

Estimation of Pan Evaporation Coefficient using Neuro-Genetic Approach

¹Pakorn Ditthakit and ²Chaiyuth Chinnarasri

¹School of Engineering and Resources Management, Walailak University,
Nakhon Si Thammarat 80160, Thailand

²Department of Civil Engineering, Faculty of Engineering,
Water Resources Engineering Research Center,

King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

Abstract: **Problem statement:** The pan evaporation coefficient (K_p) is used to convert pan Evaporation (E_p) to reference Evapotranspiration (ETo) due to its simplicity and suitability for locations with limited availability of meteorological data. **Approach:** This study presents the use of neuro-genetic approach for estimating K_p for Class A pan and Colorado Sunken pan under green and dry fetch conditions. **Results:** Representative values were used to represent the category data, i.e., wind run and relative humidity. It was found that the genetic algorithm helped automatically search for the optimal structure of the back-propagation network, replacing the very tedious trial and error approach. **Conclusion:** A comparative analysis showed that the neural-genetic approach fairly outperformed previous proposed K_p equations for both green and dry fetch conditions.

Key words: Informatics technology, statistical analysis, pan evaporation, colorado sunken, comparative analysis, indicator regression

INTRODUCTION

Hydrological extreme events are typically defined as floods and droughts. Floods are associated with high rainfall and may cause dam break events (Tingsanchali and Chinnarasri, 2001) while drought is associated with lack of precipitation and high evaporation. For modeling purpose and decision making, it is therefore a need for a study to estimate hydrological data for areas with lack of data and different climatic conditions (Hosseini *et al.*, 2011).

Pan Evaporation (E_p) has become a widespread method 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 (Phene and Campbell, 1975; Trajkovic, 2009). To convert E_p to ETo , a pan coefficient (K_p) is necessary. Although the FAO-24 K_p Table 1 provides K_p values, several K_p equations have been developed for estimating K_p values (Cuenca, 1989; Snyder, 1992; Allen, 1998; Raghuwanshi and Wallender, 1998; Abdel-Wahed and Snyder, 2008). Those equations were developed for two types of pan, i.e., Class A and Colorado sunken pans and for two conditions, i.e., a

pan placed in a short green cropped area and a pan placed in a dry fallow area.

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

Cuenca (1989) modified K_p equation as proposed by Frevert *et al.* (1983) by rounding off the coefficients of the equation. Snyder (1992) used the representative values to represent the category data of wind run and relative humidity and applied a least-squares regression approach for predicting K_p values.

To develop a K_p equation for Class A pan placed in short green cropped areas, Raghuwanshi and Wallender (1998) applied the indicator regression technique, which is a widely accepted approach for developing a relationship between categorical and quantitative data.

Allen (1998) and Abdel-Wahed and Snyder (2008) proposed K_p equations for an FAO Class A pan placed in a dry fallow area.

Corresponding Author: Chaiyuth Chinnarasri, Department of Civil Engineering, Faculty of Engineering, Water Resources Engineering Research Centre, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

Table 1: Pan coefficients for different pan siting and environments (Cuenca, 1989)

Wind speed (km/day)	Windward side distance	Class A pan under green fetch condition (case I) RH mean (%)			Class A pan under dry fetch condition case II) RH mean (%)			Colorado sunken pan under green fetch condition (case III) RH mean (%)			Colorado sunken pan under dry fetch condition (case IV) RH mean (%)		
		Low	Medium	High	low	Medium	High	Low	Medium	High	Low	Medium	High
		< 40	40-70	> 70	< 40	40-70	> 70	< 40	40-70	> 70	< 40	40-70	> 70
Light < 175 (< 2 m sec ⁻¹)	1	0.55	0.65	0.75	0.70	0.8	0.85	0.75	0.75	0.80	1.10	1.10	1.10
	10	0.65	0.75	0.85	0.60	0.70	0.80	1.00	1.00	1.00	0.85	0.85	0.85
	100	0.70	0.80	0.85	0.55	0.65	0.75	1.10	1.10	1.10	0.75	0.75	0.80
	1000	0.75	0.85	0.85	0.50	0.60	0.70	-	-	-	0.70	0.70	0.75
Moderate 175-425 (2-5 m sec ⁻¹)	1	0.50	0.60	0.65	0.65	0.75	0.80	0.65	0.70	0.70	0.95	0.95	0.95
	10	0.60	0.70	0.75	0.55	0.65	0.70	0.85	0.85	0.90	0.75	0.75	0.75
	100	0.65	0.75	0.80	0.50	0.60	0.65	0.95	0.95	0.95	0.65	0.65	0.70
	1000	0.70	0.80	0.80	0.45	0.55	0.60	-	-	-	0.60	0.60	0.65
Strong 425-700 (5-8 m sec ⁻¹)	1	0.45	0.50	0.60	0.60	0.65	0.70	0.55	0.60	0.65	0.80	0.80	0.80
	10	0.55	0.60	0.65	0.50	0.55	0.65	0.75	0.75	0.75	0.65	0.65	0.65
	100	0.60	0.65	0.70	0.45	0.50	0.60	0.80	0.80	0.80	0.55	0.60	0.65
	1000	0.65	0.70	0.75	0.40	0.45	0.55	-	-	-	0.50	0.55	0.60
Very strong > 700 (> 8 m sec ⁻¹)	1	0.40	0.45	0.50	0.50	0.60	0.65	0.50	0.55	0.60	0.70	0.75	0.75
	10	0.45	0.55	0.60	0.45	0.50	0.55	0.65	0.70	0.70	0.55	0.60	0.65
	100	0.50	0.60	0.65	0.40	0.45	0.50	0.70	0.75	0.75	0.50	0.55	0.60
	1000	0.55	0.60	0.65	0.35	0.40	0.45	-	-	-	0.45	0.50	0.55

Allen (1998) also proposed two K_p equations for Colorado sunken pans surrounded by green and dry fetch conditions. All of the existing K_p equations are the function of the daily mean Relative Humidity (RH,%), daily mean wind speed at 2 m height (U₂, km/day) and fetch distance (F, m). The main differences are the selected representative values of each range of wind run categories and relative humidity categories.

The neuro-genetic approach, a hybrid of neural networks and genetic algorithms, is relatively new for predicting pan coefficient values. The main function of the neuro-genetic approach is still based on neural networks, while the genetic algorithm helps automatically search for the optimal structure of the network. Only two related research projects for the estimation of reference evapotranspiration factors were found in the literature. Trajkovic *et al.* (2000) applied a radial basis function network to estimate the FAO Blaney-Criddle factor. Trajkovic *et al.* (2001) estimated the FAO Penman factor using a radial basis function network.

This study presents the application of the neuro-genetic approach to estimating pan evaporation coefficient values for Class A pans and Colorado Sunken pans under green and dry fetch conditions. The statistical performance comparison was undertaken between this neuro-genetic approach and the previous proposed equations.

MATERIALS AND METHODS

The neuro-genetic approach is a hybrid model between Artificial Neural Networks (ANNs) and Genetic Algorithms (GAs).

The main function of this model is based on ANNs while a GA helps to automatically find a suitable structure for the network, i.e., the decision variables are the number of hidden layers, the number of nodes in each hidden layer and the transfer functions used. A neuro-genetic process includes two sub-processes: A Neural Network (NN) process (an inner iterative loop) and a genetic algorithm process (an outer iterative loop). Detailed information for neuro-genetic processes can be seen in Fig. 1.

A typical Back Propagation (BP) neural network structure, which includes one input layer, one or more hidden layers and one output layer, is shown in Fig. 2. Each layer is made up of several neurons and the layers are interconnected by sets of weights. The neurons in the input layer receive input directly from the input variables. The neurons in the hidden and output layers receive input from the interconnections. Neurons operate on the input and transform it to produce an analogue output. The transformation is performed in two stages. First, the input from each neuron is multiplied by weights and a weighted sum is taken. Next, an activation function such as the sigmoid function converts such a weighted sum to be the output of each neuron, which becomes the input to neurons of the succeeding layer.

The training of a BP neural network involves two stages (Kumar *et al.*, 2002). In the first stage (forward pass), the input signals propagate from the network input to the output. The calculation of the output is carried out, layer by layer, in the forward direction. The output of one layer is the input to the following layer.

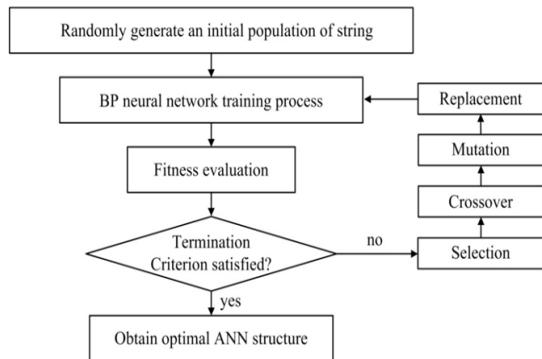


Fig. 1: Neuro-genetic process

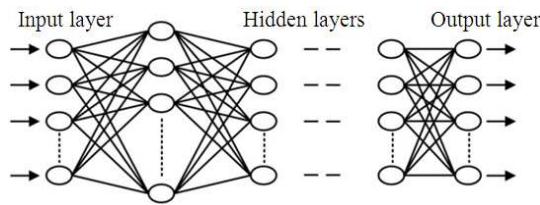


Fig. 2: Typical BP neural network structure

In the second stage (backward pass), the calculated error signals propagate backward through the network, where they are used to adjust the weights of the connections in order to improve a predefined performance measure. Initially, since the weights to the interconnections are randomly generated, the difference between the predicted and desired output values can be large. Learning therefore involves iteratively adjusting the connection weights to minimize these differences.

There are two parameters of concern for a BP neural network: the learning rate (α) and the momentum rate (β). The learning rate controls the incremental change in the interconnection weights during iterative training as a percentage of the difference between the desired or target output and the NN computed output. Thus, a high learning rate generally results in a larger weight change and faster learning. The momentum rate is a means to increase the rate of learning and avoid the possibility of getting trapped in local optima. The momentum rate is a multiplication factor to the change in the previous interconnection weights (Sivapragasam and Muttill, 2005).

A Genetic Algorithm (GA) is a stochastic optimization model based on Darwin's concept of natural selection. GA is a robust method for searching for the optimal solution of complex problem, although it may not always necessarily obtain the best possible solution (Goldberg, 1989). In a GA, the population of strings is used to represent possible problem solutions.

Each string consists of a number of blocks, which represent the individual variables of the problem. The variables represented in the string can be processed in an evaluation function, or fitness function, which is in effect the objective function.

GA consists of three basic operations (selection, crossover and mutation) involved in manipulating strings and moving to a new generation. The selection operator is that through which strings are selected for inclusion in the reproduction process and for participation in the next generation. The fittest strings have the highest probability of being used in reproduction. The crossover operator permits the exchange of genes between pairs of chromosomes in a population. This operator offers the possibility of good genetic material from different individual strings being combined to create an even fitter individual. The mutation operator permits new genetic material to be introduced to a population. As shown in Fig. 1, replacement is used to refill the population that is dropped due to poor chromosomes during the selection process.

RESULTS

To estimate Kp values based on the neuro-genetic approach, the data sets from Table 1 were used. This table shows Kp values based on the FAO-24 Kp table for Class A pans under green fetch conditions (case I), Class A pans under dry fetch conditions (case II), Colorado sunken pans under green fetch conditions (case III) and Colorado sunken under dry fetch conditions (case IV).

The Kp values depends on the fetch distance (F), which is quantitative data and two qualitative (or categorical) data, i.e., the daily mean Relative Humidity (RH) and daily mean wind speed (U). Hence, the structure of the neural networks comprised three nodes of input layer, i.e., In (F), RH and U and one node of output layer, i.e., Kp data.

The total data sets are 48 for cases I, II and IV and the total data sets are 36 for case III. Those data sets were randomly divided into two parts, that is, 90% for training and 10% for testing networks. Four categorical data of daily mean wind speed of <175, 175-425, 425-700 and >700 km day⁻¹ were assigned values of 1, 2, 3 and 4, respectively. In addition, three categorical data of relative humidity of <40, 40-70 and >70% were assigned values of 1, 2 and 3, respectively.

Back-Propagation (BP) was selected as the supervised learning method and a Genetic Algorithm (GA) helped in finding a suitable network, i.e., the number of hidden layers, the number of nodes in each hidden layer and the transfer functions to be used for each node.

Table 2: Comparison of statistical indices in estimating K_p values using different methods

Methods	R ²	RMSQ	MAE (%)	MARE (%)	MXARE (%)	DEV (%)	NE > 2%
Class A pan under green fetch condition (case I)							
Cuenca (1989)	0.9601	0.0327	2.69	4.41	13.25	3.31	27
Snyder (1992)	0.9745	0.0262	2.12	3.29	9.84	2.48	21
Allen (1998)	0.9822	0.0235	2.72	4.07	11.46	2.85	26
Raghuvanshi and Wallender (1998)	0.9796	0.0235	1.88	2.97	8.46	2.22	21
Neuro-genetic (present study)	0.9901	0.0167	1.40	2.29	6.18	1.63	15
Class A pan under dry fetch condition (case II)							
Allen (1998)	0.9849	0.0383	3.11	5.02	12.90	3.20	32
Abdel-Wahed and Snyder (2008)	0.9868	0.0194	1.59	2.87	10.00	2.18	16
Neuro-Genetic (Present study)	0.9877	0.0188	1.45	2.42	8.10	2.05	17
Colorado sunken pan under green fetch condition (case III)							
Allen (1998)	0.9840	0.0545	4.40	5.20	11.59	3.01	27
Neuro-Genetic (Present study)	0.9921	0.0201	1.72	2.27	7.64	1.62	13
Colorado Sunken pan under dry fetch condition (case IV)							
Allen (1998)	0.9851	0.0425	3.32	4.48	11.74	3.12	29
Neuro-Genetic (present study)	0.9890	0.0246	1.84	2.62	7.77	2.31	20

NeuroGenetic Optimizer (demo version) was selected as the learning tool. The default parameters of neural networks and genetic algorithm as provided by the NeuroGenetic Optimizer software package were selected in this study as follows. The input and output data scaling were between -1 and 1 and 0.1 and 0.9, respectively. The multiple hidden layers with maximum of 2 hidden layers and the maximum number of nodes in each hidden layer of 128 were selected. Three transfer functions (Tan-Hyperbolic, Logistic and Linear transfer functions) were used. The initial weights were randomly assigned between -0.3-0.3. Learning rates of between 0.1-0.4 for the first and second hidden layers and between 0.1-0.2 for the output layer were set. A momentum between 0.1-0.2 was set.

A population size (the number of chromosomes in the population) of 200 was used. The percent selection method with 50% selection was chosen. That is, the networks with fitness greater than average will be selected to survive while those less than average will be dumped into the bit bucket. Since this system uses a selection technique that drops poor chromosomes, the population needs to be refilled every generation. In this study, a cloning technique, whereby the survivors of the selection process are cloned to refill the population, was used. The tail swap mating method (also known as crossover) was selected. Hence, the system picks a cut point and exchanges genetic material between the cut point and the end of the string of the parents, essentially swapping tails. A random exchange mutation method with a probability of 0.25 was selected. With this method, two points in a given chromosome string are randomly selected and exchanged their values with a probability of 0.25.

Seven statistical indices, including the determination coefficient (R²), 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%) were used to evaluate the performance in estimating K_p values. The R² 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 K_p values and should be as small as possible (optimally 0).

DISCUSSION

The statistical indices in estimating K_p values using neuro-genetic and previous equations for all study cases are presented in Table 2. For case I, the neuro-genetic approach gave a higher performance (R² = 0.9901, RMSQ = 0.0167, MAE (%) = 1.40, MARE (%) = 2.29, MXARE (%) = 6.18, DEV (%) = 1.63 and NE> 2% = 15) in estimating K_p values than other methods. For case II, although the values of R² (0.9877) and NE> 2% (17) of neuro-genetic approach were lower than those of the equation proposed by Abdel-Wahed and Snyder (2008) other statistical indices (RMSQ = 0.0188, MAE (%) = 1.45, MARE (%) = 2.42, MXARE (%) = 8.10 and DEV (%) = 2.05) were better.

For cases III and IV, the neuro-genetic approach obviously outperformed the equations proposed by Allen (1998). It gave R² = 0.9921, RMSQ = 0.0201, MAE (%) = 1.72, MARE (%) = 2.27, MXARE (%) =

7.64, DEV (%) = 1.62 and NE> 2% = 13 for case III and $R^2 = 0.9890$, RMSQ = 0.0246, MAE (%) = 1.84, MARE (%) = 2.62, MXARE (%) = 7.77, DEV (%) = 2.31 and NE> 2% = 20 for case IV.

The structures of BP neural networks (# input layer nodes-1st hidden layer nodes-2nd hidden layer nodes- # output layer nodes-the hidden layers being grouped by type of transfer function) obtained by the GA for cases I-IV, respectively, were (3-17 nodes with Lo, 5 with T, 5 with Li-50 with Lo, 28 with T, 46 with Li-1), (3-46 with Lo, 4 with T, 24 with Li- 95 with Lo, 27 with T, 6 with Li-1), (3-5 with Lo, 6 with T, 5 with Li-15 with Lo, 2 with T, 2 with Li-1) and (3-15 with Lo, 49 with T, 44 with Li-47 with Lo, 24 with T, 3 with Li-1), respectively. Lo, T and Li stand for Logistic, Tan-Hyperbolic and Linear transfer functions, respectively and the GA always selected Tan-hyperbolic for the output node.

CONCLUSION

In this study, a Neuro-genetic approach was successfully applied to estimating K_p values for the cases of Class A pans under green fetch conditions, Class A pans under dry fetch conditions, Colorado sunken pans under green fetch conditions and Colorado sunken pans under dry fetch conditions. A comparative analysis showed that this approach fairly outperformed the previous proposed K_p equations in estimating K_p values.

ACKNOWLEDGEMENT

The researchers would like to thank the Thailand Research Fund for providing financial support under grant number BRG5280001 for the present study.

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.
- Allen, R.G., 1998. Crop evapotranspiration: Guidelines for Computing Crop Water Requirements. 1st Edn., Food and Agriculture Organization of the United Nations, Rome, ISBN: 9789251042199, pp: 300.
- Cuenca, R.H., 1989. Irrigation system design: An Engineering Approach. 1st Edn., Prentice-Hall, Englewood Cliffs, NJ., ISBN: 9780135061633, pp: 552.
- Frevert, D.K., Hill, R.W. and B.C. Braaten, 1983. Estimation of FAO evapotranspiration coefficients. *J. Irrig. Drain. Engrg.*, 109: 265-270.
- Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. 1st Edn., Addison-Wesley, Reading, MA., ISBN: 9780201157673, pp: 412.
- Hosseini, M., M.S.M. Amin, A.M. Ghafouri and M.R. Tabatabaei, 2011. Application of soil and water assessment tools model for runoff estimation. *Am. J. Applied Sci.*, 8: 486-494. DOI: 10.3844/ajassp.2011.486.494
- Kumar, M., N.S. Raghuwanshi, R. Singh, W.W. Wallender and W.O. Pruitt, 2002. Estimating evapotranspiration using artificial neural network. *J. Irrig. Drain. Engrg.* 128: 224-233.
- Phene, C.J. and R.B. Campbell, 1975. Automating pan evaporation measurements for irrigation control. *Agric. Meteorol.*, 15: 181-191. DOI: 10.1016/0002-1571(75)90003-5
- Raghuwanshi, N.S. and W.W. Wallender, 1998. Converting from pan evaporation to evapotranspiration. *J. Irrig. Drain. Eng.*, 124: 275-277.
- Sivapragasam, C. and N. Muttill, 2005. Discharge rating curve extension ? A new approach. *Water Resources Mana.*, 19: 505-520.
- Snyder, R.L., 1992. Equation for evaporation pan to evapotranspiration conversions. *J. Irrig. Drain. Eng.*, 118: 977-980.
- Tingsanchali, T. and C. Chinnarasri, 2001. Numerical modelling of dam failure due to flow overtopping. *Hydrol. Sci. J.*, 46: 113-130. DOI: 10.1080/02626660109492804
- 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 factor by RBF network. *Scientific J. Facta Univ. Series Arch. 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.