kc_{Penman} and Kc_{linear}. The correlation coefficient (*r*) of Kc_{Penman} and Kc_{ANNs} was 0.9997, while Kc_{Penman} and Kc_{linear} was 0.4770. The results revealed that ANNs model is better than the multiple regression equations according to Chowdhary and Shrivastava (2010) and Kotsiantis *et al*. (2008). Therefore, ANNs model was selected for Kc prediction under climate change scenarios of this study.

3.3. The results of Kc prediction

The results of Kc_{corn} prediction under three scenarios are shown in Fig. 5. However, there were not significantly different among three scenarios. Then, the evapotranspiration (ET) pattern was chosen to predict the Kc changes. ET value was calculated by Kc value multiply with the reference crop evapotranspiration (ET_o). The results of ET_{corn} value were shown in Fig. 6.

Corn growth stage was divided into four stages including initial stage (April), crop development stage (May), mid-season stage (June), and late season stage

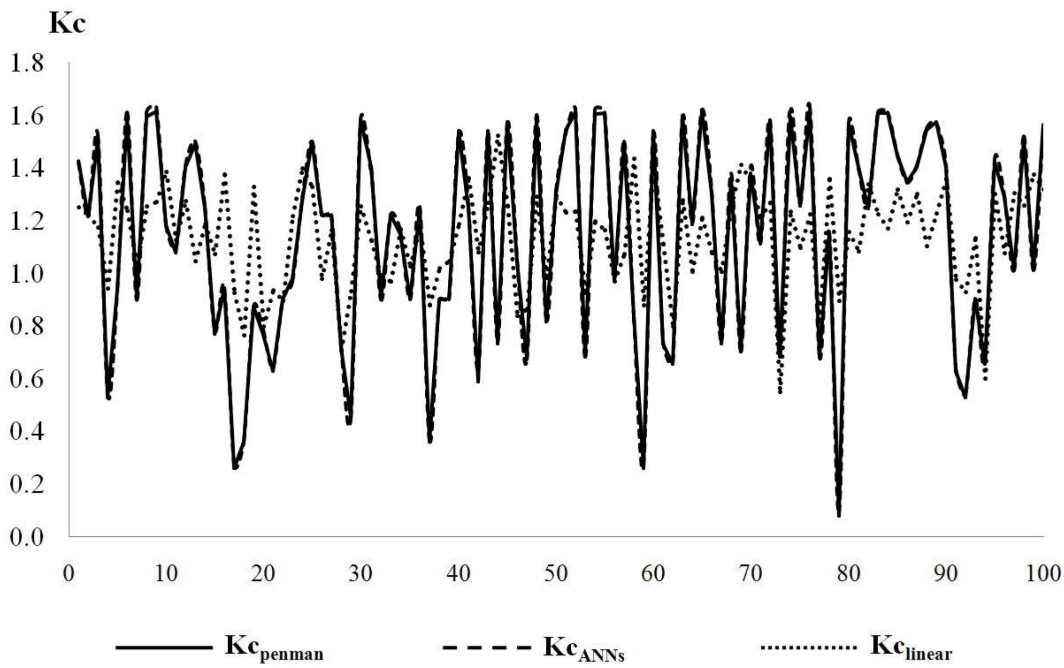

Figure 4. The comparison among Kc_{penman}, Kc_{ANNs}, and Kc_{linear}

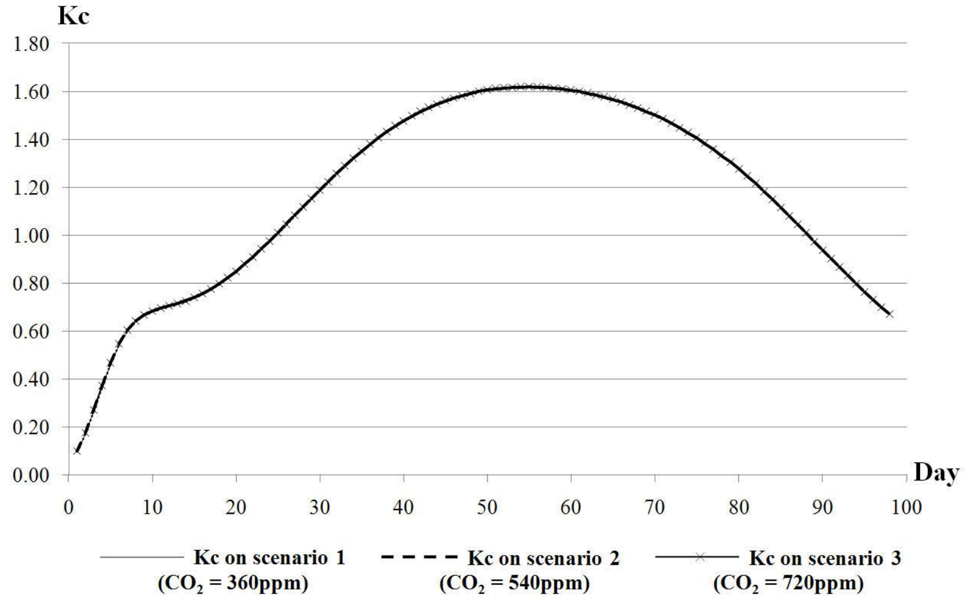

Figure 5. Kc_{corn} under three climate change scenarios

(July). The comparison between ET of scenario 1 and scenario 2 were not much different. ET values were slightly decreased about 3.34%. The ET values in initial stage, crop development stage, mid-season stage, and late season stage of scenario 2 were decreased about 0.11%, 6.60%, 1.94%, and 2.91%, respectively. Maximum and minimum temperatures were slightly changed, precipitation were increased in April and May but decreased in June and July. Solar radiation and humidity were not different. Wind speed was unstable. It has been concluded that ET_{corn} on climate scenario 2 was slightly decreased because total precipitations on seasonal were increased. The results revealed that the difference between ET of Scenario 1 and 2 indicated not much different in most growth stages. But in crop development stage during May, the ET values of Scenario 1 seemed to be higher than Scenario 2. It has been concluded that ET values of Scenario 2 was slightly higher than Scenario 1 in crop development stage.

The comparison between ET of scenario 1 and scenario 3 were high different. ET values were increased 16.13%. The ET values in initial stage, crop development stage, mid-season stage, and late season stage of scenario 3 were increased 6.09%, 14.86%, 21.12%, and 20.40%, respectively. Maximum and minimum temperatures were increased. Precipitation was increased in April and July but decreased in May and June. The solar radiation was increased during April to June. Wind speed was unstable with high value. The results showed the difference between ET of Scenario 1

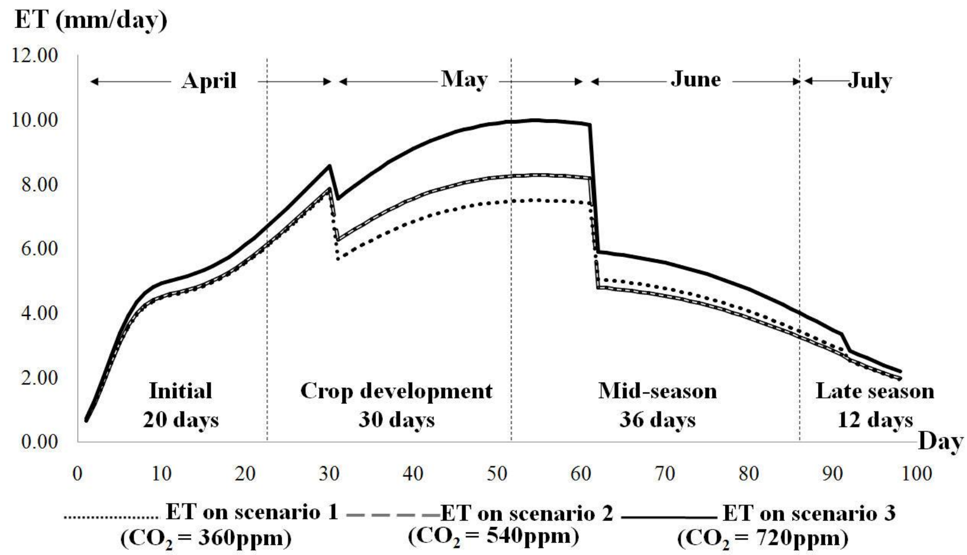

Figure 6. ET_{corn} and corn growth stage under three climate change scenarios

and 3 was slightly different in all stages. ET values of Scenario 3 seemed to be higher than those of Scenario 1. It has been concluded that ET of Scenario 3 was higher than Scenario 1 in all growth stages especially in crop development stage. The reasons should be the increasing of temperature, solar radiation, and wind speed in April and June.

3.4. Guidelines for water management

The results of this study revealed that Kc will be changed according to the atmospheric CO_2 concentration. The water management guidelines in the future to cope with the climate change which predicted by data mining technique can be developed as described as following.

(1) The results indicated precipitation will be increased in May and decreased in June. Corn requires enough water to maintain product quality and yield. Therefore, farmers should be shift crop season to be earlier.

(2) Water management organizations should overhaul the water reservation management and revise the irrigation system related to new Kc values and ETo under climate scenarios in the future.

Acknowledgement

This research was supported by the Center for Toxicology, Environment Health and Management of Toxic Chemicals under Science & Technology Postgraduate Education and Research Development Office (PERDO) of the Ministry of Education research grant. The authors would like to thank the Department of Environmental Management Technology, Faculty of Environment and Resource Studies, Mahidol University, Royal Irrigation Department, the Thai Meteorological Department, and the Southeast Asia START Regional Center for their technical supports.

References

Akinbile CO, Sangodoyin AY. Estimation crop coefficient model for upland rice (NERICA) under sprinkler irrigation system. African Journal of Agricultural Research 2010; 5: 436-41.

Allen RG, Pereira LS, Raies D, Smith M. Crop evapotranspiration, guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, United Nations, Rome. 1998.

Berry MJA, Linoff G. Data Mining Techniques. John Wiley & Sons., NY, USA. 1997.

Blum A. Neural Networks in C++. Wiley., NY, USA. 1992.

Chapman P, Kerber R, Clinton J, Khabaza T, Reinartz T, Wirth R. The CRISP-DM Process Model. Discussion Paper. 1999.

Chinwanno S. The impact of climate change in Thailand. The Southeast Asia START Regional Center (SEA START RC). Chulalongkorn University. 2007.

Chowdhary A, Shrivastava RK. Reference Crop Evapotranspiration estimation using Artificial Neural Networks. International Journal of Engineering Science and Technology Science and Technology 2010; 2: 4205-12.

Field crops research Institute. A guide book for field crops production. Department of Agriculture, Bangkok, Thailand. 2001.

Irrigation Water Management Research Group. Crop coefficient of 40 varieties. Irrigation Water Management Division., Office of Hydrology and Water Management., Royal Irrigation Department., Bangkok, Thailand. 2010.

Kotsiantis S, Kostoulas A, Lykoudis S, Argiriou A, Menagias K. Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature Values. International Journal of Mathematical, Physical and Engineering Sciences. 2008; 1: 16-20.

Mohan S, Arumugam N. Irrigation crop coefficients for lowland rice. Irrigation and Drainage System 1994; 8: 159-76.

Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. (*Eds*: Rumelhart DE, McClelland JL). Parallel Distributed Processing. MIT Press, Cambridge, MA. 1986.

Swingler K. Applying Neural Networks. A Practical Guide. Academic Press., London. 1996.

The Institution of Professional Engineers New Zealand (IPENZ). Climate change and the greenhouse effect. IPENZ National Office., New Zealand. 2001.

United Nations Environment Programme (UNEP). The Role of Ecosystem Management in Climate Change Adaptation and Disaster Risk Reduction. Copenhagen Discussion Series. 2009.

Vudhivanich V, Udakarn P. Study of Evapotranspiration and Crop Coefficient of Grape. Kasetsart University. Bangkok. 2001.

Received 12 August 2011
Accepted 20 September 2011

Correspondence to
Dr. Jongdee To-im
Faculty of Environment and Resource Studies,
Mahidol University,
Thailand
Email: jongdeetoim@gmail.com