

Reference Evapotranspiration Modeling Using Heuristic Computing Model in Distinct Climate Stations

Ala Tahsin¹ & Jazuli Abdullahi²

¹Department of Civil Engineering, Faculty of Engineering, Tishk International University, Iraq

²Department of Civil Engineering, Faculty of Civil and Environmental Engineering, Near East University, Nicosia, Cyprus

Correspondence: Jazuli Abdullahi, Near East University, Nicosia, Cyprus.

Email: jazulibinabdallah@gmail.com

Doi: 10.23918/eajse.v6i1p89

Abstract: Reference evapotranspiration (ET_0) plays important roles in environmental, hydrological and agricultural studies and its accurate prediction is significant in water resources management and water productivity increase. This study focused on evaluating the ability of support vector regression (SVR) model for modeling ET_0 in arid and semiarid climate stations of Iraq. For comparison, multiple linear regression (MLR) and calibrated Hargreaves and Samani (HS) empirical models were also applied. Daily meteorological data from Basra and Erbil stations including minimum, maximum and mean temperatures, relative humidity, wind speed, precipitation, solar radiation and surface pressure were collected for two consecutive years (2017 – 2018) and used as inputs to the models. FAO 56 Penman-Monteith was used as the benchmark ET_0 . Root mean square error (RMSE) and Nash Sutcliffe efficiency criterion (NSE) were the performance evaluation criteria employed. The results revealed that, all the applied models led to reliable results, but SVR model provided the best performance with NSEs of 0.9949, 0.9871 and RMSEs of 0.0009, 0.0016 in the validation phase for Basra and Erbil stations, respectively. The general results implied that SVR model could be employed successfully for estimation of ET_0 in arid and semiarid climate stations of Iraq.

Keywords: Support Vector Regression, Penman-Monteith, Semiarid, Station, Climate

1. Introduction

Evapotranspiration (ET) is among the primary and essential components of hydrologic cycle. However, estimation of ET with acceptable accuracy is a crucial factor to consider in the fields of irrigation scheduling, water resources management, water balance studies, etc. Moreover, ET is considered equivalent to crop water requirement as an index in plant growth process. In practice, since to determine ET is difficult for each crop, firstly, reference evapotranspiration (ET_0) is calculated and thereafter using the ET_0 , crop evapotranspiration is calculated (Mehdizadeh, 2018).

The only direct method for ET measurement is by lysimeter. Nonetheless, maintenance and insulation costs make its use difficult. Based on meteorological data, many empirical methods have been developed for the estimation of ET_0 (Mehdizadeh, 2018) such as Hargreaves and Samani (1985) (HS) method. For example, Djaman et al. (2019) evaluated the potentials of HS and 34 other equations for ET_0 modeling in a semiarid dry climate under limited data. Allen et al. (1998) recommended the use of FAO Penman-Moneith (FAO-56-PM) as the sole method for estimating ET_0 .

Despite being reliable tools for evaluating the precise physics of a phenomenon, conceptual and physical based models have limitations in practice. However, when physical understanding is of less priority to accurate predictions, utilizing black box models can be more successful. Multiple linear regression (MLR) is a classical method that determines relationship linearly between dependent and

one or multiple independent variables (Nourani et al., 2019a). Toward modeling processes, such kinds of linear models lose their merit in many fields that are subjected to nonlinearity, dynamism and high complexity in both temporal and spatial scales.

Recently, applications of artificial intelligence (AI) models such as support vector machine (SVR) are widely implemented which yielded several publications. Ferreira et al. (2019) proposed a new approach for estimation of ET_0 with limited climate data using artificial neural network (ANN) and support vector machine (SVM). Granata (2019) performed a comparative evaluation of SVR and other 3 machine learning methods for ET modeling in humid subtropical climate. Nourani et al. (2019a) employed SVR, HS, MLR and other AI and empirical models for ET_0 modeling in several climatic regions. A thorough evaluation of the current literature indicates that there was no study conducted that utilized the application of AI, regression and empirical models for ET_0 modeling in Iraq. Hence, the aim of this study was to investigate the abilities of SVR, HS and MLR models for the estimation of ET_0 in Basrah and Erbil stations of Iraq. This was done first by performing sensitivity analysis to determine the appropriate input parameters. Then SVR, MLR and HS models were trained and validated for 5 different input combinations, and finally, compared their performances.

2. Materials and Methods

2.1 Study Area and the Used Data

Iraq is historically called Mesopotamia and is situated in western Asia at $29^{\circ}5' - 37^{\circ}22'N$ latitudes and $38^{\circ}45' - 48^{\circ}45'E$ longitudes. Iraq's climate is considered to be cold in winter, dry and hot in summer due to its location (Nourani et al., 2019b). Study by Sarlak and Agha (2018) indicates that arid regions (e.g. Baghdad, Basra, and Rutbah) and semiarid regions (e.g. Erbil) covered around 97% of Iraq's climate. Basra has a monsoon climate and is located in a desert-type environmental zone. It constitutes very hot summers more especially between the months of July and August with maximum and mean temperatures around $45^{\circ}C$ and $37.4^{\circ}C$, respectively. The potential evapotranspiration of Basra province exceeds 2,450 mm/year on average, while average rainfall is below 100 mm/year annually (Jabbar & Zhou, 2013). Erbil is located in northern Iraq and is considered Kurdistan's capital city. It experiences cool and rainy winters, warm and dry summers (Rasul et al., 2015). Figure 1 presents map of Iraq and study locations.

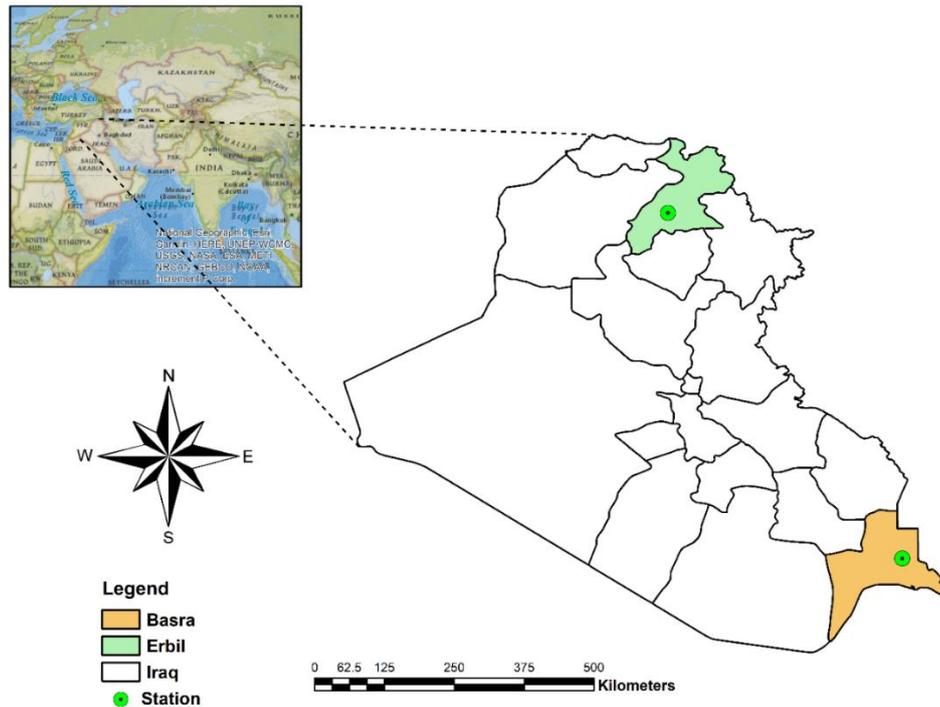


Figure 1: Regions location and study stations

Daily meteorological data including minimum, maximum and mean temperatures (T_{\min} , T_{\max} , T_{mean}), precipitation (P), surface pressure (P_s), relative humidity (R_H), wind speed (U_2) and solar radiation (R_s) were collected from January 2017 – December 2018 (730 number of observations) and used for the ET_0 modeling. As HS method requires the use of extraterrestrial radiation (R_a) as input in addition to T_{\min} , T_{\max} and T_{mean} , R_a was calculated according to Allen et al. (1998) as.

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r [\omega_s \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \sin(\omega_s)] \quad [1]$$

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

$$\delta = 0.409 \sin\left(\frac{2\pi J}{365} - 1.39\right) \quad [3]$$

$$\omega_s = \arccos[-\tan(\varphi) \tan(\delta)] \quad [4]$$

Where R_a has a unit of $\text{MJ}/\text{m}^2/\text{day}$, G_{sc} is solar constant given as $0.0820 \text{ MJ}/\text{m}^2/\text{min}$, d_r is Earth-Sun inverse relative distance, ω_s is sunset hour angle (rad), φ represents latitude (rad), and δ is declination angle (rad), J ranges between 1 and 365 or 366 (1 January – 31 December) and it signifies the number of the day in the year.

The data were divided into 75% for training and 25% for validation. The statistical description of the used data is given in Table 1.

Table 1: Used data descriptive statistics

Station	Parameter	unit	Minimum	Maximum	Mean	St. deviation
Basra	P	mm/day	0.00	21.23	0.35	1.64
	R _H	%	7.79	89.59	29.72	18.46
	T _{max}	°C	9.94	52.07	35.10	10.78
	T _{min}	°C	-2.22	34.93	20.34	9.36
	T _{mean}	°C	4.60	42.89	27.29	10.29
	U ₂	m/s	0.89	8.62	3.28	1.34
	R _S	MJ/m ² /day	1.91	29.09	19.55	6.36
Erbil	P _S	kpa	99.17	102.69	100.86	0.84
	P	mm/day	0.00	28.30	1.11	3.20
	R _H	%	7.55	94.11	42.93	23.36
	T _{max}	°C	1.47	46.78	26.18	11.78
	T _{min}	°C	-6.58	29.41	13.21	9.22
	T _{mean}	°C	-2.21	37.06	19.22	10.60
	U ₂	m/s	0.90	5.18	1.93	0.58
R _S	MJ/m ² /day	0.68	30.71	17.55	8.52	
P _S	kpa	91.99	94.54	93.22	0.56	

As seen in Table 1, temperature is high in both stations. This indicates the vulnerability of the stations to arid and semiarid climates and as temperature increases with increase in aridity index, Basra (arid station) has higher T_{max} (up to 52.07 °C) than Erbil (semiarid station, 46.78 °C). However, many of the features that distinguishes arid and semiarid climates are seen in Table 1, including less P and higher U₂ in arid climate due to the dryness of the land. R_H has the highest deviation from the mean with 18.46 % and 23.36 % for Basra and Erbil stations owing to large difference between its minimum and maximum values.

At initial stage before training of the models, the data were normalized to fall between the range of 0 and 1 to reduce data redundancy and increase data integrity in accordance to Elkiran et al. (2018) as;

$$a_{norm} = \frac{a_i - a_{min}}{a_{max} - a_{min}} \quad [5]$$

Where a_{norm} is the normalized value, a_i is the actual value, a_{max} is the maximum value and a_{min} is the minimum value.

2.2 Performance Evaluation of the Models

To determine the performance of the models, Nash-Sutcliffe efficiency criteria (NSE) or determination coefficient and root mean square error (RMSE) were used as (Abdullahi et al., 2017).

$$NSE = 1 - \frac{\sum_{i=1}^N (M_i - \hat{M}_i)^2}{\sum_{i=1}^N (M_i - \bar{M})^2} \quad [6]$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (M_i - \hat{M}_i)^2}{N}} \quad [7]$$

Where N is the number of observations, \bar{M} , \hat{M}_i , M_i and N are respectively the mean of the observed values, predicted value, observed value and number of observations? RMSE has values between 0 and ∞ whereas NSE values ranges between $-\infty$ to 1. The accuracy of the model increases as RMSE approaches 0 and NSE towards 1 (Nourani et al., 2019b).

2.3 Proposed Methodology

In this study, SVR model, conventional MLR and HS empirical model were used for modeling ET_0 in Basra and Erbil stations of Iraq. First, sensitivity analysis was applied to determine the effective input parameters. Based on the results of the input selection technique, 5 different models with distinct input combinations were developed using SVR and MLR techniques with ET_0 as a function of the independent variables as.

$$ET_0^b = f(T_{min}^b, T_{mean}^b, T_{max}^b, R_H^b, U_2^b, P_S^b, R_S^b) \quad [8]$$

$$ET_0^e = f(T_{min}^e, T_{mean}^e, T_{max}^e, R_H^e, U_2^e, P_S^e, R_S^e) \quad [9]$$

Where the superscript alphabet indicates station, b for Basra and e for Erbil.

For the HS model which has fixed inputs irrespective of the results of the sensitivity analysis, the HS equation is given by (Nourani et al., 2019a).

$$ET_0 = 0.000939R_a(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} \quad [10]$$

The overall study methodology is given in Figure 2. It should be noted that, for proper comparison of the performances of the applied models, same methodology was applied for both study stations.

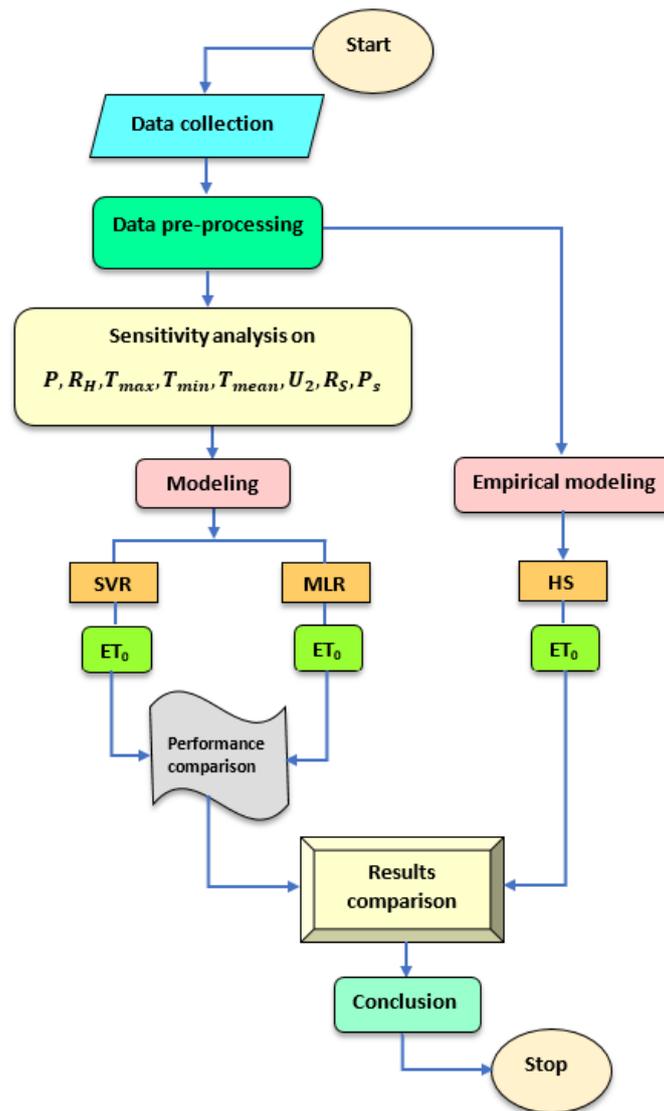


Figure 2: Schematic diagram of the proposed methodology

2.4 FAO Penman-Monteith Equation (FAO-56-PM)

From the original equation of Penman-Monteith, canopy resistance and aerodynamic equations, the FAO-56-PM has been derived (Allen et al., 1998) as.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad [11]$$

Where ET_0 is the reference evapotranspiration (mm/day), Δ is slope vapor pressure curve (kpa/ $^{\circ}$ C), R_n is net radiation at the crop surface (MJ/m²/day), G is soil heat flux density (MJ/m²/day), T is air temperature at 2 m height ($^{\circ}$ C), U_2 is wind speed at 2 m height (m/s), e_s is saturation vapor pressure (kpa), e_a is actual vapor pressure (kpa), $e_s - e_a$ is saturation vapour pressure deficit (kpa), γ is psychrometric constant (kpa/ $^{\circ}$ C).

2.5 Support Vector Regression (SVR)

Based on SVM concept, the SVR model was developed to be used for non-linear regression issues. Contrary to several other black box forecasting methods, SVR as one of the SVM based methods, consider operational risk as the objective function to be minimized instead of minimizing the error between computed and observed values. In SVR, at first, a linear regression is fitted on the data and then the outputs go through a non-linear kernel to catch the non-linear pattern of the data. Given a set of training data $\{(x_i, d_i)\}_i^N$ (d_i is the actual value, x_i represents the input vector and N is the data number), the SVR function is generally given as (Wang et al., 2013):

$$y = f(x) = w\varphi(x_i) + b \quad [12]$$

where $\varphi(x_i)$ is mapped non-linearly from input vector x , which indicates feature spaces. By minimizing the objective function and assigning positive values for the slack parameters of ξ and ξ^* , regression parameters b and w may be determined (Wang et al., 2013).

$$\text{Minimize: } \frac{1}{2} \|w\|^2 + C[\sum_i^N (\xi_i + \xi_i^*)]$$

$$\text{Subject to: } \begin{cases} w_i\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ d_i - w_i\varphi(x_i) + b_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \end{cases} \quad i=1,2,\dots,N$$

where $\frac{1}{2} \|w\|^2$ implies weights vector norm and C represents regularized constant, which determine the arrangement between the regularized term and the empirical error. ε is the tube size, which is equivalent to accuracy approximation positioned within the training data points. By defining Lagrange multipliers α_i and α_i^* , dual quadratic optimization problem could be resulted from the mentioned optimization problem. After dealing with the quadratic optimization problem, vector w can be determined as (Wang et al., 2013).

$$w^* = \sum_{i=1}^N (\alpha_i - \alpha_i^*)\varphi(x_i) \quad [13]$$

Finally, SVR is given as (Wang et al., 2013):

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)K(x, x_i) + b \quad [14]$$

$k(x_i, x_j)$ is the kernel function into feature space which performs the non-linear mapping and b is bias term. Gaussian Radial Basis Function (RBF) is the one commonly used kernel function, given as:

$$k(x_1, x_2) = \exp(-\gamma\|x_1 - x_2\|^2) \quad [15]$$

where, γ is the kernel parameter.

Figure 3 shows the SVM structure

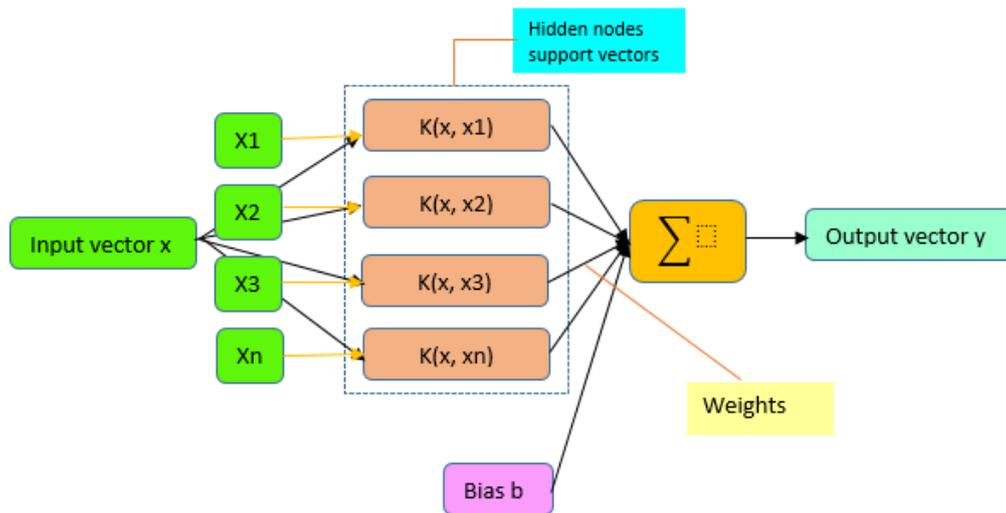


Figure 3: The structure of SVM model (Ghorbani et al., 2018)

2.6 Multi-Linear Regression

Multiple linear regression (MLR) is a classical technique that mathematically modeled the linear relationship that exists between dependent and predictor variables. In general, the n predictor variables and y dependent variable may have relation via (Nourani et al., 2019a):

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_ix_i + \xi \quad [16]$$

Where x_i represents the i^{th} predictor value, b_0 is the constant of regression, and b_i is the i^{th} predictor coefficient and ξ represents the error term.

3. Results and Discussion

One of the most important aspects of any AI based modeling is the selection of appropriate input combination, as contrary to that may lead to complications and error results (Abdullahi & Elkiran, 2017). To overcome such problems, in this study, sensitivity analysis was applied to determine the dominant inputs. Though, the effect of decrease or increase of ET_0 due to change in meteorological parameters through non dimensional sensitivity analysis has already been applied, but as shown by previous studies, the effect of each variable on ET_0 can be determined through single-input single-output AI based sensitivity analysis (Nourani et al., 2019a). Hence, this study performed single-input single output SVR based sensitivity analysis to ascertain the dominant inputs. The results of the sensitivity analysis are given in Table 2.

Table 2: Results of the applied sensitivity analysis

Station	Parameter	Training		Validation	
		NSE	RMSE ^a	NSE	RMSE ^a
Basra	T _{max}	0.693	0.147	0.660	0.122
	T _{min}	0.677	0.151	0.662	0.121
	T _{mean}	0.694	0.147	0.670	0.120
	P	0.070	0.255	0.034	0.205
	R _H	0.655	0.156	0.592	0.133
	P _S	0.763	0.129	0.613	0.130
	U ₂	0.612	0.165	0.486	0.150
	R _S	0.641	0.159	0.611	0.130
	Erbil	T _{max}	0.843	0.072	0.828
T _{min}		0.750	0.091	0.786	0.077
T _{mean}		0.807	0.080	0.809	0.073
P		0.171	0.165	0.131	0.156
R _H		0.792	0.083	0.727	0.087
P _S		0.455	0.134	0.475	0.121
U ₂		0.343	0.160	0.227	0.135
R _S		0.592	0.116	0.542	0.113

^a RMSE has no unit as the data were normalized

As seen in Table 2, being situated in the severe climates which are characterized with high temperature and less amount of precipitation, the 3 categories of temperature are more dominant parameters to ET₀. In both training and validation phases for the two stations, it can be observed that almost all the parameters have significant effect on ET₀ to some certain extent. The only exception is P, which performed the least with lowest NSE and highest RMSE 0.034 and 0.205 in the validation phase. This could be because, P is inversely proportional to temperature, meaning that where there is high temperature (such as in arid and semiarid climates) P is at its minimum amount and the direct impact of temperature on ET₀ makes it difficult for P to estimate ET₀ in such climate conditions. Owing to this circumstance, inclusion of P could reduce the accuracy and efficiency of the modeling. Therefore, P is discarded as input for the ET₀ modeling.

Based on the sensitivity analysis results, 5 different models were developed. Table 3 shows the results of the ET₀ modeling. The models for all stations were created using RBF kernel. The RBF kernel's tuning parameters are fewer than two sigmoid and polynomial kernels. Moreover, when smoothness in assumptions is considered, better performance is achieved by RBF kernel in SVR modeling (Sharghi et al., 2018). Hence, for effective ET₀ modeling in this study, the RBF kernel's parameters were used. Figure 4 shows the box plots of the predicted ET₀ by all models.

Table 3: Results of the applied models

Station	Model	Number	Input	Structure	Training		Validation		
					NSE	RMSE	NSE	RMSE	
Basra	SVR	M1	$T_{max}, T_{min}, T_{mean}$	RBF	0.8960	0.0137	0.8941	0.0192	
		M2	$T_{max}, T_{min}, T_{mean}, R_H$	RBF	0.9040	0.0129	0.9020	0.0174	
		M3	$T_{max}, T_{min}, T_{mean}, P_S$	RBF	0.9030	0.0136	0.8963	0.0176	
		M4	$T_{max}, T_{min}, T_{mean}, U_2$	RBF	0.9951	0.0006	0.9949	0.0009	
		M5	$T_{max}, T_{min}, T_{mean}, R_S$	RBF	0.9115	0.0116	0.9084	0.0166	
	MLR	M1	$T_{max}, T_{min}, T_{mean}$	3-1	0.8809	0.0157	0.8806	0.0216	
		M2	$T_{max}, T_{min}, T_{mean}, R_H$	4-1	0.8914	0.0143	0.8872	0.0204	
		M3	$T_{max}, T_{min}, T_{mean}, P_S$	4-1	0.9072	0.0157	0.8809	0.0168	
		M4	$T_{max}, T_{min}, T_{mean}, U_2$	4-1	0.9848	0.0027	0.9781	0.0029	
		M5	$T_{max}, T_{min}, T_{mean}, R_S$	4-1	0.8989	0.0144	0.8905	0.0183	
	HS	-	$T_{max}, T_{min}, T_{mean}, R_a$	4-1	0.8211	0.0324	0.6084	0.0515	
	Erbil	SVR	M1	$T_{max}, T_{min}, T_{mean}$	RBF	0.9609	0.0041	0.9556	0.0049
			M2	$T_{max}, T_{min}, T_{mean}, R_H$	RBF	0.9608	0.0040	0.9561	0.0049
			M3	$T_{max}, T_{min}, T_{mean}, P_S$	RBF	0.9572	0.0040	0.9562	0.0054
			M4	$T_{max}, T_{min}, T_{mean}, U_2$	RBF	0.9882	0.0011	0.9871	0.0016
M5			$T_{max}, T_{min}, T_{mean}, R_S$	RBF	0.9609	0.0041	0.9556	0.0049	
MLR		M1	$T_{max}, T_{min}, T_{mean}$	3-1	0.9627	0.0047	0.9473	0.0049	
		M2	$T_{max}, T_{min}, T_{mean}, R_H$	4-1	0.9618	0.0047	0.9491	0.0048	
		M3	$T_{max}, T_{min}, T_{mean}, P_S$	4-1	0.9624	0.0047	0.9489	0.0047	
		M4	$T_{max}, T_{min}, T_{mean}, U_2$	4-1	0.9775	0.0028	0.9689	0.0029	
		M5	$T_{max}, T_{min}, T_{mean}, R_S$	4-1	0.9627	0.0047	0.9473	0.0049	
HS	-	$T_{max}, T_{min}, T_{mean}, R_a$	4-1	0.7347	0.0334	0.5515	0.0413		

^a RMSE has no unit as the data were normalized

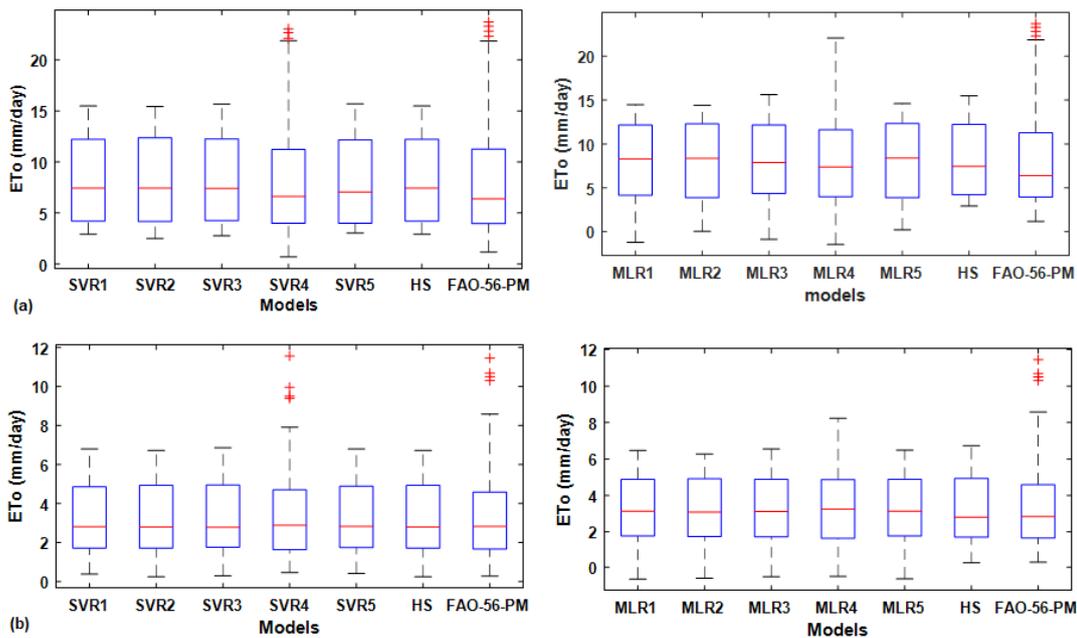


Figure 4: Computed ET₀ by all models for (a) Basra station (b) Erbil Station

In Table 3, the RBF for SVR structure indicates the kernel function employed for the model construction, while x-y for MLR and HS models indicate the number of inputs and output. For better performance comparison, the HS model was calibrated in to training and validation as SVR and MLR models. As demonstrated in Table 3, different performance is achieved by different models with different input combinations. The effectiveness of the input variables to ET₀ led to reliable results even with the use of 3 inputs.

For Basra station, both SVR and MLR models produced good performance with 3 or 4 inputs. Though, HS model performance could be acceptable, but its accuracy is inferior to the other models in both training and validation phases. This could be attributed to one or all of the following reasons; (i) SVR and MLR models have the advantage of selecting or incorporating the most effective variables to ET₀ which is significant aspect in the modeling, whereas HS model has fixed inputs irrespective of the variables performance towards ET₀. (ii) Apart from temperatures (T_{max} , T_{mean} and T_{min}), R_a is also included as input for HS model. R_a being the sun's intensity at the top of the earth's atmosphere constantly varies due to earth's elliptical orbit, which results in varied earth-sun distance. This variation in R_a coupled with severe climate condition may deter the performance of HS model in arid and semiarid climates. (iii) The climate of the station may have significant effect on HS model. For instance, in frozen days of winter, the ET₀ time series gets small values having less complexity without significant fluctuations. In this period, an empirical model may sufficiently lead to reliable results, but in the case of arid and semiarid climates, which have hotter days, the fluctuations and complexity of ET₀ process might be large thereby leading to inefficiency of HS model. Figure 5 shows time series of the observed and predicted values for the best models for Basra station.

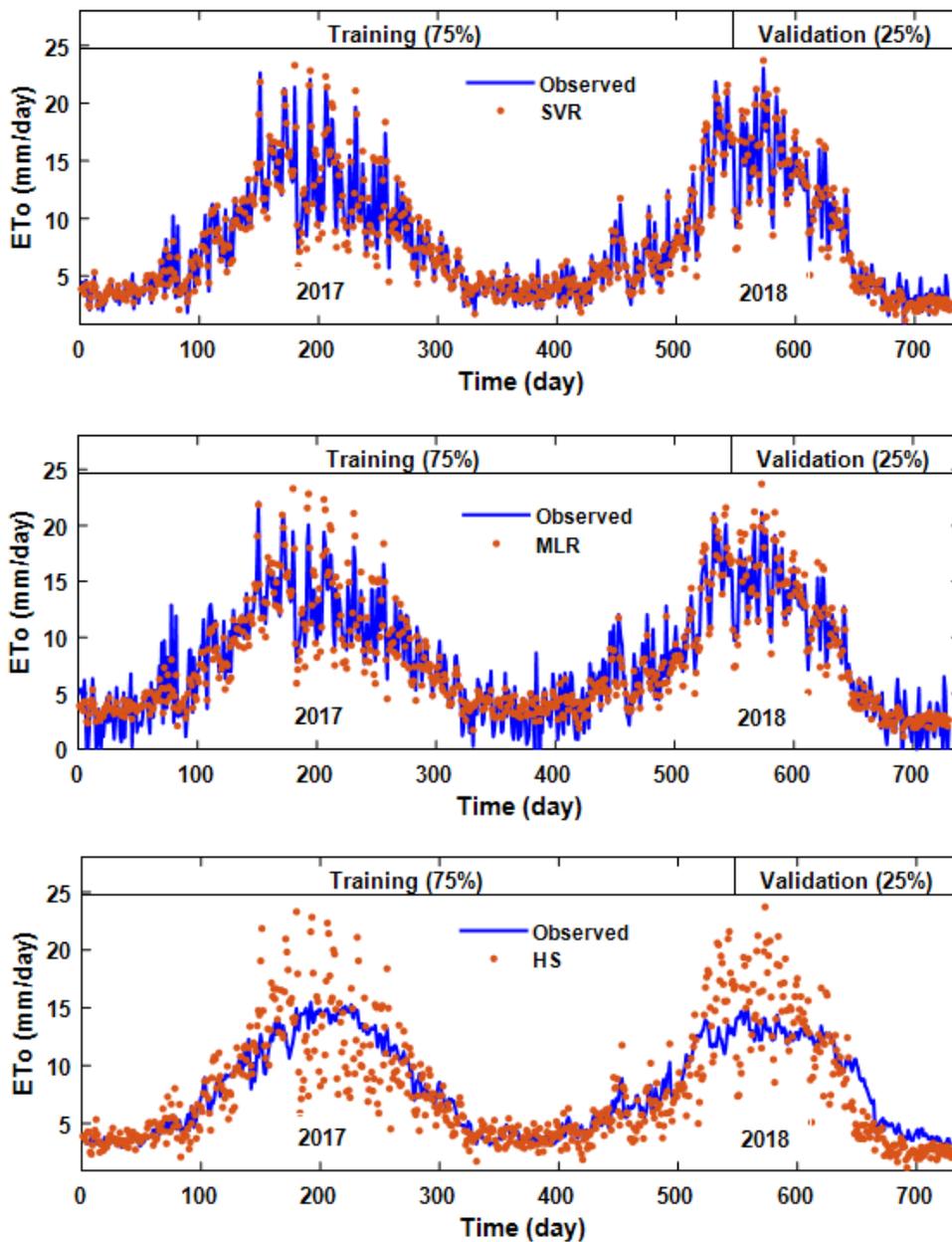


Figure 5: Computed versus observed time series for the best models in Basra station

For Erbil station in Table 3, the performance of SVR and MLR models are superior to HS model due to its inability to perform well in an extreme weather condition. However, owing to its capability of dealing with nonlinear and stochastic nature of ET_0 process, SVR is found to be the most accurate model. The results in Table 3 also show that, despite the impact of temperature in prediction of ET_0 in Erbil as shown by sensitivity analysis, ET_0 being a complex process depends on many climatic factors which their inputs are needed for efficient ET_0 modeling. Nevertheless, the results imply that with application of temperatures (T_{min} , T_{mean} , T_{max}) only, a successful modeling of ET_0 is achievable. Figure 6 shows time series of the observed and predicted values for the best models for Basra station.

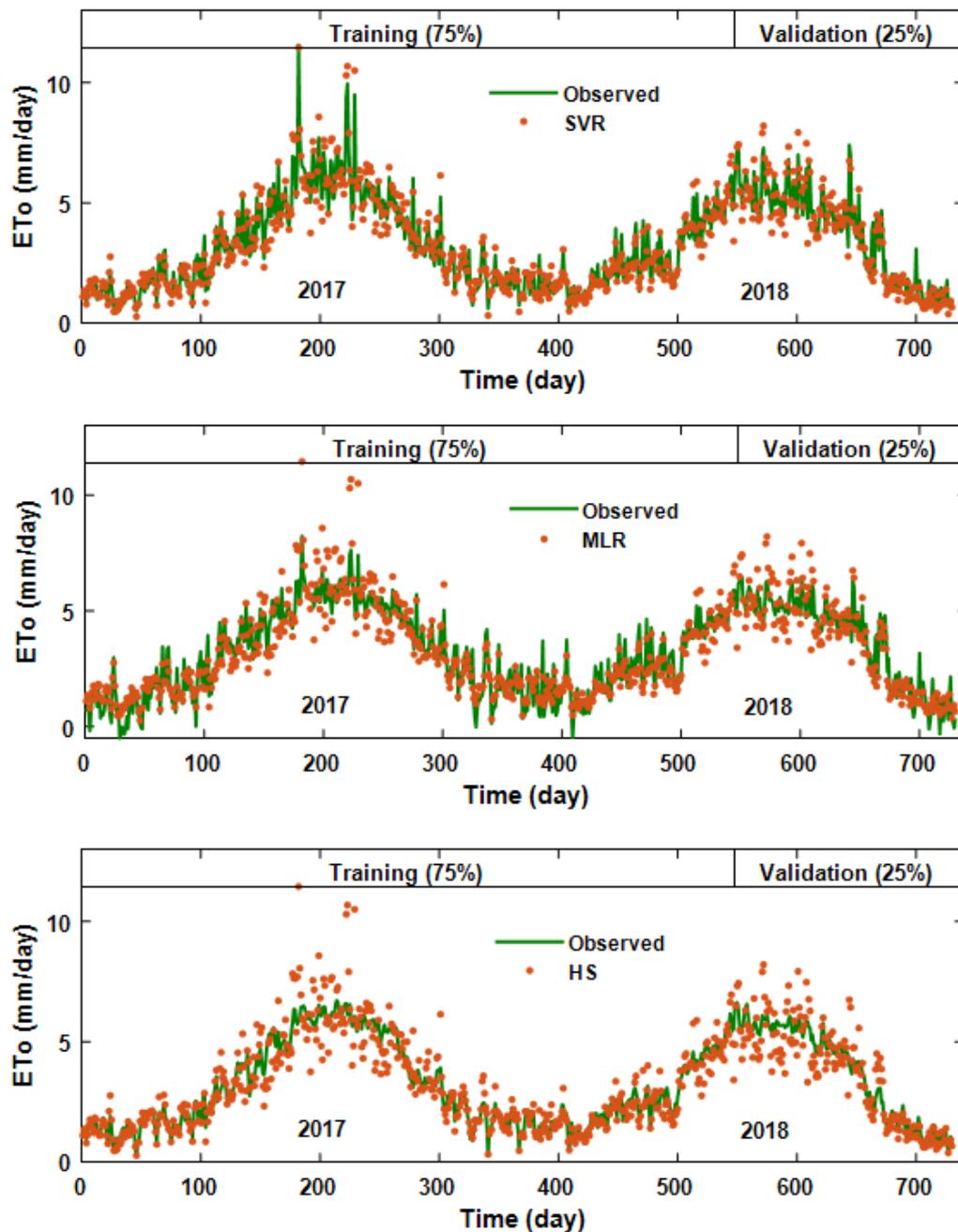


Figure 6: Computed versus observed time series for the best models in Erbil station

Comparing the performances of the models in Basra and Erbil stations, it can be deduced in Table 3 that, all the models have similar performances due to the locations of the stations to arid and semiarid climate stations, which are characterized by hot temperatures and scarce precipitations. The models performed better in Erbil station than Basra station. This might be because extremity of climate increases with increase in aridity index, also, stochastic and nonlinear nature of climate increases as aridity index increases due to more fluctuations. Thus, dealing with the mentioned problems is more challenging in Basra (Arid) than Erbil (semiarid) stations. As seen for both stations, model 4 that constitutes T_{max} , T_{min} , T_{mean} and U_2 provided the best results. This justifies the claim by Nourani et al.

(2019b) that, sole application of U_2 in hydro-climatic modeling may not give reliable performance but its inclusion with other climate parameters may significantly increase the modeling performance. With reference to Table 2, it can be observed that with exception of P, U_2 is the least sensitive parameter to ET_0 modeling, but its inclusion with other parameters led to best performance of models in Table 3. Another notable observation, which is worthy to mention is that, though almost all models' performances are better in Erbil than Basra station, the M4 results show a contrasting performance. Looking at Table 1 in the descriptive statistics, it can be understood that Basra has the highest maximum U_2 (8.62 m/s) than Erbil (5.18), also in Table 2, Basra has the highest NSE (0.486) than Erbil (0.227). These two statistics are showing that the U_2 has a direct effect to the modeling of ET_0 , the higher the U_2 the more impact it has in ET_0 modeling and vice versa.

4. Conclusions

This study ascertained the capabilities of SVR (as AI) model, MLR and empirical HS models in modeling ET_0 in Basra (arid) and Erbil (semiarid) climate stations. Sensitivity analysis was performed to determine the dominant inputs, later; the models were applied in order to achieve the study objective.

The obtained results showed that due to the locations of the stations to arid and semiarid climates, temperatures including T_{min} , T_{max} and T_{mean} were the most effective parameters. The results also showed that all the applied models can lead to reliable performance but SVR model produced the best performance due to its ability to deal with nonlinear and complex behavior of ET_0 . However, the results indicated that all the developed models are capable of producing results but, inclusion of U_2 for M4, significantly improved the performance of ET_0 modeling. This study employed local data analysis approach to train and validate the models, future studies should consider external data analysis approach to see how effective the models could perform. Also, the use of other models different from those utilized in this study and incorporation of additional stations would be useful for future studies.

References

- Abdullahi, J., & Elkiran, G. (2017). Prediction of the future impact of climate change on reference evapotranspiration in Cyprus using artificial neural network. *Procedia Computer Science*, 120, 276-283.
- Abdullahi, J., Elkiran, G., & Nourani, V. (2017). Application of Artificial Neural Network to predict reference evapotranspiration in Famagusta, North Cyprus. In *11th International Scientific Conference on Production Engineering Development and Modernization of Production* (pp. 549-554).
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO, Rome*, 300(9), D05109.
- Djaman, K., O'Neill, M., Diop, L., Bodian, A., Allen, S., Koudahe, K., & Lombard, K. (2019). Evaluation of the Penman-Monteith and other 34 reference evapotranspiration equations under limited data in a semiarid dry climate. *Theoretical and Applied Climatology*, 137(1-2), 729-743.
- Elkiran, G., Nourani, V., Abba, S. I., & Abdullahi, J. (2018). Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river. *Global Journal of Environmental Science and Management*, 4(4), 439-450.
- Ferreira, L. B., da Cunha, F. F., de Oliveira, R. A., & Fernandes Filho, E. I. (2019). Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—a new approach. *Journal of Hydrology*.
<https://doi.org/10.1016/j.jhydrol.2019.03.028>

- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management*, 217, 303-315.
- Ghorbani, M. A., Deo, R. C., Yaseen, Z. M., Kashani, M. H., & Mohammadi, B. (2018). Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theoretical and Applied Climatology*, 133(3-4), 1119-1131.
- Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, 1(2), 96-99.
- Jabbar, M. T., & Zhou, J. X. (2013). Environmental degradation assessment in arid areas: a case study from Basra Province, southern Iraq. *Environmental Earth Sciences*, 70(5), 2203-2214.
- Mehdizadeh, S. (2018). Estimation of daily reference evapotranspiration (ET_o) using artificial intelligence methods: Offering a new approach for lagged ET_o data-based modeling. *Journal of Hydrology*, 559, 794-812.
- Nourani, V., Elkiran, G., & Abdullahi, J. (2019a). Multi-station artificial intelligence-based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *Journal of Hydrology*, 577, 123958.
- Nourani, V., Elkiran, G., Abdullahi, J., & Tahsin, A. (2019b). Multi-region modeling of daily global solar radiation with artificial intelligence ensemble. *Natural Resources Research*, 1-22.
- Rasul, A., Balzter, H., & Smith, C. (2015). Spatial variation of the daytime surface urban cool island during the dry season in Erbil, Iraqi Kurdistan, from Landsat 8. *Urban Climate*, 14, 176-186.
- Sarlak, N., & Agha, O. M. M. (2018). Spatial and temporal variations of aridity indices in Iraq. *Theoretical and Applied Climatology*, 133(1-2), 89-99.
- Sharghi, E., Nourani, V., & Behfar, N. (2018). Earthfill dam seepage analysis using ensemble artificial intelligence-based modeling. *Journal of Hydroinformatics*, 20(5), 1071-1084.
- Wang, W. C., Xu, D. M., Chau, K. W., & Chen, S. (2013). Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *Journal of Hydroinformatics*, 15(4), 1377-1390.