American Journal of Environmental Sciences 8 (2): 95-103, 2012
ISSN 1553-345X

Estimation of Pan Coefficient using M5 Model Tree

[1]Pakorn Ditthakit and [2]Chaiyuth Chinnarasri
[1]School of Engineering and Resources,
Walailak University, Nakhon Si Thammarat 80160, Thailand
[2]Water Resources Engineering and Management Research Center (WAREE),
Department of Civil Engineering, Faculty of Engineering,
King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

Abstract: Problem statement: Pan Evaporation has extensively been used for estimating reference Evapotranspiration (ETo) due to its simplicity, low cost, ease of data interpretation and application and suitability for locations with limited availability of meteorological data. With this method, the pan coefficient (Kp) is a key element to be determined as well as the pan Evaporation (Ep) data. **Approach:** This study presents the development of new pan coefficient (Kp) equations for Class A pan and Colorado sunken pan under green and dry fetch conditions by using M5 model tree based on soft computing technique. The Kp values were taken from FAO-24 Kp table for the development of Kp equations. **Results:** The results of the study indicate the usefulness and applicability of the M5 model tree in developing Kp equations. Those proposed equations based on the M5 model tree gave better performance in estimating Kp values than the previous Kp equations as well as the new Kp equations developed by indicator regression technique. **Conclusion:** M5 model tree gave more accuracy in estimating Kp values. The new proposed Kp equations can be reliably used.

Key words: Estimating equation, indicator regression technique, M5 model tree, soft computing

INTRODUCTION

Accurate and reliable reference Evapotranspiration (ETo) estimation is an essential hydrological parameter for optimum water resources planning and farm irrigation scheduling. In recent years, several methods for estimating ETo from meteorological data have been proposed. For example, Turc (1961) proposed equation for estimating ETo using three meteorological data, including incoming solar radiation, mean daily air temperature at 2 m height and mean daily relative humidity. Priestley and Taylor (1972) developed equation for ETo estimation depending on daily mean air temperature, net radiation, heat flux density to the ground and atmospheric pressure. Hargreaves and Samani (1985) proposed ETo estimating equation using three meteorological data, including extraterrestrial radiation, maximum daily air temperature and minimum daily air temperature at 2 m height.

The FAO Penman-Monteith method is now recommended as a reference standard method for computing ETo. This meteorological-based method is complex and requires a significant number of meteorological data, i.e., air temperature, humidity, radiation and wind speed data. Often, the meteorological data are missing or incomplete due to instrument failure, contamination by measurement errors. For this reason, the pan Evaporation (Ep) has become widespread method due to its simplicity, low cost, ease of data interpretation and application and suitability for locations with limited availability of meteorological data (Phene and Campbell, 1975; Stanhill, 2002; Trajkovic, 2009). Indeed, the ETo can be determined as the product of a pan coefficient (Kp) and Ep.

Two types of evaporation pan, i.e., class A and Colorado sunken pans are commonly used. Colorado sunken pans are sometimes preferred in crop water requirement studies due to giving better ETo estimation; however, its maintenance is more difficult and leaks are not visible. Two cases of evaporation pan sitting are: (1) the pan is sited on a short green (grass) cover and surrounded by fallow soil and (2) the pan is sited on fallow soil and surrounded by a green crop.

Several Kp equations have been suggested based on the original and the FAO-24 Kp tables using linear, nonlinear and indicator regression techniques or combinations thereof. Frevert *et al*. (1983); Cuenca (1989); Snyder (1992) and Raghuwanshi and Wallender (1998) developed regression equations for predicting the Kp values for the FAO Class A pan placed in short

Corresponding Author: Chaiyuth Chinnarasri, Water Resources Engineering and Management Research Center (WAREE), Department of Civil Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

green cropped area based on FAO-24 Kp table. Cuenca (1989) modified Kp equation as proposed by Frevert *et al*. (1983) by rounding off the coefficients of equation. Snyder (1992) used the representative values to represent the category data of wind run and relative humidity and applied least-squares regression approach for predicting Kp values.

To develop Kp equation, Raghuwanshi and Wallender (1998) applied indicator regression technique, which is widely accepted approach for developing a relationship between categorical and quantitative data. Three Kp equations were developed based on the original data table (Allen and Pruitt, 1991; Orang, 1998; Grismer *et al*., 2002). Allen and Pruitt (1991) used stepwise and multivariate general linear regression procedures for FAO Class A pan placed in short green cropped area. Orang (1998) used linear regression technique and interpolation between fetch distances. Grismer *et al*. (2002) proposed an equation namely a modified Snyder (1992) equation. Allen (1998) and Abdel-Wahed and Snyder (2008) proposed Kp equation for the FAO Class A pan placed in dry fallow area.

Allen (1998) proposed two Kp equations for Colorado sunken pans surrounded by green and dry fetch conditions. Many attempts have been made for evaluation of pan coefficient in different regions and climates (Grismer *et al*., 2002; Irmak *et al*., 2002; Snyder *et al*., 2005; Ghare and Porey, 2008; Gundekar *et al*., 2008; Khoob, 2009).

M5 model tree, one of soft computing techniques, is gaining popularity in data analysis in sevearl branches of the science as well as water resources engineering problems. It mimics human mind dealing with imprecision, uncertainty, partial truth and approximation to achieve tractability, robustness and low solution cost (Mitra and Acharya, 2003). It is, hence, appropriate for solving hard tasks which an exact solution cannot be determined.

The examples of applying M5 model tree for water-related problems, i.e., rainfall-runoff modeling (Solomatine and Dulal, 2003), flood forecasting (Solomatine and Xue, 2004), water level-discharge relationship (Bhattacharya and Solomatine, 2005) and sedimentation modeling (Bhattacharya and Solomatine, 2006). The soft computing techniques are relatively new for predicting pan coefficient values. From literature reviews, few researches were found. For example, Trajkovic *et al*. (2000) applied radial basis function network to estimate FAO Blanney-Criddle b factor. Trajkovic *et al*. (2001) estimated the FAO Penman c factor using radial basis function network. Ditthakit and Chinnarasri (2011) presented the application of neuro-genetic approach for estimating pan evaporation coefficient for class a pan and Colorado sunken pan under green and dry fetch conditions. Although using neuro-genetic approach could be obtained higher performance in estimaitng Kp values when compared to previous methods, the explicit equations were not revealed.

It is, therefore, interested to investigate the performance of M5 model tree for pan coefficient estimation. The explicit equation would be determined in form of if-the rule.The purpose of this study is to apply M5 model tree for developing new pan coefficient equations for class A and Colorado sunken pans under green and dry fetch conditions. The indicator regression were also applied herein to determine pan coefficient equations for Class A pan placed in dry fallow area and Colorado sunken pan placed in short green cropped area and dry fallow area. The performance comparisons between the new proposed equations and previous equations is also presented and discussed.

MATERIALS AND METHODS

M5 model tree: M5 model tree was first introduced by Quinlan (1992). The mode is based on a divide-and-conquer approach for developing a relationship between independent and dependent variables. Unlike decision tree which is used for categorical data, It can be used for both qualitative (categorical) and quantitative data (Quinlan, 1986; 1992; Mitchell, 1997). This model is analogous to piece-wise linear functions with the combination of linear regression and regression tree concepts (Witten and Frank, 2005). The linear regression approach represents the relation of data set with a linear regression equation. For the regression tree approach, the data set is split up into subsets (also called leaves, child nodes, or sub-trees) and their relations at subsets (or leaves) are represented with averaged numeric values.

The regression tree is much larger and more complex than the regression equation. Like regression tree approach, the model tree make a splitting the data set into subsets (or leaves), but the relations of data set at its leaves are represented with linear regression equations, instead of averaged numeric values. The model tree can hence represent more sophisticated relations than either linear regression or regression trees and it is smaller in structure and more comprehensible than the regression tree.

In applying M5 model tree for nominal (or categorical) attribute like Kp equations development, all nominal attributes are transformed into binary variables that are then treated as numeric before constructing a model tree. For each nominal attribute,

the average class value corresponding to each possible value in the enumeration is calculated from the training instances and the values in the enumeration are sorted according to these averages. Then, if the nominal attribute has k possible values, it is replaced by k-1 synthetic binary attributes, the i th being 0 if the value is one of the first i in the ordering and 1 otherwise. Thus all splits are binary: they involve either a numeric attribute or a synthetic binary one, treated as a numeric attribute.

Building M5 model tree consists of three different stages (Quinlan, 1992; Solomatine and Xue, 2004; Pal, 2006). The first stage involves splitting of the data into subsets to create a decision tree. The splitting criterion is based on treating the standard deviation of the class values that reach a node as a measure of the error at that node and calculating the expected reduction in this error as a result of testing each attribute at that node. The formula for computing the Standard Deviation Reduction (SDR) is Eq. 1:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (1)$$

where, T represents a set of examples that reaches the node; T_i represents the subset of examples that have the i th outcome of the potential set and sd represents the standard deviation.

As a result of the splitting process, the standard deviation values of the data set in child nodes (sub-trees, or lower nodes) are less than those of parent nodes (higher nodes). After examining all the possible splits, the one that maximizes the expected error reduction was chosen. However, this division often produces a large tree-like structure that lead to overfit structure or poor generalizer. To avoid this problem, in second stage the overgrown tree is pruned and then the pruned sub-trees are replaced with linear regression functions. The pruning process concerns with merging some of the lower sub-trees into one node. Finally, the smoothing process is performed to compensate for the sharp discontinuities that will inevitably occur between adjacent linear models at the leaves of the pruned trees, particularly for some models constructed from a smaller number of training examples.

In smoothing, the adjacent linear equations are updated in such a way that the predicted outputs for the neighboring input vectors corresponding to the different equations are becoming close in value. This process substantially increases the accuracy of prediction (Quinlan, 1992; Witten and Frank, 2005). Example of M5 model tree algorithm is shown in Fig. 1. In Fig. 1a, it is the splitting of the input space $X_1 \times X_2$ (independent variables) by M5 model tree algorithm with 6 linear regression models at its leaves, labeled LM1 through LM6. Each model is a linear regression model in general form of y = a0+a1×1+a2×2, which a0, a1 and a2 are linear regression coefficients. In Fig. 1b, it is the details of its relations in form of tree diagram, in which LM_1 to LM6 are in leave level.

Existing pan coefficient (kp) equations: To evaluate the performance of the new proposed pan coefficient (Kp) equations for Class A and Colorado sunken pans, the different existing Kp equations as listed in Table 1 are used as benchmark. All equations in Table 1 were developed based on FAO-24 Kp table. The Kp values are the function of daily mean relative humidity, RH (%), daily mean wind speed at 2 m height, U_2 (km^{-1}) and fetch distance, F (m). The representative techniques for some equations can be concluded as follows.

Cuenca (1989) used representative (mean) values of each range of wind run (50, 300, 550 and850 km^{-1}, representing the wind run categories of<175, 175-425, 425-700 and >700 km^{-1}, respectively) and relative humidity (30, 55 and80%, representing the relative humidity categories of 40, 40-70≥70%, respectively). Snyder (1992) used representative (mean) values of each range of wind run (175, 300, 562 and700 km^{-1}, representing the wind run categories of <175, 175-425, 425-700 and>700 km^{-1}, respectively) and relative humidity (40, 55 and70%, representing the relative humidity categories of ≤40, 40-70, ≥70%, respectively).

In Eq. 4, which was proposed by Raghuwanshi and Wallender (1998), X_1 represents ln (F); X_2, X_3 and X_4 represent wind run categories of 175-425, 425-700 and 700 km^{-1} , respectively; and X_5 and X_6 represent relative humidity categories of 40-70 and 70%, respectively. The values of variables X_2, X_3, X_4, X_5 and X_6 are equal to 0 when the category do not present, or equal to 1 when the category present. The unit of daily mean wind speed at 2 m height, U_2 in the Eq. 5-9 is m sec^{-1}. According to Abdel-Wahed and Snyder (2008), the mean wind speeds 2, 3.5, 6.5 and8 m sec^{-1} were chosen to represent the wind run categories <2, 2-58 and, 5->8 m sec^{-1} and the mean of relative humidity 30, 55 and 75% were selected to represent the relative humidity categories of ≤40, 40-70 ≥70%.

Development of Kp equations using m5 model tree: To develop Kp equations based on M5 model tree for Class A pan placed in short green cropped area (Case I), Class A pan placed in dry fallow area (Case II), Colorado sunken pan placed in short green cropped area (Case III) and Colorado sunken pan placed in dry fallow area (Case IV), data sets based on Kp FAO-24 table (Allen, 1998) were used.

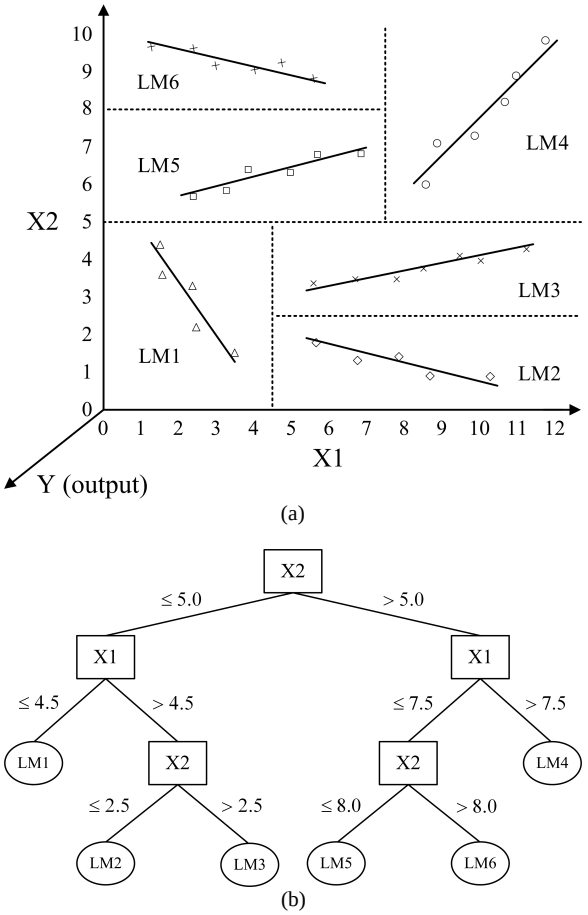

Fig. 1: Example of M5 model tree algorithm with 6 linear regression models

Table 1: The existing Kp equations

Author (year)	Kp equations	
Class A pan placed in short green cropped area (Case I)		
Cuenca (1989)	$Kp = 0.475 - 2.4\times10^{-4}U_2 + 5.16\times10^{-3}RH + 1.18\times10^{-3}F - 1.6\times10^{-5}\,RH^2 - 1.01\times10^{-6}F^2 - 8.0\times10^{-9}\,RH^2\,U_2 - 1.0\times10^{-8}\,RH^2F$	(2)
Snyder (1992)	$Kp = 0.482 - 0.000376\,U_2 + 0.024\ln(F) + 0.0045\,RH$	(3)
Raghuwanshi and Wallender (1998)	$Kp = 0.5944 + 0.0242X_1 - 0.0583\,X_2 - 0.1333X_3 - 0.2083\,X_4 + 0.0812\,X_5 + 0.1344\,X_6$	(4)
Allen (1998)	$Kp = 0.108 - 0.0286\,U_2 + 0.0422\ln(F) + 0.1434\ln(RH) - 0.000631\,[\ln(F)]^2\ln(RH)$	(5)
Class A pan placed in dry fallow area (Case II)		
Allen (1998)	$Kp = 0.61 + 0.00341\,RH - 0.000162\,U_2\,RH - 0.00000959\,U_2\,F + 0.00327\,U_2\ln(F) - 0.00289\,U_2\ln(86.4\,U_2) - 0.0106\ln(86.4\,U_2)\ln(F) + 0.00063\,[\ln(F)]^2\ln(86.4\,U_2)$	(6)
Abdel-Wahed and Snyder (2008)	$Kp = 0.62407 - 0.02660\ln(F) + 0.00028U + 0.00326\,RH$	(7)
Colorado sunken pan placed in short green cropped area (Case III)		
Allen (1998)	$Kp = 0.87 + 0.119\ln(F) - 0.0157\,[\ln(86.4\,U_2)]^2 - 0.0019[\ln F]-^2\ln(86.4\,U_2) + 0.013\ln(86.4\,U_2)\ln(RH) - 0.000053\ln(86.4\,U_2)\ln(F)RH$	(8)
Colorado sunken pan placed in dry fallow area (Case IV)		
Allen (1998)	$Kp = 1.145 - 0.080U_2 + 0.000903(U_2)^2\ln(RH) - 0.0964\ln(F) + 0.0031U_2\ln(F) + 0.0015[\ln(F)]^2\ln(RH)$	(9)

From this table, the fetch distance is quantitative data and the daily mean relative humidity and daily mean wind speed are qualitative (or categorical) data. The Kp value is dependent variable and other data, i.e., wind speed, fetch distance and mean relative humidity are independent variables.

Table 2 presents the total number of samples of 48 for Cases II, II and IV and the total number of samples of 36 for the Case III. Columns 2, 8 and 9 are input variables or independent variables and Columns 10-13 are output variables or dependent variables for cases I to IV, respectively. The ten-fold cross validation was selected for model verification. The novel proposed Kp equations based on M5 model tree algorithm was built with the help of Weka learning tool (version 3.6.0), which is public domain software (Witten and Frank, 2005).

Development of Kp equations using indicator regression: To evaluate the performance of M5 model tree, the indicator regression technique was used herein to develop Kp equations for Class A pan placed in dry fallow area (Case II), Colorado sunken pan placed in short green cropped area (Case III) and Colorado sunken pan placed in dry fallow area (Case IV). This technique was used for developing Kp equation for Class A pan placed in short green cropped area (Case I) by Raghuwanshi and Wallender (1998). Indicator regression (Draper and Smith, 1981; Milton and Arnold, 1994; Raghuwanshi and Wallender, 1998) is technique which use for developing a relationship between independent and dependent variables based on multiple linear regression method. This technique has been widely used in the areas of transportation engineering and social sciences. The advantage of this technique over others is independent variables can be both qualitative (categorical) and quantitative data. Indicator regression technique uses n-1 indicator (dummy) variables to represent categorical variables consisting of n classes. In this study, categorical data of daily mean relative humidity (RH) and daily mean wind speed (U_2) include 3 and 4 classes, respectively. Hence, the total of six dependent variables, that is, one for fetch distance (F), two for daily mean relative humidity (RH) and three for daily mean wind speed (U_2) are required.

The multiple regression equation can be expressed as Eq. 10:

$$Kp = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 \quad (10)$$

Where:

Kp = The pan coefficient;
βi = Regression coefficients;
X_1 = Natural logarithm (ln) of fetch distance in m;
X_2, X_3, X_4 = Wind run categories of 175-425, 425-700 and 700 km^{-1}, respectively
X_5, X_6 = Relative humidity categories of 40-70 and 70%, respectively

Table 2: Data preparation for Kp equations development

		Indicator regression					M5 model tree		U_2		RH	
Num of data	In (F) X1	X2	X3	X4	X5	X6	U_2	RH	Kp^{I}	Kp^{II}	Kp^{III}	Kp^{IV}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	0.0	0	0	0	0	0	<175	<=40	0.55	0.70	0.75	1.10
2	0.0	0	0	0	1	0	<175	40-70	0.65	0.80	0.75	1.10
3	0.0	0	0	0	0	1	<175	>=70	0.75	0.85	0.80	1.10
4	0.0	1	0	0	0	0	175-425	<=40	0.50	0.65	0.65	0.95
5	0.0	1	0	0	1	0	175-425	40-70	0.60	0.75	0.70	0.95
6	0.0	1	0	0	0	1	175-425	>=70	0.65	0.80	0.70	0.95
7	0.0	0	1	0	0	0	425-700	<=40	0.45	0.60	0.55	0.80
8	0.0	0	1	0	1	0	425-700	40-70	0.50	0.65	0.60	0.80
9	0.0	0	1	0	0	1	425-700	>=70	0.60	0.70	0.65	0.80
10	0.0	0	0	1	0	0	>700	<=40	0.40	0.50	0.50	0.70
11	0.0	0	0	1	1	0	>700	40-70	0.45	0.60	0.55	0.75
12	0.0	0	0	1	0	1	>700	>=70	0.50	0.65	0.60	0.75
13	2.3	0	0	0	0	0	<175	<=40	0.65	0.60	1.00	0.85
14	2.3	0	0	0	1	0	<175	40-70	0.75	0.70	1.00	0.85
15	2.3	0	0	0	0	1	<175	>=70	0.85	0.80	1.00	0.85
16	2.3	1	0	0	0	0	175-425	<=40	0.60	0.55	0.85	0.75
17	2.3	1	0	0	1	0	175-425	40-70	0.70	0.65	0.85	0.75
18	2.3	1	0	0	0	1	175-425	>=70	0.75	0.70	0.90	0.75
19	2.3	0	1	0	0	0	425-700	<=40	0.55	0.50	0.75	0.65
20	2.3	0	1	0	1	0	425-700	40-70	0.60	0.55	0.75	0.65
21	2.3	0	1	0	0	1	425-700	>=70	0.65	0.65	0.75	0.65
22	2.3	0	0	1	0	0	>700	<=40	0.45	0.45	0.65	0.55
23	2.3	0	0	1	1	0	>700	40-70	0.55	0.50	0.70	0.60
24	2.3	0	0	1	0	1	>700	>=70	0.60	0.55	0.70	0.65
25	4.6	0	0	0	0	0	<175	<=40	0.70	0.55	1.10	0.75
26	4.6	0	0	0	1	0	<175	40-70	0.80	0.65	1.10	0.75
27	4.6	0	0	0	0	1	<175	>=70	0.85	0.75	1.10	0.80
28	4.6	1	0	0	0	0	175-425	<=40	0.65	0.50	0.95	0.65
29	4.6	1	0	0	1	0	175-425	40-70	0.75	0.60	0.95	0.65
30	4.6	1	0	0	0	1	175-425	>=70	0.80	0.65	0.95	0.70
31	4.6	0	1	0	0	0	425-700	<=40	0.60	0.45	0.80	0.55
32	4.6	0	1	0	1	0	425-700	40-70	0.65	0.50	0.80	0.60
33	4.6	0	1	0	0	1	425-700	>=70	0.70	0.60	0.80	0.65
34	4.6	0	0	1	0	0	>700	<=40	0.50	0.40	0.70	0.50
35	4.6	0	0	1	1	0	>700	40-70	0.60	0.45	0.75	0.55
36	4.6	0	0	1	0	1	>700	>=70	0.65	0.50	0.75	0.60
37	6.9	0	0	0	0	0	<175	<=40	0.75	0.50	-	0.70
38	6.9	0	0	0	1	0	<175	40-70	0.85	0.60	-	0.70
39	6.9	0	0	0	0	1	<175	>=70	0.85	0.70	-	0.75
40	6.9	1	0	0	0	0	175-425	<=40	0.70	0.45	-	0.60
41	6.9	1	0	0	1	0	175-425	40-70	0.80	0.55	-	0.60
42	6.9	1	0	0	0	1	175-425	>=70	0.80	0.60	-	0.65
43	6.9	0	1	0	0	0	425-700	<=40	0.65	0.40	-	0.50
44	6.9	0	1	0	1	0	425-700	40-70	0.70	0.45	-	0.55
45	6.9	0	1	0	0	1	425-700	>=70	0.75	0.55	-	0.60
46	6.9	0	0	1	0	0	>700	<=40	0.55	0.35	-	0.45
47	6.9	0	0	1	1	0	>700	40-70	0.60	0.40	-	0.50
48	6.9	0	0	1	0	1	>700	>=70	0.65	0.45	-	0.55

Remarks: [I]: Class A pan placed in short green cropped area, [II]: Class A pan placed in dry fallow area [III]: Colorado sunken pan placed in short green cropped area and [IV]: Colorado sunken pan placed in dry fallow area

The values of variables X_2, X_3, X_4, X_5 and X_6 are equal to 0 when the category do not present, or equal to 1 when the category present. Table 2 shows assigned values for these variables. Columns 2-7 are input variables or independent variables and Columns 11-13 are output variables or dependent variables for cases 2-4, respectively. The multiple linear regression is used to determine regression coefficients.

RESULTS

The new developed Kp equations based on M5 model tree algorithm and indicator regression for all cases are presented Table 3.

In Eq. 11, the expression U = 425-700, 175-425, <175 can be interpreted as follows: if U_2 is either 425-700, 175-425, or <175, then substitute 1; otherwise, substitute 0. The other expressions in equations 12, 14 and 16 can be interpreted in the similar way. The meaning of variables in Eqs. 13, 15 and 17 can be explained as follow. X_1 represents ln(F); X_2, X_3 and X_4 represent wind run categories of 175-425, 425-700 and 700 km^{-1}, respectively; and X_5, X_6 represent relative humidity categories of 40-70, 70%, respectively. The values of variables X_2-X_5 and X_6 are equal to 0 when the category not present, or equal to 1 when the category present.

Considering the developed Kp equations based on M5 model tree, an equation is found for I, II as presented in Eq. 11, 12, respectively. The set of equations are found for cases III, IV, i.e., three rules for case III (Eq. 14) and two rules for case IV (Eq. 16). This may be because the values of standard deviation of the Kp values for Class A pans are less than those of Colorado sunken pans. The standard deviation of the Kp value for cases I-IV are 0.118, 0.121, 0.160 and 0.157, respectively. In addition, the range (the difference between maximum and minimum) of the Kp value for Class A pans (case I, II) are less than those of Colorado sunken pans (case III, IV). Those values are 0.450, 0.500, 0.600 and 0.650 for cases I-IV respectively. To evaluate the efficiency of the new proposed Kp equations, the comparison was done by using seven statistical indices, including determination coefficient (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Relative Error (MARE), Maximum Absolute Relative Error (MXARE), standard deviation of absolute relative error (DEV) and the number of samples with an error greater than 2% (NE>2%). The R^2 measures the degree to which two variables are linearly related and should optimally be one. The RMSE is a measure of the residual standard deviation and should be as small as possible (optimally 0). The MAE, MARE and MXARE measure the difference between actual and estimated Kp values and should be as small as possible (optimally0).

The actual Kp values are obtained from Allen (1998) for all cases.

Table 4 summarizes the statistical indices of the new proposed Kp equations and different existing equations for all study cases as well as Kp estimation using neuro-genetic approach as proposed by Ditthakit and Chinnarasri (2011).

Table 3: The new developed Kp equations based on M5 model tree and indicator regression

Method	Kp equations	
Class A pan placed inshort green cropped area (Case I)		
M5 model tree	Kp = 0.0243In (F) + 0.075U_2=425-700, 175-425, <175+ 0.075 U_2=175-425,<175 +0.0583U_2=<175+ 0.0812RH = 40-70,> = 70+0.0531RH = > = 70+0.386	(11)
Class A pan placed in dry fallow area (Case II)		
M5 model tree	Kp = -0.0266In (F) + 0.0667 U_2=425-700, 175-425,<175 + 0.0708 U_2=175-425,<175 + 0.0625 U_2=<175+0.0781RH =40-70,> = 70+0.0687RH = > = 70+0.5002	(12)
Indicator regression	Kp = 0.7000-0.0271X_1-0.0630X_2-0.01340X_3 -0.2000X_4+0.0797X_5+0.1500X_6	(13)
Colorado sunken pan placed in short green cropped area (Case III)		
M5 model tree	Rule: 1 If U_2=175-425,<175 <= 0.5 and In(F)>1.15 Then Kp = 0.0422 In(F) +0.0625 U_2 = 425-700, 175-425, <175+0.053 U_2 = 175-425,<175 + 0.0556 U_2=<175+0.0259RH = 40-70, > =70+ 0.548 Rule: 2 If In(F) <= 1.15 Then Kp = 0.0362 In(F) + 0.05 U_2 = 425-700,175 -425, <175+0.0833 U_2 = 175-425,<175 0.1049 U_2 = <175+0.0424RH = 40-70, >=70+0.5218 Rule: 3 Kp = 0.0399 In(F) + 0.1417 U_2 = <175+0.7708	(14)
Indicator regression	Kp = 0.8358+0.0527X_1-0.1500X_2 -0.2572X_3-0.3072X_4+0.0214X_5+0.0382X_6	(15)
Colorado sunken pan placed in dry fallow area (Case IV)		
M5 model tree	Rule: 1 If In (F) > 1.15 Then Kp = -0.0303In (F) + 0.0512 U_2 =425-700,175-425, <175+ 0.0831 U_2=175-425,<175 + 0.1037 U_2 = <175 + 0.0376RH = 40-70,> = 70+0.6634 Rule: 2 Kp = 0.0667 U_2=425-700,175-425, <175+ 0.15 U_2=175-425,<175 + 0.15 U_2=<175 + 0.7333	(16)
Indicator regression	Kp = 0.9333-0.0362X_1-0.1000X_2-0.1833X_3 -0.2333X_4+0.0167X_5+0.0500X_6	(17)

Table 4: Summary statistical indices of various Kp equations for performance comparisons

Method	R^2	RMSQ	MAE (%)	MARE (%)	MXARE (%)	DEV (%)	NE > 2%
Class A pan placed in short green cropped area (Case I)							
M5 model tree (Eq. 11)	0.9796	0.0235	1.88	2.97	8.44	2.22	21
Cuenca (1989) (Eq. 2)	0.9601	0.0327	2.69	4.41	13.25	3.31	27
Snyder (1992) (Eq. 3)	0.9745	0.0262	2.12	3.29	9.84	2.48	21
Raghuwanshi and Wallender (1998) (Eq. 4)	0.9796	0.0235	1.88	2.97	8.46	2.22	21
Allen (1998) (Eq. 5)	0.9822	0.0235	2.72	4.07	11.46	2.85	26
Neuro-Genetic	0.9901	0.0167	1.40	2.29	6.18	1.63	15
Class A pan placed in dry fallow area (Case II)							
M5 model tree (Eq. 12)	0.9870	0.0192	1.58	2.83	9.53	2.10	16
indicator regression (Eq. 13)	0.9870	0.0192	1.58	2.83	9.58	2.10	16
Allen (1998) (Eq. 6)	0.9849	0.0383	3.11	5.02	12.90	3.20	32
Abdel-Wahed and Snyder (2008) (Eq. 7)	0.9868	0.0194	1.59	2.87	10.00	2.18	16
Neuro-Genetic	0.9877	0.0188	1.45	2.42	8.10	2.05	17
Colorado sunken pan placed in short green cropped area (Case III)							
M5 model tree (Eq. 14)	0.9906	0.0218	1.67	2.33	6.99	2.04	12
Indicator regression (Eq. 15)	0.9699	0.0385	3.35	4.37	11.20	2.56	26
Allen (1998) (Eq. 8)	0.9840	0.0545	4.40	5.20	11.59	3.01	27
Neuro-Genetic	0.9921	0.0201	1.72	2.27	7.64	1.62	13
Colorado sunken pan placed in dry fallow area (Case IV)							
M5 model tree (Eq. 16)	0.9883	0.0241	1.81	2.83	10.56	2.63	18
Indicator regression (Eq. 17)	0.9536	0.0468	3.85	5.34	13.48	3.45	35
Allen (1998) (Eq. 9)	0.9851	0.0425	3.32	4.48	11.74	3.12	29
Neuro-Genetic	0.9890	0.0246	1.84	2.62	7.77	2.31	20

DISCUSSION

Obviously, neuro-genetic approach can estimte Kp value for all cases better than others. However, unlike using M5 model tree, it cannot be obtained the explicit equation. In case I, the R^2 value of M5 model tree (0.9796) is less than that of Allen (1998) equation (0.9822). The comparable result between M5 model tree (Eq. 11) and Raghuwanshi and Wallender (1998)'s equation (Eq. 4) is found. The values of absolute relative error (%) obtained from both methods are very close. Both gave less values of RMSQ (0.0235), MAE (%) (1.88), MARE (%) (2.97), DEV (%) (2.22) and NE>2% (21) in comparison to other existing equations. M5 model tree (Eq.11) gives less MXARE value than Raghuwanshi and Wallender (1998)'s equation (Eq.4), 8.44<8.46. It may be concluded that for Class A pan placed in short green cropped area the new Kp equation as presented in Eq.11 gives a promising performance in estimating Kp value and can be used as an alternative Kp equation.

In case II, all statistical indices for M5 model tree Eq. 12 and indicator regression Eq. 13 method are almost the same and better than those of Allen (1998) Eq. 6 and Abdel-Wahed and Snyder (2008) Eq. 7. The values of absolute relative error (%) obtained from M5 model tree Eq. 12 and indicator regression Eq.13 are very close. The M5 model tree gives a little higher performance in estimating Kp values than Abdel-Wahed and Snyder (2008) method. The MXARE (%) value using M5 model tree Eq.12 is less than that using indicator regression Eq.13, 9.53<9.58. This shows that the performance in estimating Kp values are improved when using M5 model tree. It could suggest two new proposed equations (Eqs. 12 and 13) instead of the previous equations (Eqs. 6 and 7) for this study case.

For case III, the Kp equation based on M5 model tree as expressed in Eq. 14 is rather more cumbersome to apply than other equations due to having three rules in a equation. However, the M5 model tree outperforms indicator regression (Eq. 15) and Allen (1998) equation (Eq. 8) in estimating Kp values. Obviously, M5 model tree gives all statistical indices better than indicator regression and Allen (1998) methods. Hence, this new Kp equation as shown in Eq. 14 can be satisfyingly used to estimate Kp value for this study case.

As explained previously, daily mean wind speed (U_2) is the categorical data and fetch distance is quantitative data. If the value of this categorical data is less than 0.5 (the average class value of binary variable (0, 1) for this case), it means not present in this category; otherwise present in this category. For instance, in Rule 1 of Eq. 14, the meaning of the expression If U=175-425, <175<= 0.5 and In(F)>1.15 can be explained as If $U_2 \neq$ 175-425 and $U_2 \neq$ <175 and ln(F)>1.15.

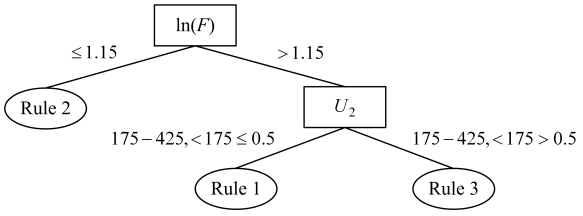

Fig. 2: Illustration of applying M5 model for estimating Kp value for Case III

The quantitative data like fetch distance can be directly interpreted ln(F)> 0.15 means the value of natural logarithm of fetch distance, F is more than 1.15meter). For Rule 3, it can give more explaination as follows. If In(F)>1.15 then Kp = 0.0399 In(F) + 0.1417 U_2=<175 + 0.7708. To illustrate of applying M5 model for estimating Kp value, Fig. 2 was drawn as similar to Fig. 1b here.

In case IV, the Kp equation developed by M5 model tree (Eq. 16) is more complex than the other equations. However, the satiable results are obtained with all statistical indices for M5 model tree much better than indicator regression and Allen (1998) equation. Most values of absolute relative error (%) obtained from M5 model tree (Eq. 16) are less than those obtained from other methods. It may suggest the new Kp equation as presented in Eq.16 to estimate Kp value for Colorado sunken pan placed in dry fallow area.

CONCLUSION

To estimate reference evaporation via evaporation pan based method, the pan coefficient (Kp) have to be obtained in addition to pan evaporation data. In this study, the M5 model tree based soft computing approach have been successfully applied for estimating pan coefficient values for both Class A and Colorado sunken Pans under green and dry fetch conditions. In comparison to the existing Pan Coefficient (Kp) equations as well as the new proposed equations based on indicator regression technique, M5 model tree gave more accuracy in estimating Kp values from Kp FAO-24 tables for all cases. Although this method does not outperform neuro-genetic approach in Kp estimtation, it gave the explicit equations, unlike neuro-genetic aproach. Thus, it indicates the application and usefulness of this technique in developing Kp equations. Based on statistical indices, the new proposed Kp equations can be reliably used.

ACKNOWLEDGEMENT

This research study is partly supported by Walailak University and Thailand Research Fund (TRF) under grant no. BRG 5280001.

REFERENCES

Abdel-Wahed, M.H. and R.L. Snyder, 2008. Simple equation to estimate reference evapotranspiration from evaporation pans surrounded by fallow soil. J. Irrig. Drain. Eng., 134: 425-429. DOI: 10.1061/(ASCE)0733-9437(2008)134:4(425)

Allen, R.G. and W.O. Pruitt, 1991. FAO-24 reference evapotranspiration factors. J. Irrig. Drain. Eng., 117: 758-773. DOI: 10.1061/(ASCE)0733-9437(1991)117:5(758)

Allen, R.G., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. 1st Edn., FAO Irrigation and Drainage Paper 56, Food and Agricultural Organization of the United Nations, Rome, ISBN-10: 9251042195, pp: 300.

Bhattacharya, B. and D.P. Solomatine, 2005. Neural networks and M5 model trees in modelling water level-discharge relationship. Neurocomputing, 63: 381-396. DOI: 10.1016/j.neucom.2004.04.016

Bhattacharya, B. and D.P. Solomatine, 2006. Machine learning in soil classification. Neural Netw., 19: 186-195. DOI: 10.1016/j.neunet.2006.01.005

Cuenca, R.H., 1989. Irrigation System Design: An Engineering Approach. 1st Edn., Prentice-Hall, Englewood Cliffs, N.J., ISBN: 9780135061633. pp: 552.

Ditthakit, P. and C. Chinnarasri, 2011. Estimation of pan evaporation coefficient using Neuro-Genetic approach. Am. J. Environ. Sci., 7: 397-401. DOI: 10.3844/ajessp.2011.397.401

Draper, N. and H. Smith, 1981. Applied Regression Analysis. 2nd Edn., John Wiley and Sons, Inc., New York, ISBN-10: 0471029955, pp: 709.

Frevert, D.K., R.W. Hill and B.C. Braaten, 1983. Estimation of FAO evapotranspiration coefficients. J. Irrig. Drain. Engrg., 109: 265-270. DOI: 10.1061/(ASCE)0733-9437(1983)109:2(265)

Ghare, A.D. and P.D. Porey, 2008. Estimation of reference evapotranspiration of nagpur region using simplified approach. Peroceedings of the 1st International Conference on Emerging Trends in Engineering and Technology, Jul. 16-18, IEEE Xplore Press, Nagpur, Maharashtra, pp: 1022-1028. DIO: 10.1109/ICETET.2008.74

Grismer, M.E., M. Orang, R. Snyder and R. Matyac, 2002. Pan evaporation to reference evapotranspiration conversion methods. J. Irrig. Drain. Eng., 128: 180-184. DOI: 10.1061/(ASCE)0733-9437(2002)128:3(180)

Gundekar, H.G., U.M. Khodke, S. Sarkar and R.K. Rai, 2008. Evaluation of pan coefficient for reference crop evapotranspiration for semi-arid region. Irrig. Sci., 26: 169-175. DOI 10.1007/s00271-007-0083-y

Hargreaves, G.H. and Z.A. Samani, 1985. Reference crop evapotranspiration from temperature. Applied Eng. Agric., 1: 96-99.

Irmak, S., D.Z. Haman and J.W. Jones, 2002. Evaluation of Class A pan coefficients for estimating reference evapotranspiration in humid location. J. Irrig. Drain. Engrg., 128: 153-159. DOI: 10.1061/(ASCE)0733-9437(2002)128:3(153)

Khoob, A.R., 2009. An evaluation of common pan coefficient equations to estimate reference evapotranspiration in a subtropical climate (north of Iran). Irrig. Sci., 27: 289-296. DOI: 10.1007/s00271-009-0145-4

Milton, J.S. and J.C. Arnold, 1994. Introduction To Probability And Statistics: Principles And Applications For Engineering And Computing Sciences. 3rd Edn., McGraw-Hill, Inc., New York, ISBN-10: 0070426236. pp: 736.

Mitchell, T.M., 1997. Machine Learning. 1st Edn., McGraw-Hill, New York, ISBN: 0070428077, pp: 414.

Mitra, S. and T. Acharya, 2003. Data Mining: Multimedia, Soft Computing and Bioinformatics. 1st Edn., John Wiley and Sons, Inc., Hoboken, New Jersey, ISBN-10: 0471460540, pp: 409.

Orang, M. 1998. Potential accuracy of the popular non-linear regression equations for estimating pan coefficient values in the original and FAO-24 table. Unpublished, California Department of Water Resources Report, Sacramento, California.

Pal, M., 2006. M5 model tree for land cover classification. Int. J. Remote Sens., 27: 825-831. DOI:10.1080/01431160500256531

Phene, C.J. and R.B. Campbell, 1975. Automating pan evaporation measurements for irrigation control. Agric. For. Meteor., 15:181-191. DOI: 10.1016/0002-1571(75)90003-5

Priestley, C.H.B. and R.J. Taylor, 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev., 100: 81-92. DOI: 10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Quinlan, J.R., 1986. Introduction of decision trees. Mach. Learn., 1: 81-106. DOI: 10.1007/BF00116251

Quinlan, J.R., 1992. Learning with continuous classes. Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence, Hobart, Australia, Nov. 16-18, World Scientific, Singapore, pp: 343-348.

Raghuwanshi, N.S. and W.W. Wallender, 1998. Converting from pan evaporation to evapotranspiration, J. Irrig. Drain. Eng., 124: 275-277. DOI: 10.1061/(ASCE)0733-9437(1998)124:5(275)

Snyder, R.L., 1992. Equation for Evaporation Pan to Evapotranspiration Conversions, J. Irrig. Drain. Eng., 118: 977-980. DOI: 10.1061/(ASCE)0733-9437(1992)118:6(977)

Snyder, R.L., M. Orang, S. Matyac and M.E. Grismer, 2005. Simplified estimation of reference evapotranspiration from pan evaporation data in California, J. Irrig. Drain. Eng., 131: 249253. DOI: 10.1061/(ASCE)0733-9437(2005)131:3(249)

Solomatine, D.P. and K.N. Dulal, 2003. Model trees as an alternative to neural networks in rainfall—runoff modelling. Hydrol. Sci J., 48: 399-411. DOI: 10.1623/hysj.48.3.399.45291

Solomatine, D.P. and Y. Xue, 2004. M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. J. Hydrol. Engrg., 9: 491-501. DOI: 10.1061/(ASCE)1084-0699(2004)9:6(491)

Stanhill, G., 2002. Is the Class A pan evaporationstill the most practical and accurate meteorological method for determining irrigation water requirements? Agric. Meteor., 112: 233-236.

Trajkovic, S., 2009. Comparison of radial basis function networks and empirical equations for converting from pan evaporation to reference evapotranspiration, Hydrol. Process., 23: 874-880. DOI: 10.1002/hyp.7221

Trajkovic, S., B. Todorovic and M. Stankovic, 2001. Estimation of FAO Penman c factor by RBF network. Sci. J. FACTA UNIVERSITATIS: Series Archit. Civil Eng., 2: 185-191.

Trajkovic, S., M. Stankovic and B. Todorovic, 2000. Estimation of FAO blaney-criddle b factor by RBF network. J. Irrig. Drain. Eng., 126: 268-270. DOI: 10.1061/(ASCE)0733-9437(2000)126:4(268)

Turc, L., 1961. Estimation of irrigation water requirements, potential evapotranspiration: A simple climatic formula evolved up to date. Ann. Argon., 12: 13-14.

Witten, I.H. and E. Frank, 2005. Data Mining: Practical Machine Learning Tools and Technique. 3rd Edn., Morgan Kaufmann Publishers, San Francisco, USA., ISBN-10: 0120884070, pp: 664.