



Modeling reference evapotranspiration using a novel regression-based method: radial basis M5 model tree

Ozgur Kisi^{1,2} · Behrooz Keshtegar³ · Mohammad Zounemat-Kermani⁴ · Salim Heddami⁵ · Nguyen-Thoi Trung^{6,7}

Received: 4 June 2020 / Accepted: 3 May 2021 / Published online: 15 May 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

In the current study, an ability of a novel regression-based method is evaluated in modeling daily reference evapotranspiration (ET_0), which is an important issue in water resources management and planning. The method was developed by hybridizing radial basis function and M5 model tree and called as radial basis M5 model tree (RM5Tree). The new model results were compared with traditional M5 model tree (M5Tree), response surface method (RSM), and two neural networks (multi-layer perceptron neural networks, MLPNN & radial basis function neural network, RBFNN) with respect to several statistical indices. Daily climatic data (relative humidity, RH , solar radiation, SR , wind speed, air temperature, T) recorded at three stations in Turkey, Mediterranean Region, were used. The effect of each weather data on ET_0 was also investigated by utilizing three different input scenarios with various combinations of input variables. On the whole, the RM5Tree provided the best results (Nash and Sutcliffe efficiency, $NES > 0.997$) followed by the MLPNN ($NES > 0.990$), and M5Tree ($NES > 0.945$) in modeling daily ET_0 . The SR was observed as the most effective input parameter on ET_0 which was followed by the T and RH . However, the findings of the third modeling scenario revealed that taking into account of all variables would considerably increase models' accuracies for the three stations.

1 Introduction

Evapotranspiration (ET) is one of main components in hydrological cycle and accurate estimation of ET has vital importance in design and management of the irrigation systems, water resources studies, and other similar cases. Knowing the ET rate of a plant can help determining the accurate amount of water required for irrigation, which will subsequently lead to increased productivity. Failure in determining the accurate ET rate may lead to the overestimation of plants' water requirement, which will consequently cause adverse effects such as waterlogged lands, soil

nutrients washout, as well as contamination of groundwater resources. On the other hand, underestimation of plants' water requirement will incur moisture tension on them, followed consequently by a reduced crop yield. In this regard, equations such as FAO Penman, FAO Penman-Monteith, Blaney-Criddle, etc. can be used to study the reference evapotranspiration (ET_0) of plants (Feng et al. 2017a, b; Chen et al. 2019a, b). However, despite their acceptable performance in most of the cases, using these methods needs the access to a large amount of input data, which is not of course always possible regarding the conditions within various regions.

✉ Behrooz Keshtegar
bkeshtegar@uoaz.ac.ir

Nguyen-Thoi Trung
nguyenthotrung@tdtu.edu.vn

¹ Department of Civil Engineering, School of Technology, Ilia State University, Tbilisi, Georgia

² Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

³ Department of Civil Engineering, Faculty of Engineering, University of Zabol, P.B. 9861335856, Zabol, Iran

⁴ Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

⁵ Faculty of Science, Agronomy Department, Hydraulics Division, University 20 Août 1955, Route El Hadaik, BP 26, Skikda, Algeria

⁶ Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁷ Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

In such conditions, using indirect approaches such as soft computing techniques and artificial intelligence based (AI-based) models can be a proper alternative modeling solution (Zounemat-Kermani et al. 2020). Lots of researchers have applied artificial intelligence (AI)-based models including artificial neural networks (e.g., MLPNN & RBFNN), fuzzy logic concepts (e.g., ANFIS), regression, and classification tree models (e.g., M5 tree & CHAID) and machine learning approaches (support vector machine, SVM & support vector regression, SVR) for simulating and modeling ET_0 (Kisi and Öztürk 2007; Shiri et al. 2014, 2015; Gocić et al. 2015; Yin et al. 2017; Mehdizadeh et al. 2017; Dou and Yang 2018; Zounemat-Kermani et al. 2019; Chia et al. 2020; Chen et al. 2020; Yamaç and Todorovic 2020; Adnan et al. 2021). In the following paragraphs, some of the most recent pertinent studies about modeling ET_0 using AI-based models are presented. The neuro-fuzzy and artificial neural network (ANN) were compared for modeling ET_0 with two input combinations to select suitable input data via Shiri et al. (2015) and they reported that the local training can be applied to validate the alternative modeling by using AI. Kisi (2016) used three different models consisting multivariate adaptive regression splines, MARS, M5Tree, and least square SVR for approximating the ET_0 . In general, it was found that the M5Tree performed superior to the other modeling approaches applied. Rahimikhoob (2016) compared the ability of artificial neural network and M5Tree for estimating ET_0 of an arid area. The study did not report the complete superiority of the utilized models.

Khoshravesh et al. (2017) analyzed performances of multivariate fractional polynomial model, Bayesian, and robust regressions to estimate ET_0 in arid climates. Outcomes of the study showed that accuracy of multivariate fractional polynomial model is better than other two models. Daily ET_0 predictions were investigated using extreme learning machine (ELM) and generalized regression neural network (GRNN) with input data of temperature at 6 stations of China by Feng et al. (2017a, b). It is conducted that ELM is a robust and accurate model compared to GRNN. Three different AI-based models (ANFIS-GP, fuzzy genetic model, and M5Tree) were used and executed in modeling monthly ET_0 values by Wang et al. in 2017. It was reported that the fuzzy genetic model performed better in comparison to the ANFIS-GP and M5Tree models. The hybrid genetic algorithm (GA) and SVM as an AI-based models were used for simulating daily ET_0 of semi-arid environment in northwest China by Yin et al. (2017) and it was compared with SVM and ANN based on eight different combinations of climatic input data. Based on their results, the SVR combined by GA model had superior performances. Antonopoulos and Antonopoulos (2017) compared ANN-based multi-layer feed forward back-propagation with several empirical models for calculation of daily ET_0 in northern Greece and showed that the ANN with a sigmoid

transfer function in hidden layer can provide more accurate predictions than the empirical models.

In another study, Keshtegar et al. (2018) explored the ability of subset ANFIS, ANFIS, ANNs, and M5Tree. They claimed that the subset ANFIS is superior to other applied methods in modeling daily ET_0 . Gavili et al. (2018) compared the ability of ANN, ANFIS, and gene expression programming (GEP) in modeling daily ET_0 . The results attained from the AI-based models were compared with those of the empirical models. Comparing the results, it was revealed that the AI-based models provided better accuracies compared to empirical models. Dou and Yang (2018) investigated and compared the feasibilities and abilities of four AI-based models using ELM, ANFIS, ANN, and SVR for prediction of daily ET_0 for four sites in China. It conducted that the AI-based ANFIS and ELM models can produce better performances compared to the ANN and SVM while the ELM model was considerably reduced computational time in modeling process. Keshtegar et al. (2019) compared the ability of RSM and polynomial chaos expansion (PCE) in modeling ET_0 . They reported that the PCE model was more accurate approach to estimate daily ET_0 . The abilities of four learning algorithms as multilayer perceptron-based deep feed-forward ANN, gradient-boosting machine, random forest regression (RF) using M5Tree model, and generalized linear model were compared for ET_0 estimation for the Punjab Northern India stations by Saggi and Jain (2019). According to the extracted results from this study, the deep feed-forward ANN performed better than the other models. Sanikhani et al. (2019) employed six AI-based models namely multilayer perceptron ANN, GRNN, radial basis neural networks, integrated ANFIS with subtractive clustering and grid partitioning, and GEP for modeling ET_0 with small number of input climatic data. In general, it was reported that all the applied models have highly practical and reliable performances for investigated stations. Heddad et al. (2018) applied and compared three evolving connectionist (ECoS) models namely, offline-based dynamic evolving neural-fuzzy inference systems named DENFIS-OF, (ii) online-based dynamic evolving neural-fuzzy inference systems named DENFIS-ON, and (iii) the evolving fuzzy neural network called (EFuNN), for modeling daily ET_0 in the northern region of the Algeria. According to the obtained results, the best accuracy was obtained using the DENFIS-OF model. Tao et al. (2018) proposed a hybrid model called adaptive neuro-fuzzy inference systems (ANFIS) with firefly algorithm (ANFIS-FA) for predicting daily ET_0 at Burkina Faso and reported that the hybrid ANFIS-FA provides higher accuracy compared to the standard ANFIS. Karbasi (2018) employed the Gaussian process regression (GPR) for forecasting daily

Table 1 Daily statistical characteristics of the dependent (T, W, SR & RH) and independent (ET_0) parameters in the three different meteorological stations in Turkey

Parameter	Station	No. data	Min	Max	Mean	Std.	Cv	Correlation*
Air temperature (C)	Adana	9959	0.40	35.40	19.93	7.25	0.36	0.76
	Antalya	7743	1.70	37.50	18.49	7.10	0.38	0.75
	Isparta	8734	-10.45	29.35	12.02	7.93	0.66	0.82
Wind speed (m/s)	Adana	9959	0.00	33.30	1.49	0.87	0.58	0.27
	Antalya	7743	0.00	13.00	2.80	1.55	0.56	0.27
	Isparta	8734	0.00	10.90	1.94	1.33	0.69	0.07
Solar radiation (langley)	Adana	9959	3.00	726.15	342.95	152.02	0.44	0.92
	Antalya	7743	3.60	797.40	411.84	179.18	0.44	0.83
	Isparta	8734	7.70	731.80	321.00	140.14	0.44	0.91
Relative humidity (%)	Adana	9959	13.00	97.70	65.67	13.87	0.21	-0.32
	Antalya	7743	14.00	98.00	63.69	16.84	0.26	-0.68
	Isparta	8734	19.30	96.70	61.16	13.62	0.22	-0.79
ET_0 (mm)	Adana	9959	-0.36	11.43	3.04	1.63	0.54	1.00
	Antalya	7743	-0.31	15.02	4.08	2.40	0.59	1.00
	Isparta	8734	-0.37	8.19	2.67	1.67	0.62	1.00

*Correlations have been between the independent values and dependent value for each station

ET_0 and demonstrated that the wavelet decomposition significantly improved the performances of the models. Fang et al. (2018) employed the RF, SVM, and MLR for predicting monthly ET_0 in China, and demonstrated that the SVM was more accurate. The accuracy of models to predict the ET_0 is one of challenges in hydrology field to manage irrigation systems and water resources.

More recently, Zhu et al. (2020) employed a hybrid extreme learning machine (ELM) with the particle swarm optimization (PSO) model for daily ET_0 prediction. They claimed that the PSO-ELM model offered the best accuracy among other applied models such as ANN and random forest (RF) models. Nagappan et al. (2020) attempted to predict ET_0 for irrigation scheduling using machine learning methods like deep learning neural network (DLNN) and RBNN. It was

found that the DLNN model acted better in the prediction process.

In a similar study, Ferreira and da Cunha (2020) showed that deep learning performed slightly better than ANN and RF in predicting ET_0 . Salam and Islam (2020) compared various data-driven models in ET_0 prediction. They utilized standard SVM model as well as ensemble learning models for the prediction process including the bagging random tree (RT), RF, and random subspace (FS) models. The findings showed that the RT model performed superior followed by the RF, RS, and SVM.

Generally, the AI methods as machine learning approaches are used to provide the accurate prediction of ET_0 due to flexible ability for providing the nonlinear relations. However, the AI approaches have some

Fig. 1 The location of the stations in Mediterranean region of Turkey



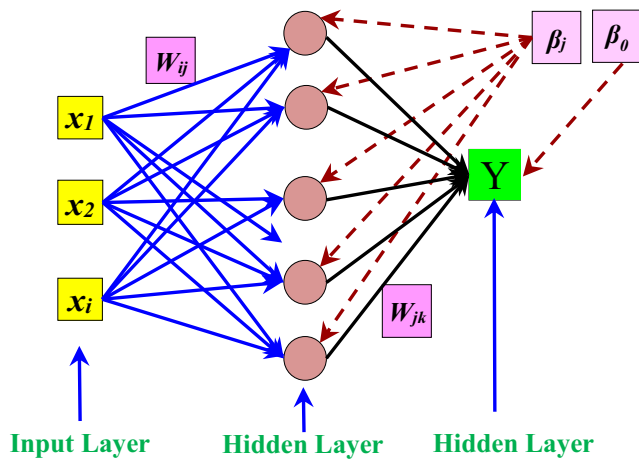


Fig. 2 General structure of the MLPNN model

limitations as (i) input variables highly affect their predictions, (ii) some control parameters are required to train the models, and (iii) training process to provide a model is time consuming. The regression-based models are the efficient modeling approaches with simple regressed process. However, the regression-based data driven approaches have some drawbacks including (i) the regressed function is important for accurately predictions, (ii) the highly nonlinearity of the input data are neglected in the regression process, and (iii) the linear cross-correlation between the input-output data is used in the modeling process. Consequently, the efficient and accurate modeling approach using the regression-based models should be developed that is free from a complex training process.

The nonlinear mapping with efficient regression region can provide a flexible nonlinear response with efficient modeling process for machine learning models. The input data for training of the M5Tree models can be controlled based on the nonlinear maps using radial basis function. Thus, the nonlinearity of input variables of the response can be considered applying the radial map. Therefore, it can improve the nonlinear functions for accurate ET_0

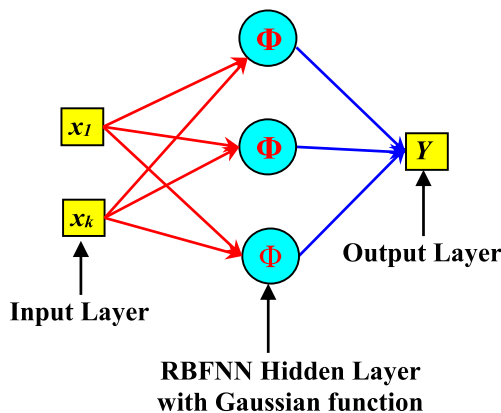


Fig. 3 Architecture of the established structure radial basis function neural network (RBFNN)

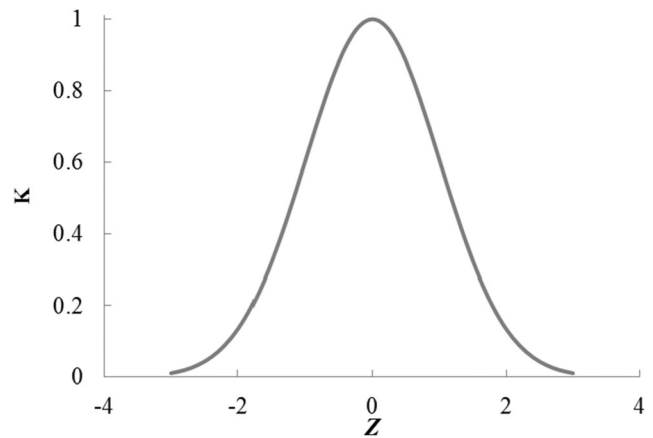


Fig. 4 Schematic view of RBF (K) for $C=0$ and $\epsilon=0.5$

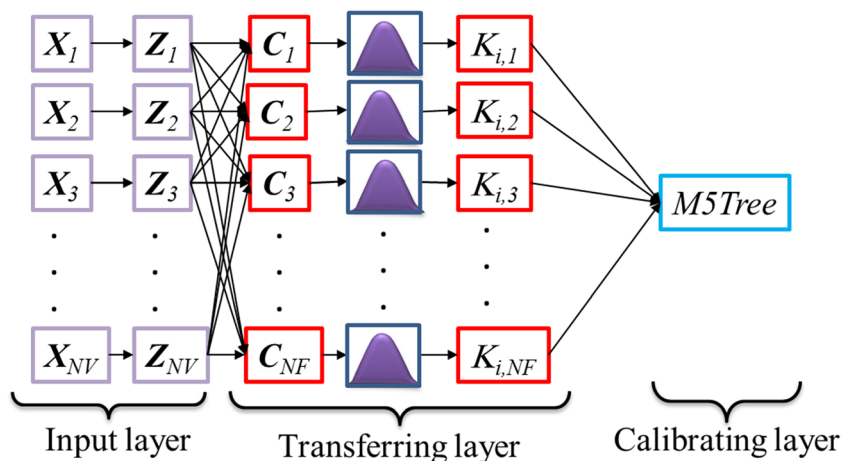
predictions using M5Tree. In this study, the nonlinear forms of the input data are used to improve M5Tree-based regression model. By this way, the input data are better controlled by providing a nonlinear cross-correlation between input-output data set. The proposed model was tested using data from three climate stations in Turkey (Isparta, Antalya & Adana stations). Afterwards, the performance of the RM5tree model was also compared with two neural networks methods, multi-layer perceptron, and radial basis function neural network and two regression methods, response surface method and M5Tree. The results showed that the proposed model has the fixable ability for nonlinear response compared to M5Tree models and by increasing the input data, the accuracy of RM5Tree models was significantly improved compared to other studied models.

2 Materials and methodology

2.1 Case study

The study used daily weather data comprising relative humidity (RH), solar radiation (SR), air temperature (T), and wind speed (W) from Adana (longitude $35^{\circ} 19' E$, latitude $37^{\circ} 00' N$ with an altitude of 27 m) Antalya (latitude: $36^{\circ} 42' N$, longitude: $30^{\circ} 44' E$ with an altitude of 47 m) and Isparta (longitude: $30^{\circ} 34' E$, latitude: $37^{\circ} 47' N$ with an altitude of 997 m) stations with Mediterranean Region, Turkey (Fig. 1). Data cover a period from 01 January of 1972 to 31 December of 2002 for Adana station, from 01 January of 1973 to 31 October of 2002 for Antalya station, and from 01 September of 1978 to 31 October of 2002 for Isparta station. There are no missing values in the used data. Data were obtained from TSMS (Turkish State Meteorological Service). Table 1 shows the statistic characteristics of the dataset in terms of minimum (Min), maximum (Max), standard deviation (Std.), mean, coefficient of variation (CV), and correlation between the input

Fig. 5 Schematic structure of RM5Tree model



parameters and the output parameter (ET_0). Table 1 implies that the Adana and Antalya stations are more similar in terms of temperature ranges. Isparta is the only station that recorded minus temperatures in the dataset used in this study. In addition, the solar radiation is the most correlated parameter followed by air temperature with ET_0 in all of the three stations. In the applications, data splitting rule of 65–35% was applied to train and test the studied models.

2.2 Modeling approaches

2.2.1 Response surface method

The RSM is commonly implemented for modeling nonlinear relations as below (Hill and Hunter 1966):

$$E\hat{T}_o = a_0 + \sum_{i=1}^{NV} a_i x_i + \sum_{i=1}^{NV} \sum_{j=i}^{NV} a_{ij} x_i x_j \quad (1)$$

where, $E\hat{T}_o$ is the predicted ET_0 , NV is the number of input variables x including the mean temperature, i.e., T_{mean} ($^{\circ}C$), solar redation, i.e., SR (langley), relative humidity, i.e., RH_{mean} (%), and wind speed, i.e., W (m/s). a_0, a_i and a_{ij} are unknown coefficients for polynomial terms of Eq. (2).

Table 2 Modeling scenarios based on the input variables for the applied models

Scenario	No. of input parameters	Category	Input parameter(s)
I	1	i	T_{mean}
		ii	W
		iii	SR
		iv	RH
II	2	v	T_{mean} , SR
	3	vi	T_{mean} , SR, RH
III	4 (all)	vii	T_{mean} , SR, RH, W

Generally, the unknown coefficients are calibrated based on the ordinary least square estimator as follows (Keshtegar and Kisi 2017; Ahmadi et al. 2020):

$$a = [P(X)^T P(X)]^{-1} [P(X)^T ET_o] \quad (2)$$

Where, $P(X)$ is the polynomial basic function which is determined based on input data in training stage (65% total of data). More details can be acquired from Keshtegar and Heddami (2017), Keshtegar et al. (2021) and Lu et al. (2020):

2.2.2 Multilayer perceptron artificial neural networks

Artificial neural networks (ANN) are black box models possessing the capabilities to produce a suitable response from an external stimulus, and they are composed of two items: the neurons and the weights (Fei et al. 2020; Sayari et al. 2021). The ANN models are constructed in two distinguished phases: the forward and the backward phases, these two are successively achieved during the backpropagation training algorithm. The information available in the predictor variables is transferred from input neurons to the hidden neurons via the weights, and then summed to get an estimate of the total stimulus of each hidden unit (Ozonoh et al. 2020). The hidden neurons send the collected information to the output neuron through an activation function, generally the sigmoid (Shahabinejad et al. 2020). Finally, the output neuron provides a response, then, compared to the desired value, and the error expected is calculated. A *multi-layer perceptron artificial neural network* (MLPNN) having an input, one hidden and one output layers is a well-known ANN architecture (Fig. 2). Such an ANN model was employed in the present study, and trained with supervised Levenberg Marquart (LM) learning algorithm. According to Fig. 1, the relationship between N possible input variables (x_i : climatic variables) and one output variable (ET_0) was created as follows (Zhu et al. 2021):

Table 3 Comparative results of different modeling approaches for Isparta Station in training and testing phases

	Models	Structure	Training					Testing							
			MAE (mm)	RMSE (mm)	NES	R ²	d	Max (RE)	MAE (mm)	RMSE (mm)	NES	R ²	d	Max (RE)	
Scenario I	Category i, Input variables, T _{mean} (°C)														
	M5Tree	–	0.675	0.862	0.733	0.856	0.918	4.291	0.782	0.975	0.659	0.825	0.898	3.409	
	RSM	–	0.717	0.896	0.712	0.844	0.909	4.547	0.806	0.999	0.642	0.823	0.895	3.746	
	RM5Tree	RF=5	0.674	0.861	0.733	0.856	0.918	4.296	0.782	0.975	0.659	0.825	0.898	3.409	
		RF=10	0.688	0.873	0.726	0.852	0.915	4.326	0.779	0.968	0.664	0.828	0.900	3.505	
		RF=20	0.688	0.873	0.726	0.852	0.915	4.348	0.779	0.968	0.664	0.828	0.899	3.480	
		RF=50	0.688	0.873	0.726	0.852	0.915	4.349	0.779	0.968	0.664	0.828	0.899	3.482	
	Category ii, Input variables, W(m/s)														
	M5Tree	–	1.342	1.583	0.101	0.317	0.412	6.060	1.401	1.688	-0.022	0.267	0.472	4.996	
	RSM	–	1.422	1.639	0.035	0.188	0.257	5.786	1.408	1.658	0.014	0.164	0.279	4.556	
	RM5Tree	RF=5	1.341	1.579	0.105	0.323	0.415	6.049	1.406	1.694	-0.029	0.261	0.466	4.986	
		RF=10	1.342	1.583	0.100	0.317	0.412	6.064	1.402	1.690	-0.025	0.264	0.471	4.990	
		RF=20	1.340	1.581	0.103	0.320	0.416	5.963	1.400	1.688	-0.022	0.267	0.472	4.985	
		RF=50	1.340	1.581	0.103	0.320	0.415	6.011	1.402	1.690	-0.024	0.265	0.471	4.984	
	Category iii, Input variables, SR(langley)														
	M5Tree	–	0.423	0.594	0.873	0.935	0.965	3.841	0.724	1.147	0.528	0.868	0.904	7.286	
	RSM	–	0.505	0.672	0.838	0.915	0.954	4.230	0.601	0.814	0.762	0.914	0.945	3.912	
	RM5Tree	RF=5	0.495	0.666	0.841	0.917	0.955	4.333	0.530	0.686	0.831	0.927	0.957	3.242	
		RF=10	0.572	0.734	0.806	0.898	0.944	3.973	0.582	0.720	0.814	0.914	0.951	2.866	
		RF=20	0.415	0.586	0.881	0.937	0.975	4.041	0.525	0.679	0.835	0.927	0.958	3.253	
RF=50		0.415	0.586	0.881	0.937	0.975	4.042	0.524	0.678	0.835	0.927	0.958	3.255		
Category iv, Input variables, RH _{mean} (%)															
M5Tree	–	0.740	0.958	0.671	0.819	0.894	3.574	0.851	1.099	0.566	0.762	0.850	4.202		
RSM	–	0.753	0.970	0.662	0.814	0.891	3.664	0.854	1.099	0.566	0.762	0.847	4.116		
RM5Tree	RF=5	0.740	0.958	0.670	0.819	0.894	3.573	0.851	1.099	0.566	0.762	0.850	4.205		
	RF=10	0.741	0.959	0.670	0.818	0.894	3.547	0.850	1.098	0.567	0.763	0.850	4.215		
	RF=20	0.741	0.959	0.670	0.818	0.894	3.550	0.850	1.098	0.567	0.763	0.850	4.212		
	RF=50	0.741	0.959	0.670	0.818	0.894	3.550	0.850	1.098	0.567	0.763	0.850	4.208		
Scenario II	Category v, Input variables, T _{mean} (°C), SR(langley)														
	M5Tree	–	0.214	0.331	0.961	0.980	0.990	2.766	0.501	0.729	0.809	0.935	0.956	9.957	
	RSM	–	0.333	0.458	0.925	0.962	0.980	3.736	0.473	0.600	0.871	0.966	0.971	2.601	
	RM5Tree	RF=5	0.159	0.256	0.976	0.988	0.994	2.555	0.459	0.566	0.891	0.967	0.974	2.614	
		RF=10	0.149	0.243	0.979	0.989	0.995	2.274	0.479	0.618	0.863	0.965	0.969	2.597	
		RF=20	0.138	0.236	0.980	0.990	0.995	2.377	0.494	0.660	0.844	0.961	0.966	3.198	
		RF=50	0.154	0.254	0.977	0.988	0.994	2.561	0.455	0.551	0.891	0.967	0.974	2.571	
	Category vi, Input variables, T _{mean} (°C), SR (langley), RH _{mean} (%)														
	M5Tree	–	0.130	0.212	0.984	0.992	0.996	3.098	0.378	0.516	0.904	0.959	0.976	2.923	
	RSM	–	0.266	0.370	0.951	0.975	0.987	2.621	0.342	0.445	0.929	0.978	0.983	1.832	
	RM5Tree	RF=5	0.115	0.186	0.988	0.994	0.997	1.831	0.327	0.422	0.936	0.973	0.984	1.616	
		RF=10	0.098	0.161	0.991	0.995	0.998	1.742	0.279	0.398	0.943	0.978	0.987	2.967	
		RF=20	0.094	0.150	0.992	0.996	0.998	1.743	0.230	0.331	0.961	0.986	0.991	2.869	
		RF=50	0.084	0.134	0.994	0.997	0.998	1.522	0.215	0.320	0.963	0.986	0.991	3.013	
Scenario III	Category vii, Input variables, T _{mean} (°C),SR(langley), RH _{mean} (%),W(m/s)														
	M5Tree	–	0.048	0.068	0.998	0.999	1.000	0.590	0.067	0.105	0.996	0.998	0.999	1.330	
	RSM	–	0.062	0.087	0.997	0.999	0.999	0.699	0.069	0.096	0.997	0.999	0.999	0.632	
	RM5Tree	RF=5	0.048	0.069	0.998	0.999	1.000	1.156	0.072	0.112	0.996	0.998	0.999	1.025	
		RF=10	0.049	0.069	0.998	0.999	1.000	0.622	0.083	0.143	0.993	0.996	0.998	1.380	
		RF=20	0.039	0.054	0.999	0.999	1.000	0.378	0.043	0.063	0.999	0.999	1.000	0.526	
		RF=50	0.038	0.053	0.999	1.000	1.000	0.372	0.038	0.051	0.999	1.000	1.000	0.327	

The bold numbers are the best statistics among others; RMSE and MAE values are in mm

$$T = \varphi_2 \left[\beta_0 + \sum_{j=1}^n w_{jk} \left[\varphi_1 \left(\sum_{i=1}^n x_i w_{ij} + \beta_j \right) \right] \right] \tag{3}$$

where x_i is an input variable, w_{ij} is the weight between the input i and the hidden neuron j , β_j is the bias of the hidden neuron j , φ_1 is the activation sigmoid function, w_{jk} is the weight of connection of neuron j in the hidden layer to

unique neuron k in the output layer, β_0 is the bias of the output neuron k (Wang et al. 2020;)

2.2.3 Radial basis function neural network

Radial basis function neural network (RBFNN) belongs to the category of feedforward neural network (FFNN).

Table 4 Comparative results of different modeling approaches for Antalya station in training and testing phases

	Models	Structure	Training					Testing						
			MAE (mm)	RMSE (mm)	NES	R ²	d	Max(RE)	MAE (mm)	RMSE (mm)	NES	R ²	d	Max(RE)
Category i, Input variables, T _{mean} (°C)														
Scenario I	M5Tree	–	1.109	1.407	0.644	0.803	0.882	6.043	1.238	1.561	0.594	0.780	0.873	8.160
	RSM	–	1.156	1.443	0.626	0.791	0.874	6.423	1.240	1.549	0.600	0.783	0.875	8.260
	RM5Tree	RF=5	1.110	1.393	0.649	0.818	0.875	6.364	1.234	1.548	0.600	0.784	0.876	8.349
		RF=10	1.104	1.388	0.651	0.819	0.876	6.298	1.224	1.540	0.604	0.785	0.875	8.387
		RF=20	1.104	1.388	0.651	0.819	0.876	6.276	1.225	1.541	0.604	0.786	0.876	8.366
		RF=50	1.102	1.387	0.652	0.820	0.877	6.292	1.226	1.544	0.602	0.784	0.875	8.377
Category ii, Input variables, W(m/s)														
Scenario I	M5Tree	–	1.788	2.200	0.131	0.361	0.475	9.819	1.849	2.321	0.101	0.364	0.445	10.529
	RSM	–	1.819	2.220	0.115	0.338	0.438	9.060	1.859	2.323	0.099	0.368	0.427	10.481
	RM5Tree	RF=5	1.766	2.160	0.127	0.357	0.462	9.723	1.851	2.320	0.102	0.368	0.439	10.904
		RF=10	1.765	2.157	0.129	0.360	0.467	9.955	1.850	2.316	0.106	0.372	0.446	10.326
		RF=20	1.765	2.157	0.129	0.360	0.467	9.805	1.849	2.316	0.105	0.372	0.446	10.545
		RF=50	1.766	2.156	0.130	0.362	0.470	9.843	1.849	2.316	0.105	0.372	0.447	10.175
Category iii, Input variables, SR(langley)														
Scenario I	M5Tree	–	0.869	1.211	0.737	0.858	0.919	6.968	1.034	1.441	0.654	0.810	0.890	8.680
	RSM	–	1.005	1.342	0.677	0.822	0.895	8.122	0.963	1.330	0.705	0.841	0.903	7.319
	RM5Tree	RF=5	0.951	1.272	0.678	0.823	0.896	8.206	0.950	1.312	0.713	0.845	0.907	7.095
		RF=10	0.949	1.271	0.678	0.824	0.896	8.183	0.955	1.324	0.708	0.842	0.904	7.209
		RF=20	0.948	1.270	0.679	0.824	0.896	8.152	0.952	1.320	0.709	0.843	0.905	7.239
		RF=50	0.947	1.270	0.679	0.824	0.896	8.145	0.953	1.319	0.710	0.843	0.905	7.263
Category iv, Input variables, RH _{mean} (%)														
Scenario I	M5Tree	–	1.411	1.665	0.502	0.708	0.812	5.792	1.548	1.841	0.434	0.665	0.783	7.775
	RSM	–	1.467	1.707	0.476	0.690	0.798	6.580	1.584	1.851	0.429	0.660	0.772	7.257
	RM5Tree	RF=5	1.398	1.638	0.497	0.705	0.803	5.800	1.547	1.821	0.446	0.673	0.781	6.806
		RF=10	1.389	1.627	0.504	0.710	0.807	5.850	1.547	1.828	0.443	0.670	0.783	7.229
		RF=20	1.386	1.626	0.505	0.711	0.808	5.902	1.545	1.826	0.443	0.671	0.785	7.134
		RF=50	1.386	1.625	0.506	0.711	0.809	5.885	1.546	1.827	0.443	0.671	0.784	7.114
Category v, Input variables, T _{mean} (°C), SR(langley)														
Scenario II	M5Tree	–	0.446	0.683	0.916	0.957	0.978	4.808	0.966	1.338	0.701	0.864	0.926	6.964
	RSM	–	0.775	1.042	0.805	0.897	0.943	5.841	0.832	1.122	0.790	0.893	0.943	7.013
	RM5Tree	RF=5	0.362	0.597	0.936	0.967	0.983	5.845	0.864	1.146	0.781	0.885	0.938	7.379
		RF=10	0.293	0.493	0.956	0.978	0.989	5.156	0.826	1.098	0.799	0.900	0.947	6.202
		RF=20	0.292	0.495	0.956	0.978	0.989	5.448	0.818	1.101	0.798	0.903	0.948	5.755
		RF=50	0.273	0.465	0.961	0.980	0.990	5.448	0.817	1.099	0.799	0.903	0.948	5.809
Category vi, Input variables, T _{mean} (°C), SR (langley), RH _{mean} (%)														
Scenario II	M5Tree	–	0.142	0.242	0.989	0.995	0.997	2.720	0.370	0.607	0.939	0.970	0.984	3.699
	RSM	–	0.257	0.409	0.970	0.985	0.992	3.689	0.297	0.469	0.963	0.982	0.991	3.224
	RM5-Tree	RF=5	0.138	0.253	0.988	0.994	0.997	2.774	0.399	0.580	0.944	0.973	0.985	4.305
		RF=10	0.124	0.214	0.992	0.996	0.998	3.312	0.309	0.470	0.963	0.982	0.990	3.391
		RF=20	0.111	0.192	0.993	0.997	0.998	2.068	0.294	0.462	0.964	0.983	0.991	3.180
		RF=50	0.101	0.164	0.995	0.998	0.999	1.694	0.300	0.469	0.963	0.983	0.991	3.051
Category vii, Input variables, T _{mean} (°C),SR(langley), RH _{mean} (%),W(m/s)														
Scenario III	M5Tree	–	0.050	0.074	0.999	1.000	1.000	0.688	0.064	0.110	0.998	0.999	0.999	1.165
	RSM	–	0.066	0.097	0.998	0.999	1.000	0.832	0.074	0.108	0.998	0.999	1.000	0.870
	RM5Tree	RF=5	0.049	0.073	0.999	1.000	1.000	0.728	0.067	0.113	0.998	0.999	0.999	1.405
		RF=10	0.050	0.072	0.999	1.000	1.000	0.855	0.068	0.104	0.998	0.999	1.000	1.077
		RF=20	0.037	0.054	0.999	1.000	1.000	0.333	0.045	0.068	0.999	1.000	1.000	0.562
		RF=50	0.031	0.044	1.000	1.000	1.000	0.297	0.038	0.052	1.000	1.000	1.000	0.428

The bold numbers are the best statistics among others; RMSE and MAE values are in mm

Contrary to the well-known MLPNN, the RBFNN possess only one hidden layer with large number of neurons, and each one implements a radial basis function, generally the Gaussian function (Tenenbaum et al. 2020). The first input layer transfers the predictor

variables to the hidden layer directly and the only output neuron linearly combines the weighted results of all hidden units (Fig. 3). The Gaussian activation function can be expressed as follows (Chen et al. 2019a, b; Pham et al. 2020):

Table 5 Comparative results of different modeling approaches for Adana station in training and testing phases

	Models	Structure	Training					Test						
			MAE (mm)	RMSE (mm)	NES	R ²	d	Max (RE)	MAE (mm)	RMSE (mm)	NES	R ²	d	Max (RE)
Category i, Input variables, T _{mean} (°C)														
Scenario I	M5Tree	–	0.809	1.016	0.595	0.771	0.860	4.625	0.875	1.102	0.580	0.762	0.850	7.161
	RSM	–	0.830	1.034	0.580	0.762	0.852	4.575	0.876	1.100	0.581	0.763	0.848	7.276
	RM5Tree	RF=5	0.804	1.007	0.586	0.769	0.860	4.622	0.861	1.084	0.594	0.771	0.853	7.209
		RF=10	0.803	1.006	0.587	0.770	0.860	4.669	0.864	1.091	0.588	0.767	0.852	7.260
		RF=20	0.802	1.005	0.587	0.770	0.860	4.660	0.867	1.094	0.586	0.766	0.852	7.259
		RF=50	0.805	1.007	0.586	0.769	0.859	4.595	0.864	1.088	0.591	0.769	0.852	7.229
Category ii, Input variables, W(m/s)														
Scenario I	M5Tree	–	1.181	1.454	0.171	0.414	0.540	6.191	1.468	1.798	– 0.118	0.234	0.483	7.819
	RSM	–	1.196	1.466	0.157	0.397	0.511	5.930	1.554	4.421	– 5.762	– 0.004	0.133	9.890
	RM5Tree	RF=5	1.193	1.462	0.162	0.402	0.515	6.090	1.548	1.906	– 0.257	0.240	0.494	9.487
		RF=10	1.180	1.453	0.172	0.415	0.541	6.191	1.469	1.796	– 0.116	0.225	0.479	9.198
		RF=20	1.214	1.488	0.132	0.367	0.499	5.925	1.607	1.979	– 0.355	0.239	0.493	10.665
		RF=50	1.180	1.453	0.172	0.414	0.541	6.191	1.477	1.814	– 0.138	0.225	0.482	9.500
Category iii, Input variables, SR(langley)														
Scenario I	M5Tree	–	0.418	0.576	0.870	0.933	0.964	4.085	0.640	0.779	0.790	0.928	0.946	4.475
	RSM	–	0.465	0.625	0.847	0.920	0.957	4.382	0.629	0.745	0.808	0.937	0.950	4.878
	RM5Tree	RF=5	0.414	0.556	0.872	0.933	0.971	4.392	0.628	0.744	0.808	0.937	0.950	4.914
		RF=10	0.413	0.556	0.873	0.933	0.971	4.375	0.630	0.746	0.807	0.938	0.950	4.569
		RF=20	0.413	0.556	0.873	0.933	0.971	4.385	0.632	0.748	0.806	0.938	0.950	4.495
		RF=50	0.414	0.557	0.873	0.933	0.971	4.402	0.626	0.743	0.809	0.937	0.950	4.979
Category iv, Input variables, RH _{mean} (%)														
Scenario I	M5Tree	–	1.169	1.405	0.225	0.475	0.582	6.056	1.308	1.591	0.124	0.429	0.604	6.465
	RSM	–	1.282	1.490	0.129	0.359	0.476	6.807	1.344	1.583	0.132	0.393	0.519	8.842
	RM5Tree	RF=5	1.166	1.376	0.194	0.464	0.567	5.550	1.313	1.552	0.167	0.449	0.594	6.712
		RF=10	1.136	1.360	0.213	0.486	0.597	6.203	1.303	1.569	0.148	0.439	0.599	6.215
		RF=20	1.164	1.373	0.197	0.467	0.568	5.903	1.312	1.550	0.169	0.453	0.597	6.447
		RF=50	1.163	1.373	0.198	0.468	0.569	5.795	1.311	1.550	0.168	0.452	0.597	6.686
Category v, Input variables, T _{mean} (°C), SR(langley)														
Scenario II	M5Tree	–	0.246	0.386	0.942	0.970	0.985	3.486	0.601	0.762	0.799	0.929	0.950	4.726
	RSM	–	0.393	0.538	0.886	0.942	0.969	3.960	0.561	0.657	0.850	0.951	0.962	5.101
	RM5Tree	RF=5	0.395	0.541	0.885	0.941	0.969	3.891	0.554	0.652	0.853	0.950	0.962	5.237
		RF=10	0.172	0.286	0.968	0.984	0.992	2.696	0.559	0.654	0.852	0.952	0.962	4.943
		RF=20	0.164	0.275	0.970	0.985	0.992	2.374	0.553	0.651	0.849	0.953	0.962	4.776
		RF=50	0.156	0.267	0.972	0.986	0.993	2.770	0.572	0.674	0.843	0.952	0.961	4.476
Category vi, Input variables, T _{mean} (°C), SR (langley), RH _{mean} (%)														
Scenario II	M5Tree	–	0.154	0.241	0.977	0.989	0.994	1.610	0.419	0.581	0.883	0.954	0.971	4.726
	RSM	–	0.238	0.339	0.955	0.977	0.988	2.435	0.313	0.424	0.938	0.980	0.985	3.061
	RM5Tree	RF=5	0.122	0.196	0.985	0.992	0.996	1.659	0.331	0.445	0.932	0.971	0.983	3.389
		RF=10	0.098	0.164	0.989	0.995	0.997	2.517	0.305	0.420	0.939	0.979	0.985	3.117
		RF=20	0.080	0.124	0.994	0.997	0.998	1.380	0.167	0.325	0.963	0.984	0.991	5.310
		RF=50	0.083	0.142	0.992	0.996	0.998	1.802	0.273	0.389	0.948	0.979	0.987	3.069
Category vii, Input variables, T _{mean} (°C), SR(langley), RH _{mean} (%), W(m/s)														
Scenario III	M5Tree	–	0.079	0.135	0.993	0.996	0.998	1.860	0.175	0.399	0.945	0.973	0.986	9.234
	RSM	–	0.037	0.055	0.999	0.999	1.000	0.652	0.071	0.533	0.902	0.952	0.976	11.300
	RM5Tree	RF = 5	0.045	0.071	0.998	0.999	1.000	0.973	0.095	0.165	0.991	0.996	0.998	2.779
		RF = 10	0.056	0.086	0.997	0.999	0.999	0.791	0.106	0.199	0.986	0.993	0.997	7.233
		RF = 20	0.035	0.054	0.999	0.999	1.000	0.977	0.042	0.134	0.994	0.997	0.998	4.784
		RF = 50	0.034	0.050	0.999	1.000	1.000	0.645	0.038	0.122	0.996	0.998	0.998	3.023

The bold numbers are the best statistics among others; RMSE and MAE values are in mm

$$\varphi(x_k, \mu_i) = \exp\left(-\frac{1}{2\sigma_i^2} \sum_{m=1}^M (x_k - \mu_i)^2\right) \quad (4)$$

Where $\mu_i = [\mu_{i1}, \mu_{i2}, \dots, \mu_{iM}]$ is the center of Gaussian function, $x_k = [x_{k1}, x_{k2}, \dots, x_{kN}]$ is a training sample and

σ^2 is the width of the RBFNN neuron, also called the spread width. During the training process, the optimal number of RBFNN neurons, the values of the centers, the weights, and biases were determined by minimizing the mean squared errors between observed and modeled values of ET₀ (Zhou et al. 2012; Bonanno et al. 2012)

Table 6 Comparing the results of the best RM5Tree model with the best structures of well-known Radial basis neural network (RBFNN) and Multi-Layer Perceptron Neural Network (MLPNN) in the testing phases for the three meteorological stations

Station	Best Scenario	Best structure	Statistics			
			MAE (mm)	RMSE (mm)	d	NSE
Isparta	III	RM5Tree(4,50,1)	0.038	0.051	1.000	0.999
		RBNN(4,50,1)	0.239	0.338	0.990	0.959
		MLPNN(4,7,1)	0.085	0.116	0.999	0.995
Antalya	III	RM5Tree(4,50,1)	0.038	0.052	1.000	1.000
		RBNN(4,50,1)	0.302	0.540	0.987	0.951
		MLPNN(4,7,1)	0.125	0.193	0.998	0.994
Adana	III	RM5Tree(4,50,1)	0.054	0.080	0.999	0.998
	II	RBNN(4,50,1)	0.333	0.468	0.981	0.924
	III	MLPNN(4,8,1)	0.063	0.171	0.997	0.990

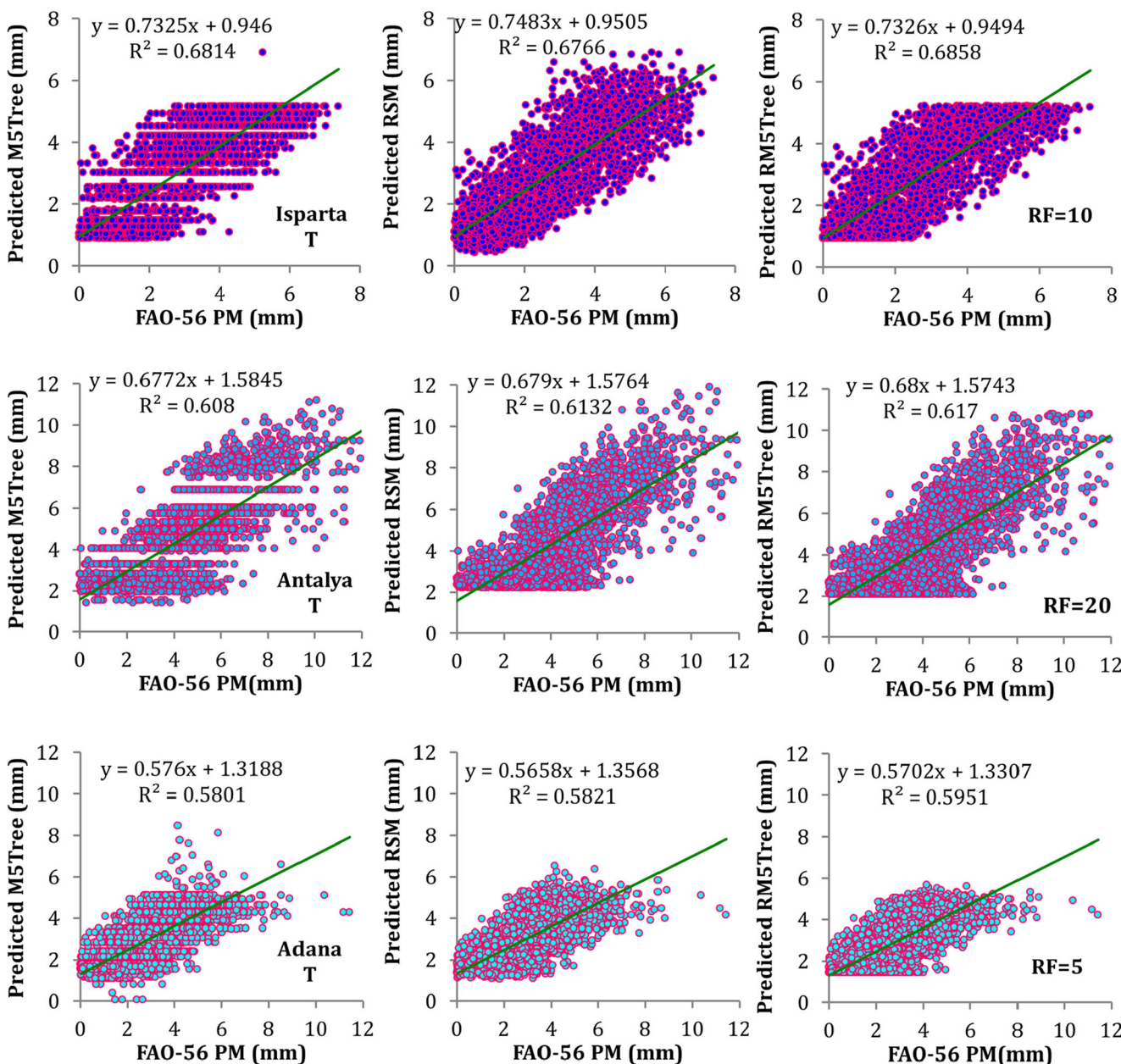


Fig. 6 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of mean temperature (T) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

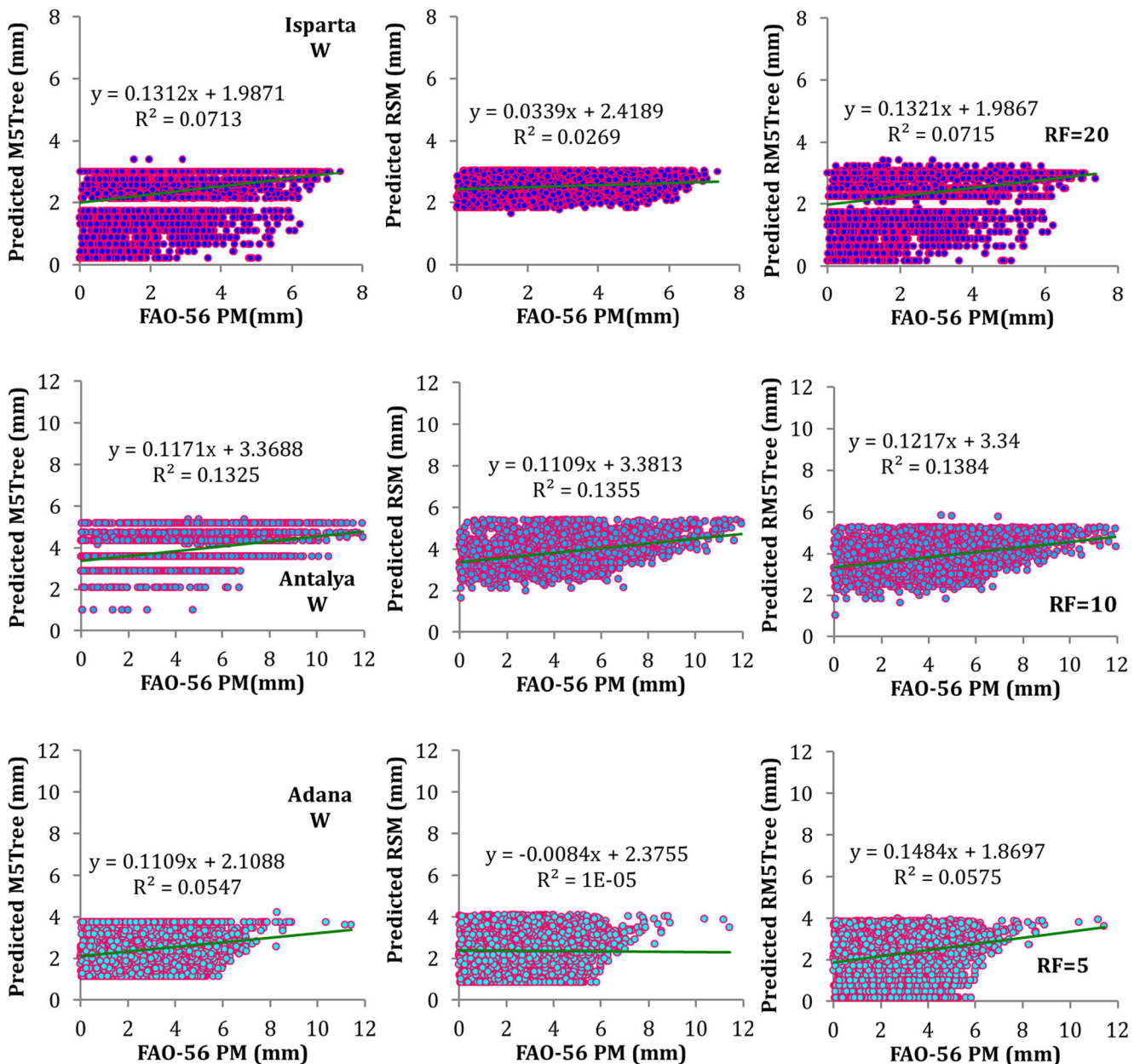


Fig. 7 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of wind speed (W) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

2.2.4 M5 tree model

M5 tree model is subset basis data mining and machine learning method. The tree-based methods are indeed a part of data mining methods, the output of which resulted from application of the input and output data will be a model with tree structure (Solomatine and Xue 2004; Zahiri et al. 2020). The tree models are fundamentally based on the decision and dominance method. Substituting the linear regression equation at the nodes is a method executed in the M5 model, which is capable to predict or estimate the numerical variables. Structure of a decision tree is similar to a

tree constituted of the root, branches, nodes, and leaves. A tree model is built up in two steps. Accordingly, in the first step, the decision tree is designed by data splitting. The split criterion in M5 model is to maximize the reduction of standard deviation (SDR) of the data at the offspring node. When no reduction of standard deviation of the data at the offspring node is possible, its parent node will not be split and, thus, reach the end node or leaf. The following formula is used to calculate SDR:

$$SDR = sd(T) - \sum_{i=1}^N \frac{T_i}{T} sd(T_i) \quad (5)$$

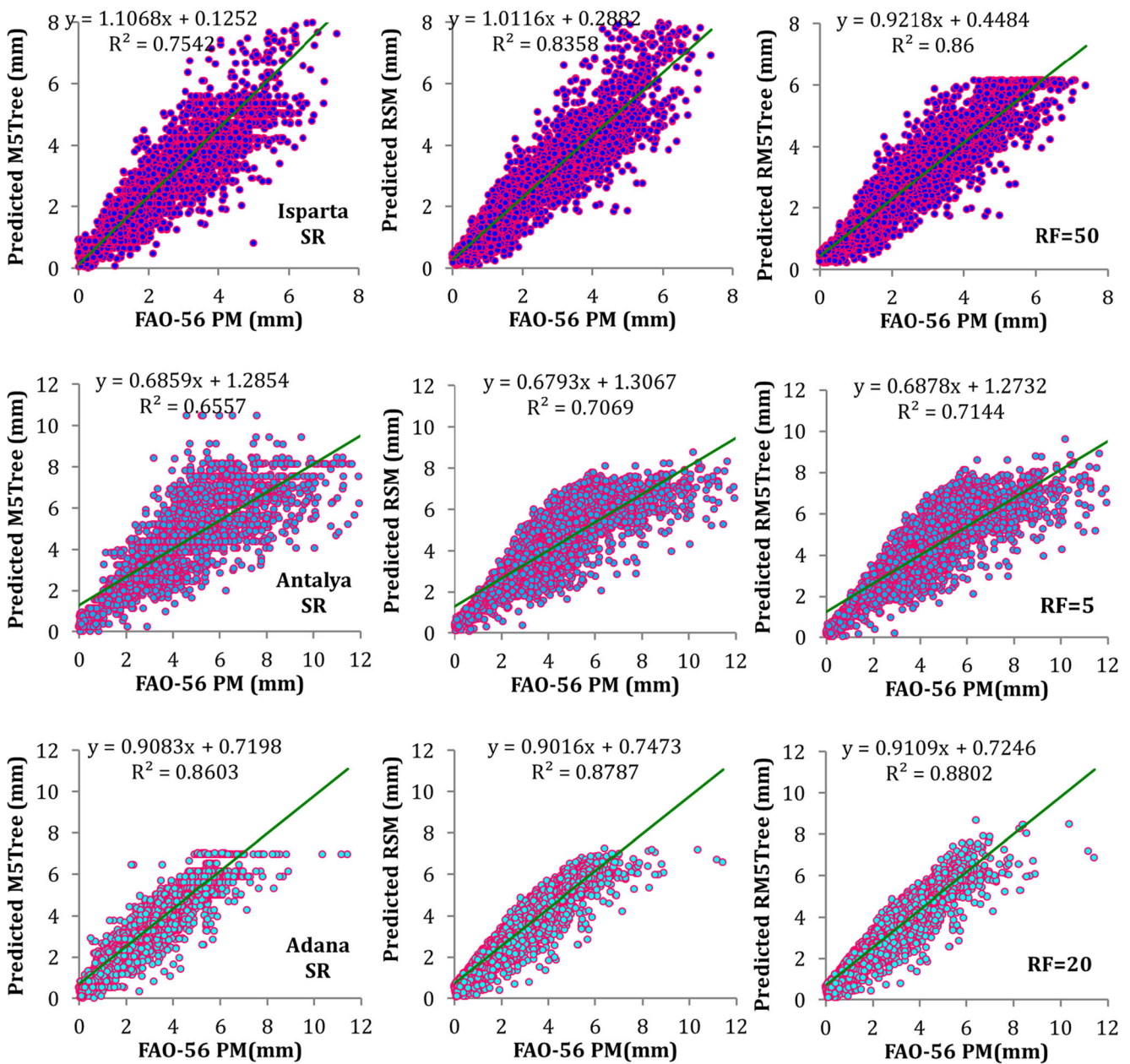


Fig. 8 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of solar radiation (SR) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

where T represents a set of the samples entering on each node, T_i represents a subset of the samples with the i^{th} result of the potential test, and Sd is standard deviation of the input data, which can be calculated as follows:

$$Sd(T) = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N (y_i)^2 - \frac{1}{N} \left(\sum_{i=1}^N y_i \right)^2 \right)} \quad (6)$$

Y_i is a numerical value of the target attribute of sample i and N indicates the number of data. Since the process of branching (classification) at offspring nodes has less standard deviation than the parent nodes, they have more accurate results and are

featured with higher homogeneity. Once all the possible classifications are examined, the M5 model selects the one with minimum expected error. However, the second step in designing a tree model is to shrink the overgrown and overfitted tree through pruning the branches and replacing them with linear regression functions (Rahimikhoob 2016).

3 Radial M5 model tree

To enhance the accuracy of ET_0 estimations, radial basis M5Tree (RM5Tree) is introduced. The input data set is

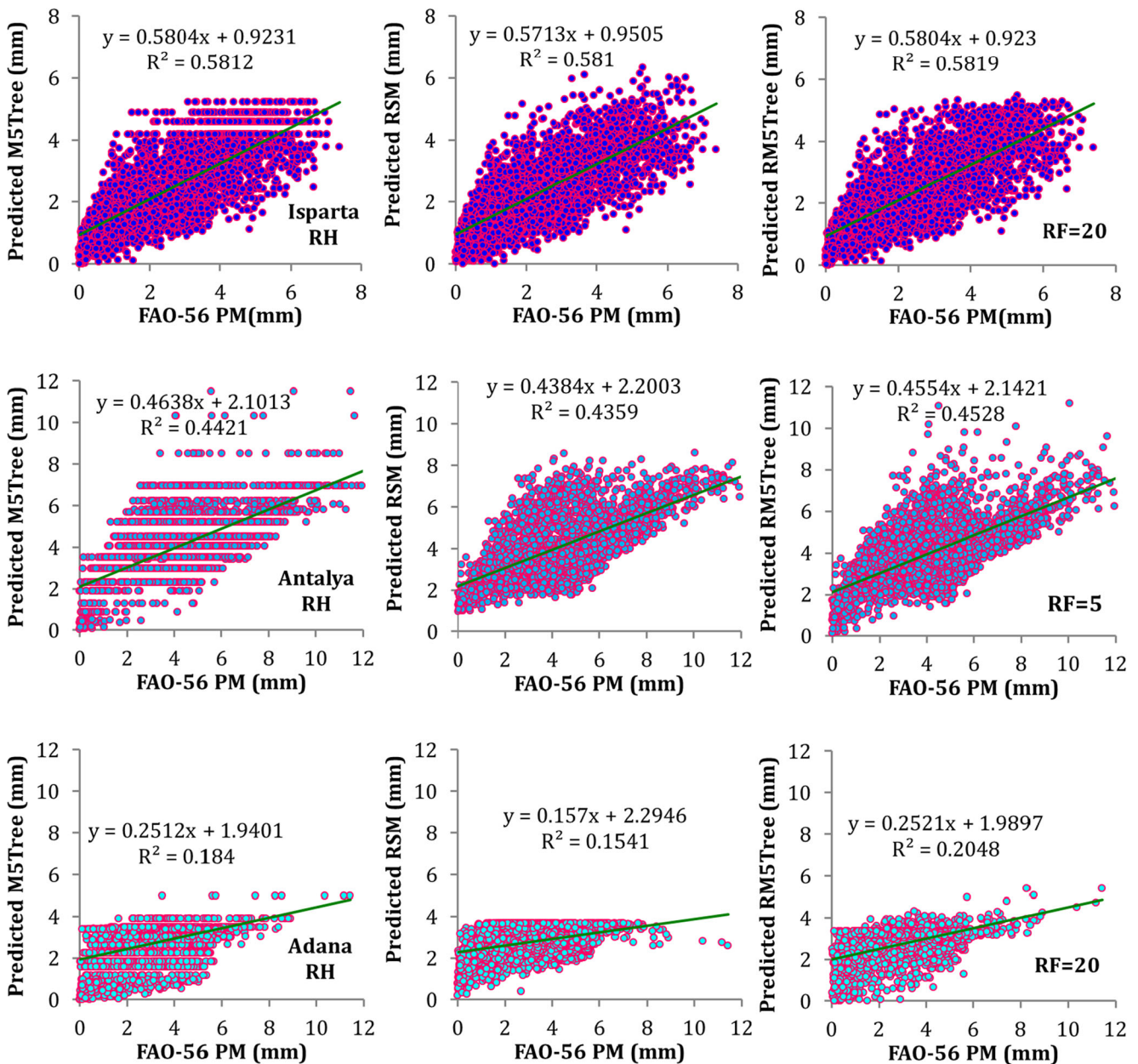


Fig. 9 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of relative humidity (RH) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

controlled by applying the radial basis function (RBF) in feature space. By transferring input data from original to radial map, the RBF is applied in RM5Tree as follows (Chen et al. 1991; Xiao et al. 2020; Zhang et al. 2020a):

$$K_{ij} = \varphi(\|Z_i - C_j\|, \varepsilon) = \exp(-\varepsilon \|Z_i - C_j\|^2) \quad i = 1, \dots, NV \quad j = 1, \dots, RF \quad (7)$$

where RF is number of radial sets with shape factor of ε , and C represents the center of RBF. Z is normalized map, which is given as follows (Zhang et al. 2019; Keshtegar and Kisi 2018):

$$Z = \frac{X - \mu_x}{\sigma_x} \quad (8)$$

where, μ_x and σ_x are respectively mean and SD of input data x . The radial transformations which are given by Eq. (7) with $C=0$ and $\varepsilon=0.5$ are shown in Fig. 4 that shows a nonlinear map. Thus, there can be utilized a new data set to train a model by transforming original data set from NV (X -space) to RF (radial-space).

Two parameters of shape and location as center points applied in RBF are selected as $\varepsilon=0.5$ and $C=[X_{min} \ X_{max}]$ which are randomly given from the domain of input dataset with RF

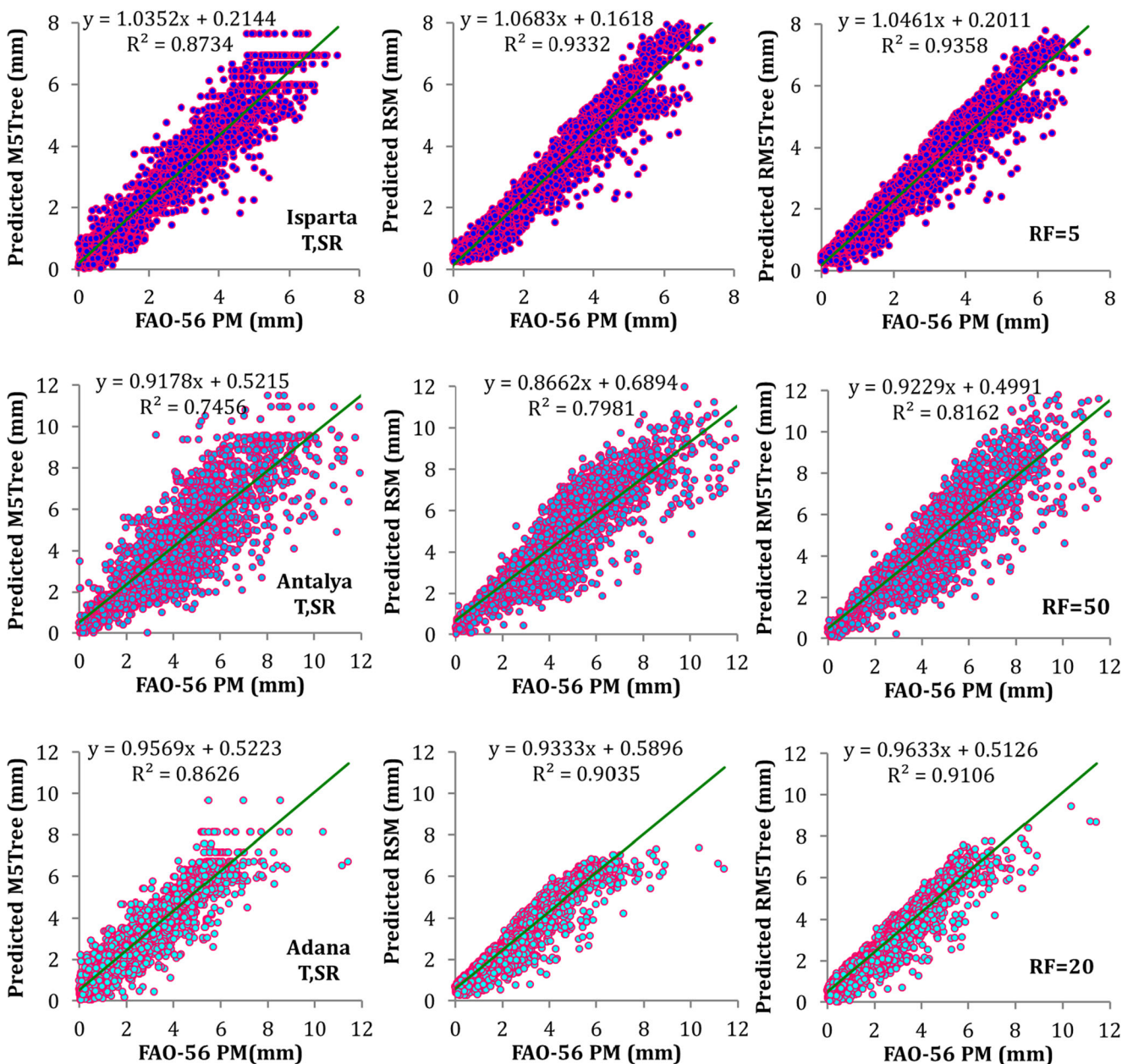


Fig. 10 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of mean temperature (T) and solar radiation (SR) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

as 5, 10, 20, and 50 in this study. The schematic structure of RM5Tree is presented in Fig. 5. This model involves three layers as input, transferring, and modeling layers. Using Eq. (7), the input dataset is normalized in the first (input) layer, while the RF-dataset is provided by transferring data to the second layer as follows:

- 1) Create RF- center point from domain of each input data, randomly.
- 2) Transfer the input data set in layer 1 into radial space by using RBF given in Eq. (7) based on the RF- center point as follows:

$$Z = \begin{bmatrix} z_{1,1} & z_{1,2} & \dots & z_{1,NV} \\ z_{2,1} & z_{2,2} & \dots & z_{2,NV} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N,1} & z_{N,2} & \dots & z_{N,NV} \end{bmatrix} \rightarrow K = \begin{bmatrix} K_{1,1} & K_{1,2} & \dots & K_{1,RF} \\ K_{2,1} & K_{2,2} & \dots & K_{2,RF} \\ \vdots & \vdots & \ddots & \vdots \\ K_{N,1} & K_{N,2} & \dots & K_{N,RF} \end{bmatrix} \quad (9)$$

where, N is number of data in the training stage as 65% of total data, number of input variables and number of radial input data and K_{ij} $i = 1, 2, \dots, N$ $j = 1, 2, \dots, RF$. The radial input data is used in the training phase of M5Tree models. Therefore, the applied nonlinear map using Gaussian function and the number of center points improve the accuracy of M5Tree models (Keshtegar et al. 2018).

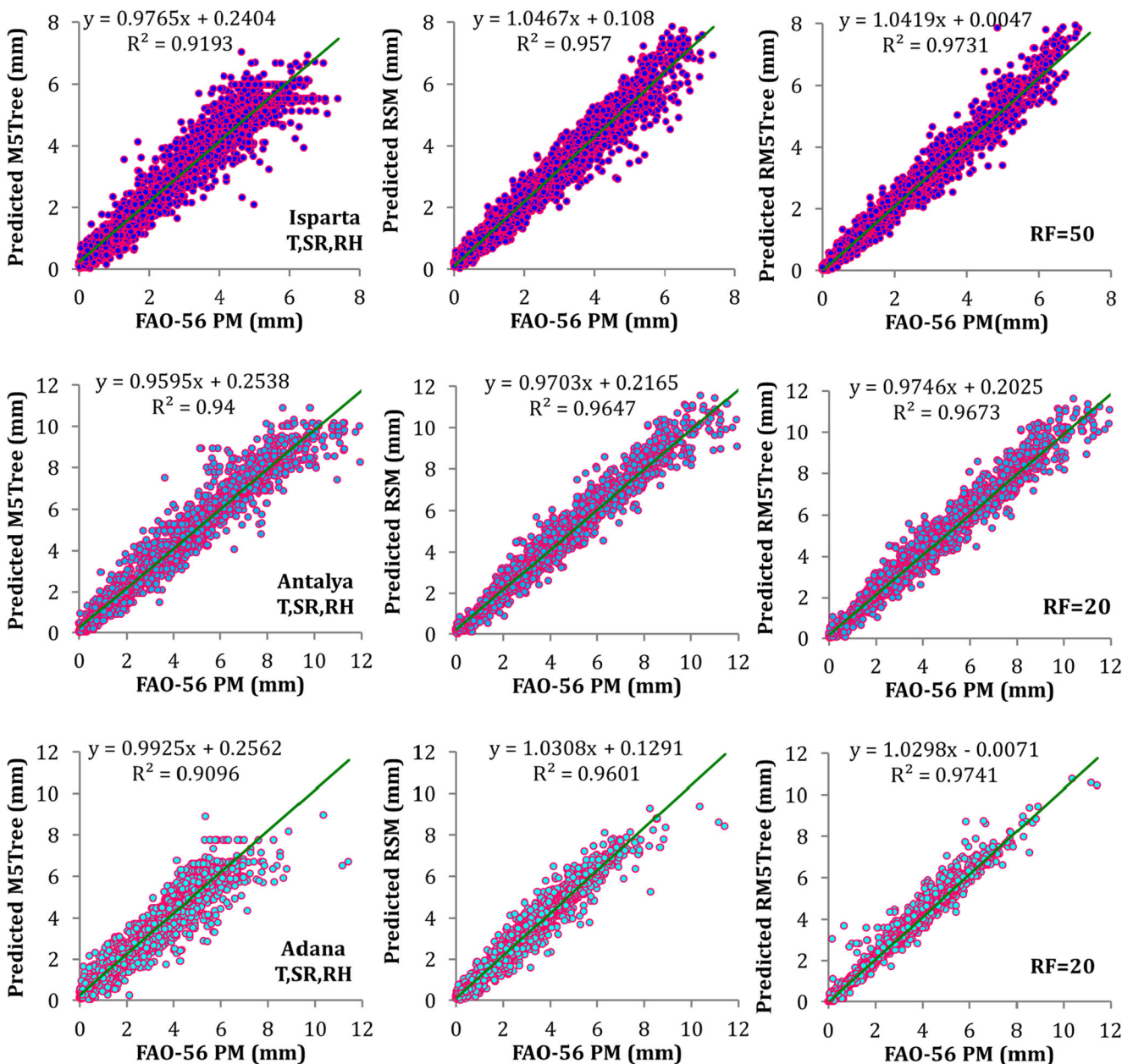


Fig. 11 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of mean temperature (T), solar radiation (SR) and relative humidity (RH) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

4 Application of the models

4.1 Modeling scenarios

Based on the results of Table 1 given for the correlation coefficients between the independent variables (T , RH , SR & W) and dependent variable (ET_0), three different modeling scenarios for constructing the machine learning methods (M5Tree, Radial M5Tree, MLPNN, RBFNN & RSM) are considered. These scenarios are tabulated in Table 2; in the first scenario, just one input parameter is considered for modeling ET_0 including (i) T_{mean} ; (ii) W ; (iii) SR ; (iv) RH .

The second scenario takes into account the most correlated parameters including (v) T_{mean} , SR and, (vi) T_{mean} , SR , RH and, (vii) T_{mean} , SR , RH . Finally, the third scenario has all of the independent parameters as (vii) T_{mean} , SR , RH , W .

4.2 Evaluation of the models

The models' accuracies were compared according to the mean absolute error (MAE), determination coefficient (R^2), root mean square error ($RMSE$), agreement index (d), and Nash and Sutcliffe efficiency (NES) statistics (Xiao et al. 2019; Zhang et al. 2020b).

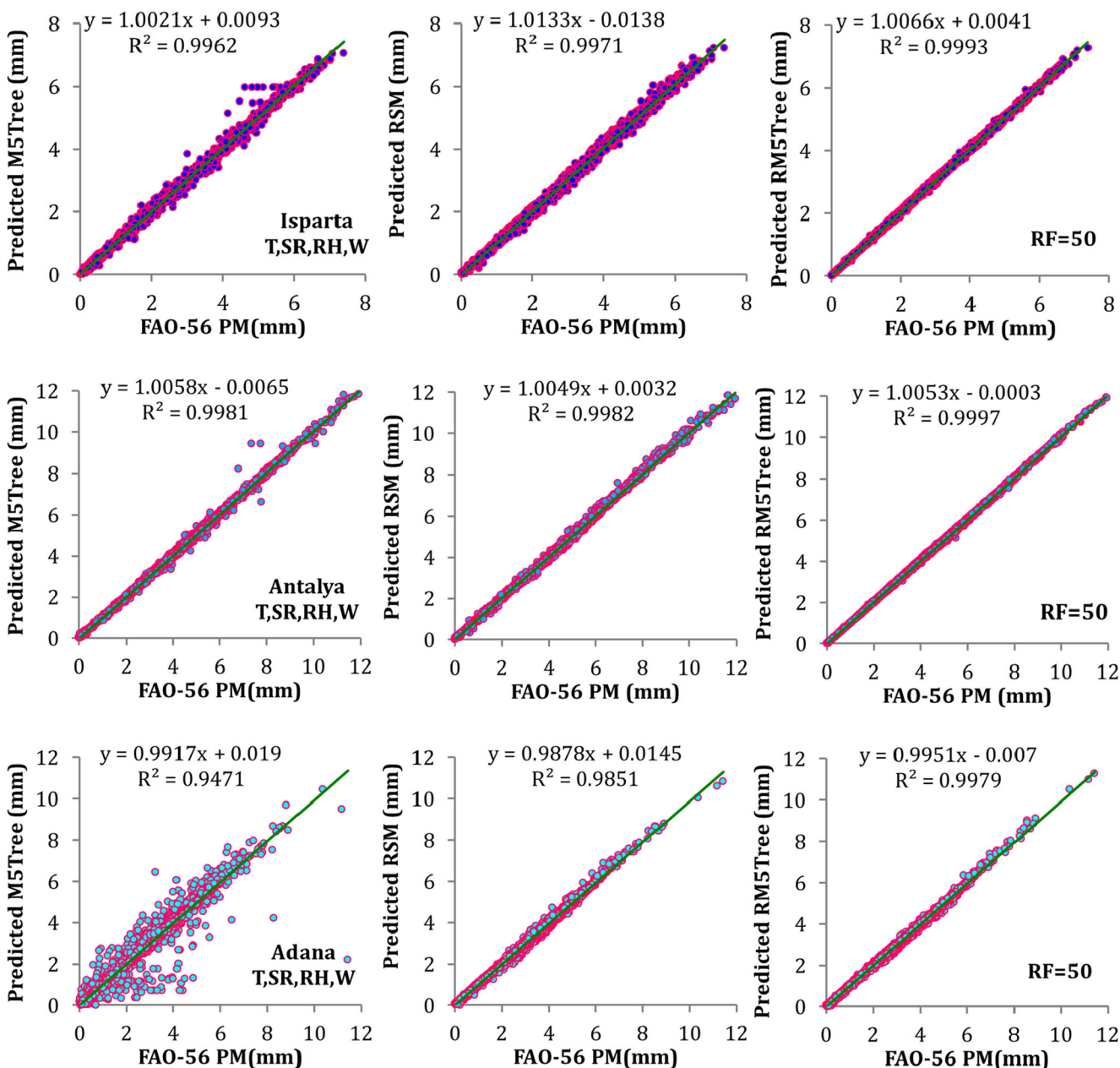


Fig. 12 Scatterplot of the M5Tree, RSM, and RM5Tree models based on the input data of mean temperature (T), solar radiation (SR), relative humidity (RH) and wind speed (W) in test (35% from all data) period for Isparta, Antalya, and Adana stations (RF: number of radial function)

$$MAE = \frac{1}{N} \sum_{i=1}^N |(ET_0)_i - (ET_p)_i| \tag{10}$$

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N [(ET_0)_i - (ET_p)_i]^2} \tag{11}$$

$$d = 1 - \frac{\sum_{i=1}^N [(ET_0)_i - (ET_p)_i]^2}{\sum_{i=1}^N [|(ET_0)_i - \overline{ET_0}| + |(ET_p)_i - \overline{ET_0}|]^2}, 0 < d \leq 1 \tag{12}$$

$$NES = 1 - \frac{\sum_{i=1}^N [(ET_0)_i - (ET_p)_i]^2}{\sum_{i=1}^N [(ET_0)_i - \overline{ET_0}]^2}, -\infty < NES \leq 1 \tag{13}$$

$$R^2 = \left[\frac{\frac{1}{N} \sum ((ET_0)_i - \overline{ET_0}) ((ET_p)_i - \overline{ET_p})}{\sqrt{\frac{1}{N} \sum ((ET_0)_i - \overline{ET_0})^2} \sqrt{\frac{1}{N} \sum ((ET_p)_i - \overline{ET_p})^2}} \right]^2, 0 < R^2 \leq 1 \tag{14}$$

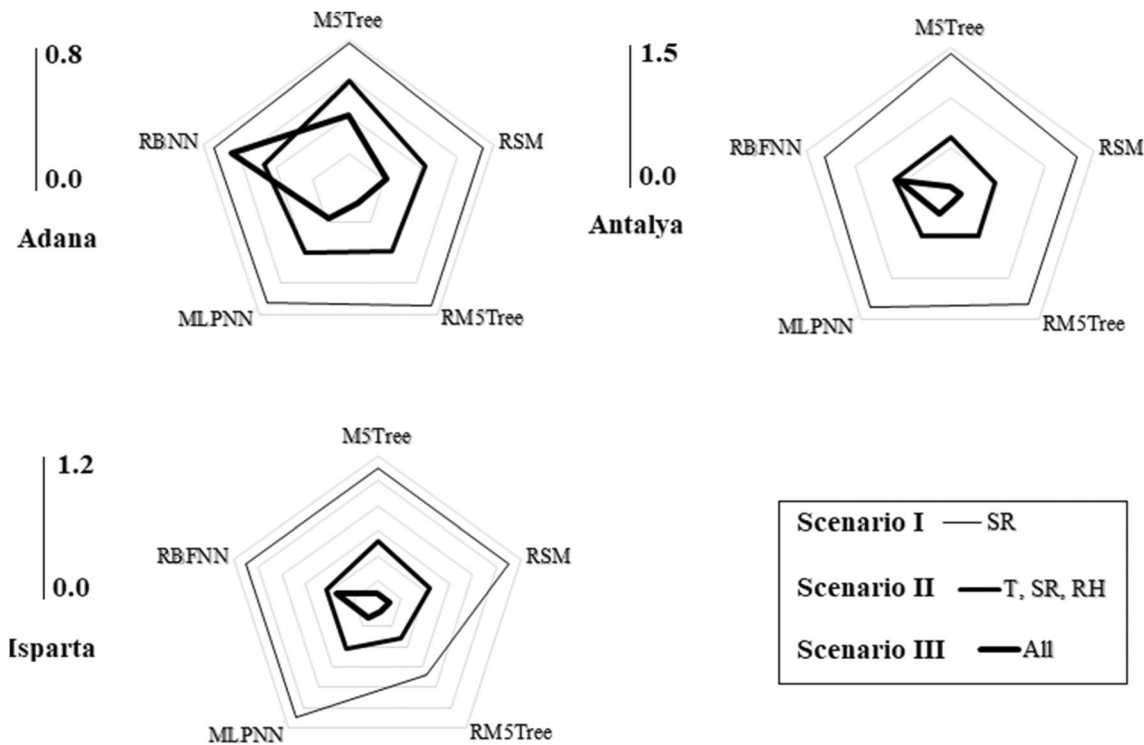


Fig. 13 Radar chart for the best calculated values of RMSE (mm) for the applied models using the three input scenarios

In which, N is the number of data, ET_0 , ET_p , $\overline{ET_0}$, $\overline{ET_p}$ are the FAO-56 PM ET_0 , predicted ET_0 mean ET_0 , and mean predicted ET_0 , respectively.

5 Results and discussion

5.1 Result analysis for the Isparta Station

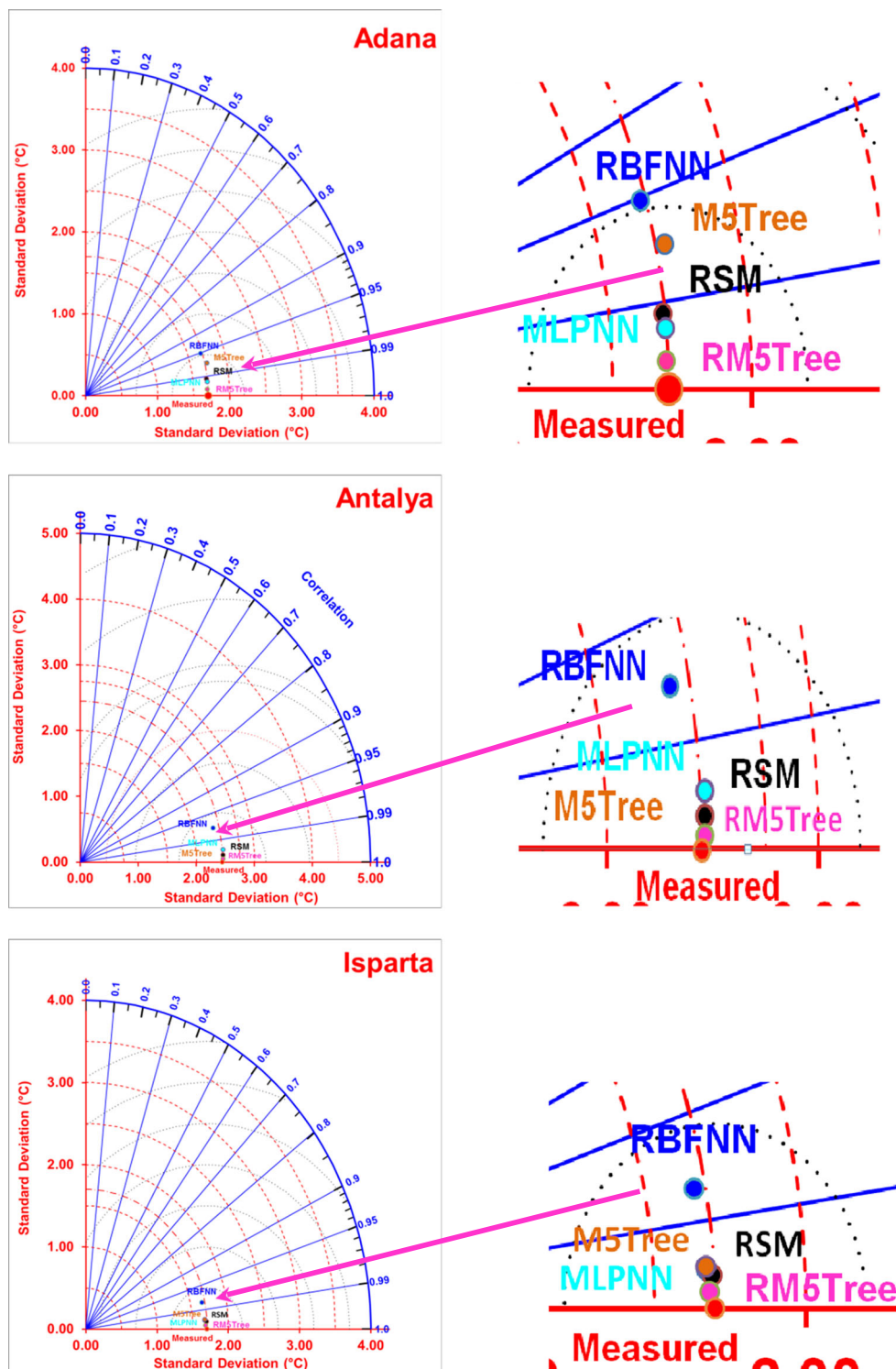
The final results of the investigated AI-based models (RSM, M5Tree, and RM5Tree) in terms of training and testing results for Isparta Station can be seen in Table 3. It can be seen that the RM5Tree model performs superior to the M5Tree and RSM models with respect to various criteria in all input combinations (Scenario III). In testing phase, the RMSE is improved (d) as accuracy (tendency) factors using proposed RM5Tree by about 42% (6%) and 15% (2%) for Scenario I, 75% (15%) and 60% (3%) for Scenario II, and 105% (1%) and 90% (1%) for Scenario III compared to M5tree and RSM models, respectively. Considering Scenario (I) implies that among the single input variables, SR is the most effective parameter on ET_0 followed by T_{mean} and RH , respectively while W has the least effect. This result was actually expected according to the calculated correlation coefficients in Table 3. In the second scenario (II), including the combination of SR (T_{mean}) parameter with T_{mean} (SR) considerably improves the models' accuracy. For example, it improved the MAE, RMSE, and NES of RM5Tree by 42% (13%), 43% (19%), and 34% (7%), respectively. Adding

RH parameter to T_{mean} and SR inputs also increases the accuracy of the employed models. For example, the values of MAE and RMSE of RM5Tree were decreased from 0.455 and 0.551 mm to 0.215 and 0.32 mm by 52% and 42%, respectively. In scenario III — even though W seems to be the least effective parameter from the first four input combinations — adding W parameter to other three inputs considerably increases the MAE and RMSE of the models (MAE and RMSE of the RM5Tree increased by 82% and 84%, respectively). According to the results of scenario (III), the accuracy of M5Tree model with respect to MAE and RMSE was improved by 43% and 51% using RM5Tree, respectively.

5.2 Result analysis for the Antalya Station

Table 4 reports the comparative results of the models in estimating ET_0 of Antalya Station. Similar to Isparta, RM5Tree model outperforms the other models. From the first scenario (categories i to iv), the effective parameters (from most to least) in modeling ET_0 are SR , T_{mean} , RH , and W . The accuracy of the RM5tree with respect to MAE, RMSE, and NES is improved up to 10% (15%), 48% (39%), and 50% (27%) by adding the SR input, respectively. Similarly, importing RH to T_{mean} and SR inputs decreases the MAE and RMSE of RM5Tree by 73% and 42%, respectively. Moreover, including W input in T_{mean} , SR and RH combination considerably increases the RM5Tree accuracy (MAE and RMSE are decreased by 87% and 89%) in scenario (III).

Fig. 14 Taylor diagram displaying a statistical comparison of the proposed models with FAO-56 PM (mm). The green circles correspond to circumferences of equal centered normalized root-mean-square (NRMS) difference between measured and calculated ET_0 , the blue lines correspond to lines of equal correlation coefficients, and dotted red circles correspond to circumferences of equal standard deviations



5.3 Result analysis for the Adana Station

Table 5 compares the training and testing statistics of the three methods for Adana Station. Similar to the Isparta, in this station, the RM5Tree model gave the best accuracy in modeling ET_0 with respect to various evaluation statistics. According to

the single input combinations in scenario (I), the most effective variables on ET_0 is SR followed by T_{mean} and RH . Using SR (T_{mean}) parameter with T_{mean} (SR) input improves the RM5Tree accuracy with respect to MAE, RMSE and NES by 35% (11%), 40% (12%) and 43% (5%) in the test period, respectively. Including RH variable as an input factor to the

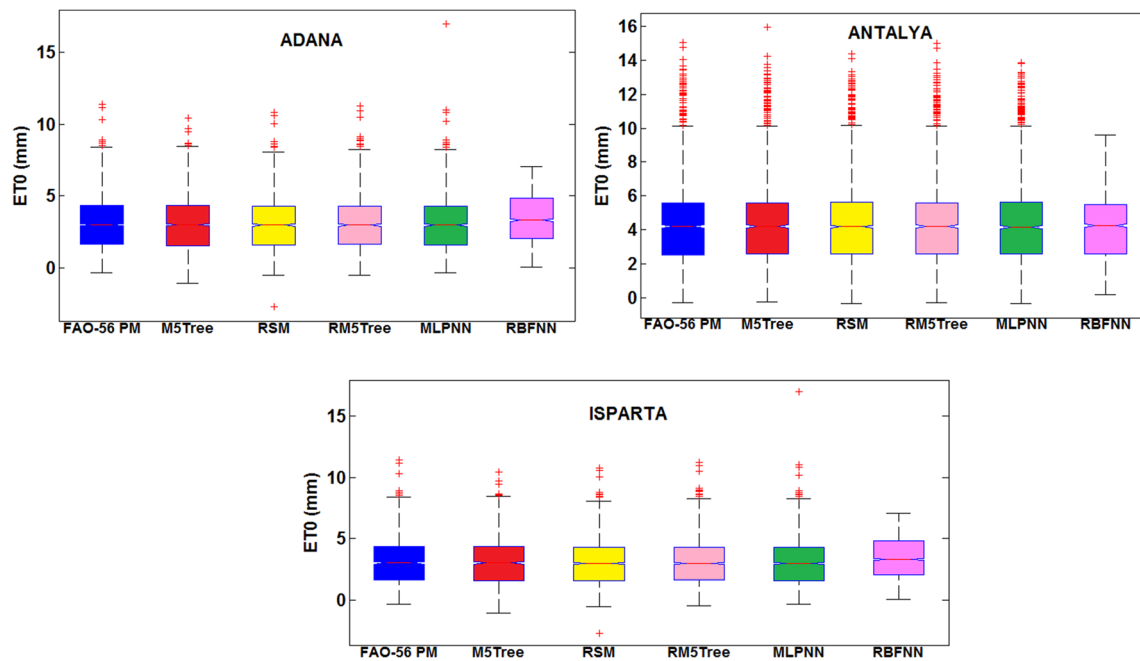


Fig. 15 Box plots of FAO-56 PM and calculated values of ET_0 in the test phase of all stations. The box stretches from the 25th percentile to the 75th percentile. The median is shown as a red line, and the whiskers correspond to the most extreme data points

RM5Tree comprising T_{mean} and SR inputs decreases the MAE and RMSE of the model from 0.553 and 0.671 mm to 0.167 mm and 0.325 mm by 70% and 50%. Similarly, importing W parameter to three inputs (T_{mean} , SR and RH) in the third scenario considerably increases RM5Tree accuracy, from 0.167 mm to 0.038 (MAE) and from 0.325 mm to 0.122 mm (RMSE), respectively.

5.4 Discussion

The ET_0 estimates of the M5Tree, RSM, and RM5Tree models are compared in Figs. 6, 7, 8, 9, 10, 11, and 12 for each station and each input combination. The effect of each variable on ET_0 can be seen from the figures visually. Comparison of Figs. 6 and 7 indicates that the effective ranks of the variables (from the most to the least) are SR , T_{mean} , RH and W . Jain et al. (2008) also reported the same trend for the effective parameters (SR , T_{mean} , RH , W and lastly dew point temperature) by using hourly data of ET_0 for a few stations in the Reynolds Creek Experimental Watershed in South-western Idaho, USA. In addition, the effect of each variable on ET_0 can also be observed from Figs. 8 and 9. Comparison of Figs. 10 and 11 shows the considerable effect of W variable even though this cannot be seen when W is used as input alone. One input model cannot catch the relationship between W and ET_0 . All these indicate the necessity of this variable in accurately modeling of ET_0 . It should be noted that the M5Tree model estimates are not accurate in Adana compared to other stations and methods. The reason of this might be the fact that the relationship between inputs and output is more non-linear in Adana compared to others and the M5tree model having linear

nature might not adequately map this highly non-linear relationship. Table 6 compares the results of the best RM5tree model with two of the most prevailing AI-based models of MLPNN and RBNN (multi-layer perceptron neural network and radial basis neural network). It can be concluded that all the AI-based models acted better by considering all the input variables considering scenario III (with the exception for the RBNN in Adana Station). Although the MLPNN model gave better results than the RBNN models but it could not surpass the performance of the proposed RM5tree model. Having a better diagnostic analysis of the efficiency of the all AI-based models (M5Tree, RM5Tree, RSM, MLPNN & RBNN), the results of the best input category in scenarios I, II, and III in terms of RMSE (mm) are shown in Fig. 13 using radar charts. Obviously, the smaller size of stars with lower values for RMSE would indicate the better performance of the models. It can be easily seen that involving all the variables (T , SR , RH , W in scenario III) would result in lower values of RMSE (with an exception for the RBNN in Adana Station). This major finding is supported by the outcomes of different AI-based model in a similar study done by Kisi (2006). Further evaluation was achieved using the Taylor diagram to check the performances of the models as presented in Fig. 14. At all stations RM5Tree performs better than the other models, it is clear evidence from the results of Fig. 14 that the proposed approach improves the accuracy of the M5Tree model. Finally, to further compare the accuracy of the models, all the results using the best input combination for each model has been considered using the Box plot as plotted in Fig. 15. Box plots corresponding to the test data in Fig. 15 clearly show that the accuracy of the RM5Tree model was higher than the other models.

6 Conclusion

In the presented work, the applicability of a new method which is developed by combining radial basis function and M5Tree methods was investigated in modeling ET_0 . The new method was compared with standard M5Tree, RSM, MLPNN, and RBNN using daily climatic data from three stations located in Turkey. Various input combinations of available data were tried to see the effect of each input variable on ET_0 . The following conclusions were derived from the applications.

- i- The comparison of methods revealed that the new proposed method, RM5Tree, provided better ET_0 estimates than the MLPNN, RBNN, M5Tree, and RSM. The accuracy of M5Tree models was considerably improved (more than 30% with respect to MAE and RMSE) by using RM5Tree.
- ii- The results obtained based on different input combinations indicated that the most effective variable on models' accuracy in estimating ET_0 was solar radiation followed by the air temperature, relative humidity, and wind speed. However, it was also observed that using wind speed together with other three inputs considerably increases models' efficiency (more than 80% with respect to MAE and RMSE).
- iii- The study showed that the proposed RM5Tree model could be utilized as a better alternative to the M5Tree in modeling daily ET_0 .
- iv- This ability of this method can be compared with other stations or this method can be applied for other hydrological problems in future.

Acknowledgements The authors would like to thank TSMS (Turkish State Meteorological Service) for providing the data for this study.

Author contribution Ozgur Kisi: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. Behrooz Keshtegar: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. Mohammad Zounemat-Kermani: Conceptualization, Investigation, Writing - review & editing. Salim Heddami: Conceptualization, Writing - original draft & editing. Nguyen-Thoi Trung: Conceptualization, Writing - review & editing.

Funding This work was supported by University of Zabol under Grant No: UOZ-GR-9618-1 and UOZ-GR-9719-1 and Iran National Science Foundation (INSF) under project No: 97023031.

Data availability The daily weather data from Adana, Antalya and Isparta stations, Turkey. The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Code availability The codes of the modeling methods applied in this current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval The authors have the responsibility about the methodology and the novelty, research, ethical standards, and professional guidelines in this contribution.

Consent to participate We as the research team in this current contribution have voluntarily agreed to participate in this research study.

Consent for publication We would like to give consent for the publication of identifiable details including text, material and methods, figures, and tables to be published in the Journal.

Conflict of interest The authors declare no competing interests.

References

- Adnan RM, Heddami S, Yaseen ZM, Shahid S, Kisi O, Li B (2021) Prediction of potential evapotranspiration using temperature-based heuristic approaches. *Sustainability* 13(1):297
- Ahmadi AA, Arabbeiki M, Ali HM, Goodarzi M, Safaei MR (2020) Configuration and optimization of a minichannel using water-alumina nanofluid by non-dominated sorting genetic algorithm and response surface method. *Nanomaterials* 10(5):901
- Antonopoulos VZ, Antonopoulos AV (2017) Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Comput Electron Agric* 132:86–96
- Bonanno F, Capizzi G, Graditi G, Napoli C, Tina GM (2012) A radial basis function neural network based approach for the electrical characteristics estimation of a photovoltaic module. *Appl Energy* 97:956–961
- Chen S, Cowan CF, Grant PM (1991) Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans Neural Netw* 2:302–309
- Chen X, Li FW, Wang YX, Feng P, Yang RZ (2019a) Evolution properties between meteorological, agricultural and hydrological droughts and their related driving factors in the Luanhe River basin, China. *Hydrol Res* 50:1096–1119. <https://doi.org/10.2166/nh.2019.141>
- Chen Z, Yang X, Liu X (2019b) RBFNN-based non-singular fast terminal sliding mode control for robotic manipulators including actuator dynamics. *Neurocomputing* 362:72–82
- Chen Z, Zhu Z, Jiang H, Sun S (2020) Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J Hydrol* 591:125286
- Chia MY, Huang YF, Koo CH (2020) Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. *Comput Electron Agric* 175:105577
- Dou X, Yang Y (2018) Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Comput Electron Agric* 148:95–106
- Fang W, Huang S, Huang Q, Huang G, Meng E, Luan J (2018) Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information. *J Hydrol* 561:764–779
- Fei CW, Li H, Liu HT, Lu C, An LQ, Han L, Zhao YJ (2020) Enhanced network learning model with intelligent operator for the motion reliability evaluation of flexible mechanism. *Aerosp Sci Technol* 107:106342
- Feng Y, Gong D, Mei X, Cui N (2017a) Estimation of maize evapotranspiration using extreme learning machine and generalized regression neural network on the China Loess Plateau. *Hydrol Res* 48(4):1156–1168

- Feng Y, Peng Y, Cui N, Gong D, Zhang K (2017b) Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput Electron Agric* 136:71–78
- Ferreira LB, da Cunha FF (2020) Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Comput Electron Agric* 178:105728
- Gavili S, Sanikhani H, Kisi O, Mahmoudi MH (2018) Evaluation of several soft computing methods in monthly evapotranspiration modelling. *Meteorol Appl* 25:128–138
- Gocić M, Motamedi S, Shamshirband S, Petković D, Ch S, Hashim R, Arif M (2015) Soft computing approaches for forecasting reference evapotranspiration. *Comput Electron Agric* 113:164–173
- Heddad S, Watts MJ, Houichi L, Djemili L, Sebbar A (2018) Evolving connectionist systems (ECoSs): a new approach for modeling daily reference evapotranspiration (ET 0). *Environ Monit Assess* 190(9): 516. <https://doi.org/10.1007/s10661-018-6903-0>
- Hill WJ, Hunter WG (1966) A review of response surface methodology: a literature survey. *Technometrics* 8(4):571–590
- Jain SK, Nayak PC, Sudheer KP (2008) Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrol Process* 22(13):2225–2234
- Karbasi M (2018) Forecasting of multi-step ahead reference evapotranspiration using Wavelet-Gaussian Process Regression Model. *Water Resour Manag* 32(3):1035–1052
- Keshtegar B, Heddad S (2018) Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. *Neural Comput Appl* 30: 2995–3006
- Keshtegar B, Kisi O (2017) Modified response-surface method: new approach for modeling pan evaporation. *J Hydrol Eng* 22(10):1–14
- Keshtegar B, Kisi O (2018) RM5Tree: radial basis M5 model tree for accurate structural reliability analysis. *Reliab Eng Syst Saf* 180:49–61
- Keshtegar B, Kisi O, Arab HG, Zounemat-Kermani M (2018) Subset modeling basis ANFIS for prediction of the reference evapotranspiration. *Water Resour Manag* 32(1):1101–1116
- Keshtegar B, Kisi O, Zounemat-Kermani M (2019) Polynomial chaos expansion and response surface method for nonlinear modelling of reference evapotranspiration. *Hydrol Sci J* 64(6):720–730
- Keshtegar B, Bagheri M, Fei C-W, Lu C, Taylan O, Thai D-K (2021) Multi-extremum-modified response basis model for nonlinear response prediction of dynamic turbine blisk. *Eng Comput*. <https://doi.org/10.1007/s00366-020-01273-8>
- Khoshravesh M, Sefidkouhi MAG, Valipour M (2017) Estimation of reference evapotranspiration using multivariate fractional polynomial, Bayesian regression, and robust regression models in three arid environments. *Appl Water Sci* 7(4):1911–1922
- Kisi O (2006) Generalized regression neural networks for evapotranspiration modelling. *Hydrol Sci J* 51(6):1092–1105
- Kisi O (2016) Modeling reference evapotranspiration using three different heuristic regression approaches. *Agric Water Manag* 169:162–172
- Kişî Ö, Öztürk Ö (2007) Adaptive Neurofuzzy computing technique for evapotranspiration estimation. *J Irrig Drain Eng* 133(4):368–379
- Lu C, Fei CW, Liu HT, Li H, An LQ (2020) Moving extremum surrogate modeling strategy for dynamic reliability estimation of turbine blisk with multi-physics fields. *Aerosp Sci Technol* 106:106112
- Mehdizadeh S, Behmanesh J, Khalili K (2017) Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput Electron Agric* 139:103–114
- Nagappan M, Gopalakrishnan V, Alagappan M (2020) Prediction of reference evapotranspiration for irrigation scheduling using machine learning. *Hydrol Sci J* 65(16):2669–2677
- Ozonoh M, Oboirien BO, Higginson A, Daramola MO (2020) Performance evaluation of gasification system efficiency using artificial neural network. *Renew Energy* 145:2253–2270. <https://doi.org/10.1016/j.renene.2019.07.136>
- Pham BT, Phong TV, Nguyen HD, Qi C, Al-Ansari N, Amini A et al (2020) A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes, and logistic model tree for flash flood susceptibility mapping. *Water* 12(1):239
- Rahimikhoob A (2016) Comparison of M5 model tree and artificial neural network's methodologies in modelling daily reference evapotranspiration from NOAA satellite images. *Water Resour Manag* 30(9):3063–3075
- Saggi MK, Jain S (2019) Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. *Comput Electron Agric* 156:387–398
- Salam R, Islam ARMT (2020) Potential of RT, Bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *J Hydrol* 590:125241. <https://doi.org/10.1016/j.jhydrol.2020.125241>
- Sanikhani H, Kisi O, Maroufpoor E, Yaseen ZM (2019) Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios. *Theor Appl Climatol* 135(1-2):449–462
- Sayari S, Mahdavi-Meymand A, Zounemat-Kermani M (2021) Irrigation water infiltration modeling using machine learning. *Comput Electron Agric* 180:105921
- Shahabinejad H, Vosoughi N, Saheli F (2020) Matrix effects corrections in prompt gamma-ray spectra of a PGNAA online analyzer system using artificial neural network. *Prog Nucl Energy* 118:103146. <https://doi.org/10.1016/j.pnucene.2019.103146>
- Shiri J, Sadraddini AA, Nazemi AH, Kisi O, Landaras G, Fard AF, Marti P (2014) Generalizability of gene expression programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *J Hydrol* 508:1–11
- Shiri J, Marti P, Nazemi AH, Sadraddini AA, Kisi O, Landaras G, Fakheri Fard A (2015) Local vs. external training of neuro-fuzzy and neural networks models for estimating reference evapotranspiration assessed through k-fold testing. *Hydrol Res* 46(1):72–88
- Solomatine DP, Xue Y (2004) M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *J Hydrol Eng* 9(6):491–501
- Tao H, Diop L, Bodian A, Djaman K, Ndiaye PM, Yaseen ZM (2018) Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: regional case study in Burkina Faso. *Agric Water Manag* 208:140–151. <https://doi.org/10.1016/j.agwat.2018.06.018>
- Tenenbaum RA, Taminato FO, Melo VS (2020) Fast auralization using radial basis functions type of artificial neural network techniques. *Appl Acoust* 157:106993. <https://doi.org/10.1016/j.apacoust.2019.07.041>
- Wang L, Kisi O, Hu B, Bilal M, Zounemat-Kermani M, Li H (2017) Evaporation modelling using different machine learning techniques. *Int J Climatol* 37(S1):1076–1092
- Wang Y, Liu H, Yu Z, Tu L (2020) An improved artificial neural network based on human-behaviour particle swarm optimization and cellular automata. *Expert Syst Appl* 140:112862. <https://doi.org/10.1016/j.eswa.2019.112862>
- Xiao M, Zhang J, Gao L, Lee S, Eshghi AT (2019) An efficient Kriging-based subset simulation method for hybrid reliability analysis under random and interval variables with small failure probability. *Struct Multidiscip Optim* 59:2077–2092
- Xiao M, Zhang J, Gao L (2020) A system active learning Kriging method for system reliability-based design optimization with a multiple response model. *Reliab Eng Syst Saf* 199:106935
- Yamaç SS, Todorovic M (2020) Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agric Water Manag* 228:105875

- Yin Z, Wen X, Feng Q, He Z, Zou S, Yang L (2017) Integrating genetic algorithm and support vector machine for modeling daily reference evapotranspiration in a semi-arid mountain area. *Hydrol Res* 48(5): 1177–1191
- Zahiri J, Mollaei Z, Ansari MR (2020) Estimation of suspended sediment concentration by M5 model tree based on hydrological and moderate resolution imaging spectroradiometer (MODIS) data. *Water Resour Manag* 34(12):3725–3737
- Zhang J, Xiao M, Gao L, Chu S (2019) A combined projection-outline-based active learning Kriging and adaptive importance sampling method for hybrid reliability analysis with small failure probabilities. *Comput Methods Appl Mech Eng* 344: 13–33
- Zhang J, Gao L, Xiao M (2020a) A new hybrid reliability-based design optimization method under random and interval uncertainties. *Int J Numer Methods Eng* 121:4435–4457
- Zhang Y, Gao L, Xiao M (2020b) Maximizing natural frequencies of inhomogeneous cellular structures by Kriging-assisted multiscale topology optimization. *Comput Struct* 230:106197
- Zhou Q, Ren P, Tan YL (2012) Soft measurement of paper smoothness based on time-frequency analysis of paper quantization noise. *Measurement* 45(3):493–499. <https://doi.org/10.1016/j.measurement.2011.10.023>
- Zhu B, Feng Y, Gong D, Jiang S, Zhao L, Cui N (2020) Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput Electron Agric* 173:105430
- Zhu S-P, Keshtegar B, Tian K, Trung N-T (2021) Optimization of load-carrying hierarchical stiffened shells: comparative survey and applications of six hybrid heuristic models. *Arch Comput Methods Eng*. 1–12. <https://doi.org/10.1007/s11831-021-09528-3>
- Zounemat-Kermani M, Kisi O, Piri J, Mahdavi-Meymand A (2019) Assessment of artificial intelligence-based models and metaheuristic algorithms in modeling evaporation. *J Hydrol Eng* 24(10):04019033
- Zounemat-Kermani M, Matta E, Cominola A, Xia X, Zhang Q, Liang Q, Hinkelmann R (2020) Neurocomputing in surface water hydrology and hydraulics: a review of two decades retrospective, current status and future prospects. *J Hydrol* 588:125085. <https://doi.org/10.1016/j.jhydrol.2020.125085>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.