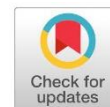


Research Article

Open Access



Estimating Reference Evapotranspiration for Shahat Region in Libya Using Genetic Programming

Osama A. Abdelatty^{*1}, Mohamed A. Momen²

***Corresponding author:** Osa-ma.abdelhamed@omu.edu.ly

Department of Soil and Water,
Faculty of Agriculture, Omar Al-Mukhtar University, Libya

² Department of Soil and Water,
Faculty of Agriculture, Omar Al-Mukhtar University, Libya

Received:

30 September 2023

Accepted:

18 March 2024

Publish online:

31 May 2024

Abstract: This study was conducted to estimate the reference evapotranspiration (ET_o) for Shahat region in Libya using the genetic programming (GP) model compared to the FAO Penman-Monteith equation (FPM56). The climatic data of Shahat Meteorological Station was used for the period from 1963 to 1999. Six different combinations of available meteorological variables were used, such as the average air temperature (T_{mean}), the average relative humidity (RH_{mean}), and the extraterrestrial radiation (R_a). The latter is calculated as a function of the location and time during the year. The GP model was trained using 70% of the climatic data and tested using the remaining 30%. The values of the statistical indicators obtained in this study showed that the root mean square error (RMSE), coefficient of determination (R^2), and Nash-Sutcliffe coefficient of efficiency (NSE) ranged between 0.26 and 0.98 ($mm.day^{-1}$); 0.67 and 0.98; 0.66 and 0.98, respectively during the testing period. Therefore, GP models represent a great option to estimate ET_o, when climatic data are scarce.

Keywords: Reference Evapotranspiration, Genetic Programming, FAO Penman-Monteith Equation, Shahat Region.

تقدير البخر نتح المرجعي لمنطقة شحات في ليبيا باستخدام البرمجة الجينية المستخلص: أجريت هذه الدراسة لتقدير البخر نتح المرجعي (ET_o) لمنطقة شحات في ليبيا باستخدام البرمجة الجينية (GP) مقارنة بمعادلة الفاو بنمان مونتيث (FPM56). حيث تم استخدام البيانات المناخية لمحطة أرصاد شحات للفترة من عام 1963 وحتى عام 1999. تم استخدام ستة تركيبات مختلفة من متغيرات الأرصاد الجوية المتاحة مثل متوسط درجة حرارة الهواء (T_{mean})، ومتوسط الرطوبة النسبية (RH_{mean})، والإشعاع الشمسي فوق الغلاف الجوي (R_a)، هذا الأخير يتم حسابه بمعرفة الموقع والتوقيت خلال السنة. تم تدريب نموذج (GP) باستخدام 70% من البيانات المناخية وتم اختباره باستخدام 30% المتبقية. بينت قيم المؤشرات الإحصائية التي تم الحصول عليها في هذه الدراسة أن الجذر التربيعي لمتوسط مربع الخطأ ومعامل التحديد ومعامل ناش-ساتكليف للكفاءة تراوحت بين 0.26 و 0.98 (ملم. يوم⁻¹)؛ 0.67 و 0.98؛ 0.66 و 0.98، على التوالي خلال فترة الاختبار. لذلك تمثل نماذج (GP) خياراً رائعاً لتقدير (ET_o) عندما تكون البيانات المناخية شحيحة.

الكلمات المفتاحية: البخر نتح المرجعي، البرمجة الجينية، معادلة الفاو بنمان مونتيث، معادلة الفاو بنمان مونتيث، منطقة شحات.



INTRODUCTION

Evapotranspiration (ET) is an essential hydrological component for the sustainable and efficient management of agricultural water resources, optimum irrigation scheduling, hydrologic water balance and water resources planning and management (Huang et al., 2019; Wu et al., 2019; Yamaç & Todorovic, 2020).

The so-called reference evapotranspiration, denoted as ETo . The reference surface is a hypothetical grass reference crop with an assumed crop height of 0.12 m, a fixed surface resistance of 70 s.m^{-1} and an albedo of 0.23. The reference surface closely resembles an extensive surface of green, well-watered grass of uniform height, actively growing and completely shading the ground (Allen et al., 1998).

Reference evapotranspiration (ETo) is measured by techniques and relatively complex physical principles (Allen et al., 2011), and the most direct and accurate way to estimate it is by water balance in the soil using lysimeters. However, due to limitations associated with the method, the adoption of physical mathematical models has become a practical alternative to ETo estimation.

The FAO-56 Penman-Monteith (FPM56) equation is recommended by the Food and Agriculture Organization (FAO) of the United Nations as the sole standard method for the definition and computation of the reference evapotranspiration (Allen et al., 1998). The FPM56 equation requires air temperature, air humidity, wind speed, and radiation data. These elements are often not available due to the small number of weather stations available in many regions, and when these are present, they may contain insufficient data.

There has been substantial research in recent years focusing on the estimation and prediction of natural phenomena, including the estimation of ETo using machine learning models, e.g., artificial neural network (ANN), fuzzy logic (FIS), genetic programming (GP), multivariate adaptive regression splines (MARS), decision tree (DT), random forests (RFs), support vector machine (SVM), extreme learning machine (ELM), and adaptive neuro-fuzzy inference system (ANFIS) (Adamala et al., 2019), (Adeloye et al., 2012) (Chia et al., 2020), (Egipto et al., 2023), (Spontoni et al., 2023), (Raza et al., 2023), (Liu et al., 2022).

A review of the literature shows that applications of GP for modeling evapotranspiration are limited. The study of (Güven et al., 2008) applied GP for modeling daily reference evapotranspiration as a function of solar radiation, mean air temperature, wind speed and relative humidity, and compared the performance of this model with other ETo equations. They found quite satisfactory results and it can be used as an alternative to the conventional models.

The current study is an attempt to develop a genetic programming model based on different combinations of available meteorological variables such as mean air temperature, relative humidity, and extraterrestrial radiation for predicting the ETo at Shahat, Libya.

MATERIALS AND METHODS

The reference evapotranspiration (ETo) in this study was calculated using data from Shahat meteorological station located at the longitude of $21^{\circ} 51'E$, the latitude of $32^{\circ} 49'N$, and mean altitude is 621 meters above sea level. The historical data series includes average monthly maximum (T_{max}), minimum (T_{min}) and mean air temperature (T_{mean}) ($^{\circ}C$), mean relative humidity (RH_{mean}) (%), and

wind speed (U_2) (m.s^{-1}), which covered the period from 1963 to 1999. Table (1) shows the statistical parameters of meteorological variables at Shahat weather station.

Table: (1). Statistical parameters of meteorological variables at Shahat weather station

Parameter	$T_{\max}$ (°C)	$T_{\min}$ (°C)	T_{mean} (°C)	RH_{mean} (%)	U_2 (m.s^{-1})	Sun (hr)
Mean	20.9	12.3	16.6	67.8	4.7	8.0
Standard Error	0.3	0.2	0.2	0.4	0.1	0.1
Standard Deviation	5.9	4.7	5.2	9.3	1.6	2.5
Range	22.1	17.9	19.8	50.0	8.0	11.1
Maximum	31.1	21.2	25.9	89.0	10.0	13.0
Minimum	9.0	3.3	6.2	39.0	2.1	1.9
Count	444	444	444	444	444	444

The REF-ET version 4.1 program (Allen, 2000) was used to calculate the reference evapotranspiration ET_o using the Penman-Monteith equation recommended by the Food and Agriculture Organization (FAO) in Bulletin 56. (Allen et al., 1998). This Equation takes the form:

$$ET_o = \frac{\left[0.408 \times \Delta(R_n - G) + \gamma \left(\frac{900}{T + 273} U_2 (e_s - e_a) \right) \right]}{\Delta + \gamma(1 + 0.34 U_{02})} \quad (1)$$

Where:

ET_o : is the reference evapotranspiration [mm.day^{-1}];

R_n : is the net radiation at the crop surface [$\text{MJ m}^{-2} \text{day}^{-1}$];

G : is the soil heat flux density [$\text{MJ m}^{-2} \text{day}^{-1}$];

T : is the mean daily air temperature at 2 m height [°C];

U_2 : is the wind speed at 2 m height [m.s^{-1}];

e_s : is the saturation vapour pressure [kPa];

e_a : is the actual vapour pressure [kPa];

$e_s - e_a$: is the saturation vapour pressure deficit [kPa];

Δ : is the slope vapour pressure curve [$\text{kPa.}^\circ\text{C}^{-1}$]; and

γ : is the psychrometric constant [$\text{kPa.}^\circ\text{C}^{-1}$]

The extraterrestrial solar radiation (R_a) is not measured data but estimated for a certain day and location. One of the outputs of the REF-ET model version 4.1 is extraterrestrial radiation (Allen, 2000). The extraterrestrial radiation, for each day of the year and different latitudes can be estimated from the solar constant, the solar declination, and the time of the year by:

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r [\omega_s \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \sin(\omega_s)] \quad (2)$$

Where:

R_a : Extraterrestrial radiation [$\text{MJ m}^{-2} \text{day}^{-1}$],

G_{sc} : Solar constant = $0.0820 \text{ MJ m}^{-2} \text{min}^{-1}$,

d_r : Inverse relative Earth-Sun,

ω_s : Sunset hour angle [rad],

φ : Latitude [rad],

δ : Solar declination [rad],

In the northern hemisphere, the latitude, φ , is expressed as a positive value in radians, while in the southern hemisphere, it is expressed as a negative value. The remaining variables in equation (2) can be calculated using the method outlined by (Allen et al., 1998) as follows:

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi J}{365}\right) \quad (3)$$

$$\delta = 0.409 \sin\left(\frac{2\pi J}{365} - 1.39\right) \quad (4)$$

$$\omega_s = \arccos[-\tan(\varphi) \tan(\delta)] \quad (5)$$

where J represents the day number in the year, ranging from 1 (1st January) to 365 or 366 (31st December).

Genetic programming

Genetic programming (GP) is a type of evolutionary algorithm (EA) that was introduced by (Koza, 1992). It is based on the principles of natural selection and genetics. GP is a relatively recent addition to the family of EAs, which includes evolutionary programming (Fogel et al., 1966), genetic algorithms (Holland, 1975), and evolution strategies (Schwefel, 1981).

Genetic symbolic regression operates with two sets of variables, namely the functional set and the terminal set (Koza, 1994).

In this study, Genetic programming models were calculated based on the steps mentioned in a field guide to Genetic Programming (Poli et al., 2008), using the open-source program GPdotnet5.1.2 developed by (Hrnjica, 2018).

The steps followed can also be summarized as follows:

1. Determine the external terminals, which are the independent variables, such as (T_{mean} , RH_{mean} , R_a), and the internal terminals, represented by the functions (addition, subtraction, multiplication, division).
2. Determine the fitness function through which strong solutions are selected and weak solutions are excluded.
3. Determine the parameters used in the analysis, such as population size, crossing over, mutation, reproduction, number of constants, and starting method, as shown in Table (2).
4. Determine the stopping point of the program that was achieved after the program reached 500 generations, where the best fitness has not changed more since generation 300 and the execution procedures can be summarized as follows:
 1. Randomly initiate populations.
 2. Evaluate the fitness of the population
 3. Iterate until the solution convergence:
 - a. Choose parents from the population:
 - b. Generate a new population through crossover.
 - c. Apply mutation to the new population.
 - d. Compute the fitness of the new population.

The function set consisted of addition, subtraction, multiplication, and division. The data was divided into two parts: 70% of the data for training and 30% for testing. The data was also normalized

using the Min-Max Normalization method. The statistical parameters of the climatic variables used in this study are shown in Table (3), and the used models for several scenarios are shown in Table (4).

Table: (2). Genetic programming parameters:

GP parameter	Value
Population	500
Size	Root Mean Square Error (RMSE)
Fitness	Half and Half
Initialization	20
Selection	Rank selection
Method	0.9
Probability of GP operations	0.05
Crossover	0.2
Mutation	0-1
Reproduction	5
Random constant	
Interval	
Number of random constants	

Table: (3). Statistical parameters of the climatic variables

Statistical parameters	Climatic Variables		
	T_{mean} (°C)	RH_{mean} (%)	R_a (MJ m ⁻² day ⁻¹)
Training processes			
Maximum	25.2	89	41.46
Minimum	6.2	39	17.99
Mean	16.44	68.05	30.66
Standard Deviation	5.11	9.47	8.38
Count	311	311	311
Testing processes			
Maximum	25.9	83	41.46
Minimum	7.7	43	17.99
Mean	16.99	67.27	30.53
Standard Deviation	5.57	8.93	8.45
Count	133	133	133

Table: (4). GP models scenarios

Model	Input variables		
	T_{mean}	RH_{mean}	R_a
GP1	✓		
GP2	✓		✓
GP3		✓	
GP4		✓	✓
GP5	✓	✓	
GP6	✓	✓	✓

Performance criteria

Three performance indicators were used to evaluate the model: root mean square error (RMSE), coefficient of determination (R^2) (Kennedy & Neville, 1986), and Nash-Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970), between ET_o using FPM56 and predicted values using GP model. These statistics parameters are defined as follow:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (ET_{GP} - ET_{FPM56})^2}{n}} \quad (6)$$

$$R^2 = \frac{[\sum_{i=1}^n (ET_{GP} - \overline{ET}_{GP})(ET_{FPM56} - \overline{ET}_{FPM56})]^2}{\sum_{i=1}^n (ET_{GP} - \overline{ET}_{GP})^2 \sum_{i=1}^n (ET_{FPM56} - \overline{ET}_{FPM56})^2} \quad (7)$$

$$NSE = 1 - \frac{[\sum_{i=1}^n (ET_{FPM56} - ET_{GP})]^2}{\sum_{i=1}^n (ET_{FPM56} - \overline{ET}_{FPM56})^2} \quad (8)$$

Smaller values of RMSE and higher values of R^2 indicates higher model performance. The Nash-Sutcliffe (NSE) efficiency is used to evaluate the predictive power of the model and varies from $-\infty$ to 1, with 1 being the perfect fit between the data estimated by the model and the measured data.

Where:

ET_{FPM56} : FPM56 , (mm.day⁻¹),

ET_{GP} : Predicted evapotranspiration, (mm.day⁻¹),

$\overline{ET}_{FPM56}$: Average FPM56, (mm.day⁻¹),

$\overline{ET}_{GP}$: Average predicted evapotranspiration, (mm.day⁻¹),

n : Total number of samples.

RESULTS AND DISCUSSION:

Table (1) demonstrates the values of the statistical criteria used in this study. From these results, it can be clearly seen that when using only (T_{mean}) as input to the genetic programming model, which is referred to here as GP1, the values of RMSE and R^2 were equal to 0.77 and 0.81 respectively. Figure (3) illustrates the scatter plot of predicted ETo values by the GP1 model, compared with FPM56 during testing. When added (R_a) to GP1, which is known here as GP2, it significantly increased the performance. The RMSE decreased from 0.77 to 0.46, by 40% and R^2 increased from 0.81 to 0.93, by approximately 15%. Figure (4) illustrates the scatter plot of this relationship.

In GP3, only (RH_{mean}) was used, we notice that the model performs poorly, where RMSE increased to 0.98, and R^2 decreased to 0.67. Figure (5) illustrates the scatter plot of this relationship. GP4 added (R_a) and performed better than GP3. The performance of this model is almost equal to the performance of GP2. Figure (6) illustrates the scatter plot of this relationship.

The results improved significantly when using (T_{mean}), (RH_{mean}) and (R_a) as inputs to GP6. Figure (8) shows the scatter plot of predicted ETo values by the GP6 model, compared with FPM56 during testing. Furthermore, it can be seen from Table (5) that GP6 outperformed the other models in all performance parameters. GP6 was ranked best in the testing process. These results are in accordance with (Liu et al., 2022), (Egipto et al., 2023; Raza et al., 2023) who also indicated that machine learning models represent a great option to estimate ETo.

Table: (5). Performance criteria of the GP models during training and Testing

Model	Input variables	Training			Testing		
		RMSE (mm.day ⁻¹)	R^2	NSE	RMSE (mm.day ⁻¹)	R^2	NSE
GP1	T_{mean}	0.84	0.75	0.75	0.77	0.81	0.79
GP2	T_{mean} , R_a	0.53	0.90	0.90	0.46	0.93	0.93
GP3	RH_{mean}	0.93	0.69	0.69	0.98	0.67	0.66
GP4	RH_{mean} , R_a	0.51	0.91	0.91	0.55	0.90	0.89
GP5	T_{mean} , RH_{mean}	0.43	0.94	0.93	0.49	0.93	0.92
GP6	T_{mean} , RH_{mean} , R_a	0.37	0.96	0.95	0.26	0.98	0.98

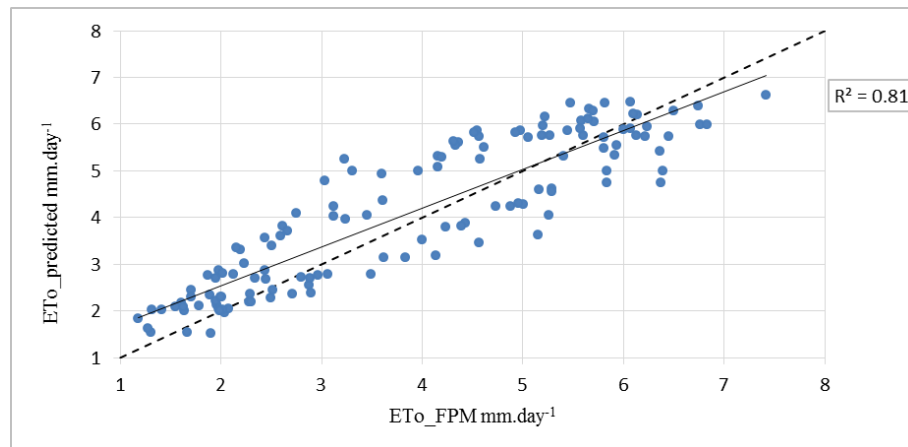


Figure: (3). scatter plot of predicted ETo values by the GP1 model, compared with FPM56 during testing.

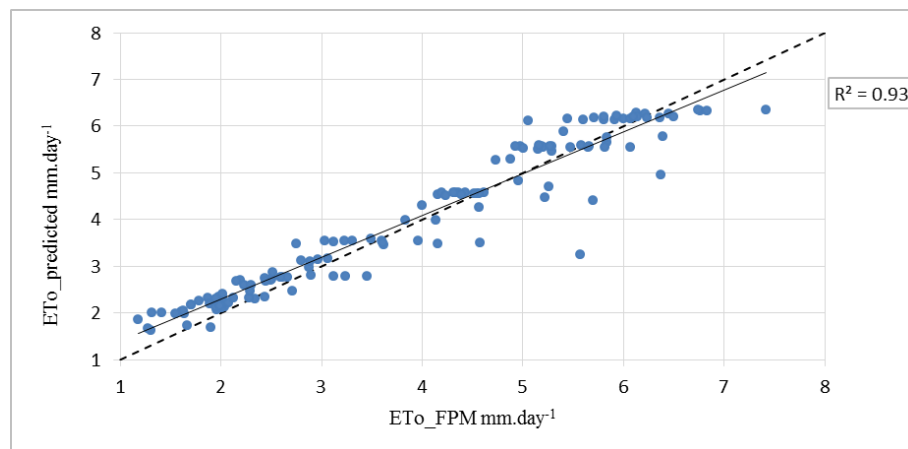


Figure: (4). scatter plot of predicted ETo values by the GP2 model, compared with FPM56 during testing.

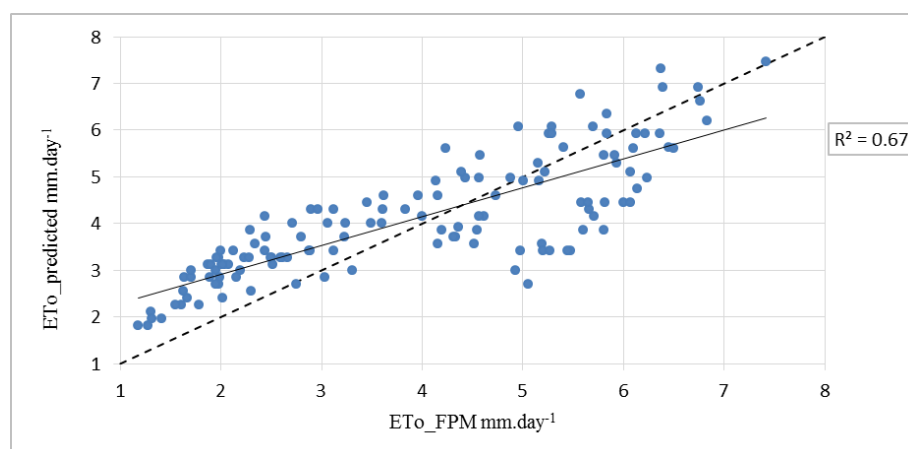


Figure (5). scatter plot of predicted ETo values by the GP3 model, compared with FPM56 during testing.

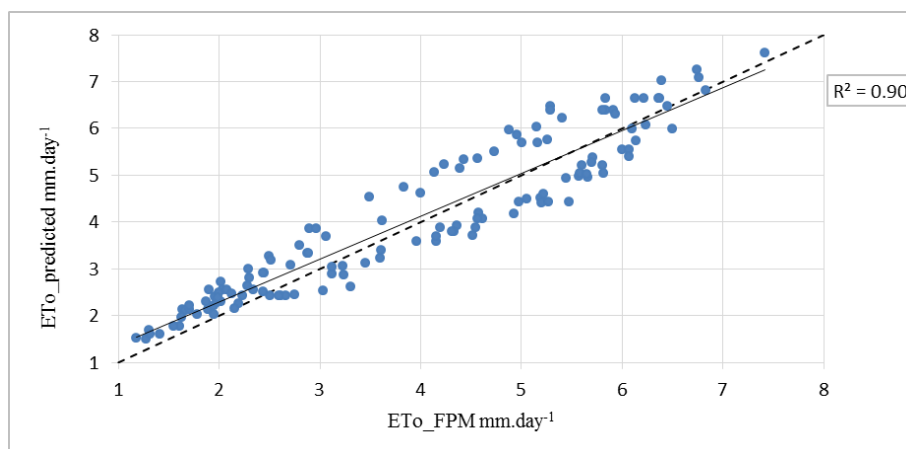


Figure: (6). scatter plot of predicted ETo values by the GP4 model, compared with FPM56 during testing.

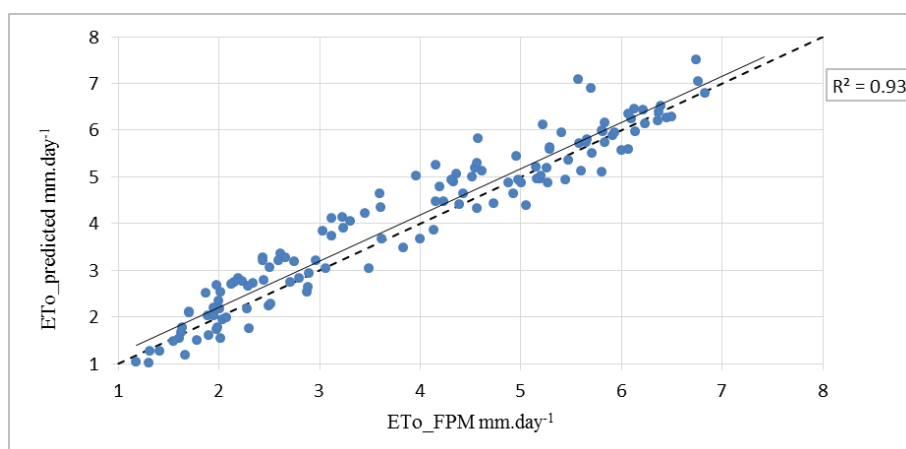


Figure: (7). scatter plot of predicted ETo values by the GP5 model, compared with FPM56 during testing.

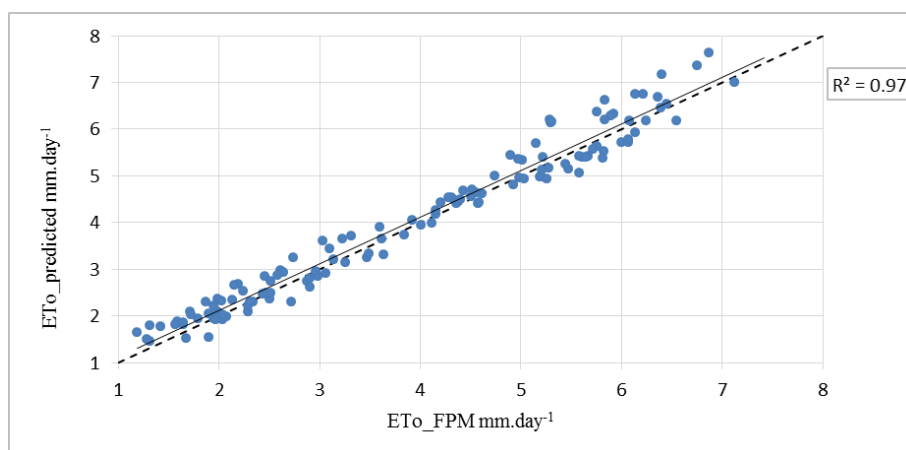


Figure: (8). scatter plot of predicted ETo values by the GP6 model, compared with FPM56 during testing.

CONCLUSION

From the results obtained in this study, It can be concluded that the performance of the GP model is a promising approach and a powerful tool that can be used to calculate reference evapotranspiration

when using (T_{mean}), (RH_{mean}) and (R_a) as inputs, especially under the deficiency of complete meteorological data required for the Penman-Monteith equation recommended by Food and Agriculture Organization to calculate the reference evapotranspiration.

Duality of interest: The authors declare that they have no duality of interest associated with this manuscript.

Author contributions: All Authors contributed equally to this manuscript.

Funding: No specific funding was received for this work,

REFERENCES

- Adamala, S., Raghuwanshi, N., Mishra, A., & Singh, R. (2019). Generalized wavelet neural networks for evapotranspiration modeling in India. *ISH Journal of Hydraulic Engineering*, 25(2), 119-131 .
- Adeloye, A. J., Rustum, R., & Kariyama, I. D. (2012). Neural computing modeling of the reference crop evapotranspiration. *Environmental modelling & software*, 29(1), 61-73 .
- Allen, R. G. (2000). REF-ET: Reference Evapotranspiration calculation software for FAO and ASCE standardized equation Version 3.1.16 for Windows. University of Idaho Research and Extension Center, Kimberly, Idaho
- Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management* , 920-899 ,(6)98
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper no. 56, Rome, Italy.
- Chia, M. Y., Huang, Y. F., Koo, C. H., & Fung ,K. F. (2020). Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agronomy*, 10(1), 101 .
- Egipto, R., Aquino, A., Costa, J. M., & Andújar, J. M. (2023). Predicting Crop Evapotranspiration under Non-Standard Conditions Using Machine Learning Algorithms, a Case Study for Vitis vinifera L. cv Tempranillo. *Agronomy*, 13(10), 2463 .
- Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). Intelligent decision making through a simulation of evolution. *Behavioral science*, 11(4), 253-272 .
- Guven, A., Aytekin, A., Yuce, M. I., & Aksoy, H. (2008). Genetic programming - based empirical model for daily reference evapotranspiration estimation. *Clean -Soil, Air, Water*, 36(10 - 11), 905-912 .
- Holland, J. (1975). Adaption in Natural and Artificial Systems. *Ann Arbor: University of Michigan, Press*, 13 .

- Hrnjica, B. (2018). *GPdotNET Tree based genetic programming tool, open source project*.
<http://github.com/bhrnjica/gpdotnet>
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., . . . Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, 574, 1029-1041 .
- Kennedy, J. B., & Neville, A. M. (1986). *Basic Statistical Methods for Engineers and Scientists* (3rd ed.) .
- Koza, J. R. (1992). On the programming of computers by means of natural selection. In *Genetic programming* .
- Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, 4, 87-112 .
- Liu, Q., Wu, Z., Cui, N., Zhang, W., Wang, Y., Hu, X., . . . Zheng, S. (2022). Genetic algorithm-optimized extreme learning machine model for estimating daily reference evapotranspiration in Southwest China. *Atmosphere*, 13 .971 (6)
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282-290 .
- Poli, R., Langdon, W., & McPhee, N. (2008). A field guide to genetic programming (Riccardo Poli, William B. Langdon, and Nicholas Freitag McPhee .
- Raza, A., Fahmeed, R., Syed, N. R., Katipoğlu, O. M., Zubair, M., Alshehri, F., & Elbeltagi, A. (2023). Performance Evaluation of Five Machine Learning Algorithms for Estimating Reference Evapotranspiration in an Arid Climate. *Water*, 15(21), 3822 .
- Schwefel, H.-P. (1981). *Numerical optimization of computer models*. John Wiley & Sons, Inc .
- Spontoni, T. A., Ventura, T. M., Palácios, R. S., Curado, L. F., Fernandes, W. A., Capistrano, V. B ., . Rodrigues, T. R. (2023). Evaluation and modelling of reference evapotranspiration using different machine learning techniques for a brazilian tropical savanna. *Agronomy*, 13(8), 2056 .
- Wu, L., Zhou, H., Ma, X., Fan, J., & Zhang, F. (2019). Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: Application in contrasting climates of China. *Journal of Hydrology*, 577, 123960 .
- Yamaç, S. S., & Todorovic, M. (2020). Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agricultural Water Management*, 228, 105875 .