November-December 2019
ISSN: 0193-4120 Page No. 858- 866

Particle Swarm Optimization of Multi-Linear Regression for Evapotranspiration Estimation Model

Nor Farah Atiqah Ahmad*[1], Sobri Harun[2], Haza Nuzly Abdull Hamed[3]
[1]*Department of Civil Engineering, Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia, Edu Hub Pagoh, 84600 Muar, Johor, Malaysia*
**Corresponding author: norfarah@uthm.edu.my*
[2]*School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
sobri@utm.my
[3]*School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
haza@utm.my

Article Info
Volume 81
Page Number: 858- 866
Publication Issue:
November-December 2019

Article History
Article Received: *3 January 2019*
Revised: *25 March 2019*
Accepted: *28 July 2019*
Publication: *25 November 2019*

Abstract

The data demanding Food and Agricultural Organization-56 Penman-Montieth model (FPM-56) is the most accurate model in estimating evapotranspiration (ET) but it is not applicable at data scarce region. This paper evaluates the performance of conventional MLR models and improved MLR by using PSO algorithms (MLR-PSO) in estimating potential evapotranspiration (ETp) by only using 2 significant parameters affecting ETp for tropical climate. In this study, 17 meteorological stations around Peninsular Malaysia were used in this study and obtained its both MLR and MLR-PSO models. These models were compared by using root mean square error (RMSE), coefficient of determination (R2) and its accuracy (Acc). The obtained results show MLR models itself has accuracy closed to 94% against FPM-56 models. Whereas optimized MLR-PSO models has improved up to 2.95% of accuracy. Out of 4 PSO algorithm, the standard c1=c2=2.0 and w=1.0 resulted better performance in 7 stations compared to others. The results proves that MLR and MLR-PSO models both useful for estimating ETp at data scarce region as it required only 2 main parameters affecting ETp.

Keywords: *Evapotranspiration model, Particle swarm optimization, Peninsular Malaysia, Multi-linear regression, water resources*

1 INTRODUCTION

Water scarcity has become the main concern to the most of study area throughout the world. For a small region like Malaysia which received approximately 2000 to 3000 mm annual precipitation, the country still facing the water scarcity issue as the distribution of precipitation is not uniform let alone throughout the globe. Therefore, the increased of competition on preserving, managing and optimizing water resources are noticeable from various field of study. From the hydrological community, the fundamental understanding behind water vulnerability is the struggling in maintaining the water balance system. ET is one of the

TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

essential processes in hydrological cycle amongst precipitation, evaporation, transpiration as well as infiltration. Unlike precipitation, physical measurement of ET is tedious nearly impossible for certain region. Although lysimeter is the prominent apparatus to measure actual ET, it consumes cost, time and labour. The FPM-56 model has been introduced by [1] and has been used globally since then. With over 10000 citation, the accuracy of this model proved to be as observed ET. However, FPM-56 also known as data demanding model which has become a major hindrance for certain region. Knowing this setback, the updated publication by [2] suggested that instead of simplify the FPM-56 model, user should estimate the missing data and use the Penman-Monteith [3] instead. Nevertheless, estimating physically-based processes can produce significant error as these processes are highly non-linear [4] and [5] suggested that the estimation for hydrological parameters should be done by established forecasting methods in order to accurately predict these parameters.

The simpler empirical models can be classified as mass-transfer-based, temperature-based, radiation-based, pan-evaporation-based and combination. The example of simpler empirical models such Turc [6], Hargreaves-Samani [7], Priestley-Taylor [8] and Makkink [9] is used at almost any region around the globe. These models though simple yet need calibration before it can be used at other region since it is a site specific model. By taking Hargreaves as example, the model shows overestimates ET_0 under humid locations humid locations [10] and underestimates under arid locations [11] in [12]. Conclusively, [13] found that the radiation-based and temperature-based models are more suitable in estimating ET in humid climate in Iran. As according to [14], mean temperature and solar radiation variables are the most influential parameters of ET_p for Peninsular Malaysia.

2 STUDY AREA AND METHODS

Daily meteorological data of maximum- (T_{max}) , minimum- (T_{min}), average temperature (T_{mean}), relative humidity (RH), solar radiation (R_s) and wind speed (u) recorded from 1987-2003 from 17 stations located around Peninsular Malaysia were used in this study as shown in Figure 1. The stations includes Alor Setar (AS), Batu Embun (BE), Bayan Lepas(BL), Cameron Highland (CH), Chuping (Chu), Kota Bharu (KB), Kuala Krai (KKrai), Kluang (Klu), Kuala Terengganu (KT), Kuantan (Ktn), Melaka (Mlk), Mersing (MS), Muadzam Syah (Mdz), Subang (Sbg), Senai (Sn), Stitiawan (Stwn) and Temerloh (TM) were chosen due to its arability of desired meteorological data.

Figure 1. Location of meteorological station in Peninsular Malaysia

2.1 FPM-56 Models

FPM-56 has been recognized in hydrologist world as the most prominent empirical model that gives close to accurate estimation of ET. However, note that this is

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

also a data demanding model which is not applicable at some region. Despite that, estimation of ET_p from FPM-56 is used as the observed data for analysis. The mathematical equation is presented as in Eq. (1).

$$FAO-PM = \frac{0.408(R_n - G) + \gamma \frac{900}{T_a + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

In which R_n is the net radiation (MJ/m^2/day), G is the soil heat flux (MJ/m^2/day), γ is the psychrometric constant (kPa/°C), e_s is the saturation vapor pressure (kPa), e_a is the actual vapor pressure (kPa), Δ is the slope of the saturation vapor pressure-temperature curve (kPa/°C), T_a is the average daily air temperature (°C) and u_2 is the average daily wind speed at 2m height (m/s). Therefore, grass height and bulk canopy resistance were assumed to be 0.12m and 70m/s respectively.

2.2 Multi-Linear Regression, MLR

MLR is a one of the oldest statistical analysis that observed the relationship between several predictor independent variables and dependent variables [15]. MLR represents a mathematical equation expressing the response variable as a function of several explanatory variables and is described as in Eq. (2). The equation describes how the mean changes with the explanatory variables.

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_x + c \quad (2)$$

Where *a* is the intercept, *b* is the slope or coefficient, *n* is number of observations and *c* is the unexplained noise in the data (error).

2.3 Particle Swarm Optimization, PSO

The application of (PSO) is no longer a stranger for computer science community. This based on the social behavior of animals theory model [16] may have not yet established but its evolutionary algorithm that can be used to find optimal solutions to numerical and qualitative problems [17] prevails to another field of study such hydrology [18, 19]. Based on the theory of animal social behavior, a certain numbers of individuals known as particles are collecting information from each other through their respective positions. Each particle has their own p_{best} and will update their position and velocity to their neighbors in order to obtain the objective function. The new velocity and position of the swarm is called as g_{best} that can be represent by using Eq. 3 and Eq. 4. This process iterates until the termination criteria is satisfied.

$$\overrightarrow{v_{new}} = \vec{v} + C_1 \times \left(\overrightarrow{p_{best}} - \vec{p}\right) + C_2 \times \left(\overrightarrow{g_{best}} - \vec{p}\right) \quad (3)$$

$$\overrightarrow{p_{new}} = \vec{p} + \overrightarrow{v_{new}} \quad (4)$$

Where $\overrightarrow{v_{new}}$, $\vec{v}$, $\overrightarrow{p_{new}}$ and $\vec{p}$ are new velocity, current velocity, new position and current position of particles respectively. Unlike the basic PSO proposed by [16] where no inertia weight (*w*) is included, [20] has introduced initial weight as it helps in balancing the both local and global search. The suggested range of initial weight is from 0.9 to 1.2 for a better performance. According to [19, 21] a large inertia weight contributes in good global search while a smaller value aid in local exploration. The practice is to use larger initial weight during the initial exploration and gradual reduction of its values as the search proceeds in further

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

iterations. The equation to update the velocity can be expressed as in Eq. (5).

$$\overrightarrow{v_{new}} = w \cdot \vec{v} + C_1 \times \left(\overrightarrow{p_{best}} - \vec{p}\right) + C_2 \times \left(\overrightarrow{g_{best}} - \vec{p}\right) \quad (5)$$

Where c_1 and c_2 are the cognitive and social coefficients respectively. In this study, 4 sets of PSO algorithms were used in this study as shown in Table 1.

Table 1. PSO Algorithm

C_1	C_2	w	Reference
2.0	2.0	1.0	[20]
2.0	2.0	0.9	[22]
2.0	2.0	0.4	[22]
1.49618	1.49618	0.7298	[23]

3 RESULTS

All the runs performed in this study were executed on a computer with a 2.4 GHz Core i5 processor. Four set of algorithm were run 10 times in order to avoid the influence of randomicity at 17 stations. The performance indicators used in this study are root mean square error (RMSE), coefficient of determination (R^2) and accuracy (Acc.)

3.1 Models Development

By taking 2 most prominent parameters affecting ET_p; R_s and T_{mean} in tropical region, 17 MLR models were developed correspond to each station used in this study. MLR models shows the accuracy between 86% to 93.8% when compares to FPM-56 (Figure 4). Further, the coefficients of MLR models were optimized using PSO algorithms in order to improvise the accuracy. The accuracy has improved ranging between 0.04% to 2.95% as tabulated

Table 2. Seven models were taken from PSO algorithm (w=1.0) followed by 5 models from w=0.9, 3 models from w=0.7928 and 2 from w=0.4. This has proved that the larger initial weight, the better performance of PSO as mentioned by [19, 21] stated that larger initial weight facilitate global search instead local exploration only.

Table 2. MLR and PSO Model for All Stations

No.	**Station**	**Model**		**Accuracy Improvement, %**
1	AS	MLR	$ET_p = 0.29T_{mean} + 0.107R_s - 5.824$	0.06
		$PSO_{w=0.7298}$	$ET_p = 0.095T_{mean} + 0.086R_s - 0.015$	
2	BE	MLR	$ET_p = 0.166T_{mean} + 0.133R_s - 3.294$	1.32
		$PSO_{w=1.0}$	$ET_p = 0.086T_{mean} + 0.133R_s - 1.026$	
3	BL	MLR	$ET_p = 0.285T_{mean} + 0.154R_s - 7.181$	0.04
		$PSO_{w=0.7928}$	$ET_p = 0.101T_{mean} + 0.111R_s - 1.310$	
4	CH	MLR	$ET_p = 0.126T_{mean} + 0.118R_s - 1.838$	2.95

Published by: The Mattingley Publishing Co., Inc.

		$PSO_{w=1.0}$	$ET_p = 0.147T_{mean} + 0.089R_s - 1.795$	
5	CHU	MLR	$ET_p = 0.356T_{mean} + 0.148R_s - 9.2$	0.08
		$PSO_{w=0.9}$	$ET_p = 0.122T_{mean} + 0.109R_s - 1.918$	
6	KB	MLR	$ET_p = 0.119T_{mean} + 0.152R_s - 2.478$	2.47
		$PSO_{w=0.9}$	$ET_p = 0.119T_{mean} + 0.116R_s - 1.775$	
7	KKrai	MLR	$ET_p = 0.154T_{mean} + 0.166R_s - 3.821$	2.02
		$PSO_{w=1.0}$	$ET_p = 0.085T_{mean} + 0.142R_s - 1.433$	
8	KLU	MLR	$ET_p = 0.117T_{mean} + 0.168R_s - 2.717$	1.94
		$PSO_{w=0.4}$	$ET_p = 0.069T_{mean} + 0.165R_s - 1.261$	
9	KT	MLR	$ET_p = 0.098T_{mean} + 0.159R_s - 2.024$	1.1
		$PSO_{w=0.9}$	$ET_p = 0.093T_{mean} + 0.120R_s - 1.149$	
10	KTN	MLR	$ET_p = 0.026T_{mean} + 0.163R_s - 0.059$	1.72
		$PSO_{w=1.0}$	$ET_p = 0.090T_{mean} + 0.141R_s - 1.426$	
11	MDZ	MLR	$ET_p = 0.117T_{mean} + 0.122R_s - 1.524$	1.96
		$PSO_{w=0.7928}$	$ET_p = 0.104T_{mean} + 0.116R_s - 1.096$	
12	MLK	MLR	$ET_p = 0.229T_{mean} + 0.15R_s - 5.499$	0.66
		$PSO_{w=1.0}$	$ET_p = 0.109T_{mean} + 0.145R_s - 1.979$	
13	MS	MLR	$ET_p = 0.11T_{mean} + 0.150R_s - 2.171$	1.03
		$PSO_{w=0.4}$	$ET_p = 0.066T_{mean} + 0.143R_s - 0.878$	
14	SBG	MLR	$ET_p = 0.192T_{mean} + 0.159R_s - 4.637$	1.09
		$PSO_{w=1.0}$	$ET_p = 0.071T_{mean} + 0.130R_s - 0.860$	

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

15	SN	MLR	$ET_p = 0.1T_{mean} + 0.168R_s - 2.197$	1.88
		$PSO_{w=1.0}$	$ET_p = 0.046T_{mean} + 0.169R_s - 0.641$	
16	STWN	MLR	$ET_p = 0.197T_{mean} + 0.162R_s - 4.906$	0.21
		$PSO_{w=0.9}$	$ET_p = 0.088T_{mean} + 0.133R_s - 1.409$	
17	TM	MLR	$ET_p = 0.169T_{mean} + 0.133R_s - 3.075$	0.69
		$PSO_{w=0.9}$	$ET_p = 0.134T_{mean} + 0.118R_s - 1.844$	
			$ET_p = 0.092T_{mean} + 0.146R_s - 1.152$	

Figure 3 shows the R^2 performance of MLR and 4 algorithms of PSO models. The R^2 is well-known performance indicator as its ability to evaluate the performance of dependent variables from its independent variables. A high correlation does not mean a good prediction but it measures a good precision instead. The precision is compared by FPM-56 and there is no significant difference shown yet the performance of R^2 for MLR models at all stations is better than PSO models. The percentage difference of R^2 performances of all models is ranging from 0.2-8.5%. Out of 17 stations, 12 stations with PSO models (w=0.9) shows better performance when compared to the other PSO models. This interpretation is similar to [19, 21] stated that the higher the initial weight helps in exploring search space globally. In other words, it gives better chance in exploring better value. Surprisingly, PSO algorithm models (w=1.0) shows poor performance among PSO algorithm as only 3 stations shows promising result.

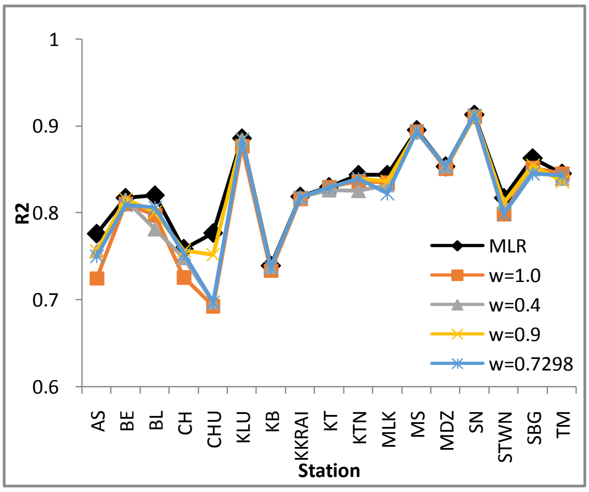

Figure 2. MLR and PSO R^2 Performance

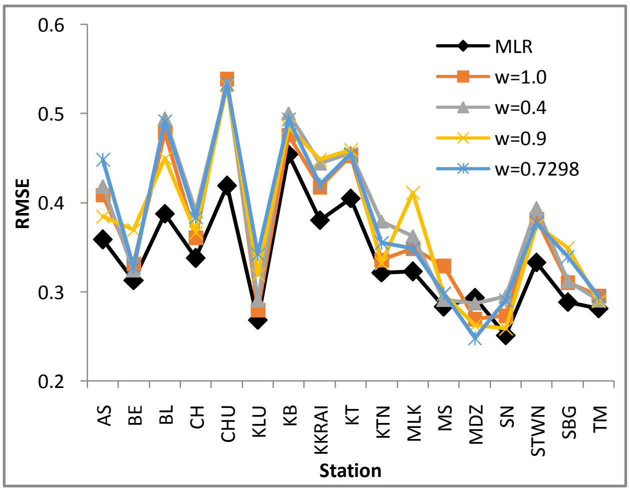

Figure 3. MLR and PSO RMSE Performance

TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

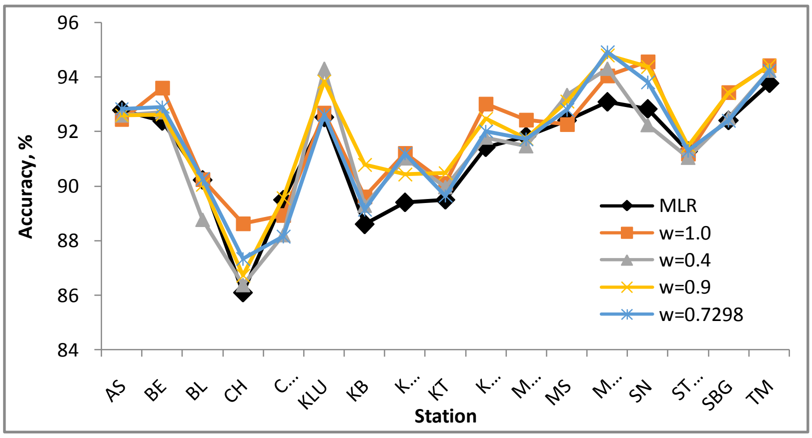

Figure 4. MLR and PSO Accuracy Performance

Figure 3 shows the RMSE performance for the MLR and PSO models. All performance fall within satisfactory value as the range different is between 1 to 9%. However from the Figure 3, the RMSE for MLR models at all station is lower than the all PSO algorithm except at Mdz station. The best RMSE for Mdz station is at PSO algorithm (w=0.7298) with percentage different of 4.6%. At this stage, based on the performance of MLR models, it is safe to conclude that the MLR models are adequate to estimate ET.

Figure 4 shows in-depth the accuracy performance for each PSO algorithm at each station. The algorithm with w=1.0 proves to be the best among the rest as the different accuracy percentage in 7 stations is higher with this algorithm. Followed by w=0.9, w=0.7298 and w=0.4. Algorithm of w=0.4 shows poor performance with only 2 stations showed higher different accuracy. It can be concluded that the performance of PSO is affected by initial weight. As the w decrease, it lowers the performance of PSO altogether.

4 CONCLUSION

This study evaluates the performance of ET_p estimation models both obtained by using MLR and optimized MLR using PSO. The models were developed by using 2 main parameters that have significant effect on ET_p; solar radiation and average temperature for Peninsular Malaysia region. Three statistical evaluators R^2, RMSE and Accuracy were applied to examine the performance. The results shows that all optimized MLR models improved up to 2.95% and the accuracy of MLR models itself were high as 94%. Standard PSO algorithm (c1=c2=2.0, w=1.0) stands out compared the other algorithm. MLR models from 7 stations shows better accuracy in ET_p estimation by using this algorithm which followed by w=0.9, w=0.7298 and w=0.4. In light of these results, it is safe to use the MLR model in estimating ET_p especially for data scarce region. Adaptation of PSO in optimized the models helps making the models performed better.

ACKNOWLEDGEMENT

The authors would like to express their gratitude towards Ministry of Education Malaysia for their financial support through FRGS/1/2018/ICT02/UTM/02/1 (R.J130000.7851.5F069) grant.

TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

REFERENCES

[1] Allen, R.G., et al., *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56.* FAO, Rome, 1998. **300**: p. 6541.

[2] Pereira, L.S., et al., *Crop Evapotranspiration Estimation With FAO56: Past And Future.* Agricultural Water Management, 2015. **147**(0): p. 4-20.

[3] Monteith, J., *Evaporation and environment: 19th Symposia of the Society for Experimental Biology, Cambridge, 1965.* University Press, Cambridge.

[4] Kisi, O., *Evapotranspiration modelling from climatic data using a neural computing technique.* Hydrological Processes: An International Journal, 2007. **21**(14): p. 1925-1934.

[5] Almedeij, J., *Modeling pan evaporation for Kuwait by multiple linear regression.* The Scientific World Journal, 2012. **2012**.

[6] Turc, L., *Evaluation des besoins en eau d'irrigation, évapotranspiration potentielle.* Ann. Agron., 1961. **12**: p. 13-49.

[7] Hargreaves, G.H. and Z.A. Samani, *Reference crop evapotranspiration from temperature.* Applied engineering in agriculture, 1985. **1**(2): p. 96-99.

[8] Priestley, C. and R. Taylor, *On the assessment of surface heat flux and evaporation using large-scale parameters.* Monthly weather review, 1972. **100**(2): p. 81-92.

[9] Makkink, G., *Testing the Penman formula by means of lysimeters.* Journal of the Institution of Water Engineerrs, 1957. **11**: p. 277-288.

[10] Itenfisu, D., et al., *Comparison of reference evapotranspiration calculations as part of the ASCE standardization effort.* Journal of Irrigation and Drainage Engineering, 2003. **129**(6): p. 440-448.

[11] Droogers, P. and R.G. Allen, *Estimating reference evapotranspiration under inaccurate data conditions.* Irrigation and drainage systems, 2002. **16**(1): p. 33-45.

[12] Pandey, V., P.K. Pandey, and A.P. Mahanta, *Calibration and Performance Verification of Hargreaves Samani Equation in a Humid Region.* Irrigation and Drainage, 2014. **63**(5): p. 659-667.

[13] Tabari, H., M.E. Grismer, and S. Trajkovic, *Comparative analysis of 31 reference evapotranspiration methods under humid conditions.* Irrigation Science, 2013. **31**(2): p. 107-117.

[14] Ahmad, N.F.A., et al., *Sensitivity Analysis of a FAO Penman Monteith for Potential Evapotranspiration to Climate Change.* Jurnal Teknologi, 2017. **79**(7).

[15] Sharma, S., et al., *Deriving spatially distributed precipitation data using the artificial neural network and multilinear regression models.* Journal of Hydrologic Engineering, 2012. **18**(2): p. 194-205.

[16] Eberhart, R. and J. Kennedy. *A New Optimizer Using Particle Swarm Theory*. in *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*. 1995. IEEE.

[17] Tapoglou, E., et al., *Groundwater-level forecasting under climate change scenarios using an artificial neural network trained with particle swarm optimization.* Hydrological Sciences Journal, 2014. **59**(6): p. 1225-1239.

[18] Jiang, Y., et al., *Application of particle swarm optimization to parameter calibration of Xin'anjiang model.* Eng. J. Wuhan Univ, 2006. **39**: p. 14-17.

[19] Gill, M.K., et al., *Multiobjective Particle Swarm Optimization For Parameter Estimation In Hydrology.* Water Resources Research, 2006. **42**(7).

[20] Shi, Y. and R. Eberhart. *A modified particle swarm optimizer*. in *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*. 1998. IEEE.

[21] Chaturvedi, K.T., M. Pandit, and L. Srivastava, *Particle swarm optimization with time varying acceleration coefficients*

November-December 2019
ISSN: 0193-4120 Page No. 858- 866

for non-convex economic power dispatch. International Journal of Electrical Power & Energy Systems, 2009. **31**(6): p. 249-257.
[22] Jiang, Y., et al., *An improved particle swarm optimization algorithm.* Applied Mathematics and Computation, 2007. **193**(1): p. 231-239.
[23] Poli, R., J. Kennedy, and T. Blackwell, *Particle swarm optimization.* Swarm intelligence, 2007. **1**(1): p. 33-57.