Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Estimation of Potential Evapotranspiration using Multiple Linear Regression and Particle Swarm Optimization

Nor Farah Atiqah Ahmad[1,*], Sobri Harun[2], Haza Nuzly Abdull Hamed[3], Suhaila Sahat[1], Siti Nooraiin Mohd Razali[1], Masiri Kaamin[1], Khairun Nissa Mat Said[4], Mohd Khairul Idlan Muhammad[2]

[1] Department of Civil Engineering, Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia, Pagoh Education Hub, 84600 Pagoh, Johor, Malaysia
[2] Department of Water and Environmental Engineering, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[3] Department of Applied Computing, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[4] Department of Geotechnic and Transportation, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

ARTICLE INFO

Article history:
Received 29 November 2024
Received in revised form 24 February 2025
Accepted 14 March 2025
Available online 28 March 2025

Keywords:
Potential evapotranspiration; multiple linear regression; particle swarm optimization

ABSTRACT

The study of estimating potential evapotranspiration (ET_p) in managing water resources for domestic, agricultural and irrigation purposes in Malaysia is considered lacking. Not only study on ET_p provides valuable insight on water cycle management, but also helps in predicting the water demand and optimize the agricultural practices. Hence, this study evaluates the performance of developed ET estimation model using MLR algorithm and optimized by PSO algorithm for five main region in Peninsular Malaysia using historical data from 1987 till 2003. The developed ET_p models (MLR-ET) were then optimized (MLR-PSO) using PSO algorithm. Results show that the performance of MLR-ET for R^2 ranging from 0.897 to 0.987 where MBE value is from 0.04 to 0.19, PE range is between 1.8 to 5.3 and RMSE is range between 0.20 to 0.30. Whereas for MLR-PSO models, performance (R^2) ranging from 0.942 to 0.993 MBE value was from 0.12 to 0.20, PE range is between 0.4 to 4.0 and RMSE ranged between 0.14 to 0.33. The performance of RMSE, MBE and PE indicators for MLR-ET were better than MLR-PSO yet the performance of R^2 for MLR-PSO models were better. Nonetheless, both MLR-ET and MLR-PSO models are reliable as the results of performance indicators lies within appropriate limits.

1. Introduction

The issue of water scarcity in Malaysia surprisingly has spike the attention not only by the government but also among hydrologists. Water scarcity leads to the water rationing and in 2014, Selangor faced the worst water rationing event that affected 300,000 households of the state [1]. Seven out of eight dams in Selangor has gone dry due to El Nino from December 2013 till February 2014. This was considered a wakeup call for researchers in all aspects; economics, water resources,

* *Corresponding author.*
E-mail address: norfarah@uthm.edu.my

https://doi.org/10.37934/ard.126.1.8190

environmental, ecologists and even in social and health field since 29 death cases were reported during the dry spell. In designing and managing water resources, knowledge on magnitude and variation of evapotranspiration (ET) is required. ET is one of the process in hydrological cycle is defined as the combination of water evaporates to the air and transpiration from plants. The accurate estimation of ET helps in irrigation design [2,3] and scheduling as well as domestic water resources management [4,5].

The importance of ET are constantly mentioned and various study related with ET can easily be found especially related to the application of ET estimation models. Although the Food and Agricultural Penman-Monteith 56 [6] (herein after FAO-PM) model has been decided as the most accurate ET estimation model, this model is not applicable at data scarce region like Malaysia. This situation can be cause the model to work improperly. As mentioned by [7] each of the estimation models developed for specific climatic regions and hence to ensure satisfactory result, few models needs to be testified before they can be used at own region.

For the purpose of this study, multiple linear regression (MLR) algorithm is chosen in order to develop ET estimation model as it is highly flexible and can provide an assessment performance of independent variable and dependent variable. In this case, independent variables comprised of meteorological parameters and dependent variable is daily ET. According to Aiken [8] a distinct trademark of MLR is ability to provide unique contribution of each individual predictor within the set also from another subset. Silva *et al.*, [9] has used the MLR algorithm to develop reference evapotranspiration at Rio Branco, Acre which classified as humid climate. Average air temperature, wind speed and solar radiation hour parameters were taken as MLR model independent variables and FAO-PM model value of ET was taken as dependent variable. Sriram *et al.*, [10] developed MLR models to estimates ET using tropical climate meteorological data in India. A straight forward study using ET values obtained from FAO-PM model as dependent variables and mean air temperature, minimum humidity, maximum humidity and wind speed parameters as independent variables, the developed MLR models shows an outstanding performance (R2 = 0.995). Perugu *et al.*, [11] developed MLR models to estimate the daily reference evapotranspiration at five regions of Andhra Pradesh, India. The models were developed using four meteorological parameters that influenced evapotranspiration at the region such air temperature (T), sunshine hour (S), relative humidity (RH) and wind speed (W). Through this study, not only authors agree that MLR model is are convenient to be used in estimating ET, it was found that sunshine parameter plays significant effect on ET followed by wind speed, temperature and relative humidity under humid region.

Particle swarm optimization (PSO) has been proposed by Eberhart *et al.*, [12] which is based on the theory of animal social behavior; bird flocking. This swarm based meta-heuristic optimization technique has been used in almost every field of study for more than 95,000 as according to SCOPUS on 2022. Compared to other evolutionary algorithms, not only PSO optimization covers both global best and local best experienced by the swarm within the solution space but also it is an easy recognition and high efficiency in solving optimization problems [13]. The specialty of PSO in developing solution for continuous variables and now with discrete variables makes it one of the favorable optimization method for hydrologists.

Technically PSO have five control parameters known as acceleration coefficient (C_1 and C_2), velocity clamping limit (V), swarm size (N), number of iteration (n) and inertia weight (ω). Since there are no specific optimal parameters defined, researchers take the safe approach using the values recommended by [14].

The main objectives of this study are; (1) to develop ET estimation model using MLR and evaluate its performance and (2) to optimize the MLR-ET model using PSO algorithm. As actual ET value is not measured, the FAO-PM model is used for ET estimation which taken as benchmark data and has been used to compare with the performance of MLR-ET and MLR-PSO model.

2. Methodology

Malaysia is dominated by the humid tropic climate according to Köppen-Geiger system and is influenced directly by the South-West (SW) monsoon from early June till late September, North-East (NE) monsoon from November to March and transition season from April to May and September to October. The SW monsoon brings wet spell for area that is not blocked by Sumatra Island such as the range of Titiwangsa [15]. Whereas, NE monsoon is expected to bring wet season in the east coast and northern areas of Peninsular Malaysia [16]. During transition periods, wind fluctuated and temperature rise especially during the afternoon. There are 43 meteorological stations in Malaysia where 28 stations located in Peninsular Malaysia, eight and seven stations are in Sarawak and Sabah respectively. Out of 28 stations in Peninsular Malaysia, 17 stations were selected in this study as shown in Figure 1. These stations were chosen based on the data availability and covers five main region for Peninsular Malaysia.

Daily historical meteorological data were obtained from the Malaysia Meteorological Department (MMD) from 1 January 1987 to 31 December 2003 comprised of maximum (T_{max}), minimum (T_{min}) and mean (T_{mean}) air temperature, wind speed (u), relative humidity (RH), solar radiation (R_s) and pan-evaporation (E_{pan}) parameters. However, due to many stations fail to measured E_{pan} this data is eliminated from this study. The obtained wind speed data from MMD was measured in a height at 10m above the ground which later transformed into 2m height wind speed following the calculation in [6] for FAO-PM model. As this study aim to developed estimation model and not prediction model, the historical data is valid to be used and evaluate the performance of developed model with the FAO-PM model.

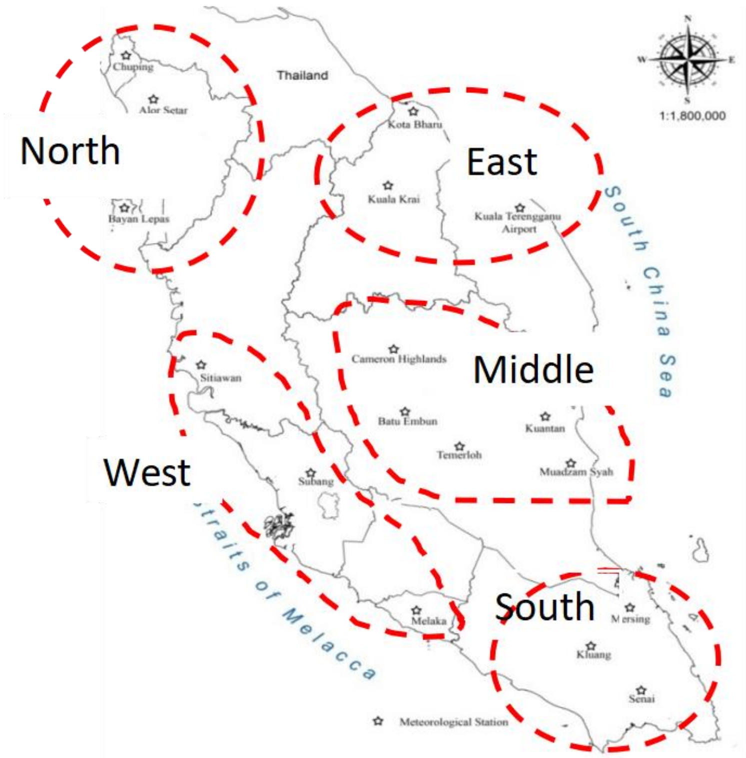

Fig. 1. Selected meteorological station in Peninsular Malaysia with its main region

2.1 FAO-PM Model

The estimated value from FAO-PM model is taken as benchmark values since observed ET was not measured in this study. The calculation to estimate the value of ET was followed the manual as in [6]. The model can be written as in Eq. (1):

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

Where, R_n is the net radiation (MJ/m^2/day), G is the soil heat flux (MJ/m^2/day), γ is the sychrometric 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), Ta 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 Multiple Linear Regression

There are many problems in hydrology that can be solved using statistical modeling especially using regression method. Regression method is performed to learn the relationship of two variables, to remove a portion of the variation in one variable for better understanding of another variable and to estimate the values of one variable based on knowledge of another variable [17]. Generally, regression method can be categorized either simple linear or multiple linear regression. 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.

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + c \tag{2}$$

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). Several assumptions need to be satisfied when dealing with MLR method as listed below;

Assumption 1: The dependent variable should be measured on continuous scale.
Assumption 2: Must have two or more independent variables which can be either continuous or categorical.
Assumption 3: Must have independence of observations where the Durbin-Watson value in the rage on 0 to 4. However the value should be close to 2 to show the strong independence.
Assumption 4: A linear relationship between the dependent variable and independent variable. If the relationship displayed is not linear then transforming the data is required.
Assumption 5: The data needs to be homoscedasticity; the variances along the line fit remain similar as moving along the line.
Assumption 6: The data must not show multicollinearity with each other. The correlation more than 0.8 may be problematic. This assumption can also be tested with variance inflation factor (VIF) and Tolerance value where it need to be below than 10 and above 0.2 respectively for the assumption to be met.

Assumption 7: There should be no significant outliers, high leverage points or highly influential points. This can be checked with Cook's value that must be below 1.0 to show there is no outliers.
Assumption 8: The residual errors are approximately normally distributed.

2.3 Particle Swarm Optimization

The governing equations applied in PSO algorithm are for updating the velocity and position for each particle as in Eq. (3) and (4) respectively. The rand() values are the vector between 0 and 1.

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

$$\overrightarrow{p_{new}} = \vec{p} + \overrightarrow{v_{new}} \tag{4}$$

Although there are five control parameters in PSO, only two parameters drives the algorithm; acceleration coefficient (C_1 and C_2), and inertia weight (ω). By taking into consideration the parameters suggested by Eberhart *et al.,* [12] the founder of PSO, three algorithms of C_1 and C_2 is set to 2.0 but with different inertia weight. However, contradict values of C_1, C_2 and ω as suggested by Poli *et al.,* [18] was adapted in this study to show the variation of PSO's algorithm and examined the result. Thus, four PSO control parameters were chosen and as tabulated in Table 1 whereas the other parameter was fixed; number of iteration = 1000, number of particle = 20, minimum and maximum clamping is 0.0 and 2.0 respectively. Figure 2 shows the summary of standard process in PSO algorithm.

Table 1
Adapted PSO Algorithm used in this study

Remark	C_1	C_2	w	Reference
PSO_1	2.0	2.0	1.0	[20]
PSO_2	2.0	2.0	0.9	[21]
PSO_3	2.0	2.0	0.4	[22]
PSO_4	1.49618	1.49618	0.7298	[18]

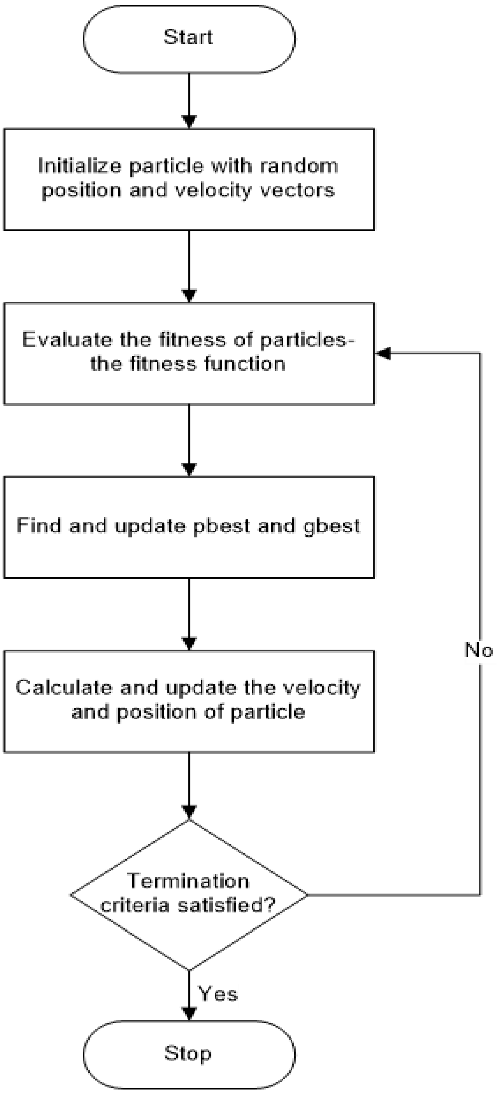

Fig. 2. Summary of PSO process [19]

2.4 Performance Indicator

By taking FAO-PM model as benchmark data, the performance for each ET empirical models developed by MLR were evaluated using coefficient of determination (R^2), root mean square error (RMSE), mean bias error (MBE) and percentage error (PE) as shown in Eq. (5) to (8) respectively.

$$R^2 = \left(\frac{\sum_{i=1}^{N} (O_1 - \bar{O})(P_1 - \bar{P})}{\left(\sum_{i=1}^{N} (O_1 - \bar{O})^2 \right)^{0.5} \left(\sum_{i=1}^{N} (P_1 - \bar{P})^2 \right)^{0.5}} \right) \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}} \tag{6}$$

$$MBE = \sum_{i=1}^{N} \frac{(\overline{P} - \overline{O})}{N} \quad (7)$$

$$PE = \left| \frac{\overline{P} - \overline{O}}{\overline{O}} \right| \times 100\% \quad (8)$$

3. Results

The study was conducted by two steps. First the development of MLR model for each region by taking the mean of T_{mean}, R_s as input and ET from FAO-PM model as output. Second, the MLR model was optimized by PSO algorithms as listed in table x. The R^2, RMSE, MBE and PE were used as performance indicator for both steps. The development of ET estimation model for this study only includes average air temperature and solar radiation variable as these two parameters have significant influence on ET in Peninsular Malaysia as obtained in sensitivity analysis done by [23].

3.1 Developed MLR Model

The developed MLR models for each station are presented in [24]. The result shows good performance of R^2 ranging from 0.744 to 0.996 which statistically significant at the 0.05 level. The R^2 value is obtained based on the comparison of ET estimated value developed by MLR model and FAO-PM model.

Table 2 shows the developed MLR-ET model for main region in Peninsular Malaysia correspond to its R2. The performance of R^2 ranging from 0.897 to 0.987 where MBE value is from 0.04 to 0.19, PE range is 1.8 to 5.3 and RMSE is range between 0.2 to 0.3. From the result, it shows that the developed MLR-ET for each region is reliable to estimate the ET values.

Table 2
Developed ET estimation model using MLR approach for each region

No.	Region	Model	R^2
1	North	$ET = 0.219T_{mean} + 0.2R_s - 5.629$	0.897
2	West	$ET = 0.111T_{mean} + 0.19R_s - 2.609$	0.985
3	East	$ET = 0.109T_{mean} + 0.181R_s - 2.398$	0.985
4	South	$ET = 0.056T_{mean} + 0.192R_s - 1.175$	0.987
5	Middle	$ET = 0.113T_{mean} + 0.203R_s - 2.603$	0.953

3.2 Particle Swarm Optimization Model

Table 3 shows the optimized MLR-PSO model for each region. In all region the value of R^2 is no less than 0.94 which means that the MLR-PSO estimation is close to the estimation ET by FAO-PM model. The MBE value is from 0.12 to 0.20, PE range is 0.4 to 4.0 and RMSE is range between 0.14 to 0.33. Each region has run by four PSO algorithm as mentioned in Table 2 and it gives different performance of algorithm. The PSO_3 shows better performance at North and Middle region whereas PSO_1 at East, West and South region. It shows that by set the value 2.0 for C_1 and C_2 parameters will have a better performance. A similar result done by Sudheer *et al*., [25] where author initially C_1 and C_2 varied from 0 to 10 which later set to 2.0 once the R value reached 0.93.

Table 3
Optimized MLR-ET model using PSO algorithm

No.	Region	Model	R^2
1	North	$ET_{p'} = 0.083T_{mean} + 0.187R_s - 1.427$	0.942
2	East	$ET_{p'} = 0.097T_{mean} + 0.111R_s - 0.793$	0.992
3	Middle	$ET_{p'} = 0.041T_{mean} + 0.217R_s - 0.835$	0.975
4	West	$ET_{p'} = 0.057T_{mean} + 0.179R_s - 1.053$	0.992
5	South	$ET_{p'} = 0.023T_{mean} + 0.221R_s - 0.709$	0.993

Figure 3 shows the comparison of performance for MLR-ET and MLR-PSO models based on R^2, RMSE, MBE and PE. Although it is aimed to obtain better performance from MLR-PSO model, the performance of RMSE, MBE and PE at North, Middle and South region shows that MLR-ET model is better. Nevertheless, the difference does not have major impact in overall performance of MLR-PSO model as the values still in the acceptable range especially for R^2 performance where MLR-PSO shows significant improvement.

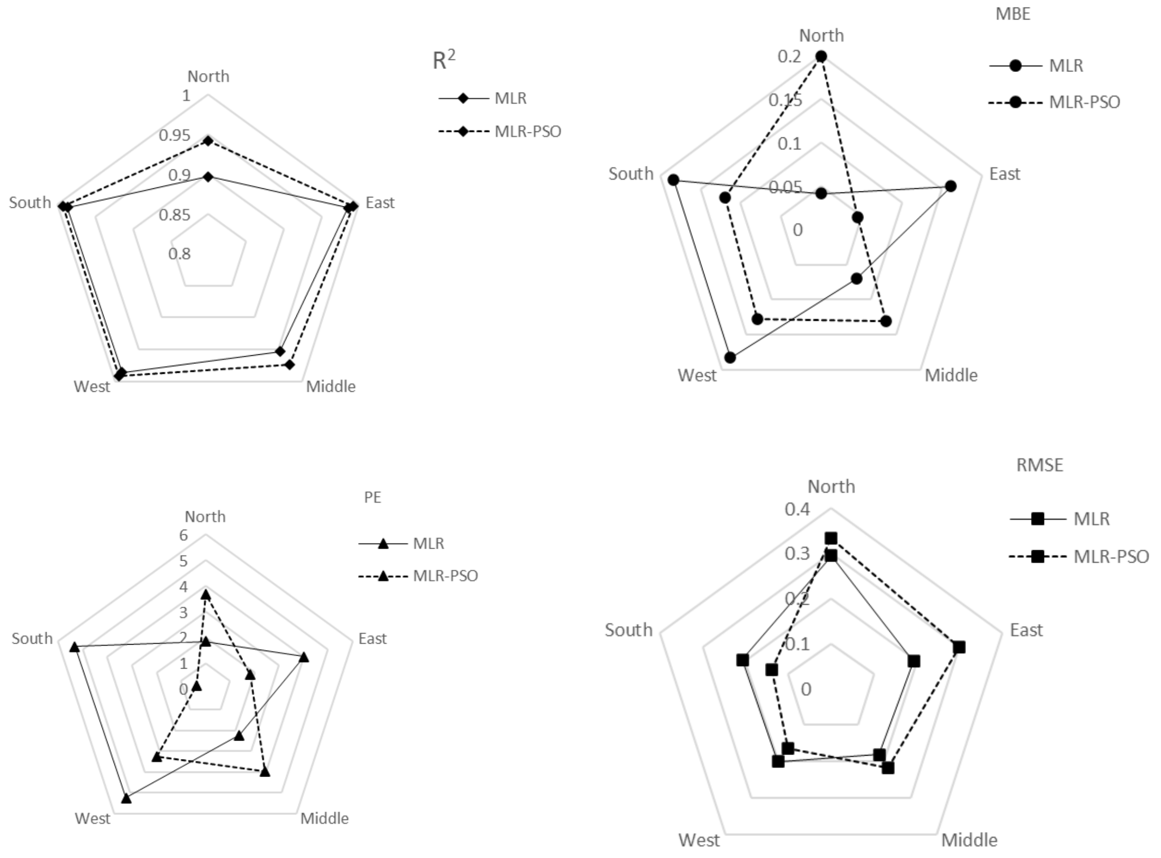

Fig. 3. The comparison of performance for MLR-Et and MLR-PSO model

4. Conclusions

This study evaluates the performance of developed ET estimation model using MLR algorithm and optimized by PSO algorithm for five main region in Peninsular Malaysia using historical data from 1987 till 2003. This model has been developed using two meteorological data that plays significant role in ET estimation; mean air temperature and solar radiation. Four statistical performance indicators were used to evaluate the performance for each model; R^2, RMSE, PE and MBE. In light of these results, it is reliable to use the MLR model in estimating ETp especially for data scarce region. The adaptation of PSO algorithm in optimizing the models, facilitates in increasing the reliability of developed MLR-ET models.

Acknowledgement
The authors would like to express their gratitude towards Universiti Tun Hussein Onn Malaysia for their financial support through TIER1 UTHM H974 grant.

References
[1] Lee Yee Mun, "300,000 more household go dry," *The Star*, 2014.
[2] Wang, Kaicun, and Robert E. Dickinson. "A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability." *Reviews of Geophysics* 50, no. 2 (2012). https://doi.org/10.1029/2011RG000373
[3] Tegos, A., N. Malamos, and D. Koutsoyiannis. "A parsimonious regional parametric evapotranspiration model based on a simplification of the Penman–Monteith formula." *Journal of Hydrology* 524 (2015): 708-717. https://doi.org/10.1016/j.jhydrol.2015.03.024
[4] Ortega-Farias, Samuel, Suat Irmak, and R. H. Cuenca. "Special issue on evapotranspiration measurement and modeling." *Irrigation Science* 28 (2009): 1-3. https://doi.org/10.1007/s00271-009-0184-x
[5] Jovic, Srdjan, Blagoje Nedeljkovic, Zoran Golubovic, and Nikola Kostic. "Evolutionary algorithm for reference evapotranspiration analysis." *Computers and electronics in agriculture* 150 (2018): 1-4. https://doi.org/10.1016/j.compag.2018.04.003
[6] Allen, Richard G., Luis S. Pereira, Dirk Raes, and Martin Smith. "Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56." *Fao, Rome* 300, no. 9 (1998): D05109.
[7] Paparrizos, Spyridon, Fotios Maris, and Andreas Matzarakis. "Sensitivity analysis and comparison of various potential evapotranspiration formulae for selected Greek areas with different climate conditions." *Theoretical and Applied Climatology* 128 (2017): 745-759. https://doi.org/10.1007/s00704-015-1728-z
[8] Aiken, Leona S., Stephen G. West, Steven C. Pitts, Amanda N. Baraldi, and Ingrid C. Wurpts. "Multiple linear regression." *Handbook of Psychology, Second Edition* 2 (2012). https://doi.org/10.1002/9781118133880.hop202018
[9] da Silva, Helder JF, Marconio S. dos Santos, Jório B. Cabral Junior, and Maria HC Spyrides. "Modeling of reference evapotranspiration by multiple linear regression." *Journal of Hyperspectral Remote Sensing* 6, no. 1 (2016): 44-58. https://doi.org/10.5935/2237-2202.20160005
[10] Sriram, A. V., and C. N. Rashmi. "Estimation of potential evapotranspiration by multiple linear regression method." *IOSR Journal of Mechanical and Civil Engineering* 11, no. 2 (2014): 65-70. https://doi.org/10.9790/1684-11246570
[11] Perugu, Mallikarjuna, Aruna Jyothy Singam, and Chandra Sekhar Reddy Kamasani. "Multiple linear correlation analysis of daily reference evapotranspiration." *Water resources management* 27 (2013): 1489-1500. https://doi.org/10.1007/s11269-012-0250-7
[12] Eberhart, Russell, and James Kennedy. "A new optimizer using particle swarm theory." In *MHS'95. Proceedings of the sixth international symposium on micro machine and human science*, pp. 39-43. Ieee, 1995. https://doi.org/10.1109/MHS.1995.494215
[13] Chen, Ke, Fengyu Zhou, Yugang Wang, and Lei Yin. "An ameliorated particle swarm optimizer for solving numerical optimization problems." *Applied Soft Computing* 73 (2018): 482-496. https://doi.org/10.1016/j.asoc.2018.09.007
[14] Shi, Yuhui, and Russell Eberhart. "A modified particle swarm optimizer." In *1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360)*, pp. 69-73. Ieee, 1998. https://doi.org/10.1109/ICEC.1998.699146

[15] Masseran, Nurulkamal, and Ahmad Mahir Razali. "Modeling the wind direction behaviors during the monsoon seasons in Peninsular Malaysia." *Renewable and Sustainable Energy Reviews* 56 (2016): 1419-1430. https://doi.org/10.1016/j.rser.2015.11.040
[16] Deni, Sayang Mohd, Jamaludin Suhaila, Wan Zawiah Wan Zin, and Abdul Aziz Jemain. "Spatial trends of dry spells over Peninsular Malaysia during monsoon seasons." *Theoretical and applied climatology* 99 (2010): 357-371. https://doi.org/10.1007/s00704-009-0147-4
[17] Helsel, Dennis R., and Robert M. Hirsch. *Statistical methods in water resources*. Vol. 49. Elsevier, 1993. [Online]. Available: http://water.usgs.gov/pubs/twri/twri4a3/
[18] Poli, Riccardo, James Kennedy, and Tim Blackwell. "Particle swarm optimization: An overview." *Swarm intelligence* 1 (2007): 33-57. https://doi.org/10.1007/s11721-007-0002-0
[19] Jahed Armaghani, Danial, Raja Shahrom Nizam Shah Bin Raja Shoib, Koohyar Faizi, and Ahmad Safuan A. Rashid. "Developing a hybrid PSO–ANN model for estimating the ultimate bearing capacity of rock-socketed piles." *Neural Computing and Applications* 28 (2017): 391-405. https://doi.org/10.1007/s00521-015-2072-z
[20] Zeugmann, Thomas, Pascal Poupart, James Kennedy, Xin Jin, Jiawei Han, Lorenza Saitta, Michele Sebag et al. "Particle swarm optimization." *Encyclopedia of machine learning* 1, no. 1 (2011): 760-766. https://doi.org/10.1007/978-0-387-30164-8_630
[21] Jiang, Yan, Tiesong Hu, ChongChao Huang, and Xianing Wu. "An improved particle swarm optimization algorithm." *Applied Mathematics and Computation* 193, no. 1 (2007): 231-239. https://doi.org/10.1016/j.amc.2007.03.047
[22] Baltar, Alexandre M., and Darrell G. Fontane. "A generalized multiobjective particle swarm optimization solver for spreadsheet models: application to water quality." (2006).
[23] Ahmad, Nor Farah Atiqah, Muhamad Askari, Sobri Harun, Abu Bakar Fadhil, and Amat Sairin Demun. "Sensitivity analysis of a FAO Penman Monteith for potential evapotranspiration to climate change." *Jurnal Teknologi (Sciences & Engineering)* 79, no. 7 (2017). https://doi.org/10.11113/jt.v79.8377
[24] N. F. A. Ahmad, S. Harun, and H. N. Abdull Hamed, "Particle Swarm Optimization of Multi-Linear Regression for Evapotranspiration Estimation Model," *TEST Engineering and Management*, (2019).
[25] Sudheer, Ch, R. Maheswaran, Bijaya K. Panigrahi, and Shashi Mathur. "A hybrid SVM-PSO model for forecasting monthly streamflow." *Neural Computing and Applications* 24 (2014): 1381-1389. https://doi.org/10.1007/s00521-013-1341-y