



Estimation of monthly evaporation values using gradient boosting machines and mode decomposition techniques in the Southeast Anatolia Project (GAP) area in Turkey

Metin Sarıgöl¹ · Okan Mert Katipoğlu²

Received: 10 December 2022 / Accepted: 27 February 2023 / Published online: 13 March 2023

© The Author(s) under exclusive licence to Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2023

Abstract

Today, the biggest issue appears to be the increase in drought in some regions brought on by global warming, which has greatly increased the significance of water management. In light of evaporation's effect on drought, this research intends to evaluate the effectiveness of hybrid machine learning (ML) models, such as the Gradient Boosting Machines (GBM) technique paired with Empirical Mode Decomposition (EMD), Robust Empirical Mode Decomposition (REMD), Ensemble Empirical Mode Decomposition (EEMD), and Variational Mode Decomposition (VMD) signal decomposition techniques, for monthly evaporation prediction models in the Southeast Anatolia Project Area. In the design of the models, 80% of the data was used for training and 20% for testing. Furthermore, tenfold cross-validation was applied to solve the overfitting problem, which negatively affected the forecast performance. In the model setup, various combinations of precipitation, average air temperature, minimum air temperature, maximum air temperature, wind speed, actual air pressure, relative humidity, and solar time variables are presented to artificial intelligence models as input. The study revealed that the GBM methodology in combination with the signal decomposition methods REMD, EMD, EEMD, and VMD generally allowed for more accurate evaporation estimations than the GBM model alone. The study's results are essential in relation to agricultural production, irrigation planning, water resources management studies, and hydrological modeling studies in the region.

Keywords Evaporation · Empirical mode decomposition · Machine learning · Variational mode decomposition · Gradient boosting machines · Southeast Anatolia Project (GAP) area

Introduction

Evaporation occurrence in nature is an essential element for the hydrological cycle and the prediction of evaporation is a significant issue for water resources, agricultural modeling

and water management (Gundalia and Dholakia 2013; Malik and Kumar 2015; Feng et al. 2018). Therefore, modeling evaporation is an especially important process, especially in regions and basins where measurements are insufficient (Dalkiliç et al. 2014). The evaporation process is not linear and has a complex structure and generally depends on heat energy and vapor pressure, which depends on atmospheric pressure, solar radiation, air temperature, relative humidity, and wind speed. It is also affected by season, geographical location, and climate type (Vicente-Serrano et al. 2018).

Direct or indirect procedures are applied to calculate and predict evaporation. While pan evaporation (E_p) measurements from direct approaches are used, methods such as Penman–Monteith techniques and energy and water budgets are used from indirect procedures. Because the evaporation process is not linear, it is not possible to use a mathematical model to accurately represent it (Rezaie-Balf et al. 2019). However, many semi-empirical and empirical methods have been established to predict evaporation (Antonopoulos and

Edited by Dr. Senlin Zhu (ASSOCIATE EDITOR) / Dr. Michael Nones (CO-EDITOR-IN-CHIEF).

Metin Sarıgöl and Okan Mert Katipoğlu have contributed equally to the publication.

✉ Okan Mert Katipoğlu
okatioglu@erzincan.edu.tr

¹ Design Department, Erzincan Uzunlu Vocational School, Erzincan Binali Yıldırım University, Erzincan, Türkiye

² Faculty of Engineering and Architecture, Department of Civil Engineering, Erzincan Binali Yıldırım University, Erzincan, Türkiye

Antonopoulos 2017; Lu et al. 2018). Nevertheless, the biggest problem in using such evaporation prediction methods is that meteorological variables are dynamic due to their nonlinearity and stochastic properties (Yaseen et al. 2020).

Due to the lengthy duration and numerous limitations of the physically based and statistical methods employed in the hydrological study, machine learning (ML) algorithms have gained prominence. Prediction models developed with ML algorithms allow simpler implementation, faster results with less input, and excellent computing efficiency (Mosavi et al. 2018; Zounemat-Kermani et al. 2020; Dazzi et al. 2021).

The use of ML algorithms in the field of hydrology has been widely increasing in recent years. Algorithms such as Gradient Boosting Machines (GBM), Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), Ensemble Empirical Mode Decomposition (EEMD), and Robust Empirical Mode Decomposition (REMD) algorithms have enjoyed wide usage. The following paragraphs present the numerous instances where these algorithms have been used.

Wang et al. (2017) researched the performances of neuro-fuzzy inference systems with grid partition (ANFIS-GP), fuzzy genetic (FG) and M5Tree to predict the monthly E_p values of stations. They found that FG models were successful in estimating E_p . On the other hand, Pammar and Deka (2017) used support vector machines (SVM) and discrete wavelet transform (DWT) for a pan evaporation prediction. In terms of accuracy, they concluded that the results of the E_p predictions of the established models were very promising. Lu et al. (2018) predicted daily E_p using empirical models together with gradient boost decision tree (GBDT), random forests (RFs), and M5 model tree (M5Tree). As a result of the research, they found the GBDT model to be the best in estimating and stable daily E_p among all models. Yaseen et al. (2020) employed the cascade correlation neural network (CCNNs), gene expression programming (GEP), classification and regression tree (CART), and SVM to forecast evaporation using meteorological variables. At the conclusion of the study, they observed that all ML algorithms performed well in forecasting evaporation, with the SVM providing the greatest performance results.

Rezaie-Balf et al. (2019) used EEMD combined with SVM and model tree (MT) to predict the monthly E_p . Ultimately, they discovered that MT and SVM algorithms combined with Ensemble Empirical Mode Decomposition (EEMD) performed better at estimating monthly E_p than MT and SVM algorithms alone. Ali Ghorbani et al. (2018) used the Quantum-Behavioral Particle Swarm Optimization (QPSO) model trained on the multilayer perceptron (MLP) algorithm to predict the daily evaporation rates and found that the hybrid MLP-QPSO algorithm outperformed the hybrid MLP-PSO and the standalone algorithm. Mohamadi et al. (2020) used firefly algorithms (FFAs) and shark

algorithm (SA), which trained MLP, ANFIS and radial basis function (RBF) algorithms to predict monthly evaporation and revealed that ANFIS-SA gave better than the other algorithms. Wu et al. (2020) researched hybrid WOAELM and FPAELM algorithms to demonstrate the applicability flower pollination algorithm (FPA) and whale optimization algorithm (WOA) of the extreme learning machine (ELM) algorithm for monthly E_p prediction. It was concluded that the hybrid FPAELM algorithm gave the best estimation result and both gave better results than other models. Jasmine et al. (2022) used the ANFIS and hybridization of ANFIS, which include the firefly algorithm (FFA), GA, and PSO algorithms, to forecast the evaporation of agricultural areas and found that ANFIS-PSO and ANFIS gave better results than ANFIS-FFA and ANFIS-GA.

To predict the monthly E_p estimation, Al-Mukhtar (2021) researched the performances of ML models, which are multivariate adaptive regression splines (MARS), bagged multivariate adaptive regression splines (BaggedMARS), conditional random forest regression (CRFR), weighted K-nearest neighbor (KKNN), K-nearest neighbor (KNN), and model tree M5. He found that the weighted KNN algorithm gave the best results. Malik et al. (2022) also researched the effectiveness of GBM and deep learning (DL) algorithms using the maximum air temperature parameter in predicting evaporation for two stations and found that the DL algorithm gave better results than the GBM algorithm.

To model pan evaporation under different climatic conditions in Iran, Dehghanipour et al. (2021) used an MLP-NN and genetic algorithm. They used the best overall relationship in the first method as the main relationship in the second method and determined the climatic correction coefficients for the other climate types with the genetic algorithm rhythm optimization model. Their study found that both methods gave accurate results in modeling pan evaporation in Iran.

In the evaporation research, the most well-known algorithms in the literature—ANFIS, SVM, and MLP—were employed, as the literature review has shown (Wang et al. 2017; Pammar and Deka 2017; Ali Ghorbani et al. 2018). The research has determined that the evaporation estimation research topics made with hybrid algorithms in the literature are few and limited to certain algorithms (Wu et al. 2020; Jasmine et al. 2022). Furthermore, evaporation estimates, especially with signal separation techniques, are limited and generally used a signal separation model (Rezaie-Balf et al. 2019).

Signal separation techniques allow understanding of the basic structure of time series and reveal patterns or trends that may be present. Therefore, predicting future values and detecting unusual data observations or anomalies is important. Also, decomposition techniques can help express these patterns and reveal their effects.

Considering this deficiency in the literature, evaporation prediction research was carried out by combining the four

signal decomposition techniques REMD, EMD, EEMD and VMD algorithms with the GBM model. In addition, the effectiveness of four different signal separation techniques, which were not used simultaneously before on the GBM model was investigated.

Consequently, for the first time in the literature, the results of monthly evaporation estimates for a region were investigated with the GBM technique, which was hybridized with REMD, EMD, EEMD and VMD signal separation techniques using eight parameters. The innovative contribution of this research to the literature is to show the final state of artificial intelligence algorithms in evaporation and to determine the evaporation prediction performance of some hybrid models by parameter optimization. Additionally, the effect of various meteorological parameters on evaporation prediction was evaluated. This revealed the type of input combination that can predict evaporation successfully.

This study aims to forecast evaporation using monthly total precipitation (P), monthly average temperature (T_{avg}), monthly minimum temperature (T_{min}), monthly maximum temperature (T_{max}), monthly average wind speed (WS), monthly average actual compression (AP), monthly average relative humidity (RH), and various combinations of monthly total solar time (ST) variables. The data was modeled using the GBM, EMD-GBM, REMD-GBM, EEMD-GBM, and VMD-GBM algorithms. The models underwent training and testing phases before being assessed using the NSE, MAE, RMSE, R2 coefficient, radar plot, and boxplot analysis. It is presumed that the models examined in the study can be used to forecast evaporation in various places.

The remainder of this article is organized as follows: "Method and material" section introduces machine learning and signal separation algorithms after showing the study area and the data used. Then the technical details of the

model performance methods used are presented. In "Results and discussion" section, the analyses of the models used for evaporation prediction and the results of the graphical analysis are compared. The performance of the best model is also evaluated and compared with similar studies. Finally, the research results are given in "Conclusion" section.

Method and material

Study data and area

This study estimated monthly evaporation values in the GAP region using the GBM model and signal preprocessing techniques. For this purpose, EMD, REMD, EEMD and VMD algorithms were used to increase the performance of a single-GBM algorithm. In the design of the algorithms, 80% of the data was used as training and 20% as testing. Additionally, tenfold cross-validation was applied to solve the overfitting problem, which negatively affects the forecast performance. The GAP Project area is the Euphrates-Tigris Basin and the Upper Mesopotamia region. It covers the provinces of Gaziantep, Adıyaman, Mardin, Diyarbakır, Kilis, Siirt, Şanlıurfa, Batman, and Şırnak. Utilizing the resources of the Southeastern Anatolia Region, raising the income level and standard of living of the local population, expanding employment opportunities in rural regions, and ensuring economic development are some of the GAP's top priorities (Gap.gov 2022). This research used meteorology station data in Siirt, Adıyaman and Diyarbakır provinces. Figure 1 shows the meteorology station's locations.

Summary information about the meteorological stations used in the study is presented in Table 1, and the parameters of meteorology stations are shown in Table 1. There



Fig. 1 Location of Siirt, Adıyaman, and Diyarbakır Meteorology Stations

Table 1 Summary information of meteorological stations

Province name	Latitude	Longitude	Station code	Station name	Station elevation (M)	Selected time period
Siirt	37.93	41.94	17,210	Siirt	895	1972–2017
Adiyaman	37.76	38.28	17,265	Adiyaman	672	1966–2017
Diyarbakır	37.90	40.20	17,280	Diyarbakır Airport	674	1963–2017

are approximately 10 meteorological observation stations in the study area. However, the 3 most reliable stations were used in the study since the meteorological data at some stations were missing and the neighboring stations did not have enough data to complete the data. Therefore, the study area is limited to 3 stations.

The statistical parameters of the meteorological data used in the study are presented in Table 2. These are skewness, kurtosis, standard deviation representing mean, maximum, minimum values and distribution parameters that provide essential information about statistical training and test data.

Machine learning and decomposition models

The performance of monthly evaporation prediction models is compared by using GBM, EMD-GBM, REMD-GBM, EEMD-GBM and VMD-GBM algorithms widely used in the literature. While selecting the ML algorithms, 80% of the data were used for training and %20 used for testing, respectively, and eight parameters were used as input. In addition, the evaporation data were used as output. In Fig. 2, the application steps of the study are shown in order. First, the input combination of the model was selected by subjecting the meteorological data to correlation analysis. Second, the lagged values of the variables with high correlation with evaporation were chosen for the model's setup. Then, the selected input values are subdivided by various decomposition techniques. In creating hybrid models, each sub-component is presented to the GBM model. In the setup of the GBM model, the number of trees parameter is 10000, interaction depth = 1, the distribution type is Gaussian, and shrinkage values are set to 0.01 default values.

Empirical mode decomposition (EMD)

This model adaptively decomposes a non-stationary signal from high frequencies to low frequencies into a set of intrinsic-mode functions (IMFs), and the decomposed signal can be expressed as:

$$x_t = \sum_{i=1}^N C_i(t) + r_N(t). \quad (1)$$

Here, $r_N(t)$: residual of signal, $C_i(t)$: i th IMF $x(t)$ (Kedardouche et al. 2016).

Ensemble empirical mode decomposition (EEMD)

Although this model has been largely used in non-stationary and/or nonlinear signal analysis, decomposition results encounter various problems due to mode mixing. The ensemble EMD method has been created to fix this problem. As this algorithm repeatedly decomposes the signal into the intrinsic modal function (IMFs) using the EMD method, finite-amplitude white noise is added to the original signal. The aggregated means of IMFs generated from each trial are also expressed as IMFs of the EEMD method. Thus, the mode mixing problem is removed by EEMD (Zhang et al. 2010).

The EEMD algorithm contains the following steps and is shown in Fig. 3 (Wu et al. 2009).

Variational mode decomposition (VMD)

This algorithm non-recursively decomposes a real-valued multicomponent f signal into semi-vertical band-limited sub-signals. Furthermore, each mode is compact around a center vibration. Thus, constrained variational problem equation can be written as:

$$\left\{ \begin{array}{l} \min \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \{u_k\}, \{w_k\} \\ \text{s.t. } \sum_{k=1}^K u_k = f. \end{array} \right. \quad (2)$$

Here u_k : decomposed band-limited IMF, w_k : frequency center of each IMF, $\{w_k\} = \{w_1, w_2, \dots, w_k\}$; $\{u_k\} = \{u_1, u_2, \dots, u_k\}$ (Wang and Markert 2016).

Robust Empirical mode decomposition (REMD)

EMD is an efficient algorithm for extracting useful information from multicomponent and modulated signals. Robust Empirical Mode Decomposition (REMD) algorithm is established by applying the elimination sifting stopping criterion (SSC) to the EMD method. As a result, the EMD algorithm

Table 2 Statistical properties of meteorological data used in the study

	P	Tavg	Tmax	Tmin	WS	AP	RH	ST	EVP
Adiyaman									
Statistic	Train								
Mean	30.39	22.71	34.74	11.37	1.84	934.26	42.16	277.46	171.99
Max	232.80	33.50	45.30	22.00	3.50	943.60	78.50	417.30	354.00
Min	0.00	7.70	18.90	-3.50	0.60	926.50	13.80	0.00	18.40
Std. Dev	39.88	6.88	6.59	6.41	0.57	4.09	13.47	87.61	84.82
Skewness	1.77	-0.24	-0.49	-0.15	0.22	0.26	0.31	-0.72	0.10
Kurtosis	3.36	-1.15	-0.82	-1.00	-0.08	-0.90	-0.73	0.34	-1.10
Test									
Mean	32.02	21.91	34.02	10.52	1.92	934.29	41.88	274.97	171.12
Max	152.90	32.70	44.50	21.70	3.60	943.60	74.50	398.80	352.20
Min	0.00	8.10	18.50	-1.30	0.80	920.50	17.50	0.00	24.50
Std. Dev	39.60	6.76	6.56	6.04	0.65	4.42	14.10	90.91	88.86
Skewness	1.40	-0.18	-0.41	0.04	0.36	-0.33	0.44	-0.44	0.11
Kurtosis	1.21	-1.08	-0.68	-0.95	-0.64	0.10	-0.53	-0.38	-1.12
Siirt									
Train									
Mean	41.06	22.05	33.60	11.15	1.43	910.59	42.37	273.01	211.04
Max	291.10	33.50	44.40	23.00	2.50	921.20	78.90	399.00	515.20
Min	0.00	6.80	16.30	-4.30	0.30	902.10	17.40	0.00	17.40
Std. Dev	53.55	7.12	6.48	6.96	0.44	4.09	14.82	89.04	118.22
Skewness	1.77	-0.27	-0.51	-0.13	0.24	0.22	0.30	-0.78	0.21
Kurtosis	3.20	-1.11	-0.74	-0.99	-0.50	-0.80	-0.95	0.31	-0.79
Test									
Mean	32.78	21.23	32.83	10.54	1.53	911.30	42.18	269.32	207.44
Max	189.60	32.10	43.30	23.00	2.70	919.10	71.00	405.90	420.90
Min	0.00	6.80	18.60	-2.80	0.40	903.30	13.60	0.00	7.90
Std. Dev	42.08	7.38	7.27	6.88	0.59	4.42	15.55	88.56	116.09
Skewness	1.65	-0.26	-0.59	-0.08	0.11	0.03	0.06	-0.35	0.12
Kurtosis	2.53	-1.12	-0.83	-0.94	-0.99	-1.26	-1.23	-0.38	-1.26
Diyarbakır									
Train									
Mean	24.26	21.49	34.54	8.32	2.58	933.97	46.28	290.76	230.50
Max	158.70	32.50	44.80	21.20	5.20	943.80	82.00	418.50	951.40
Min	0.00	6.60	18.90	-8.80	0.90	925.20	15.10	91.00	1.10
Std. Dev	33.90	7.36	6.79	6.92	0.71	4.48	17.15	81.76	139.20
Skewness	1.71	-0.27	-0.53	-0.22	0.40	0.10	0.23	-0.39	0.74
Kurtosis	2.61	-1.16	-0.86	-1.01	0.07	-0.97	-1.17	-1.05	1.52
Test									
Mean	30.78	20.72	34.10	7.01	2.66	933.95	47.82	284.37	239.36
Max	163.50	33.30	44.70	18.60	4.20	943.10	79.10	402.00	1251.60
Min	0.00	6.50	18.20	-7.80	1.40	926.10	15.10	111.00	14.00
Std. Dev	37.55	7.63	7.47	6.96	0.70	4.64	16.37	83.28	198.64
Skewness	1.41	-0.18	-0.59	-0.16	0.12	0.18	0.22	-0.18	2.80
Kurtosis	1.75	-1.09	-0.60	-0.82	-0.73	-0.97	-1.01	-1.45	11.93

Fig. 2 Flowchart of the study

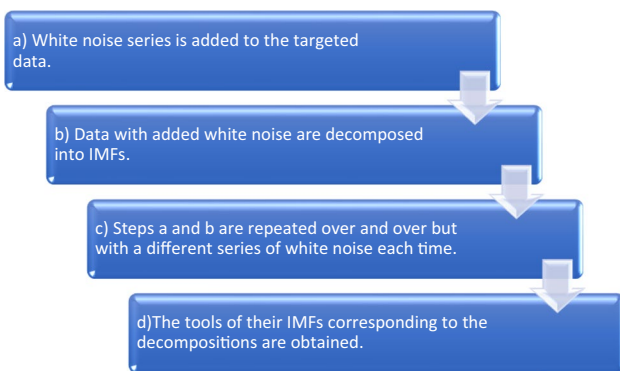
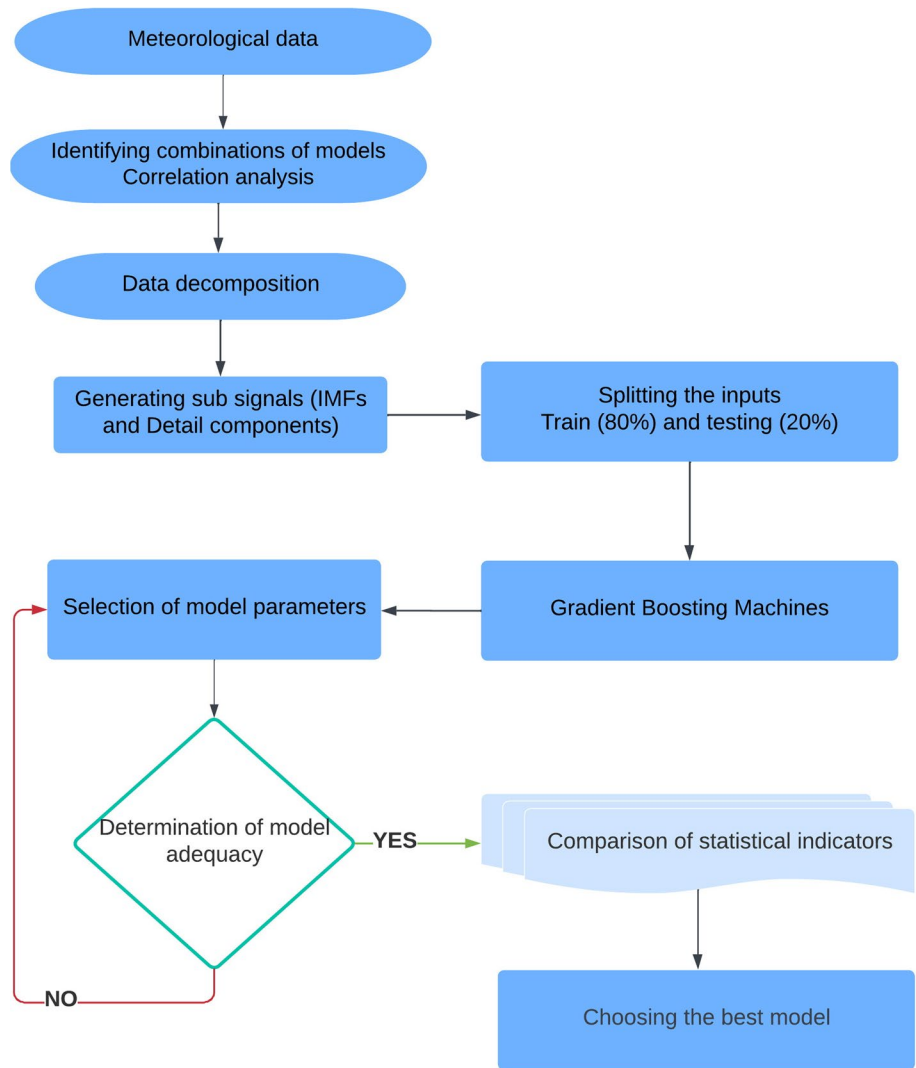


Fig. 3 EEMD algorithm steps

emerges as a method that can ease the mode mixing problem in signal decomposition and improve the demodulation performance in signal demodulation (Liu et al. 2022).

Gradient-boosted machine (GBM)

While this algorithm generates a fixed function, it starts the algorithm process by taking the first guess, then iteratively adds a decision tree at each stage until it reaches the optimal reduction in the loss function, and the GBM method can be expressed as follows:

$$f_0(x) = \sum_{m=1}^k \gamma_m h_m(x). \tag{3}$$

Here $f_k(x)$: classified function, γ_m : weight of each decision tree, x : clustered results input variable, k :total number of decision trees, h_m : numerical value that minimizes the loss function (Asante-Okyere et al. 2020).

Analysis of model performance

In this research, forecasting analyzes were made using various evaluation criteria which are R^2 , RMSE, NSE, and MAE performance measures to determine the accuracy of the proposed ML algorithms.

The root-mean-square error (RMSE) measures how far the regression line is from the data points.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (E_{pi} - E_{oi})^2}{n}} \quad (4)$$

Here; E_{pi} : Predicted value, E_{oi} : Observed value, n : number of data.

Nash–Sutcliffe Efficiency (NSE) is calculated by subtracting the ratio of the mean squared errors and written as below:

$$\text{NSE} = 1 - \left[\frac{\sum_{i=1}^n (E_{pi} - E_{oi})^2}{\sum_{i=1}^n (E_{oi} - E_o)^2} \right], \quad (5)$$

where n : number of data E_{pi} : Predicted value, E_{oi} : Observed value, E_o : Average of observed value (Başakın et al. 2021).

The determination coefficient (R^2) is the linear regression between the predicted and observed values and is expressed by the formula below (Zare and Koch 2013).

$$R^2 = \left(\frac{\sum_{i=1}^n (E_{oi} - E_{om})(E_{pi} - E_{pm})}{\sqrt{\sum_{i=1}^n (E_{oi} - E_{om})^2 \sum_{i=1}^n (E_{pi} - E_{pm})^2}} \right)^2 \quad (6)$$

Here, n : number of data, E_{oi} : Observed value, E_{pi} : Predicted value, E_{om} : average of observed value, E_{pm} : average of predicted value.

The Mean Absolute Error (MAE) shown in Eq. 7 is the mean of the absolute difference between the calculated value and actual value expressed by the formula below (Chai and Draxler 2014).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |E_{pi} - E_{oi}| \quad (7)$$

Here, n : number of data, E_{oi} : Observed value, E_{pi} : Predicted value.

Box plots provide information about data distribution, maximum and minimum quintile and median values for established models (Nhu et al. 2020; Dehghani et al. 2022).

Radar plot is a graphic in which each circle represents a certain error size and spouses-centered circles come together (Pardo et al. 2017).

Results and discussion

In the model setup, monthly total precipitation (P), monthly average temperature (T_{avg}), monthly minimum temperature (T_{min}), monthly maximum temperature (T_{max}), monthly average wind speed (WS), monthly average actual compression (AP), monthly average relative humidity (RH), and various combinations of monthly total solar time (ST) variables are presented to artificial intelligence models as input. Figures 4, 5 and 6 show the correlation matrix between the model inputs and outputs of Adiyaman, Diyarbakır and Siirt meteorological stations selected for evaporation prediction. When the correlation matrices are examined, it is seen that all meteorological variables used have a significant relationship with evaporation. Correlation coefficients and scatter diagrams are also seen in the correlation matrix. It is observed that while RH, AP and P variables have an inverse relationship with EVP, T_{avg} , T_{max} , WS, RH and ST variables have a linear relationship with EVP. When the correlation coefficients are evaluated, it is noteworthy that the variable with the highest correlation with the EVP values in all meteorology stations is T_{avg} .

Table 3 expresses the model combinations used in evaporation prediction. Based on correlation matrices, the high-variable meteorological parameters were presented as inputs to the models. Thus, it is aimed to reveal which meteorological variables are more effective in evaporation prediction. As an example, the sub-signals of T_{avg} values obtained by various decomposition techniques at Adiyaman station are visualized (Fig. 7).

The established GMB model is created according to the default values as $n.trees = 10,000$, $interaction.depth = 1$, $distribution = Gaussian$, $n.trees = 10,000$, $interaction.depth = 1$, $shrinkage = 0.01$. In the setup of hybrid models, the inputs separated into sub-signals with EMD, REMD, EEMD and VMD techniques are combined with the GBM algorithm. EMD, REMD, EEMD and VMD techniques and various IMF and residual values of T_{avg} data at Adiyaman station are examples.

The test results of hybrid GBM models combined with the established single-GBM and signal processing are shown in Table 4. M1, M2 and M3 input combinations for evaporation prediction at Adiyaman meteorological station, M3 and M7 input combinations at Diyarbakır station and M1 input combinations at Siirt station are prominent. When all models are evaluated together, the most accurate estimation results (RMSE: 3.2638, R^2 : 0.9989, MAE: 1.9455, NSE: 0.9986) were obtained with the M1 model combination and EMD-GBM hybrid model at Adiyaman station. In addition, when the model performances are evaluated, it is noteworthy that the prediction successes are generally high and close to each other. In this case,

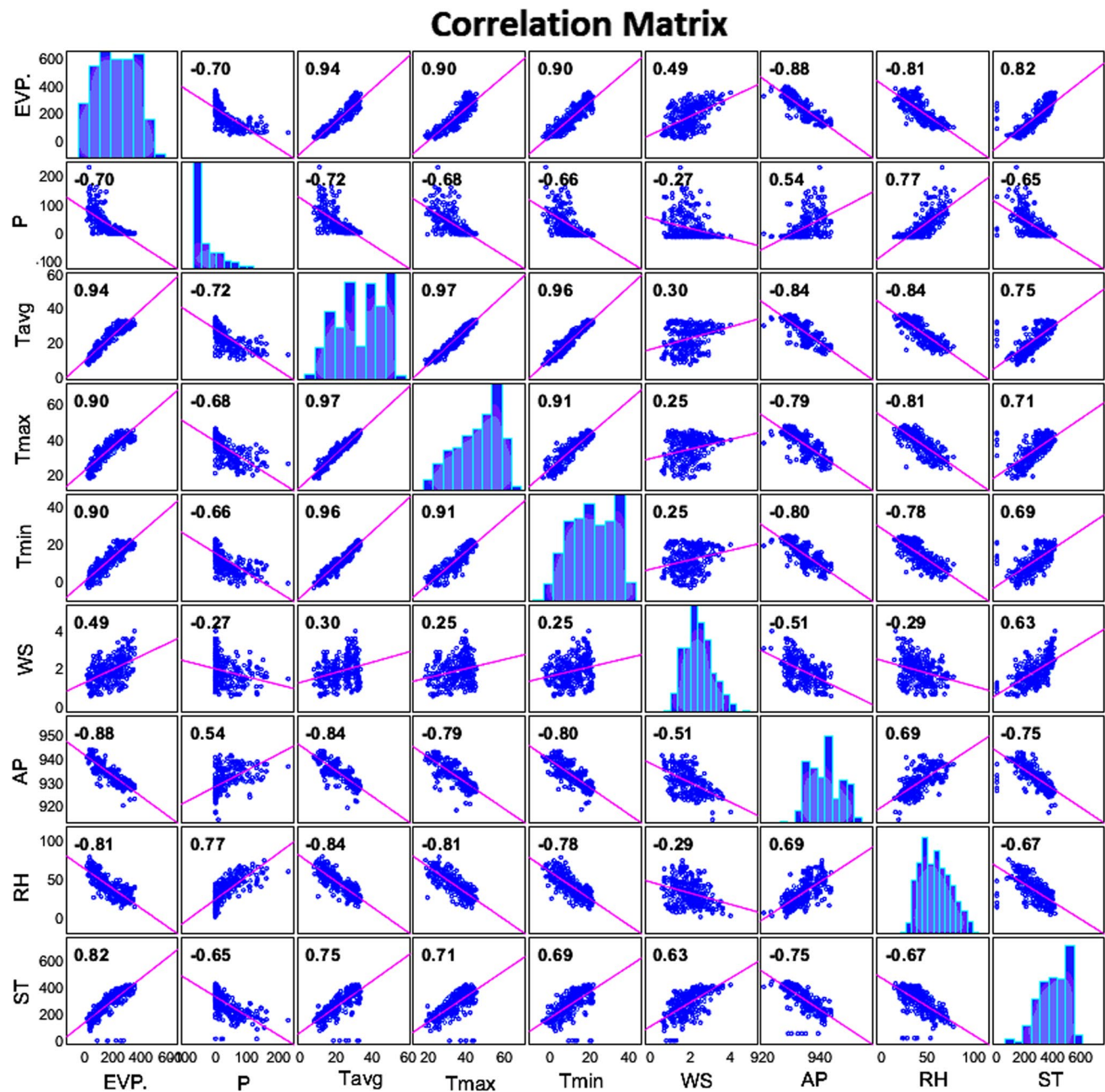


Fig. 4 Correlation matrix of Adiyaman station

it can be deduced that all installed modules can predict evaporation effectively.

Boxplots of the test results of the evaporation prediction model established in the GAP region are shown in Fig. 8. Performance evaluation is made according to the models' structure based on this boxplot and their distribution among the real data. Accordingly, it can be said that all models found in Adiyaman and Siirt meteorology observation stations have a high forecasting success because they have a similar distribution with the real data. On the other hand, the prediction performance of the Diyarbakir station is relatively

weak. When all boxplots were examined, it was revealed that the top model is the M1 model combination and EMD-GBM hybrid model at Adiyaman station.

Figure 9 shows scattering diagrams of all models established for evaporation prediction in the GAP region. When the established models are evaluated, it is noteworthy that the data in Adiyaman and Siirt stations are distributed around a 45-degree line. This shows that the evaporation estimates at Adiyaman and Siirt stations are close to the truth. However, in Diyarbakir station, it is seen that evaporation values above 400 mm cause the higher error as they

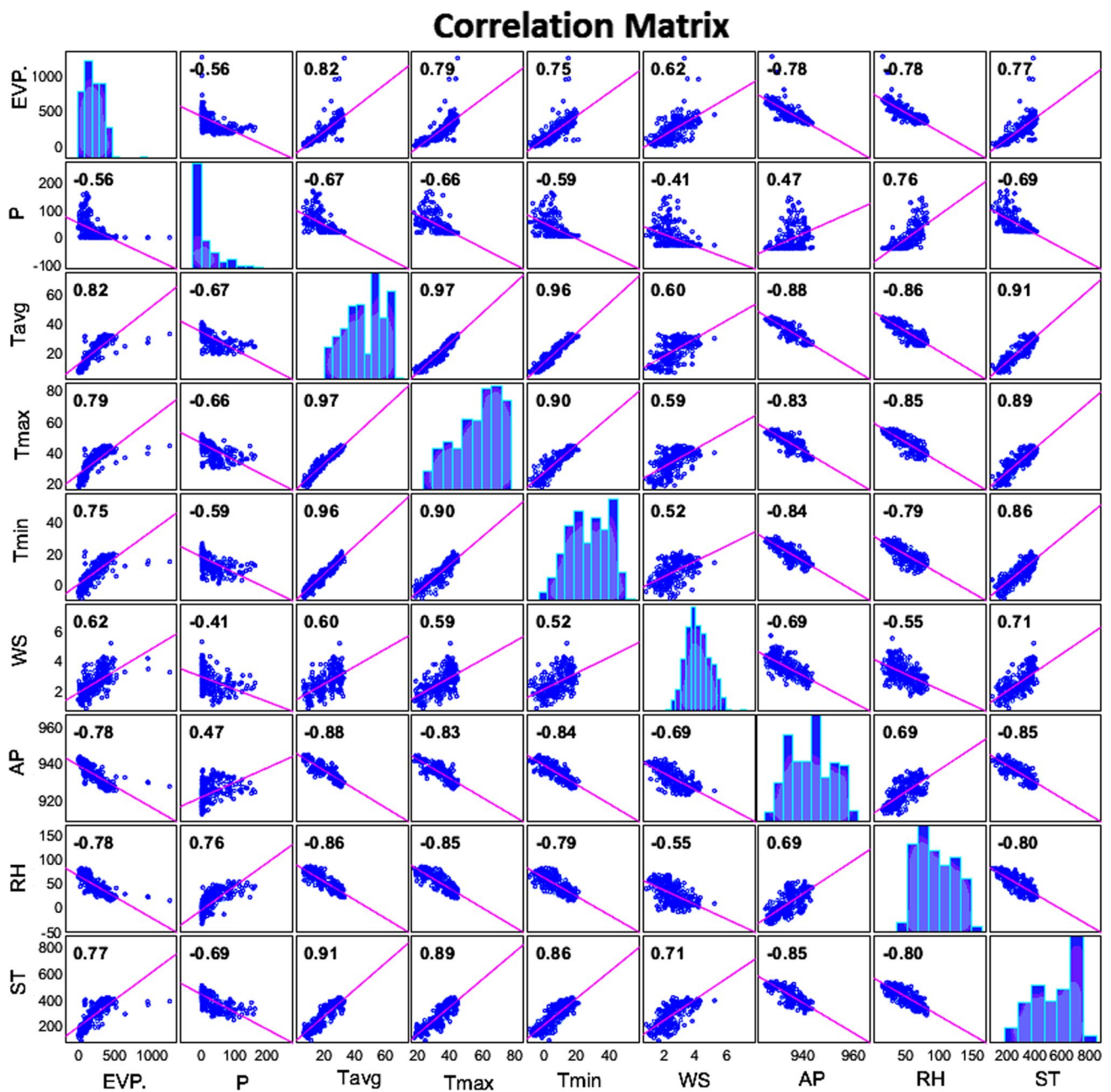


Fig. 5 Correlation matrix of Diyarbakır station

deviate from the linear line. In addition, it is noteworthy that the most successful evaporation prediction was made with EMD-GBM at Adıyaman station. In addition, when the scatter diagrams are not evaluated on a province basis, it is noteworthy that Adıyaman M3, Diyarbakır M7, and in Bitlis M1 model input combinations come to the fore. In line with these results, it was concluded that in Adıyaman, P , T_{avg} , T_{max} , T_{min} , in Diyarbakır, P , T_{avg} , T_{max} , T_{min} , WS, AP, RH, ST and in Bitlis P , T_{avg} meteorological parameters were the most effective on evaporation estimation.

Figure 10 shows the error changes of GBM and hybrid GBM models with the radar plot. Low error value was analyzed in selecting the most suitable model. Accordingly, it is seen that hybrid models built using signal separation techniques generally have lower errors than the single-GBM model. In addition, the most accurate evaporation values were obtained with the EMD-GBM models with the lowest error. In addition, the lowest error value was seen in the Adıyaman meteorology station. In addition, when the polar diagrams of the errors are evaluated on a province basis, it is

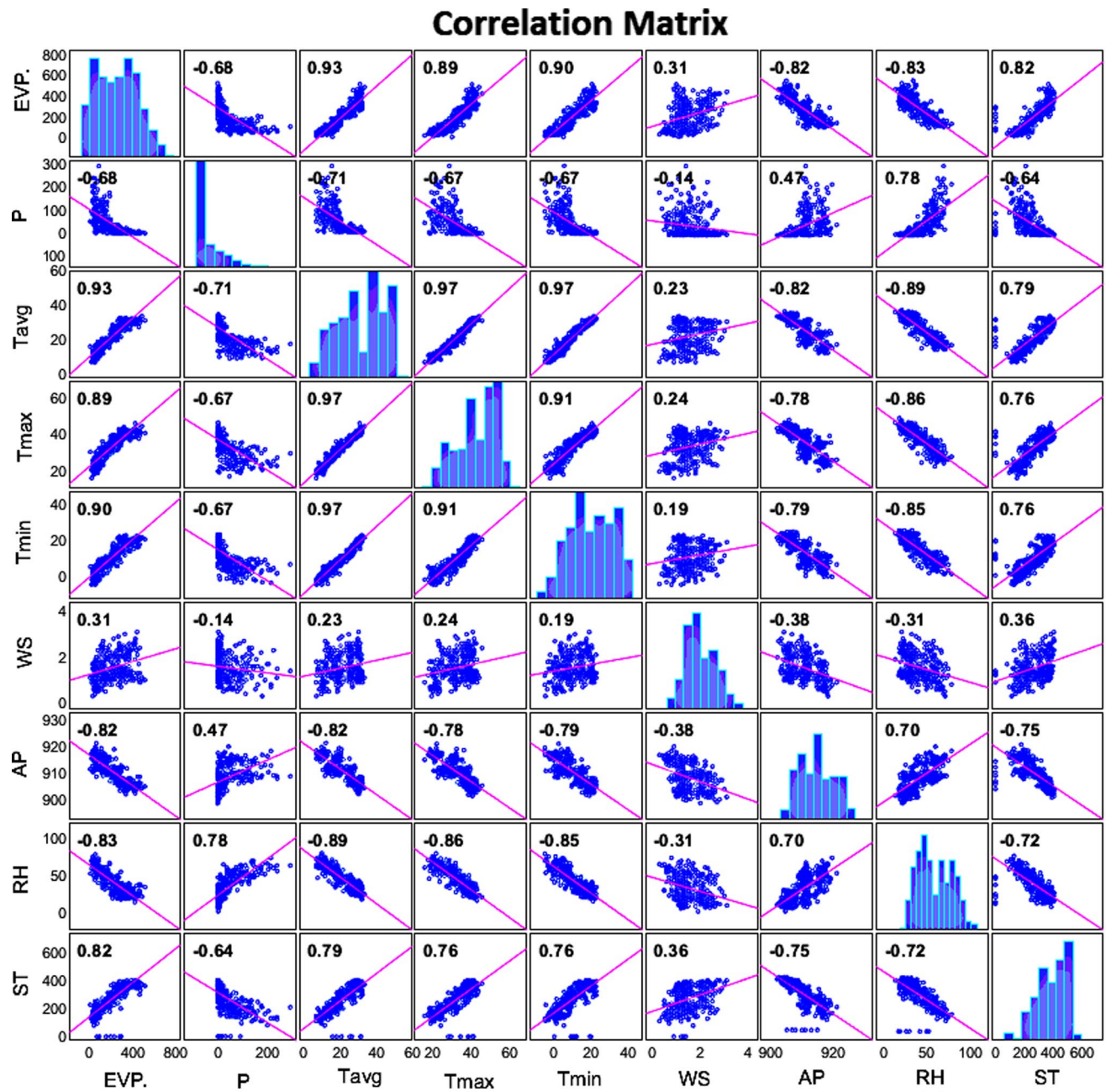


Fig. 6 Correlation matrix of Siirt station

Table 3 Installed model combinations

Model	Input	Target
M1	P, T_{avg}	EVP
M2	P, T_{avg}, T_{max}	EVP
M3	$P, T_{avg}, T_{max}, T_{min}$	EVP
M4	$P, T_{avg}, T_{max}, T_{min}, WS$	EVP
M5	$P, T_{avg}, T_{max}, T_{min}, WS, AP$	EVP
M6	$P, T_{avg}, T_{max}, T_{min}, WS, AP, RH$	EVP
M7	$P, T_{avg}, T_{max}, T_{min}, WS, AP, RH, ST$	EVP

noteworthy that the highest error values occur in Adiyaman, while the lowest error values are observed in Siirt. Based on this, it can be said that the evaporation estimates in Siirt province are closer to the truth than other stations. When the success of the models established according to the error values is compared, it is seen that the performances of the GBM and hybrid GBM models are quite similar. However, it has been observed that decomposition techniques in other locations other than Siirt generally increase the success of the GBM model.

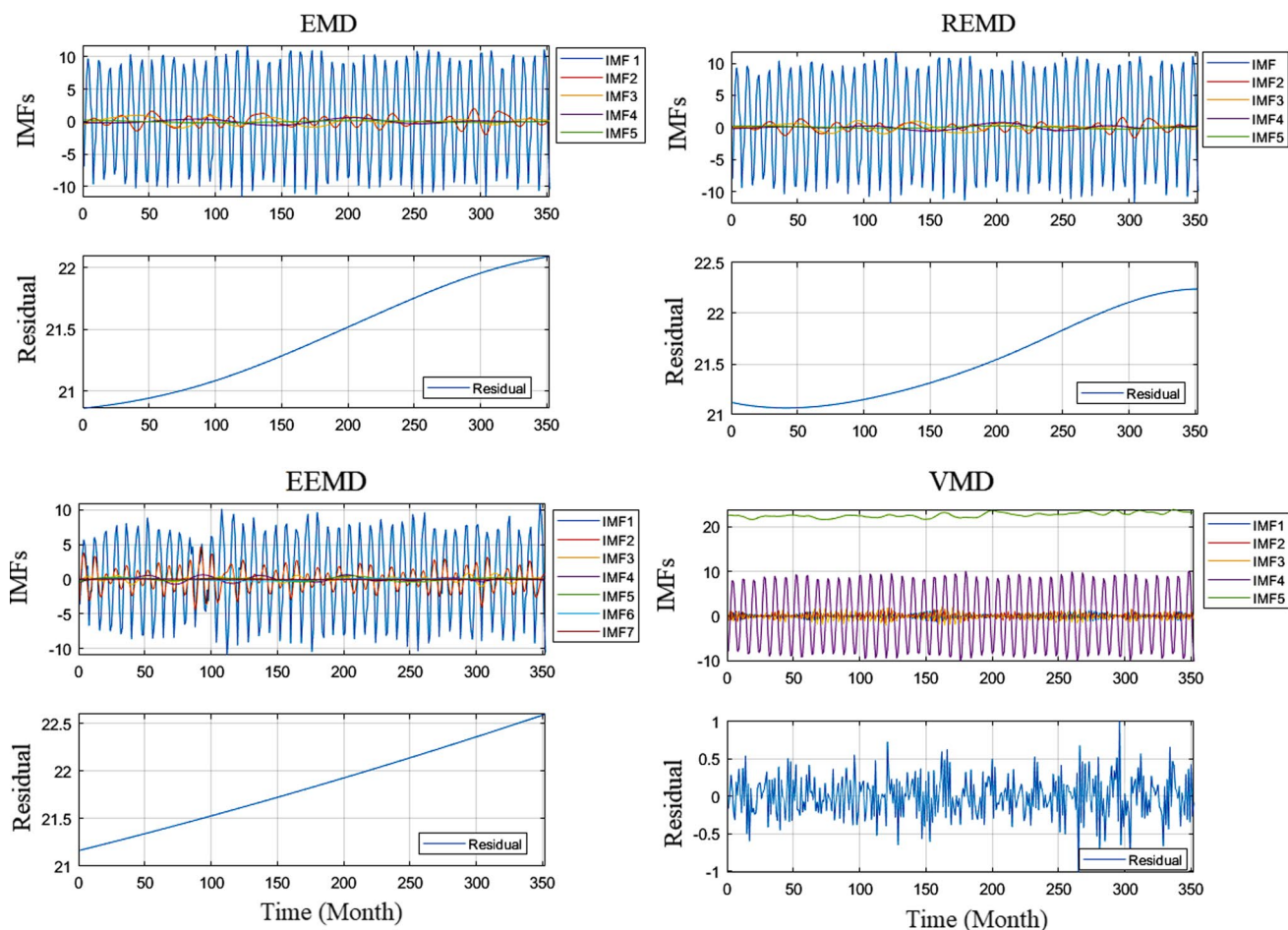


Fig. 7 The sub-signals of the T_{avg} variable produced by the signal decomposition techniques used in Adıyaman

Figure 11 shows the variation of the predicted and actual evaporation time series of the EMD-GBM model, which predicts monthly evaporation data with the highest accuracy. These curves' parallel progression and overlap support that the forecast model produces realistic forecasts.

This study aims to forecast the evaporation using eight parameters by hybrid ML algorithms, which are GBM, EMD-GBM, REMD-GBM, EEMD-GBM and VMD-GBM hybrid models. For this purpose, this study analyzed the effect of signal preprocessing techniques on the machine learning algorithm's performance. It has been concluded that the models made by combining the GBM algorithm and signal processing techniques make the evaporation prediction effective and reliable. The results of the Rezaie-Balf et al. (2019), Wu et al. (2020), Mohamadi et al. (2020), Gümüş et al. (2016) and Gümüş et al. (2021) researches are compatible with the presented study. Rezaie-Balf et al. (2019) used EEMD combined with SVM and MT models to estimate monthly E_p . Ultimately, they found that the MT and SVM algorithms combining EEMD gave better results than MT and SVM algorithms in estimating monthly P_E . The

results of this study overlap with Rezaie-Balf et al. (2019) evaporation prediction research in establishing hybrid models such as the EEMD-SVM algorithm and giving better results than other algorithms. Wu et al. (2020) researched hybrid WOAELM and FPAELM algorithms to demonstrate the applicability of FPA and WOA algorithms of ELM algorithm for monthly E_p prediction and they determined that the hybrid FPAELM algorithm gave the best estimation result and both gave better results than other models. The results of this study overlap with Wu et al. (2020) evaporation prediction research in establishing the hybrid algorithms and giving better results than other algorithms. Mohamadi et al. (2020) used FFAs and SA algorithms which trained MLP, ANFIS and RBF algorithms to predict monthly evaporation and determined that ANFIS-SA gave better than the other models. The findings obtained from the study by Mohamadi et al. (2020) coincide with establishing the hybrid models used and giving better results than other models. Gümüş et al. (2021) predicted the monthly E_p of Diyarbakır and Adıyaman using the parameters of wind speed, temperature, pressure, relative humidity, monthly clear days and sunshine

Table 4 Test results of installed single-GBM and hybrid GBM models

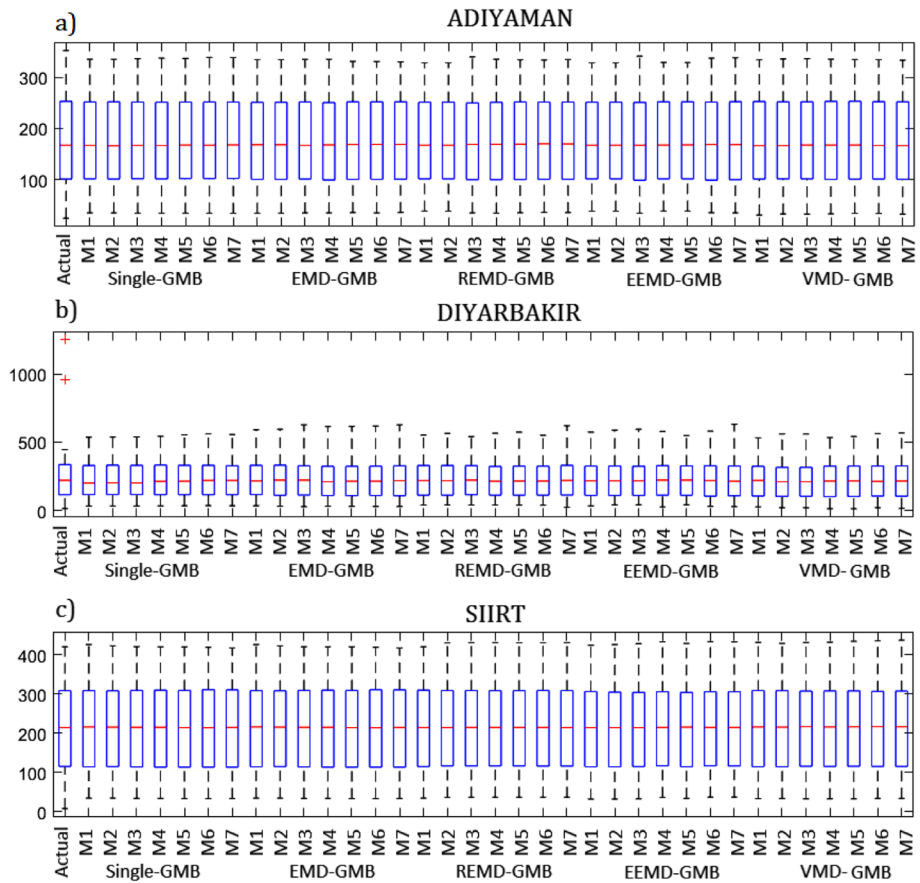
	Adiyaman				Diyarbakır				Siirt			
	RMSE	R ²	MAE	NSE	RMSE	R ²	MAE	NSE	RMSE	R ²	MAE	NSE
<i>Single-GBM</i>												
M1	3.5270	0.9985	1.5997	0.9984	107.5144	0.7602	24.7992	0.7024	3.4239	0.9991	1.4820	0.9991
M2	3.4772	0.9986	1.6243	0.9985	107.4610	0.7596	24.9940	0.7027	3.7159	0.9990	1.7183	0.9989
M3	3.5958	0.9985	1.7241	0.9983	106.6776	0.7635	24.8161	0.7070	3.6970	0.9990	1.7852	0.9989
M4	3.8405	0.9982	1.8882	0.9981	105.2992	0.7723	24.5827	0.7145	3.7886	0.9989	1.9704	0.9989
M5	4.1644	0.9979	2.0935	0.9978	104.6984	0.7767	24.6165	0.7178	3.9643	0.9988	2.0493	0.9988
M6	4.4114	0.9976	2.3278	0.9975	102.9976	0.7948	24.6165	0.7269	4.1329	0.9987	2.1424	0.9987
M7	4.4005	0.9977	2.3672	0.9975	101.1294	0.8049	24.6496	0.7367	3.8632	0.9989	1.9324	0.9988
Best Model	M2				M7				M1			
<i>EMD-GBM*</i>												
M1	3.2638	0.9989	1.9455	0.9986	97.4017	0.8097	25.4164	0.7558	4.0233	0.9988	2.2464	0.9987
M2	3.4869	0.9986	1.6390	0.9984	95.7062	0.8174	25.2950	0.7642	4.6102	0.9984	2.2021	0.9983
M3	3.5196	0.9986	2.2642	0.9984	89.8752	0.8469	22.7148	0.7920	4.2631	0.9986	2.2679	0.9986
M4	3.8209	0.9983	2.3158	0.9981	93.5964	0.8313	26.2368	0.7745	4.2227	0.9986	2.2109	0.9986
M5	3.8386	0.9983	2.3005	0.9981	92.9214	0.8357	25.9258	0.7777	4.0978	0.9987	2.2862	0.9987
M6	4.3420	0.9981	2.0569	0.9980	92.7338	0.8367	25.6086	0.7786	4.1269	0.9987	2.4119	0.9987
M7	3.7491	0.9983	2.3186	0.9982	91.6685	0.8405	25.4576	0.7837	4.0972	0.9987	2.3030	0.9987
Best Model	M1*				M3				M1			
<i>REMD-GBM</i>												
M1	4.1091	0.9983	1.9448	0.9989	102.2299	0.7933	26.3926	0.7599	4.1431	0.9987	1.9078	0.9988
M2	4.1195	0.9983	1.9476	0.9988	100.6999	0.8019	27.8536	0.7736	4.2613	0.9986	2.0367	0.9987
M3	4.0739	0.9983	1.9462	0.9988	104.0196	0.7887	25.9663	0.7380	4.2896	0.9986	2.0453	0.9987
M4	4.1771	0.9983	1.9854	0.9985	101.4054	0.8002	28.2251	0.7620	4.2987	0.9986	2.0471	0.9987
M5	4.1249	0.9983	1.9657	0.9984	100.4162	0.8049	27.9115	0.7630	4.2680	0.9986	2.0252	0.9987
M6	4.1298	0.9983	1.9544	0.9983	103.2920	0.7889	25.9847	0.7819	4.2428	0.9987	1.9990	0.9988
M7	4.1299	0.9983	1.9501	0.9984	99.4773	0.8108	26.8619	0.7797	4.2389	0.9987	2.0520	0.9988
Best Model	M3				M7				M1			
<i>EEMD-GBM</i>												
M1	4.1267	0.9983	1.9409	0.9985	102.8956	0.7934	26.2690	0.7396	4.2258	0.9987	1.9339	0.9989
M2	4.1277	0.9983	1.9472	0.9986	102.5148	0.7959	26.2805	0.7496	4.2337	0.9987	2.0163	0.9989
M3	4.0966	0.9983	1.9264	0.9986	101.8025	0.7991	26.0652	0.7541	4.2925	0.9986	2.0412	0.9987
M4	4.1682	0.9981	1.8759	0.9985	103.6336	0.7875	26.1472	0.7508	4.2899	0.9986	2.0370	0.9986
M5	4.1848	0.9981	1.9110	0.9985	103.4903	0.7885	26.2078	0.7560	4.2931	0.9986	2.0226	0.9986
M6	4.1935	0.9981	1.8815	0.9985	103.6165	0.7889	26.2111	0.7574	4.3520	0.9986	2.0202	0.9985
M7	4.1035	0.9982	1.8110	0.9985	102.4165	0.7947	26.2729	0.7913	4.3210	0.9986	2.0046	0.9987
Best Model	M3				M3				M1			
<i>VMD-GBM</i>												
M1	4.1155	0.9982	1.9376	0.9984	103.4636	0.7845	29.5162	0.7244	4.2171	0.9986	2.2038	0.9986
M2	4.1058	0.9982	1.9289	0.9984	103.4636	0.7845	29.5162	0.7244	4.5229	0.9985	2.5584	0.9984
M3	4.1037	0.9983	1.9373	0.9984	107.8430	0.7593	34.4306	0.7006	4.6316	0.9984	2.6756	0.9983
M4	4.1815	0.9982	1.9075	0.9982	105.9240	0.7707	33.7602	0.7111	4.8240	0.9983	2.8187	0.9982
M5	4.3036	0.9981	2.0111	0.9981	102.5491	0.7862	32.9457	0.7293	4.7042	0.9983	2.7945	0.9983
M6	4.2948	0.9981	2.0072	0.9981	102.5491	0.7862	32.9457	0.7293	4.8263	0.9982	3.0544	0.9982
M7	4.3124	0.9981	2.0126	0.9980	101.5292	0.7883	33.4962	0.7346	4.9637	0.9981	3.0893	0.9981

Table 4 (continued)

	Adıyaman				Diyarbakır				Siirt			
	RMSE	R^2	MAE	NSE	RMSE	R^2	MAE	NSE	RMSE	R^2	MAE	NSE
Best Model	M3				M7				M1			

Bold characters indicate the best input combination of the selected model, and * indicates the most successful model among all models

Fig. 8 Evaluation of model performances with Boxplots
a Adıyaman, **b** Diyarbakır, and **c** Siirt



intensity. They used ANN, ANFIS, and GEP algorithms with different input combinations while estimating evaporation and found that the GEP algorithm gave better results than other algorithms. This study was conducted in terms of both the region where the study was conducted and the parameters used by Gümüş et al. (2021) shows great similarities with the evaporation estimation study. Although the artificial intelligence models used in the estimation of evaporation are different, it is thought that the study will be of great importance in estimating evaporation for the region. Gümüş et al. (2016) used the climatic parameters of wind speed, monthly average temperature, humidity, pressure, sunshine intensity and duration to estimate Adana's monthly average evaporation. For the modeling process, ANFIS, ANN and GEP algorithms were used and found that the evaporation estimation of all methods used gave good results. Still, the

combination of 6 inputs gave the best results in the ANFIS method. The findings obtained from the study by Gümüş et al. (2016) evaporation prediction research in which the parameters used and the study area overlap in terms of being a close region. Although the machine learning algorithms used are different, it is evident that there will be essential data about the evaporation status of the area.

Dehghanipour et al. (2021) used a MLP-NN and genetic algorithm model to predict pan evaporation under different climatic conditions by using temperature (T), relative humidity (RH), wind speed (WS), sunshine hours (SH) parameters in Iran. It was employed the best overall relationship in the first method as the main relationship in the second method and determined the climatic correction coefficients for the other climate types with the genetic algorithm rhythm optimization model. Ultimately,

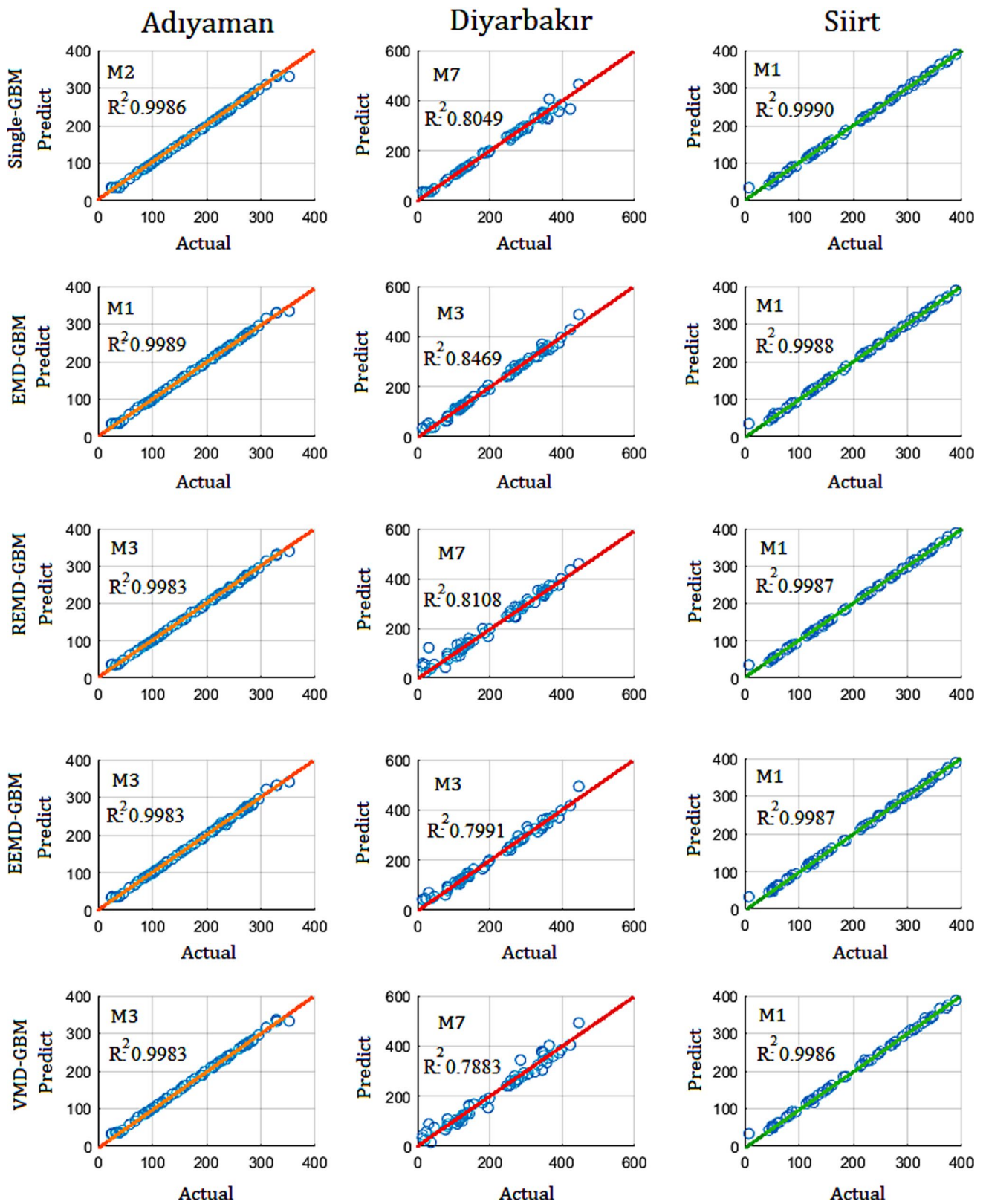


Fig. 9 Scatter diagrams of test results of top models

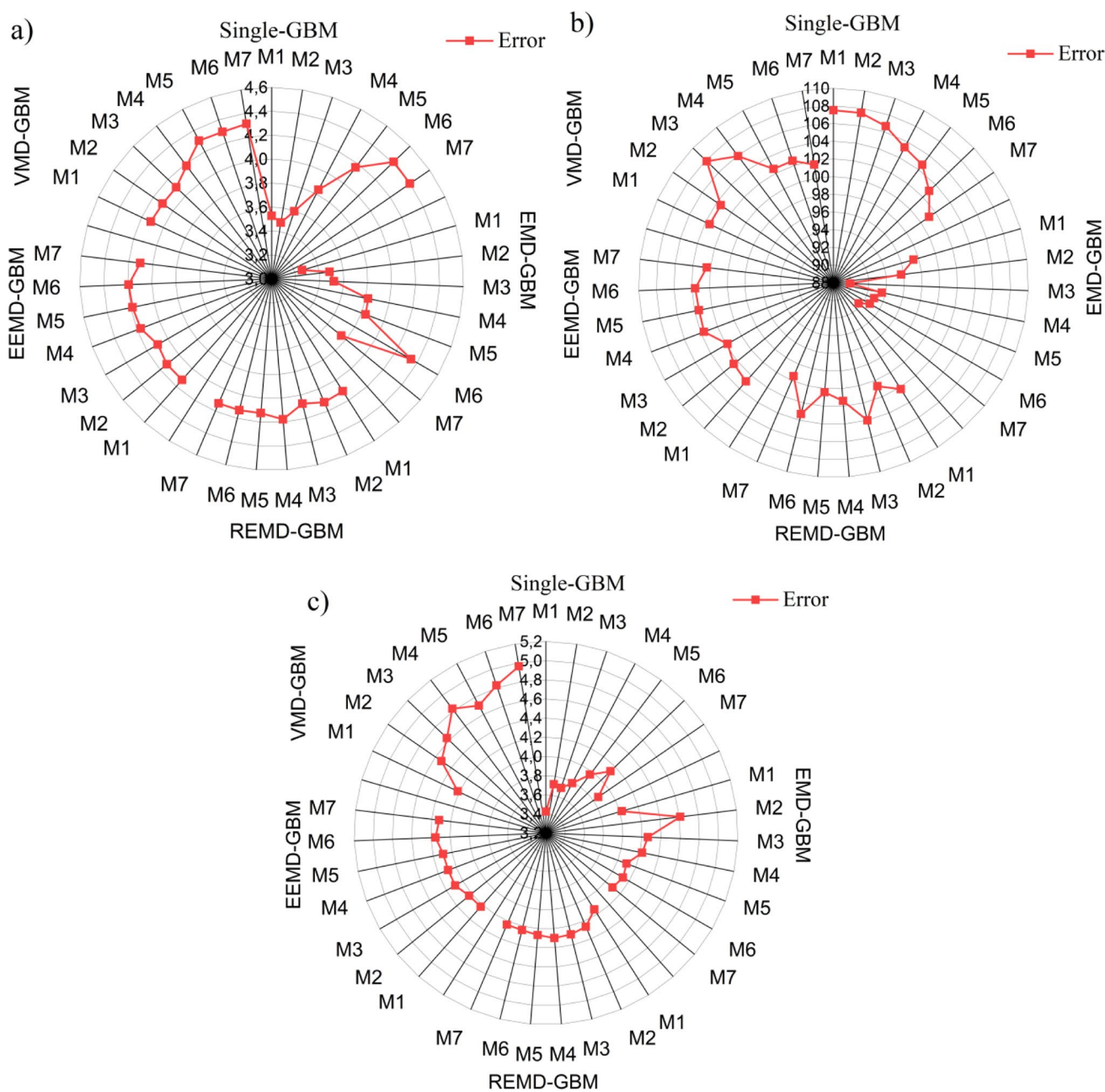


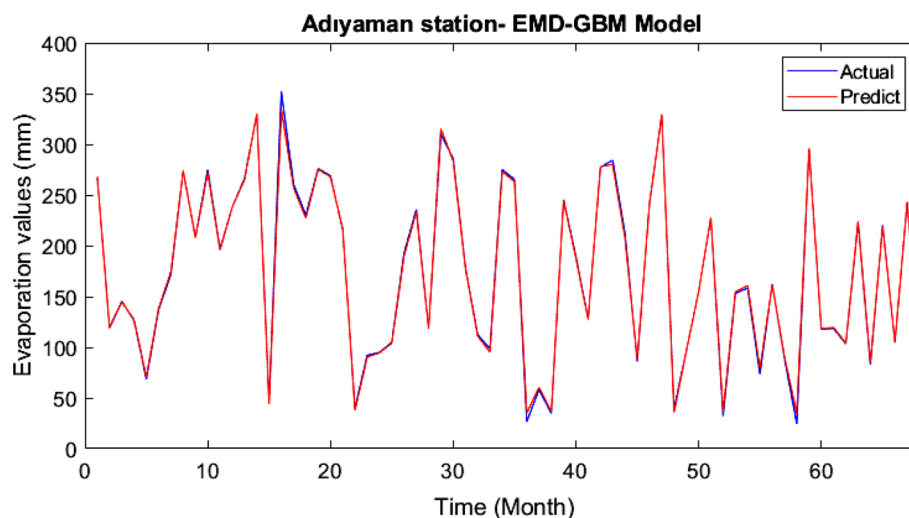
Fig. 10 Analysis of the variation of errors with radar plot a Adiyaman, b Diyarbakir, and c Siirt

they determined that both methods gave accurate results at study area. In this study, 8 parameters were used for evaporation estimation for 3 regions, while 4 parameters were used for 6 different regions in Dehghanipour et al. (2021) evaporation estimation study. The ML algorithms used when creating models are completely different from each other. Only in terms of parameters used Dehghanipour et al. (2021) study is partially similar to this study.

Conclusion

This study proposed a new hybrid model combining GBM and signal processing techniques to estimate monthly open surface evaporation values in the GAP region. In addition, the study analyzed the effect of signal preprocessing techniques on the performance of the ML algorithm. The study outputs constitute a resource for decision-makers

Fig. 11 Comparison of evaporation time series of the best model



and planners in managing water resources, irrigation planning, developing irrigation systems and construction of water structures in the GAP region. The main outputs of the research are summarized as follows:

- According to the correlation analysis, EVP values and RH, AP and P variables show a negative relationship, while T_{avg} , T_{max} , WS, RH and ST variables show a positive relationship.
- In addition, evaporation values in the GAP region were found to have the highest correlation with average temperature values.
- Evaporation values can be estimated with accuracy close to reality with the Single-GBM model.
- When various signal separation techniques such as EMD, REMD, EEMD and VMD are combined with the GBM model, generally more accurate evaporation estimates can be made than the Single-GBM model.
- The highest prediction success was obtained with the EMD-GBM hybrid model and M1 Model input combinations (P , T_{avg}) at Adiyaman meteorology station (RMSE: 3.2638, R^2 : 0.9989, MAE: 1.9455, NSE: 0.9986). Although the most successful results are obtained with the combination of EMD-GBM model and P , T_{avg} input in the study area with a semi-arid climate, it will be useful to compare the results by evaluating all input combinations in different climatic regions. Because parameters such as water resources, climate, altitude and pressure centers vary in different regions.

In future studies, evaporation prediction accuracy can be researched by combining bio-inspired algorithms such as gray wolf optimizer, whale optimization, ant colony optimization, artificial bee colony, and chaos game optimization.

Appendix

See Table 5.

Table 5 Explanations of terms used in the study

GBM	Gradient boosting machines
EMD	Empirical mode decomposition
REMD	Robust empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
VMD	Variational mode decomposition
DWT	Discrete wavelet transform
SVM	Support vector machines
QPSO	Quantum-behavioral particle swarm optimization
MLP-QPSO	Multilayer perceptron quantum-behavioral particle swarm optimization
FFAs	Firefly algorithms
SA	Shark algorithm
WOA	Whale optimization algorithm
ELM	Extreme learning machine
FPA	Flower pollination algorithm
MARS	Multivariate adaptive regression splines
BaggedMARS	Bagged multivariate adaptive regression splines
CRFR	Conditional random forest regression
KKNN	Weighted K-nearest neighbor

Acknowledgements Thanks to the state meteorology general directorate for providing the meteorological data used in the study.

Author contribution OMK contributed to the data analysis, findings, and conclusions. MS contributed with data collection, literature review, and writing methods. All authors read and approved the final manuscript.

Funding No funding was received for conducting this study.

Data availability Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The author declares no conflict of interest.

Ethical approval The manuscript complies with all the ethical requirements. The paper was not published in any journal.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Ali Ghorbani M, Kazempour R, Chau KW, Shamshirband S, Taherei Ghazvinei P (2018) Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran. *Eng Appl Comput Fluid Mech* 12(1):724–737. <https://doi.org/10.1080/19942060.2018.1517052>
- Al-Mukhtar M (2021) Modeling of pan evaporation based on the development of machine learning methods. *Theoret Appl Climatol* 146(3):961–979. <https://doi.org/10.1007/s00704-021-03760-4>
- 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. <https://doi.org/10.1016/j.compag.2016.11.011>
- Asante-Okyere S, Shen C, Ziggah YY, Rulegeya MM, Zhu X (2020) A novel hybrid technique of integrating gradient-boosted machine and clustering algorithms for lithology classification. *Nat Resour Res* 29(4):2257–2273. <https://doi.org/10.1007/s11053-019-09576-4>
- Başakın EE, Ekmekcioğlu Ö, Özger M (2021) Drought prediction using hybrid soft-computing methods for semi-arid region. *Model Earth Syst Environ* 7(4):2363–2371. <https://doi.org/10.1007/s40808-020-01010-6>
- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE). *Geosci Model Develop Discuss* 7(1):1525–1534. <https://doi.org/10.5194/gmd-7-1247-2014,2014>
- Dalkılıç Y, Okkan U, Baykan N (2014) Comparison of different ANN approaches in daily pan evaporation prediction. *J Water Resour Protection*. <https://doi.org/10.4236/jwarp.2014.64034>
- Dazzi S, Vacondio R, Mignosa P (2021) Flood stage forecasting using machine-learning methods: a case study on the Parma River (Italy). *Water* 13(12):1612. <https://doi.org/10.3390/w13121612>
- Dehghani R, Torabi Poudeh H, Izadi Z (2022) Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. *Model Earth Syst Environ* 8(2):2599–2613. <https://doi.org/10.1007/s40808-021-01253-x>
- Dehghanipour MH, Karami H, Ghazvinian H, Kalantari Z, Dehghanipour AH (2021) Two comprehensive and practical methods for simulating pan evaporation under different climatic conditions in Iran. *Water* 2021(13):2814. <https://doi.org/10.3390/w13202814>
- Feng Y, Jia Y, Zhang Q, Gong D, Cui N (2018) National-scale assessment of pan evaporation models across different climatic zones of China. *J Hydrol* 564:314–328. <https://doi.org/10.1016/j.jhydrol.2018.07.013>
- Gap.gov (2022) <http://www.gap.gov.tr/gap-nedir-sayfa-1.html> (Recived:07.12.2022)
- Gümüş V, Şimşek O, Soydan NG, Aköz MS, Yenigün K (2016) Adana istasyonunda buharlaşmanın farklı yapay zeka yöntemleri ile tahmini. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi* 7(2):309–318
- Gümüş V, Yeşiltaş Y, Şimşek O (2021) Yapay Zekâ Yöntemleri ile Adıyaman ve Diyarbakır İstasyonlarının Aylık Tava Buharlaşmalarının Tahmin Edilmesi. *Türk Doğa ve Fen Dergisi* 10(2):112–122. <https://doi.org/10.46810/tdfd.893630>
- Gundalia MJ, Dholakia MB (2013) Modelling pan evaporation using mean air temperature and mean pan evaporation relationship in middle south Saurashtra Region. *Int J Water Resour Environ Eng* 5(11):622–629. <https://doi.org/10.5897/IJWREE2013.0426>
- Jasmine M, Mohammadian A, Bonakdari H (2022) On the prediction of evaporation in arid climate using machine learning model. *Math Comput Appl* 27(2):32. <https://doi.org/10.3390/mca27020032>
- Kedadouche M, Thomas M, Tahan AJMS (2016) A comparative study between Empirical wavelet transforms and Empirical mode decomposition methods: application to bearing defect diagnosis. *Mech Syst Signal Process* 81:88–107. <https://doi.org/10.1016/j.ymssp.2016.02.049>
- Liu Z, Peng D, Zuo MJ, Xia J, Qin Y (2022) Improved Hilbert-Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. *ISA Trans* 125:426–444. <https://doi.org/10.1016/j.isatra.2021.07.011>
- Lu X, Ju Y, Wu L, Fan J, Zhang F, Li Z (2018) Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models. *J Hydrol* 566:668–684. <https://doi.org/10.1016/j.jhydrol.2018.09.055>
- Malik A, Kumar A (2015) Pan evaporation simulation based on daily meteorological data using soft computing techniques and multiple linear regression. *Water Resour Manage* 29(6):1859–1872. <https://doi.org/10.1007/s11269-015-0915-0>
- Malik A, Saggi MK, Rehman S, Sajjad H, Inyurt S, Bhatia AS, Yaseen ZM (2022) Deep learning versus gradient boosting machine for pan evaporation prediction. *Eng Appl Comput Fluid Mech* 16(1):570–587. <https://doi.org/10.1080/19942060.2022.2027273>
- Mohamadi S, Ehteram M, El-Shafie A (2020) Accuracy enhancement for monthly evaporation predicting model utilizing evolutionary machine learning methods. *Int J Environ Sci Technol* 17(7):3373–3396. <https://doi.org/10.1007/s13762-019-02619-6>
- Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):1536. <https://doi.org/10.3390/w10111536>
- Nhu VH, Shahabi H, Nohani E, Shirzadi A, Al-Ansari N, Bahrami S, Nguyen H (2020) Daily water level prediction of Zrebar Lake (Iran): a comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. *ISPRS Int J Geo-Inf* 9(8):479. <https://doi.org/10.3390/ijgi9080479>
- Pammar L, Deka PC (2017) Daily pan evaporation modeling in climatically contrasting zones with hybridization of wavelet transform and support vector machines. *Paddy Water Environ* 15(4):711–722. <https://doi.org/10.1007/s10333-016-0571-x>
- Pardo S, Dunne N, Simmons DA (2017) Using radar plots to demonstrate the accuracy and precision of 6 blood glucose monitoring systems. *J Diabetes Sci Technol* 11(5):966–969. <https://doi.org/10.1177/193229681771130>
- Rezaie-Balf M, Kisi O, Chua LH (2019) Application of ensemble empirical mode decomposition based on machine learning methodologies in forecasting monthly pan evaporation. *Hydrol Res* 50(2):498–516. <https://doi.org/10.2166/nh.2018.050>
- Vicente-Serrano SM, Bidegain M, Tomas-Burguera M, Dominguez-Castro F, El Kenawy A, McVicar TR, Giménez A (2018) A comparison of temporal variability of observed and model-based pan

- evaporation over Uruguay (1973–2014). *Int J Climatol* 38(1):337–350. <https://doi.org/10.1002/joc.5179>
- Wang Y, Markert R (2016) Filter bank property of variational mode decomposition and its applications. *Signal Process* 120:509–521. <https://doi.org/10.1016/j.sigpro.2015.09.041>
- Wang L, Kisi O, Hu B, Bilal M, Zounemat-Kermani M, Li H (2017) Evaporation modelling using different machine learning techniques. *Int J Climatol* 37:1076–1092. <https://doi.org/10.1002/joc.5064>
- Wu Z, Huang NE, Chen X (2009) The multi-dimensional ensemble empirical mode decomposition method. *Adv Adapt Data Anal* 1(03):339–372. <https://doi.org/10.1142/S1793536909000187>
- Wu L, Huang G, Fan J, Ma X, Zhou H, Zeng W (2020) Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction. *Comput Electron Agric* 168:105115. <https://doi.org/10.1016/j.compag.2019.105115>
- Yaseen ZM, Al-Juboori AM, Beyaztas U, Al-Ansari N, Chau KW, Qi C, Shahid S (2020) Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models. *Eng Appl Comput Fluid Mech* 14(1):70–89. <https://doi.org/10.1080/19942060.2019.1680576>
- Zare M, Koch M (2013) An Analysis of MLR and NLP for use in river flood routing and comparison with the Muskingum method. IAHR World Congress
- Zhang J, Yan R, Gao RX, Feng Z (2010) Performance enhancement of ensemble empirical mode decomposition. *Mech Syst Signal Process* 24(7):2104–2123. <https://doi.org/10.1016/j.ymsp.2010.03.003>
- 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>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.