

e-ISSN:2757-9077

ENGINEERING PERSPECTIVE

https://sciperspective.com

Research Paper

Hybrid VMD-EMD Enhanced Machine Learning Models for Daily Reference Evapotranspiration Forecasting

Reşat Acar[1], Okan Mert Katipoğlu[2], Burak Çırağ[1*]

[1]*Department of Civil Engineering, Faculty of Engineering, Atatürk University, Erzurum 25100, Türkiye.*
[2]*Department of Civil Engineering, Faculty of Engineering and Architecture, Erzincan Binali Yıldırım University, Erzincan 24002, Türkiye*

ABSTRACT

Getting accurate daily estimates of reference evapotranspiration (ET_0) is essential for hydrological modeling, irrigation planning, and climate research. This study evaluated the performance of four machine learning models: Adaptive Boosting (AdaBoost), Categorical Boosting (CatBoost), Deep Neural Network (DNN), and Light Gradient Boosting Machine (LGBM) for daily ET_0 prediction. Feature selection was implemented using Partial Autocorrelation (PACF) and Autocorrelation (ACF) analysis on the basis of an 8-day lag structure, while the CatBoost feature importance framework was implemented to identify the most important inputs. Moreover, the methodologies of decomposition-based preprocessing, Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) were implemented on the lagged inputs to build the hybrid models augmented with intrinsic constituents. These augmented hybrid models were created to improve the data representation as well as predictive capability. Model performance was evaluated using key statistical metrics, including RMSE, MAE, NSE, KGE, R^2, and bias-related indicators. CatBoost produced the lowest RMSE (1.57 mm/day) and MAE (1.13 mm/day), and the highest NSE (0.67), KGE (0.78), and R^2 (0.67), along with the lowest AIC score (627.15), indicating strong and reliable predictive performance. Furthermore, the VMD based CatBoost model yielded even greater improvements, highlighting the added value of decomposition enhanced learning. On the contrary, DNN showed the weakest results, with the highest RMSE (2.39 mm/day) and lowest NSE (0.23). These findings suggest that CatBoost and its hybrid variants are highly effective and reliable tools for daily ET_0 prediction, particularly in data limited environments and mountainous regions.

Keywords: Boosting model; Evapotranspiration; Feature Importance; Machine Learning; Tree algorithm

History
Received :10.03.2026
Revised :29.04.2026
Accepted :11.05.2026

Author Contacts : racar@atauni.edu.tr , okatipoglu@erzincan.edu.tr , burak.cirag@atauni.edu.tr

Cite this paper: Acar, R. Katipoğlu, O.M., Çırağ, B. (2026). Hybrid VMD-EMD Enhanced Machine Learning Models for Daily Reference Evapotranspiration Forecasting. Engineering Perspective, 6 (3), 354-373. https://doi.org/10.64808/engineeringperspective.1906862

*Corresponding Author

1. Introduction

To effectively evaluate and analyze surface energy, the carbon cycle, water balance, and the management of water resources, we must focus on evapotranspiration (ET). This process encompasses the total flow of water vapor that leaves the soil through evaporation and the vegetation through transpiration, ultimately entering the atmosphere [1-2]. Reference evapotranspiration (ET_0) is a variant of the general ET concept, representing the evaporation of water from vegetation and soil surface under standard conditions. ET_o consists of evaporation and transpiration processes. Evaporation refers to the passage of water from the surface to the atmosphere, while transpiration refers to the uptake of water by plant roots and its excretion from leaves through stomata [3-4].

Globally, about 60% of the annual precipitation finds its way back into the atmosphere through a process known as evapotranspiration (ET), as noted by researchers like Oki and Kanae [5]. Understanding and accurately estimating ET is key to grasping the underlying mechanisms of this process, which helps us tackle the uncertainties that climate change brings to the hydrological cycle [6-7]. However, measuring ET accurately is not an easy process. Techniques like the liquid flow method, lysimeter method, and vorticity correlation method can be quite labor-intensive and tricky to implement over large areas [8]. Plus, capturing ET across different spatial and temporal scales is a real challenge since it's nonlinear and heavily influenced by both environmental and human factors [9]. Recent

hydrological studies have also emphasized that the accuracy, reliability, and temporal variability of data series should be carefully evaluated before model development, particularly when time-dependent hydrological processes are analyzed [10].

There are two primary approaches to predicting evapotranspiration (ET): physical-based models and data-driven models. For instance, the Priestley-Taylor equation [11] and the Penman-Monteith model [12-13] are both examples of physical models that consider the underlying physical processes. These models are based on a variety of theoretical hypotheses. In recent years, data-driven models have become increasingly popular for calculating or predicting ET while taking into account fewer physical mechanisms [14]. Dou and Yang [15] compared the extreme learning machine (ELM) and the adaptive neuro-fuzzy inference system (ANFIS) models with artificial neural network (ANN) and support vector machine (SVM) for daily ET prediction in four different ecosystems (grassland, forest, cropland, and wetland). The best accuracy was obtained in forest ecosystems, and the lowest in agricultural areas, and hybrid ELM and uncertainty clustering-based ANFIS models were found to be more successful. The results show that advanced machine learning methods are effective in ET prediction. Rashid Niaghi et al. [16] applied four machine learning methods to determine reference evapotranspiration: Gene Expression Programming (GEP), Support Vector Machine (SVM), Multiple Linear Regression (LR), and Random Forest (RF). They concluded that Random Forest was the most effective across all combinations they examined. Sattari et al. [17] sought to evaluate machine learning techniques for estimating monthly reference evapotranspiration (ET_0) utilizing constrained meteorological data in the Çorum region of Turkey. They examined 10 different combinations of weather data using Gaussian Process Regression (GPR), SVR, Broyden-Fletcher-Goldfarb-Shanno Artificial Neural Network (BFGS-ANN), and Long Short-Term Memory (LSTM) models. The results highlight that all models predict ETo with acceptable accuracy, but the BFGS-ANN model shows the highest success. The objective of Amiri et al. [18] was to refine daily ET_0 forecasting for different climate zones. They developed a two-stage prediction model, named DIRECTORS, that employs pattern mining and random forest techniques to effectively capture the unique behaviors of each climate. The results demonstrated that DIRECTORS attained elevated accuracy and robust generalization capability across all climatic conditions. Elzain et al. [19] sought to evaluate CatBoost against various deep learning models for ET_0 prediction in a semi-humid coastal area. The pseudo-labeling method and different combinations of weather variables were used to train the models. The CatBoost-PL model was more accurate and faster than the others, especially when RH was involved. Fu and Li [20] aimed to estimate monthly evapotranspiration in the southeast margins of the Tengger Desert using a VMD-WOA-LSTM hybrid model, which combines variational mode decomposition (VMD) and the whale optimization algorithm (WOA) with LSTM. They stated that the performance of their proposed hybrid model yielded more accurate results than other machine learning-based models.

The goal of this study is to compare how well different machine learning algorithms can estimate daily reference evapotranspiration (ET_0). The study utilized data sourced from a meteorological station situated in Erzurum province, part of Turkey's Eastern Anatolia Region, noted for its significant altitude, rugged landscape, and persistent snow cover. During the modeling phase, we relied on the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) analyses to enhance the input data derived from the time series. In addition, Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) techniques were employed as preprocessing methods to extract intrinsic components of the ET_0 signal, aiming to improve the models' predictive capacity. We also used the CatBoost feature importance method to assess which variables were most important. We chose AdaBoost, CatBoost, Deep Neural Network (DNN), and Light Gradient Boosting Machine (LGBM) as the best models for making predictions. Each model was evaluated both individually and in hybrid form with the EMD and VMD components to examine the effectiveness of signal decomposition in enhancing model performance. In this respect, the study provides an applied example in which different artificial intelligence-based modeling approaches are compared in the prediction of a hydrometeorological variable with high temporal dependence and meteorological sensitivity, such as evapotranspiration. The resulting model structures are expected to provide useful information for researchers, practicing engineers, and policymakers in areas such as hydrological modeling, water resources management, agricultural irrigation planning, and climate-based decision support systems.

2. Materials and Methods

2.1. Study area and data

Erzurum is a province in eastern Türkiye and is located approximately between latitudes 39°-41° N and longitudes 40°-42° E, with its city center at an elevation of 1959 m above sea level. Erzurum province is of great hydrological importance as the snow and rainwater in Erzurum province is the source of many rivers [21]. The geographical location of the study area is given in Figure 1.

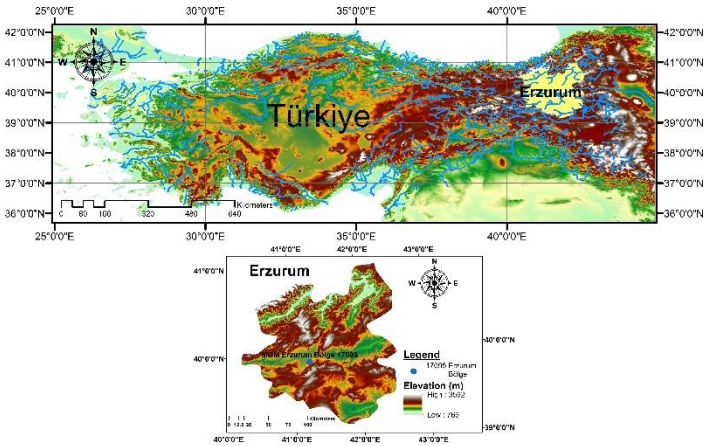
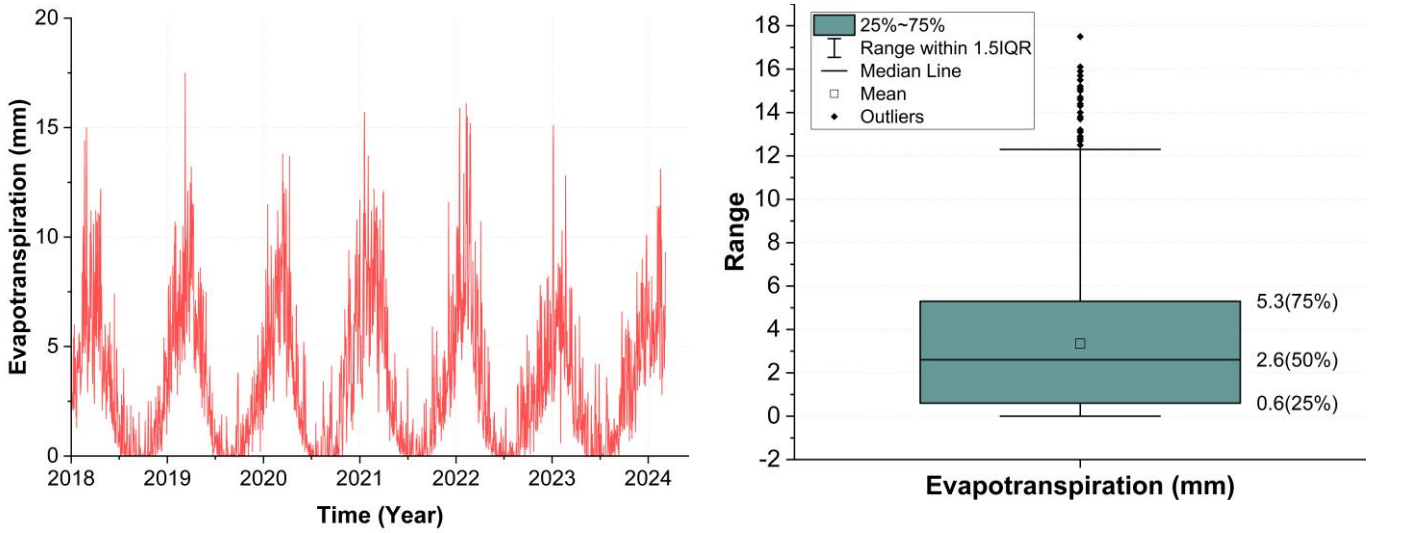


Figure 1. Geographical location of the study area

In this study, we utilized evapotranspiration data from the Erzurum Bölge meteorological station 17095, which was generously provided by the Turkish Meteorology General Directorate (MGM). The dataset includes daily measurements collected from 2018 to 2024. The daily ET_0 records were checked for missing values before model development.



(a) (b)

Figure 2. Time series (a) and box plot (b) of daily reference evapotranspiration (ET_0) values for the period 2018-2024

A limited number of missing records was detected in three short periods: 10-26 August 2020, corresponding to 17 missing daily records; 2-30 March 2022, corresponding to 29 missing daily records; and 5-16 August 2023, corresponding to 12 missing daily records. In total, 58 missing daily values were identified, representing approximately 2.27% of the complete 2018-2024 daily dataset. After removing missing and incomplete lagged observations, the final dataset consisted of approximately 2463 complete samples used for model development. The distribution and box plot of the data are shown in Figure 2.

Missing values were not filled or estimated in this study. Missing records were excluded from the modelling process. During the construction of the lagged input matrix, any sample containing a missing value in $ET_0(t)$, $ET_0(t-1)$, …, $ET_0(t-8)$, or in the target variable $ET_0(t+1)$, was removed from the analysis. Therefore, all models were trained and tested only using observed and complete ET_0 records. Since the missing-data ratio was low, this exclusion-based approach was considered appropriate and prevented the introduction of artificial information into the time-series modelling framework.

To determine the amount of water that evaporates, MGM employs the Penman-Monteith method. This method was designed to calculate evaporation by leveraging energy balance and mass transfer principles, and it has been improved to factor in aerodynamic resistance (r_a) and surface resistance (r_s). Surface resistance is the force that restricts water vapor from moving through the stomata, leaf surfaces, and soil, while aerodynamic resistance is what limits air flow over plants. These factors help us understand the dynamics of wind, turbulence, and plant pores, enabling us to accurately calculate ET_o [22-23]. Below is a general approach to calculating ET_0:

$$\lambda ET = \frac{\Delta\left(R_n - G\right) + \rho_a c_p \dfrac{\left(e_s - e_a\right)}{r_a}}{\Delta + \gamma\left(1 + \dfrac{r_s}{r_a}\right)} \quad (1)$$

where R_n is the net radiation, G is the soil heat flux, e_s - e_a represents the vapour pressure deficit of the air, ρ_a is the mean air density at constant pressure, c_p is the specific heat of air, Δ represents the slope of the saturation vapour pressure curve, γ is the psychrometric constant, and r_s and r_a denote the (bulk) surface resistance and aerodynamic resistance, respectively. This method allows direct calculation of crop evapotranspiration for different plant species by modeling energy exchange and the associated latent heat flux and evapotranspiration processes. Most of the parameters can be directly measured or calculated from meteorological data, so that water loss of vegetation over large areas can be reliably estimated [22].

The proposed modeling framework is based on an auto-regressive prediction approach, where lagged values of ET_0 are used as inputs. Therefore, the term forecasting in this study refers to time-series-based prediction rather than a fully exogenous meteorological forecasting framework.

2.2. Research design and methodology

The Light Gradient Boosting Machine (LGBM), Categorical Boosting (CatBoost), Adaptive Boosting (AdaBoost), and Deep Neural Network (DNN) represent some of the most advanced algorithms in machine learning and deep learning. They're utilized to develop precise prediction models. Over time, these algorithms have adapted to make it easier to learn from intricate, multi-dimensional datasets, resulting in improved accuracy and better generalization. While the methods based on boosting make powerful models through ensemble of weak learners, DNN enhances accuracy through learning deep representations from the data, utilizing its multi-layer architecture. These models have numerous applications such as evapotranspiration estimation, missing value completion as well as other regression and classification issues. Table 1 presents specifications of the models used for the study.

Table 1. Parameters of all models use

Model	Parameter	Default Value	Description	Explanation
LGBM	objective	regression	Objective function (regression, classification, etc.)	Choose the objective function appropriate to the data type.
	metric	MSE	Metric to be used (error measure)	Choose the metric appropriate to your problem (RMSE, MAE, etc.).
	boosting_type	GBDT	Boosting type (GBDT, DART, GOSS)	GBDT generally gives good results, DART can be used in sparse data.
	num_leaves	31	Number of leaves	More leaves, more complex model, watch out for overfitting.
	learning_rate	0.1	Learning rate	Smaller values give better generalization, require more iterations.
	max_depth	-1	Tree depth	-1 unlimited depth, use with caution.
	min_data_in_leaf	20	Minimum number of data per leaf	Prevents overfitting.
	feature_fraction	1.0	Feature sampling rate	1.0 uses all features, smaller values reduce overfitting.
	bagging_fraction	1.0	Data sampling rate	1.0 uses all data, smaller values reduce overfitting.
	lambda_l1	0.0	L1 regularization	Reduces overfitting.
	lambda_l2	0.0	L2 regularization	Reduces overfitting.
CatBoost	iterations	1000	Number of iterations	More iterations, better results, longer training time.
	learning_rate	0.03	Learning rate	Smaller values give better generalization, require more iterations.
	depth	6	Tree depth	Larger values, more complex model, watch out for overfitting.
	l2_leaf_reg	3	L2 regularization	Reduces overfitting.
	random_strength	1	Randomness coefficient	Reduces overfitting.
	bagging_temperature	1	Bagging temperature	Reduces overfitting.
	border_count	254	Number of splits	More splits, better accuracy, longer training time.
AdaBoost	verbose	True	Training output	To monitor the training process.
	n_estimators	100	Number of weak learners	More learners, better results, longer training time.
	learning_rate	1	Learning rate	Smaller values give better generalization, require more iterations.
	loss	linear	Loss function	linear, square, exponential options.
DNN	units	Variable	Number of neurons in layers	Adjusted according to the complexity of the problem.
	activation	relu	Activation function	relu, sigmoid, tanh, etc.
	optimizer	adam	Optimization algorithm	adam, sgd, rmsprop, etc.
	loss	MSE	Loss function	mse, mae, categorical_crossentropy, etc.
	epochs	1000	Number of training rounds	Determine sufficient number of iterations.
	batch_size	32	Batch size	Adjust memory and speed balance.
	learning_rate	0.001	Learning rate	Smaller values give better generalization, require more iterations.

LGBM, a machine learning algorithm based on Gradient Boosting Decision Trees (GBDT), is faster to train, more efficient to compute, and more accurate at making predictions than other tree-based learning algorithms [24-25]. It also has big data modeling that is very accurate, fast, and efficient. The model's main objective is to minimize the loss function over the given supervised training set. In doing so, it operates within the GBDT's iterative framework and follows the negative gradient step-by-step. LGBM offers two main improvements over classical GBDT: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS reduces computational cost while maintaining model accuracy by selecting samples according to the gradient magnitude. EFB, on the other hand, simplifies the feature space by converting mutually exclusive features in high-dimensional sparse data into a single feature and speeds up the histogram construction process. In this way, LightGBM offers an efficient structure in terms of both memory usage and processing time [26-27].

In 2017, Yandex engineers made CatBoost, which is a machine

learning algorithm. It is based on GBDT [28]. This algorithm is special because it can solve classification and regression problems very accurately, even when the data is noisy, has different types of data, or has complex dependencies. One of the best things about CatBoost is that it can handle categorical features well. In standard GBDT algorithms, the mean label value is often used to represent these kinds of variables. CatBoost, on the other hand, adds prior value to this method to reduce the conditional shift problem [28-29]. Moreover, the greedy algorithm combines all categorical features in the current tree with other categorical variables in the dataset to produce richer representations. CatBoost also uses an ordered boosting method to combat the gradient bias issue observed in conventional GBDT systems, so enhancing general performance by avoiding biased gradient estimates from overfitting the model [28, 30]. In these respects, CatBoost provides notable benefits over comparable algorithms in terms of both accuracy and training speed [31-32].

AdaBoost is an ensemble learning method that combines weak classifiers or regressors to make a better and more accurate prediction model [33]. This algorithm starts with a weak learner on the dataset and then uses that learner on the data many times, changing the instance weights based on the errors made at each step. Thus, each new learner focuses on the errors made in the previous steps; in particular, misclassified instances receive a higher weight, increasing the sensitivity of the model to these challenging instances. Thanks to this weighted approach, AdaBoost improves the overall performance of the model by providing bias reduction in classification and regression problems. Furthermore, weight redistribution based on the success of the learner at each step supports the dynamic nature of the algorithm and allows it to produce more generalizable models that focus on the most challenging observations [33-35]. For AdaBoost, the model was implemented using AdaBoostRegressor(n_estimators=100, random_state=42). Recent applications have shown that AdaBoost can be effectively used in engineering-based prediction problems and that model performance can be evaluated using common error metrics such as MSE, RMSE, MAE, MAPE, and correlation-based accuracy indicators [36]. The learning rate was not manually tuned; therefore, the default scikit-learn value of 1.0 was used. This configuration was adopted to provide a consistent and reproducible baseline for comparison with CatBoost, LGBM, and DNN models. Accordingly, the AdaBoost model in this study is defined with n_estimators = 100 and learning_rate = 1.0.

There are several layers in the DNN, including an input layer with several neurons, one or more hidden layers, and an output layer. Each neuron processes the input from the front layer and sends the output to the next layer. The layer-layer structure lets the network learn deep representations and complex patterns, which makes it very accurate at predicting things, especially for large and complex data sets [37-40]. DNN is a feed-forward network built on multi-layer perceptron (MLP) architecture; architectural parameters, including the number of hidden layers, neurons, learning rate, activation functions (ReLU, sigmoid, tanh), dropout, and momentum, directly affect the model performance. In the training stage, the weights and biases of the neurons are optimized by using the backpropagation algorithm and typically stochastic gradient descent (SGD) or Adam optimizer. In training, the accuracy of the model is enhanced by minimizing the loss function (i.e., mean squared error or cross-entropy) [41-44]. In this study, the DNN model was designed as a fully connected feed-forward neural network. The architecture consisted of an input layer, two hidden dense layers with 64 and 32 neurons, respectively, and one linear output neuron for $ET_0(t+1)$ prediction. ReLU activation was used in the hidden layers, whereas a linear activation function was used in the output layer. The model was trained using the Adam optimizer with a learning rate of 0.001, the MSE loss function, 1000 epochs, and a batch size of 32. A validation split of 0.20 within the training set was used to monitor model training. The model was implemented using a fixed architecture without extensive hyperparameter tuning in order to provide a consistent baseline comparison across models.

The dataset was divided using a chronological split rather than a random split. The first 70% of the lagged observations was used for model training, whereas the last 30% was kept as an independent test set. The parameter shuffle=False was used to preserve the temporal order and to avoid information leakage from future observations into the training phase. For DNN training, 20% of the training subset was used as an internal validation subset for monitoring the training process. After removing missing and incomplete lagged observations, the final dataset consisted of a reduced number of complete samples used for model development.

2.2.1. Pre-processing methods

Various preprocessing methods are commonly used to reduce noise in the data, eliminate trend components, and transform the data scale in order to improve model accuracy and minimize prediction errors [45]. Prediction accuracy of machine learning algorithms could be significantly improved by adopting preprocessing techniques that outline the underlying structure of non-stationary time series. In the current study, the initial evapotranspiration signal was decomposed into a small number of intrinsic modes using Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) before the model was trained.

Empirical Mode Decomposition (EMD) is a decomposition technique based on Huang et al. [46], that is essentially data-driven. It breaks down a signal into a complete set of modes that have certain periodicity by using a spline fitting method on the signal's crest and trough points [47]. The EMD process commences with the execution of spline fitting on the extrema (minima, maxima) points in the dataset, followed by the calculation of the mean profile considered based on the previous fittings. The difference between the original series and the mean profile is taken to obtain a zero-mean intrinsic mode function (IMF), and this process is repeated in an iterative process called sifting. A typical IMF must satisfy the conditions that the number of zero-crossings and extrema are equal or differ by at most one, and that the mean value of the envelopes defined by the local maxima and local minima is zero. This method provides an alternative to traditional approaches such as wavelet analysis and Fourier analysis in that it does not use any mathematical functional form during decomposition [46, 48-49].

VMD is a powerful and novel decomposition method used to decompose both stationary and non-stationary time series data into a finite number of intrinsic mode functions (IMFs). In this approach, each mode is compactly centered around its own central frequency, enabling the IMFs to reconstruct the original data series with high

accuracy. With these features, VMD is an innovative model that stands out in the field of decomposition [50-51]. This technique allows the number of modal decompositions to be determined, enabling the selection of the optimal decomposition number suitable for the data set. Especially for complex, nonlinear, and asymmetric serial data, the VMD algorithm produces relatively stable intrinsic modal functions that exhibit gradual frequency variations and are suitable for analyzing data volatility. The essence of VMD is based on a variation problem, and its fundamental approach is to construct and solve this problem [52-53].

In this study, VMD was implemented in MATLAB using the command vmd (current_column) for each lagged ET_0 input series. The number of modes was not manually optimized through a grid-search procedure; instead, the default decomposition settings of the MATLAB vmd() function were used. Under this implementation, each input series was decomposed into five intrinsic mode functions, IMF1-IMF5, and one residual component. Therefore, the effective number of modes used in this study was K = 5. The use of five modes was considered appropriate for the present study based on both the decomposition characteristics of the ET_0 series and the subsequent correlation-based evaluation of the extracted IMFs. A lower number of modes could merge different frequency structures into the same component, whereas a larger number of modes could unnecessarily increase the input dimensionality and the risk of overfitting, particularly under a relatively limited daily dataset. After decomposition, the extracted IMFs were evaluated using correlation analysis with the target variable $ET_0(t+1)$. The results showed that some VMD-derived components, particularly IMF5, carried strong predictive information across different lag periods. Therefore, K = 5 was treated as a balanced decomposition setting obtained from the MATLAB VMD implementation and supported by the subsequent component relevance analysis.

2.3. Performance metric indexes

In this study, important statistical metrics such as the Nash-Sutcliffe efficiency (NSE) [54], which is widely used for the evaluation of hydrological models, and the coefficient of determination (R^2), which represents the extent to which the independent variable explains the dependent variable, were employed. We also used Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Akaike Information Criterion (AIC), Kling-Gupta Efficiency (KGE) [55], Bias Factor (BF), Mean Bias Error (MBE), and Percent Bias (PBIAS) to rate how well the models worked. A lower RMSE, MBE, MAE, and PBIAS means that the model is more accurate, while R^2, NSE, and KGE values that are closer to one mean that the model is better. A lower AIC value suggests that the model fits better. These metrics create a comprehensive evaluation framework, helping to identify the model with the best predictive accuracy and generalization capability. The equations that outline these metrics can be found in the relevant sections [56-57].

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(EVP_i - EVP_i\right)^2} \quad (2)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|EVP_i - EVP_i\right| \quad (3)$$

$$AIC = 2k - 2\ln\left(L\right) \quad (4)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{m}\left(EVP_i - EVP_i\right)^2}{\sum_{i=1}^{m}\left(EVP_i - EVP\right)^2}\right] \quad (5)$$

$$KGE = 1 - \sqrt{\left(r-1\right)^2 + \left(\alpha-1\right)^2 + \left(\beta-1\right)^2} \quad (6)$$

$$R^2 = \sqrt{\frac{\sum_{i=1}^{m}\left(EVP_i - EVP_i\right)^2}{m}} \quad (7)$$

$$MBE = \left[\frac{1}{m}\sum_{i=1}^{m}\left(EVP_i - EVP_i\right)\right] \quad (8)$$

$$BF = \frac{1}{m}\sum_{i=1}^{m}\frac{EVP}{EVP_i} \quad (9)$$

$$PBIAS = 100\frac{\sum_{i=1}^{m}\left(EVP_i - EVP_i\right)^2}{\sum_{i=1}^{m}EVP_i} \quad (10)$$

In this context, m represents the total number of observations that have been made, while EVP_i denotes each of the observed values individually. $\widehat{EVP_i}$ represents the model's predicted output or estimated values, $\bar{x}$ denotes the average (mean) of the actual or observed data values, k signifies the number of parameters used within the model, L indicates the maximum likelihood, a statistical measure of how well the model fits the data, r represents the correlation coefficient, quantifying the strength and direction of the linear relationship between predicted and observed values, and β represents the flow ratio.

AIC was calculated as an error-based comparative criterion using AIC=nln(MSE)+2k, where n is the number of test observations, MSE is the mean squared error obtained on the test set, and k represents the number of input predictors used in the corresponding model. In this study, k was therefore defined according to the number of model inputs rather than the total number of internal trainable parameters of complex algorithms such as DNN, CatBoost, or LGBM. This formulation was used to provide a consistent error-complexity comparison among models under the same prediction framework. Negative AIC values observed in hybrid models are mathematically valid and arise when the ln(MSE) term becomes negative due to low prediction error values.

3. Results

To identify the input variables for the model, we utilized graphs from the Partial Autocorrelation Function (PACF) and the Autocorrelation Function (ACF), both of which are standard tools in time series analysis.

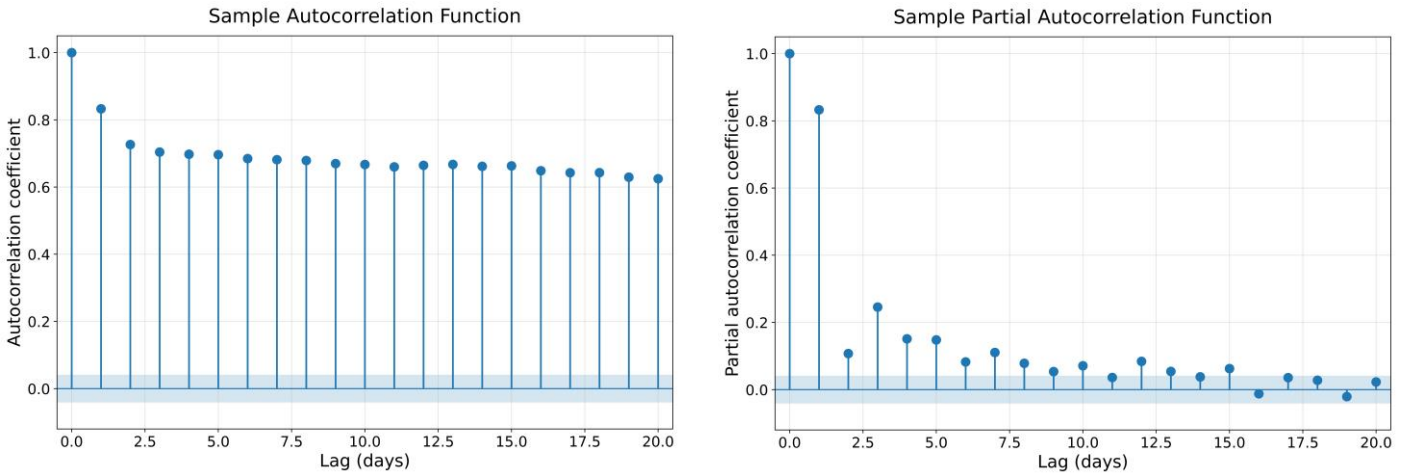


Figure 3. ACF and PACF graphs generated for feature selection

The ACF reveals how each lag in the series relates linearly to its past values. The PACF, on the other hand, shows how that specific lag directly affects the series and gives a stronger evaluation. Moreover, the selection of appropriate input components affects the forecasting performance of the models [58]. The ACF and PACF plots created to refine the feature selection within the scope of the study are given in Figure 3.

A look at the PACF graph from this study shows that the values after the 8th lag are all within the 95% confidence interval. That means that the first 8 lags of the series have a lot of autocorrelation. After that, though, the statistical significance starts to go down. In this context, only lagged values corresponding to the first 8 days are included in the model as input features. This avoids the unnecessary increase in model complexity while preserving inputs with high information content. Otherwise, the use of a large number of lagged inputs may cause the model to overfit the training data, i.e., overfitting, and reduce its generalization capacity against new data. Especially in machine learning based models, such unnecessary input redundancy may result in performance degradation. For this reason, the PACF-based feature selection approach provides a structure that is both informationally meaningful and simple.

In summary, the 8-day lag structure was selected based on the combined interpretation of ACF and PACF plots. While ACF was used to evaluate the overall serial dependence, PACF was used to identify the direct contribution of individual lags after removing the influence of shorter lags. The PACF values beyond the 8th lag was generally not statistically significant at the 95% confidence level, indicating that additional lags would provide limited independent information. Therefore, $ET_0(t)$, $ET_0(t-1)$, …, $ET_0(t-8)$ were used as predictors for $ET_0(t+1)$. This choice was also supported by CatBoost feature importance analysis, which showed that the current ET_0 value was the dominant predictor, while the previous 8 days collectively contributed to short-term temporal memory. The residual plots and density scatters generated to see the effectiveness of the model results are presented in Figure 4 and 5.

In CatBoost and LGBM models in test phase, the concentration of residual values within the ±2 band and their homogeneous distribution around zero indicate that these models produce balanced and low-bias estimates. Furthermore, the models show high overlap between estimated values and actual values in density scatter plots. Especially in the CatBoost model, high-density predictions are clustered close to the true values, indicating the strong generalization capacity of the model. In the DNN model, a single linear trend is observed in the residual plot, indicating that the model has a tendency for systematic prediction errors. This left-skewed trend indicates that the model tends to overestimate smaller values while underestimating larger ones. Such a trend is also evident in the density scatter plot, notably where the prediction deviates from actual values and congregates along a defined linear path. The results reveal that the DNN model doesn't perform well when it comes to generalizing, especially at very high or low values. On the flip side, the AdaBoost model shows a more chaotic and unstable residual distribution compared to the others. This implies that its predictive accuracy is compromised. Additionally, the density scatter plot indicates that the predictions are not entirely accurate; they are scattered and exhibit lower density. After the hybridization with EMD, the residual structures of CatBoost and LGBM remained stable, but no noticeable improvements were observed in their prediction density. Especially in EMD-CatBoost, some clustering artifacts appear in the residual distribution, suggesting overfitting risks caused by component-wise expansion of the input data. On the other hand, the VMD-based hybrid models show a big improvement in both the symmetry of the residuals and the alignment of the density. The VMD-CatBoost and VMD-LGBM models, in particular, exhibit sharply clustered points along the 1:1 line in the density scatter plots and homogeneously narrow-band residuals. These findings indicate that the VMD-based approach significantly enhances predictive sharpness and structural consistency, especially in tree-based models.

A holistic analysis of the error density, time series, and polar coordinate error graphs of the models provides the opportunity to evaluate the behavior of the models in a multidimensional way by revealing the spatial and temporal characteristics of the error distribution as well as the prediction accuracy. The graphs created in this context are given in Figure 6.

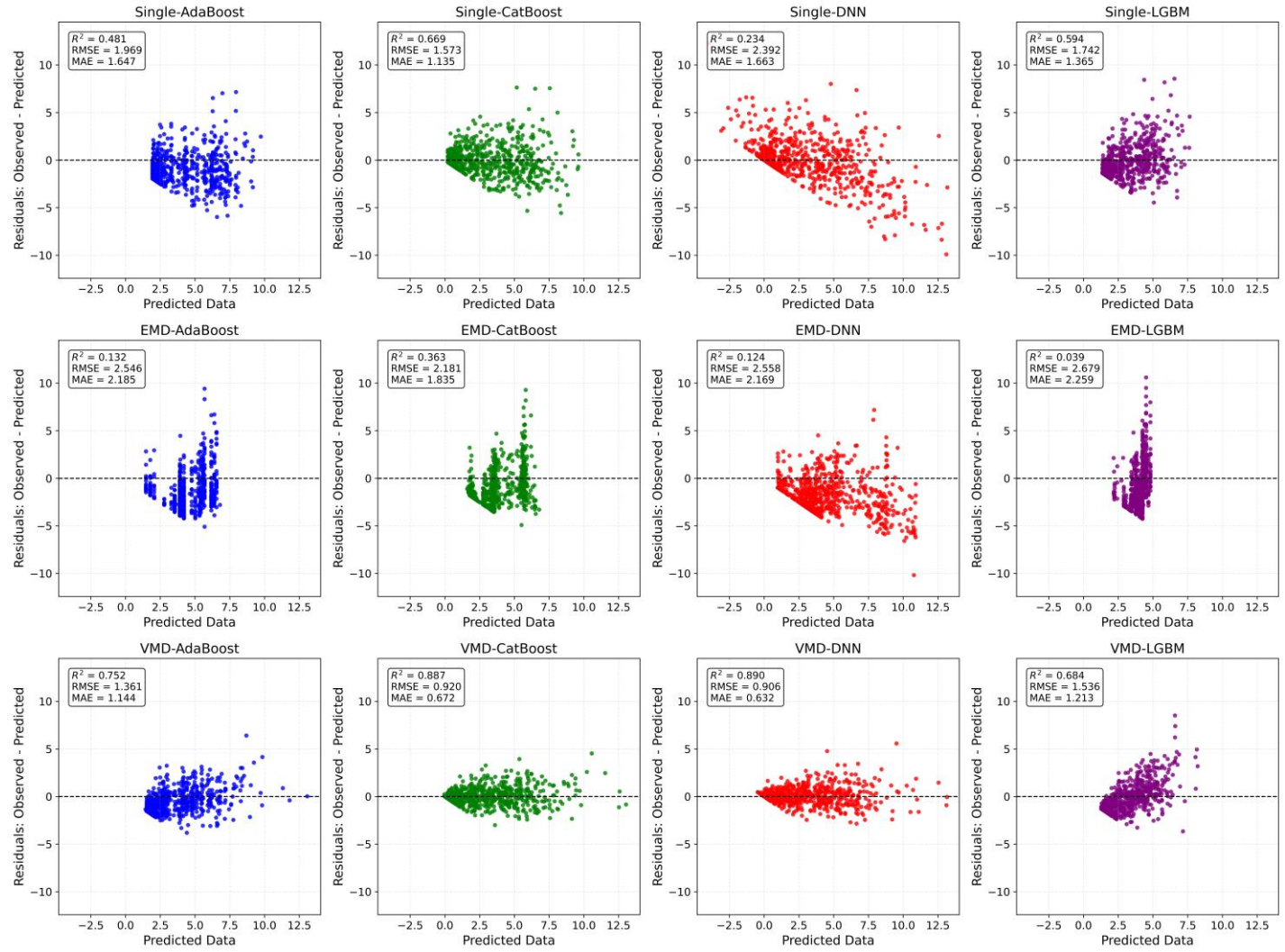


Figure 4. Residual plots of daily ET_0 for single and hybrid (EMD/VMD) models

The error density plot indicates that the error distribution of the CatBoost model and the LGBM model is sharper and symmetric around zero, indicating lower variance as well as non-systematic errors in test phase. In this plot, the horizontal axis represents error magnitude, while the density curve reflects the frequency distribution of prediction errors, allowing the identification of bias and dispersion characteristics. In the case of the distribution of CatBoost, it can be inferred from the curve of the distribution that the density is of a small range, and there are not many outliers. However, for the AdaBoost model, there is a higher error variance as shown by a wider as well as right-skewed distribution, where the model tends to deviate positively. The DNN model is rather susceptible to outliers of the errors, with both tails stretching out. The error series plot shows the error behavior of the models over time. Here, the temporal variation of errors is presented, enabling the assessment of model stability and the detection of time-dependent fluctuations or abrupt deviations. While the CatBoost and LGBM models exhibit more stable and narrow error values over time, the DNN model exhibits significant jumps and wide variations. AdaBoost, on the other hand, exhibits a high-frequency variability with increasing error levels in certain time periods. This shows that the generalization ability of the model decreases in some time periods and is weaker in terms of adaptive learning. With the integration of signal decomposition, especially in the VMD-based models, the error series of all four models exhibit visibly smoother and narrower amplitude behavior compared to their single and EMD-based counterparts. The VMD-CatBoost and VMD-LGBM models show this improvement the most. In these models, the error amplitudes stay close to zero across the whole time domain, which means that the models are better at generalizing over time and reducing noise propagation. The polar error plot is particularly important for assessing the impact of seasonal or cyclical patterns, as it shows how the errors are directionally distributed and in which angular segments they are concentrated. In this representation, the angular position corresponds to temporal progression (e.g., seasonal or periodic structure), while the radial distance represents the magnitude of prediction error. Therefore, a more compact and uniform circular pattern indicates both lower error magnitude and reduced directional bias. In this graph, the error curves of the CatBoost and LGBM models are smoother and form a narrow circle around the center, whereas the deviations in the DNN model are more irregular and amplitudinal with respect to the directions.

Similarly, AdaBoost is characterized by wider error values in certain angular segments. This suggests that the forecast errors are sensitive not only to time but also to directional (possibly seasonal or environmental) conditions, especially in the DNN and AdaBoost models. Notably, in the hybrid VMD-based models, polar error curves become even more contracted and directionally uniform, suggesting that these models not only reduce error magnitude but also correct for possible seasonal or cyclic biases more effectively. Compared to EMD-based versions, which still exhibit relatively scattered polar structures, the VMD-based hybrid models demonstrate stronger directional consistency.

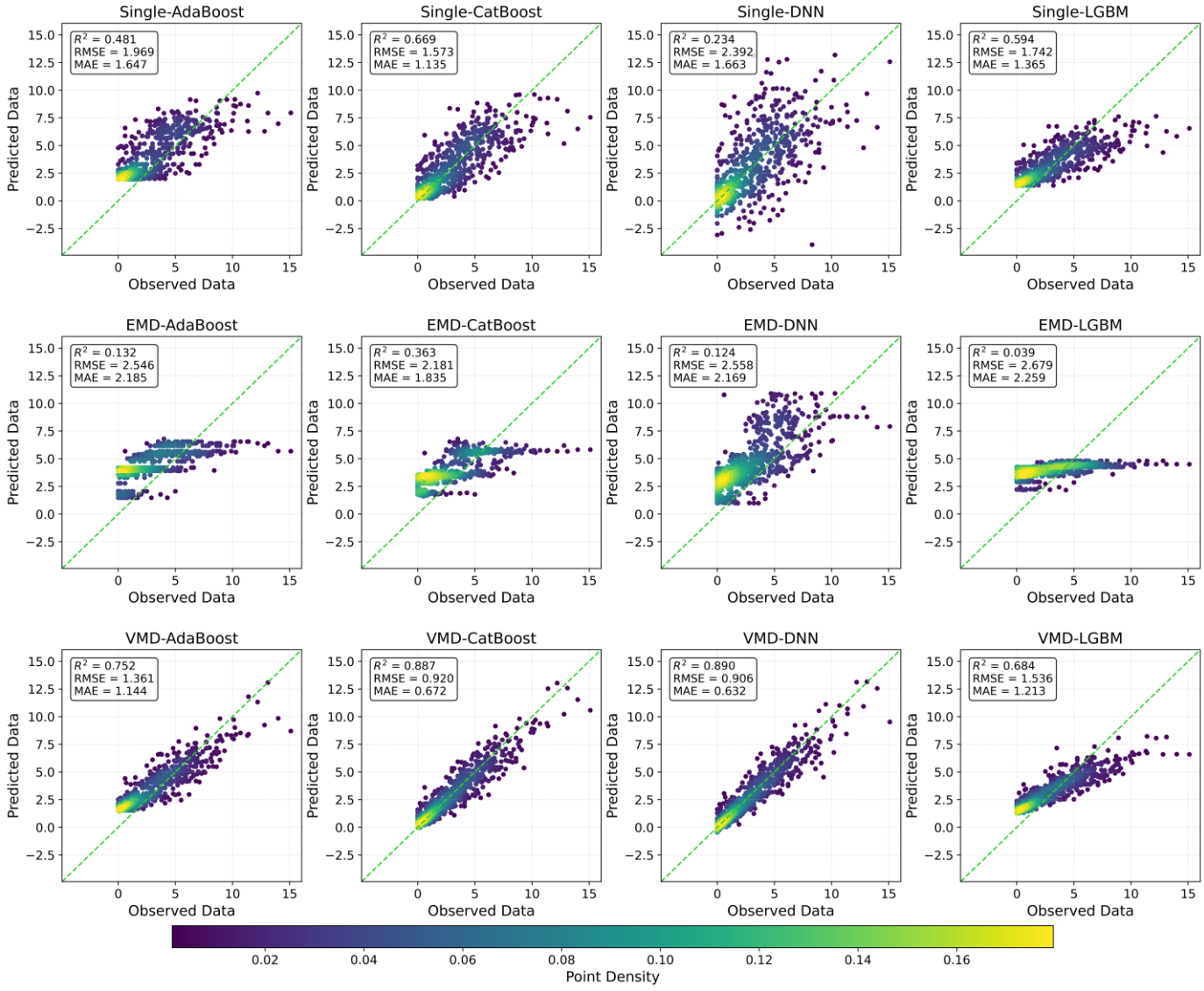


Figure 5. Density scatter plots (b) of daily ET_0 for single and hybrid (EMD/VMD) models

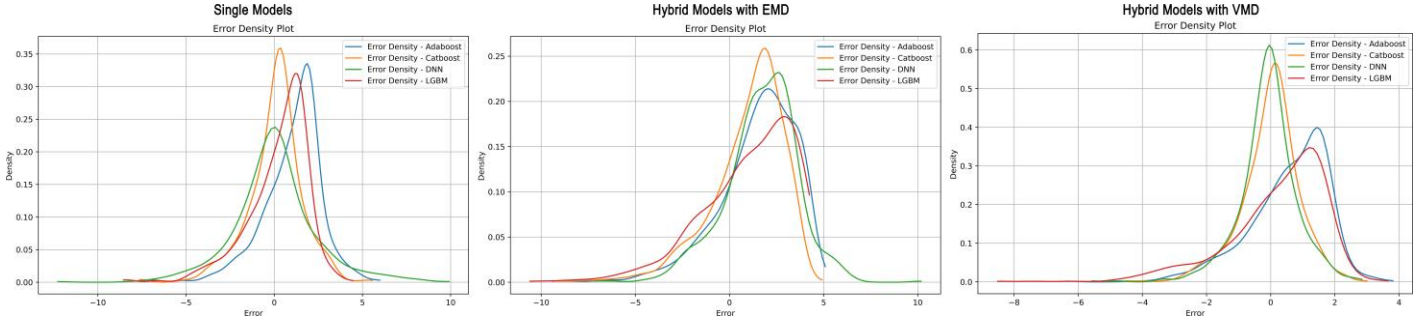
The cumulative distribution function (CDF) and prediction histograms created to evaluate the performance of the models visually reveal the extent to which the predicted values overlap with the observed data and which model produces results that are distributionally closer to the actual data structure. These comparative graphs allow to evaluate the distributional success of each model and are presented in Figure 7 and 8.

In the CDF plots of test phase, especially the cumulative curves of CatBoost and LGBM models, are very close to the distribution of observed values, indicating that these two models have high generalization success. Overlapping of the prediction and the observation curve, particularly around the median values, suggests that the above models are capable of reproducing the dominant trends in ET_0 prediction. In the DNN, the gross structure is reproduced properly, while the minimum values deviate from the observation. In AdaBoost, the prediction curve is always much farther from the observation curve, suggesting that there is a systematic bias for both low as well as high values of ET, and its generalization ability is limited. After VMD-based decomposition, the lines of prediction by CatBoost as well as that of LGBM almost coincide with the observed cumulative distributions, indicating significant distributional accuracy improvements. However, EMD-based hybrids do not

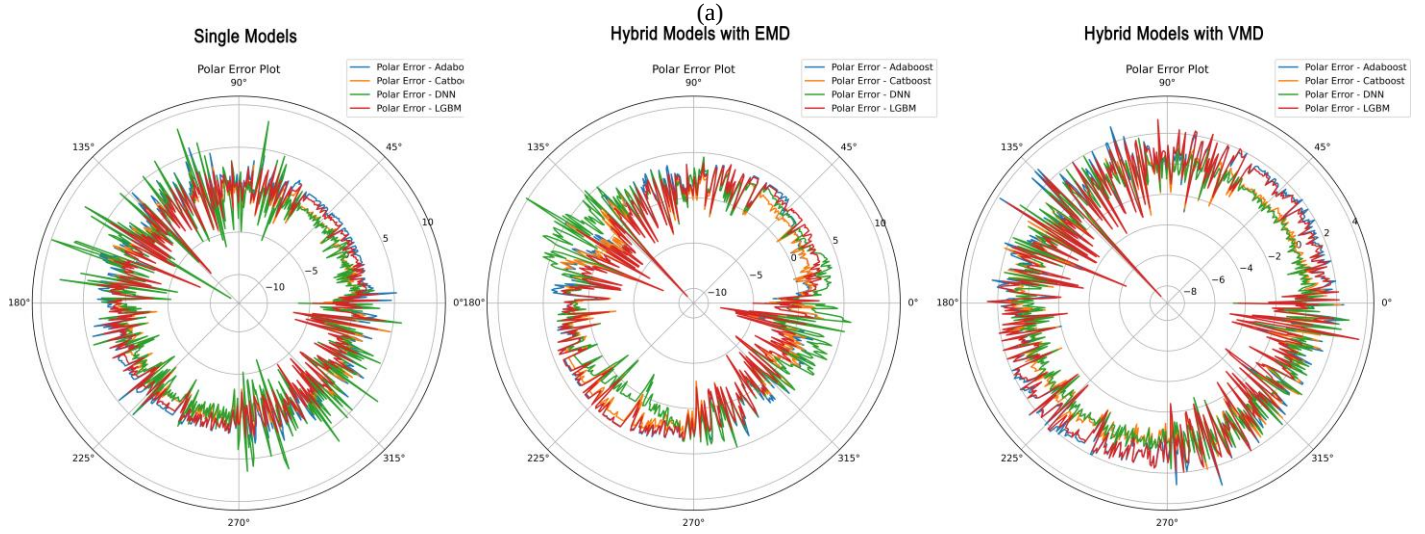
coincide with the actual distribution; even the tails, the lines of prediction deviate significantly from the observed values. This indicates that VMD helps the models generalize not only central tendencies but also extreme values better.

Table 2. Results for ET_0 estimation in the test phase (RMSE, MAE, and MBE are in mm/day)

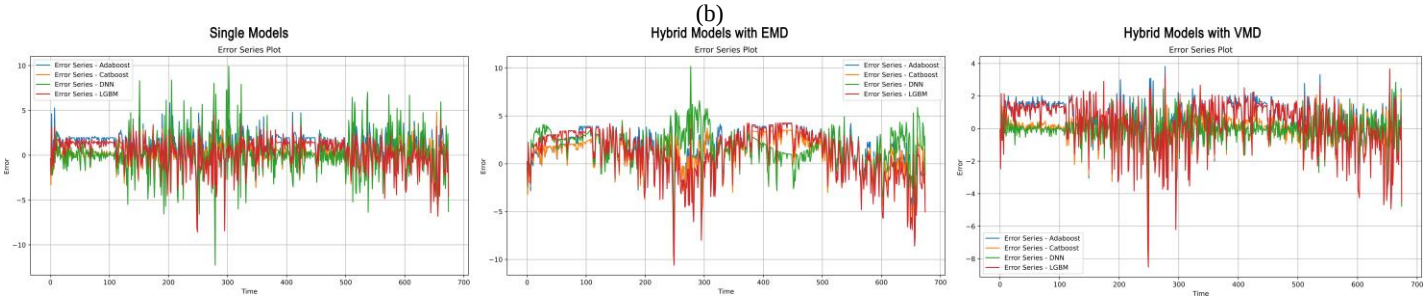
	Single models				Hybrid Models with EMD				Hybrid Models with VMD			
	AdaBoost	CatBoost	DNN	LGBM	AdaBoost	CatBoost	DNN	LGBM	AdaBoost	CatBoost	DNN	LGBM
RMSE	1.97	**1.57**	2.39	1.74	2.55	**2.18**	2.56	2.68	1.36	0.92	**0.91**	1.54
MAE	1.65	**1.13**	1.66	1.37	2.19	**1.83**	2.17	2.26	1.14	0.67	**0.63**	1.21
AIC	930.44	**627.15**	1193.40	764.97	1277.75	**1068.94**	1284.18	1346.42	432.33	-97.13	**-117.56**	595.69
NSE	0.48	**0.67**	0.23	0.59	0.13	**0.36**	0.12	0.04	0.75	**0.89**	**0.89**	0.68
KGE	0.52	**0.78**	0.64	0.52	0.21	0.34	**0.36**	0.01	0.64	0.89	**0.91**	0.53
R^2	0.63	**0.67**	0.45	**0.67**	0.47	**0.55**	0.54	0.30	0.84	**0.89**	0.89	0.82
MBE	1.06	0.06	**0.05**	0.24	1.50	**0.96**	1.70	1.01	0.58	**-0.02**	-0.08	0.22
BF	0.64	**0.98**	0.25	0.78	0.59	0.66	0.58	**0.71**	0.68	**0.87**	**0.87**	0.75
Pbias	36.01	2.10	**1.54**	8.08	51.03	**32.80**	57.85	34.46	19.79	**-0.59**	-2.81	7.61



(a)



(b)



(c)

Figure 6. Error density (a), polar error (b), and error time series (c) plots for single and hybrid (EMD/VMD) models

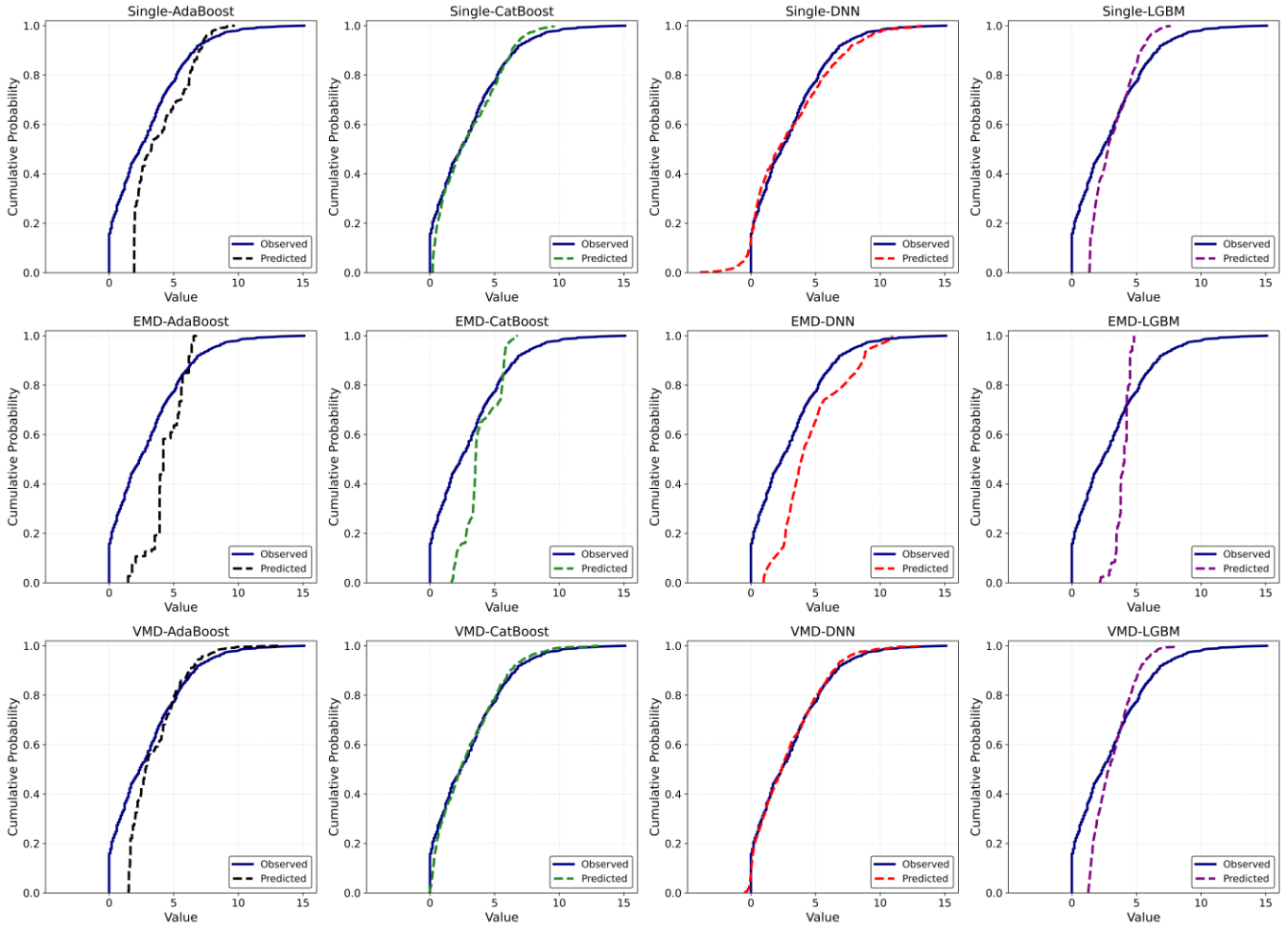


Figure 7. Cumulative distribution function (CDF) and comparative histogram of observed and predicted daily reference evapotranspiration values for each learning model

Model comparisons were evaluated on both error metrics and a correlation heatmap. The CatBoost model has the best prediction performance because it has the lowest RMSE (1.57 mm/day), MAE (1.13 mm/day), and AIC (627.15) scores and the highest NSE (0.67), KGE (0.78), and R^2 (0.67) scores. Moreover, the low MBE (0.06 mm/day) and PBIAS (2.10%) values indicate that the model is largely free from systematic biases. On the other hand, the DNN model fails to achieve sufficient generalization capacity when working with a limited amount of data, and this is frequently emphasized in the literature. Deep learning models need large datasets to achieve high performance; in this study, the relatively small amount of data caused DNN to perform lower than other machine learning based models. The AdaBoost model, on the other hand, gave moderate results, but stood out with its high PBIAS value (36.01%). With the introduction of hybrid approaches, a significant performance improvement is observed, particularly for CatBoost and LGBM. It yields the optimum performance on nearly all the metrics, i.e., the minimum RMSE (0.92 mm/day) and the minimum MAE (0.67 mm/day), as well as the maximum NSE (0.89) and KGE (0.89), which implies that the VMD decomposition significantly improves the performance of the model. Likewise, the VMD-LGBM also yields an optimum performance having high correlation along with minimum error metrics. These improvements are also verified from the correlation heatmaps. While the single-model correlations with actual data remain around 0.80, VMD-based models show correlation coefficients above 0.90 for all algorithms. Especially, the VMD-CatBoost and VMD-LGBM models show almost perfect correlation with the observed values, supporting their high predictive reliability. On the contrary, EMD-based models, although an improvement over single models, exhibit slightly lower correlation values, especially in LGBM. This pattern reaffirms that VMD decomposition provides a more informative and predictive input structure than EMD.

The feature importance graph for the CatBoost model is displayed in Figure 10a. This graph illustrates the significance of each input based on historical data used to predict ET_0 values for the next day, at time t+1. It highlights how much each feature contributes to the overall model. For this reason, the most important input is ET_0 (t), which is the ET_0 value right now. This feature is by far the most important variable, and it shows that the model uses this data the most when it tries to guess the ET_0 value for the next day. This indicates that there is a strong temporal autocorrelation (i.e., dependence on its own history) in the ET_0 values. The historical data at other time steps have relatively similar and lower levels of importance. This shows that although the model evaluates the ET_0 data of the past 8 days, these are not decisive on their own but

provide an information contribution collectively. In particular, the effect of some days, such as ET_0 (t-3) and ET_0 (t-4), is slightly higher than others, indicating the effect of short-term weather conditions and transitions. Following this initial analysis, signal decomposition methods such as Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) were applied to enhance the information structure of each lagged input. The correlation tables and feature importance graphs of the decomposed inputs (Figure 10b, 10c and Table 3 and 4) further reveal which intrinsic components significantly improve model performance.

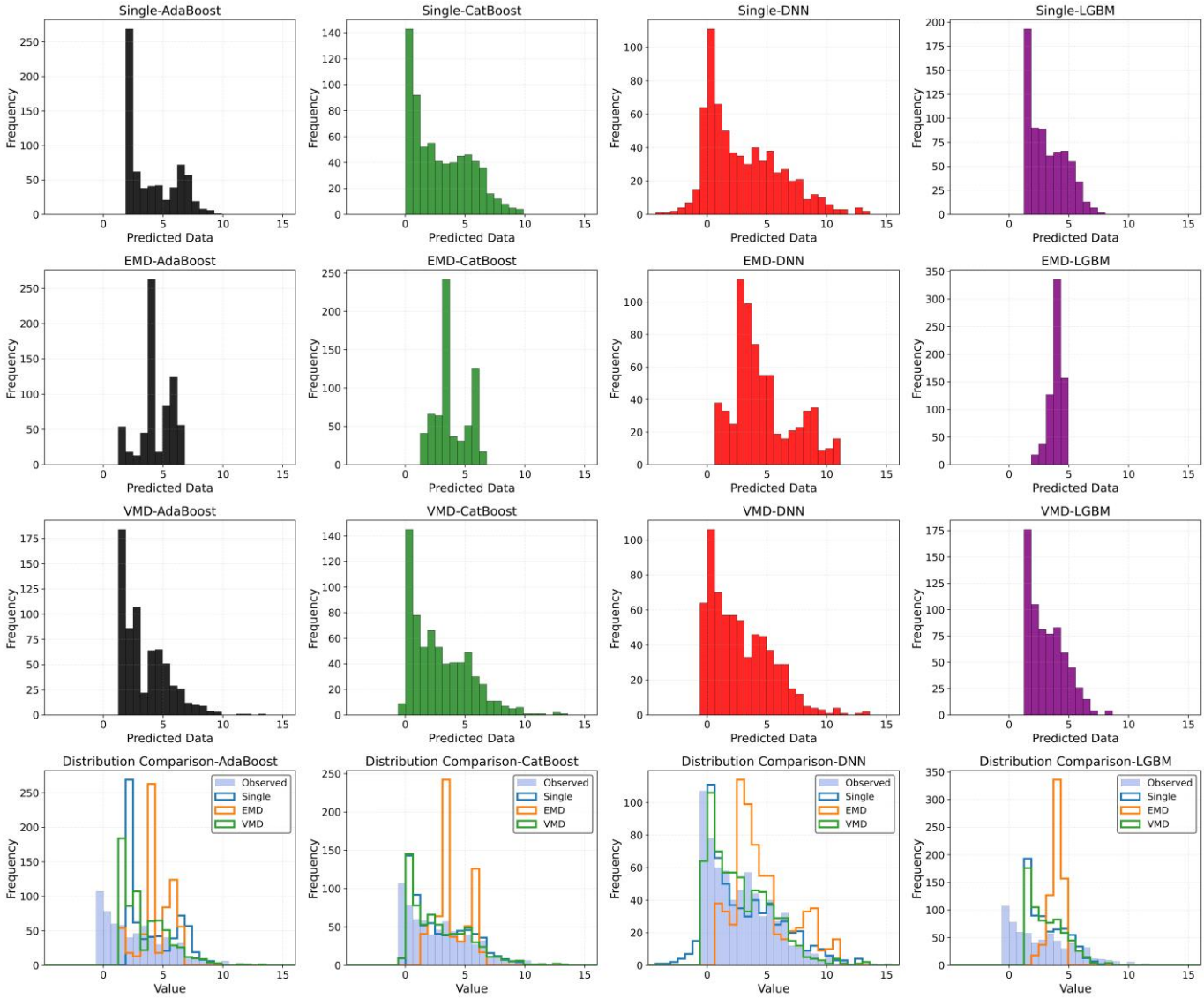


Figure 8. Cumulative distribution function (CDF) plots illustrating the distributional characteristics of observed and predicted daily reference evapotranspiration values for each learning model

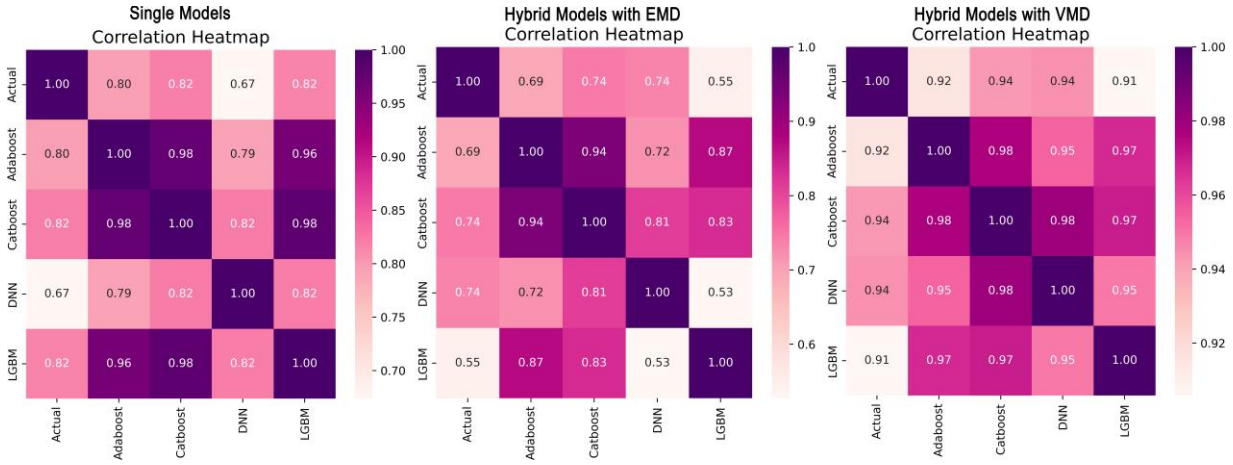


Figure 9. Correlation heatmap illustrating the relationships between observed daily reference evapotranspiration values and the predictions of each learning model.

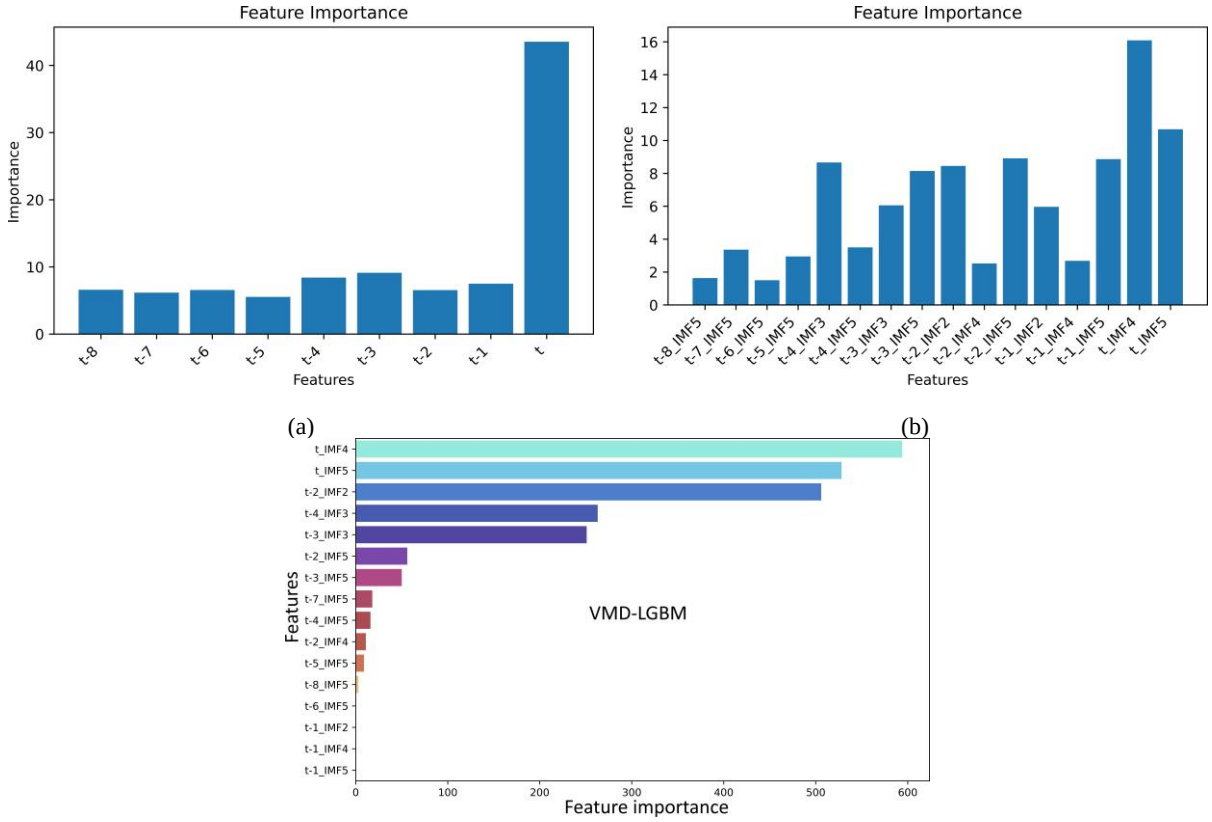


Figure 10. Feature importance analysis for ET_o (t+1) prediction: (a) Baseline CatBoost model, (b) EMD-based CatBoost model, and (c) VMD-based LGBM model

Table 3. Correlation coefficients between decomposed inputs via EMD and the output values

EVPT_t-8									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	IMF9	Residual
0.00	-0.05	-0.10	0.08	**0.73**	**0.33**	0.03	-0.01	-0.05	-0.02
EVPT_t-7									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
0.02	-0.06	-0.09	0.09	**0.73**	**0.36**	0.05	-0.01	-0.03	
EVPT_t-6									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	IMF9	Residual
0.01	-0.08	-0.08	0.11	**0.69**	**0.44**	0.12	0.03	-0.06	0.02
EVPT_t-5									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
0.02	-0.12	-0.05	0.14	**0.74**	**0.33**	0.10	0.02	-0.03	
EVPT_t-4									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
0.04	-0.15	0.00	0.14	**0.74**	**0.35**	0.10	0.02	-0.02	
EVPT_t-3									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
0.00	-0.15	0.06	0.16	**0.74**	**0.36**	0.06	0.00	-0.03	
EVPT_t-2									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
-0.06	-0.09	0.11	0.18	**0.74**	**0.36**	0.07	0.01	-0.03	
EVPT_t-1									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
-0.13	0.05	0.18	0.18	**0.74**	**0.36**	0.01	0.00	-0.03	
EVPT_t									
IMF1	IMF2	IMF3	IMF4	**IMF5**	**IMF6**	IMF7	IMF8	Residual	
0.02	0.22	0.22	0.19	**0.74**	**0.36**	0.01	0.00	-0.03	

Note: Values with a deviation of ±0.2 from the output are presented in the hybrid models. The selected components are written in bold.

Inclusion of Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) into the input framework significantly altered the learning patterns of the models together with their temporal sensitivities. The correlation analysis provided in Tables 3 and 4 shows that the IMF5 and IMF6 components derived through EMD regularly exhibit high and substantial correlations with the target ET_0 in nearly all lag periods. Likewise, in the framework of VMD, IMF5 stands significantly high on the correlation values across multiple time lags, which highlights the high predictive ability of the component. These components were selected as inputs for hybrid models based on correlation thresholds and included in subsequent training processes.

This EMD-CatBoost feature importance plot (Figure 10b) also confirms observations about correlations. IMF5 and IMF6 are the most important contributors across different time delays. This shows that the input features that come from signal decomposition play a big role in how well the model can predict. On the other hand, the VMD-LGBM feature importance plot (Figure 10c) shows that IMF4 and IMF5, mostly around the current timestep ET_0 (t), are the most important. This comment implies that VMD extracts the most significant structure related to the last temporal window, rendering the input features more concentrated as well as interpretable.

In general, the results show that decomposed components created by EMD and VMD contain more detailed and structured time information than raw lagged values. This makes the model more generalizable. IMF5 consistently shows high predictive value for both decomposition methods, which shows how important it is for ET_0 forecasting. These results also show that preprocessing based on decomposition has two benefits: it improves accuracy and makes it easier to choose the best features in data-driven modeling frameworks.

Table 4. Correlation coefficients between decomposed inputs via VMD and the output values

EVPT_t-8						**EVPT_t-7**					
IMF1	IMF2	IMF3	IMF4	**IMF5**	Residual	IMF1	IMF2	IMF3	IMF4	**IMF5**	Residual
0.03	-0.05	-0.08	-0.05	**0.82**	0.07	0.04	0.05	0.10	-0.17	**0.83**	-0.09
EVPT_t-6						**EVPT_t-5**					
IMF1	IMF2	IMF3	IMF4	**IMF5**	Residual	IMF1	IMF2	IMF3	IMF4	**IMF5**	Residual
-0.05	-0.09	0.15	-0.14	**0.83**	0.05	0.02	-0.10	0.08	-0.11	**0.84**	0.03
EVPT_t-4						**EVPT_t-3**					
IMF1	IMF2	**IMF3**	IMF4	**IMF5**	Residual	IMF1	IMF2	**IMF3**	IMF4	**IMF5**	Residual
-0.11	0.16	**-0.21**	0.04	**0.85**	0.03	0.10	-0.07	**-0.21**	0.14	**0.85**	-0.05
EVPT_t-2						**EVPT_t-1**					
IMF1	**IMF2**	IMF3	**IMF4**	**IMF5**	Residual	IMF1	**IMF2**	IMF3	**IMF4**	**IMF5**	Residual
-0.03	**-0.22**	-0.04	**0.23**	**0.86**	-0.05	-0.03	**-0.22**	-0.04	**0.23**	**0.86**	0.00
			EVPT_t								
			IMF1	IMF2	IMF3	**IMF4**	**IMF5**	Residual			
			-0.11	0.03	0.19	**0.30**	**0.86**	-0.09			

Note: Values with a deviation of ±0.2 from the output are presented in the hybrid models. The selected components are written in bold.

4. Discussion

This study evaluated the performance of models developed for daily ET_0 forecasting in Erzurum, a mountainous and snow-dominated region of Turkey. Most short-term hydrological or agricultural forecasting models require reliable ET_0 forecasts [59]. The stability of machine learning models is a critical factor to consider when predicting reliable ET_0 [60]. Machine learning approaches have provided flexible and data-driven solutions in modeling complex environmental processes, producing successful results in ETo prediction with different combinations of meteorological variables, and even achieving accuracy levels that can compete with traditional methods, even when using a limited number of inputs [61]. Previous studies have demonstrated that machine learning approaches such as ANN and ANFIS can successfully estimate ET using meteorological variables including temperature, wind speed, solar radiation, and relative humidity, with model performances evaluated based on RMSE and R^2 showing strong agreement with reference methods [62]. In addition, it has been reported that ANN models can outperform traditional empirical methods such as Hargreaves-Samani and Turc by achieving lower error values and better capturing the nonlinear dynamics of evapotranspiration processes [63]. In ET_0 forecasting, the incorporation of climatic variables as model inputs significantly influences forecast accuracy [60, 64]. This study, on the other hand, is about ET_0 forecasting in areas where there isn't much data or none at all. In this context, both numerical metrics and graphical analyses are used to evaluate the model performances; thus, not only the average error values of the models, but also the direction, stability, and distributional characteristics of the error structures are comprehensively examined. In addition to conventional lag-based inputs, the study incorporated signal decomposition techniques-namely, Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD)-to enhance temporal feature extraction. These methods allowed for the isolation of intrinsic mode functions (IMFs) from each time-lagged input, thereby enriching the input space with components showing stronger correlation to the output, as validated in Tables 3 and 4.

CatBoost showed the most balanced and successful results in terms of both graphical analysis and numerical performance metrics. In the residual plots, the residual values are symmetric around zero and concentrated in a narrow band, while in the density scatter plot, the predictions are highly concentrated around the 1:1 line. Polar plot analysis shows error orientations are consistent, but histogram and CDF plots show that the model largely retains the right-skewed character of the true data. These findings verify CatBoost attains a good representation for both error magnitude and distribution structural form. Numerical values further corroborate it, with the model recording lowest values of RMSE (1.57 mm/day) and MAE (1.13 mm/day) along with highest values of NSE (0.67) and KGE (0.78) scores, and being parsimonious with the lowest AIC value (627.15). G. Huang et al. [65] reported that CatBoost provides high prediction accuracy and computational efficiency, particularly when trained with complete meteorological inputs. Liang et al. [66] reported that machine learning models can achieve reliable performance under varying climatic and seasonal conditions, depending on input data availability. In that sense, these results are consistent with the robust performance of CatBoost as well as LightGBM models of our study as well as affirm that these models offer generalizable architectures under varying climatic settings. When combined with EMD, the CatBoost model continued to demonstrate strong results, although a slight increase in RMSE (2.18 mm/day) and MAE (1.83 mm/day) was observed. This performance degradation can be attributed to the nature of EMD decomposition, which may introduce mode mixing and generate components that do not sufficiently separate distinct frequency structures. As a result, some of the extracted IMFs may contain overlapping or redundant

information, reducing the effectiveness of feature representation for machine learning models. In contrast, VMD provides a more controlled decomposition by explicitly separating frequency bands, leading to more stable and informative components. This difference explains why EMD-based CatBoost exhibits reduced performance, whereas VMD-based CatBoost achieves significant improvement. More generally, VMD improves model performance by decomposing the original signal into band-limited intrinsic mode functions with reduced noise and minimal mode mixing, whereas EMD may produce overlapping components that introduce redundant or distorted information. This leads to a more stable feature space and enhances the learning efficiency of machine learning models. However, the IMF5 and IMF6 among all temporal lags continuously emerged as patterns of high importance (Figure 10b), showing that the decomposed signals had specific predictive patterns that supported short-term temporal learning. Crucially, the VMD-CatBoost model noticeably improved performance, achieving the minimum values for RMSE (0.92 mm/day), MAE (0.67 mm/day), and AIC (-97.13), while registering NSE and KGE values both as 0.89. This reveals that the VMD paradigm was most successful in uncovering significant patterns from the initial time series data. Overall, these findings indicate that CatBoost provides the most reliable and accurate predictions among all models when evaluated using key performance metrics such as RMSE, MAE, NSE, and KGE.

The LGBM model produced results close to CatBoost and exhibited a stable error pattern, especially in graphical evaluations. Symmetric distribution of errors in residual plots, high-density linear distribution in a density scatter plot, and directional stability in a polar plot support the reliability of the model. LGBM reveals patterns similar to the pattern of real data in the histogram and CDF plots, inferring that the predictions are not only point-value successful but distribution structure too. Numerically, even values such as RMSE (1.74 mm/day), MAE (1.37 mm/day), and NSE (0.59) fall behind CatBoost, but it is achieved that the model offers an evenly balanced prediction structure. Fan et al. [67] indicated that machine learning models such as tree-based approaches can produce satisfactory results under limited meteorological data conditions across different climatic regions. They also emphasized that the model stands out as an effective and generalizable alternative to the difficulties in data availability, especially in regional modeling. Similarly, Zhou et al. [68] compared advanced machine learning and boosting-based models for ET_0 prediction across different climatic regions, saying that both the models exhibit strong accuracy as well as decent stability throughout the tests. These results compare equally to the findings provided here. Although EMD-LGBM yielded slightly weaker results (RMSE = 2.68 mm/day), the VMD-enhanced LGBM model substantially outperformed its standard and EMD-based versions, with RMSE (1.54 mm/day), MAE (1.21 mm/day), and NSE (0.68), confirming that decomposition-driven hybridization effectively reduces model bias. Figure 10c further supports this, as IMF4 and IMF5 at time t were the dominant predictors, implying a strong localized temporal influence captured through VMD.

The DNN model performed poorly in graphical and numerical evaluations. Especially in the residual plots, the fact that the errors follow a certain linear pattern reveals that the model contains systematic bias. The directionally irregular and scattered error structure in polar plot analyses, and the concentration of predictions in the density scatter plot deviating from the actual values, indicate that the generalization capability of the model is limited. CDF and histogram plots show a wide error spread and a tendency to deviate from the true distribution. This is in line with the poor metric values such as RMSE (2.39 mm/day) and NSE (0.23). The substantial performance gap observed between EMD-DNN and VMD-DNN can be explained by the sensitivity of deep learning models to input feature quality and representation. EMD decomposition may introduce mode mixing and irregular signal components, which can distort the underlying temporal patterns and lead to unstable learning behavior in DNN models. In contrast, VMD provides a more structured decomposition by separating the signal into distinct frequency bands, resulting in smoother and more informative input features. This structured representation enhances the ability of the DNN model to capture nonlinear temporal relationships, leading to a significant improvement in predictive performance. Therefore, the dramatic increase in NSE from 0.12 to 0.89 reflects the critical role of decomposition quality in deep learning-based ET_0 estimation.

Deep learning (DL) estimation of daily ET_o exhibited lower performance compared to traditional machine learning approaches, primarily due to data limitations. This result is consistent with the findings of Ahmadi et al. [69], who systematically compared statistical and DL models such as LSTM and temporal fusion transformer (TFT) for ET_o forecasting and reported that DL models generally suffer from low data efficiency and require sufficiently long time series to achieve reliable performance. In contrast, simpler statistical models were found to perform comparably or even better under limited data conditions. In addition to model structure, the role of data preprocessing is also critical for improving model performance. Ahmadi et al. [51] demonstrated that decomposition-based approaches can significantly enhance prediction accuracy in hydrological applications. In their study, empirical mode decomposition (EMD) did not provide a substantial improvement, whereas variational mode decomposition (VMD) considerably increased model performance. This indicates that advanced decomposition techniques such as VMD can improve feature representation and support better learning, particularly for complex and data-hungry models.

On the other hand, the AdaBoost framework exhibited a predictive process defined by stable bias and high variance. The wide range of residual values in the residual plots and the significant deviation of the forecasts from the 1:1 line in the density scatter plot indicate that the model provides inconsistent results in terms of accuracy. Polar plot and error series plots also support this structure, showing that the model has an unstable error profile both directionally and temporally. In the histogram analysis, it is noteworthy that low ET_0 values are underrepresented, and the predictions are concentrated on medium and high values. In the numerical statistics, the RMSE (1.97 mm/day), MAE (1.65 mm/day), and particularly the high value of PBIAS (36.01%) show that the model generates systematically high estimates. Wu et al. [70] reported that boosting-based models may show reduced performance under limited input conditions, particularly when model complexity is not well matched to data availability and stressed that the performance of the model is poor, particularly with

fewer meteorological inputs. The results show that AdaBoost has low generalization ability under complicated and changing environmental conditions. EMD-based AdaBoost showed further degradation in performance (RMSE = 2.55 mm/day; PBIAS = 51.03%), confirming the method's vulnerability to overfitting or underfitting in decomposed data structures. In contrast, VMD-AdaBoost considerably improved results, particularly in reducing RMSE to 1.36 and PBIAS to 19.79%, though still lagging behind CatBoost and LGBM.

In general, the CatBoost and the LGBM models behaved better than the AdaBoost and DNN models according to both numerical indicators of success as well as the stability and distributional quality of the error patterns. On the other hand, the AdaBoost and DNN models exhibited shortcomings in most areas, showing that error metrics, error visualization, as well as distribution appropriateness, need to be accounted for during their choice. It supports the conclusion that signal decomposition, particularly VMD, notably optimizes the inputs to the models by isolating temporally stable, strongly correlated features. Optimized inputs result in improved performance, even among complex models or situations where there is limited data, the best-performing combined settings being the ones under the titles CatBoost-VMD as well as DNN-VMD.

5. Conclusions

Evapotranspiration is an essential hydrometeorological factor for water budgeting, irrigation planning, and climate planning. Erzurum, located in the Eastern Anatolia Region of Turkey, is characterized by a harsh continental climate, high elevation, mountainous terrain, and extended snow coverage. For a region where annual temperature variations are considerable, evapotranspiration processes are very important for crop production as well as water resources planning. This study comparatively evaluates the performance of four different machine learning models (AdaBoost, CatBoost, DNN, and LGBM) for daily reference evapotranspiration (ET_0) estimation. In the modeling process, the 8-day lag structure obtained by PACF and ACF analyses and the effective variables determined by the CatBoost feature importance method are used.

The multi-faceted graphical analyses obtained in the study clearly show that the CatBoost model produces more balanced, low-bias and stable forecasts compared to other models. In particular, the symmetric distribution of residual values around zero in the residual plots, the high density of forecasts around the 1:1 line in the density scatter plot, and the narrow and homogeneous distribution of the model's error structure in the CDF, error distribution and polar plot are important visual findings supporting CatBoost's forecast reliability. The results of the graphically run tests are supported by the low RMSE of the model (1.57 mm/day) and MAE (1.13 mm/day), the maximum NSE value of 0.67 and KGE value of 0.78, in addition to a low AIC value of 627.15, confirming that the model is highly accurate with low complexity. On the basis of the results presented above, we conclude that CatBoost is the best model for evapotranspiration estimation. Furthermore, hybrid modeling experiments using decomposed inputs from Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) demonstrated that VMD-CatBoost achieved the best overall performance, highlighting that decomposition-based feature engineering can further enhance model accuracy and stability under data-constrained conditions. This indicates that decomposition-driven hybrid models can be considered as powerful decision-support tools for ET_0 forecasting, particularly in mountainous and snow-dominated environments where meteorological data availability is limited.

This study focuses on ET_0 forecasting in regions where meteorological data are missing or unavailable. In this context, the CatBoost model stands out as an effective and reliable method for ET_0 estimation, based on the conditions of the study area. However, it should be noted that these findings are derived from a single meteorological station and may not be directly generalizable to other regions without further validation. The success of the model suggests that it has the potential to be used as a decision support system in applications such as hydrological modeling, climate-responsive irrigation management, and water resources planning, particularly in regions with similar climatic and topographic characteristics.

Acknowledgment

Special thanks to the General Directorate of Meteorology (MGM) for providing the database and data used in this study.

Conflict of Interest Statement

The authors declare no competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT Author Statement

Reşat Acar: Conceptualization, Data curation and analysis, Formal and statistical analysis, Visualization, Writing - review & editing. **Okan Mert Katipoğlu:** Conceptualization, Visualization, Data curation and analysis, Modelling and software section (programming and running), Writing - review & editing**. Burak Çırağ:** Investigation, Methodology, Visualization, Data curation and analysis, Conceptualization, Writing - original draft and Writing - review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this paper, the authors used GPT and QuillBot to improve language and readability. After using these tools, the authors thoroughly reviewed and edited the content as necessary.

References

1. Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., Kilic, A., Tu, K., Miralles, D. D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A. J., French, A., Schimel, D., ... Wood, E. F. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resources Research, 53(4), 2618–2626. https://doi.org/10.1002/2016WR020175
2. Volk, J. M., Huntington, J. L., Melton, F. S., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Ruhoff, A., Senay, G. B., Minor, B., Morton, C., Ott, T., Johnson, L., Comini de Andrade, B., Carrara, W., Doherty, C.

T., Dunkerly, C., Friedrichs, M., Guzman, A., ... Yang, Y. (2024). Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. Nature Water, 2(2), 193–205. https://doi.org/10.1038/s44221-023-00181-7

3. Raza, A., Hu, Y., Shoaib, M., Abd Elnabi, M. K., Zubair, M., Nauman, M., & Syed, N. R. (2021). A systematic review on estimation of reference evapotranspiration under PRISMA guidelines. Polish Journal of Environmental Studies, 30(6), 5413–5422. https://doi.org/10.15244/pjoes/136348
4. Khairan, H. E., Zubaidi, S. L., Muhsen, Y. R., & Al-Ansari, N. (2023). Parameter optimisation-based hybrid reference evapotranspiration prediction models: A systematic review of current implementations and future research directions. Atmosphere, 14(1), 77. https://doi.org/10.3390/atmos14010077
5. Oki, T., & Kanae, S. (2006). Global hydrological cycles and world water resources. Science, 313(5790), 1068–1072. https://doi.org/10.1126/science.1128845
6. Krishna, A. P. R. (2019). Evapotranspiration and agriculture—A review. Agricultural Reviews, 40(1), 1–11. https://doi.org/10.18805/ag.R-1848
7. Wanniarachchi, S., & Sarukkalige, R. (2022). A review on evapotranspiration estimation in agricultural water management: Past, present, and future. Hydrology, 9(7), 123. https://doi.org/10.3390/hydrology9070123
8. Dong, J., Zhu, Y., Jia, X., Shao, M., Han, X., Qiao, J., Bai, C., & Tang, X. (2022). Nation-scale reference evapotranspiration estimation by using deep learning and classical machine learning models in China. Journal of Hydrology, 604, 127207. https://doi.org/10.1016/j.jhydrol.2021.127207
9. Douna, V., Barraza, V., Grings, F., Huete, A., Restrepo-Coupe, N., & Beringer, J. (2021). Towards a remote sensing data based evapotranspiration estimation in Northern Australia using a simple random forest approach. Journal of Arid Environments, 191, 104513. https://doi.org/10.1016/j.jaridenv.2021.104513
10. Deger, İ. H. (2026). Flood frequency and temporal variability analysis of Yeşilırmak Basin under homogeneity and change point criteria. Engineering Perspective, 6(2), 185–209. https://doi.org/10.64808/engineeringperspective.1896337
11. Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Review, 100(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100%3C0081:OTAOSH%3E2.3.CO;2
12. Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193(1032), 120–145. https://doi.org/10.1098/rspa.1948.0037
13. Monteith, J. L. (1965). Evaporation and environment. In The state and movement of water in living organisms: Symposia of the Society for Experimental Biology (Vol. 19, pp. 205–234). Cambridge University Press.
14. Hu, X., Shi, L., Lin, G., & Lin, L. (2021). Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. Journal of Hydrology, 601, 126592. https://doi.org/10.1016/j.jhydrol.2021.126592
15. Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. Computers and Electronics in Agriculture, 148, 95–106. https://doi.org/10.1016/j.compag.2018.03.010
16. Rashid Niaghi, A., Hassanijalilian, O., & Shiri, J. (2021). Estimation of reference evapotranspiration using spatial and temporal machine learning approaches. Hydrology, 8(1), 25. https://doi.org/10.3390/hydrology8010025
17. Sattari, M. T., Apaydin, H., Band, S. S., Mosavi, A., & Prasad, R. (2021). Comparative analysis of kernel-based versus ANN and deep learning methods in monthly reference evapotranspiration estimation. Hydrology and Earth System Sciences, 25(2), 603–618. https://doi.org/10.5194/hess-25-603-2021
18. Amiri, M., Sharafi, S., & Ghaleni, M. M. (2025). Enhancing daily reference evapotranspiration (ETref) prediction across diverse climatic zones: A pattern mining approach with DIRECTORS model. Journal of Hydrology, 657, 133045. https://doi.org/10.1016/j.jhydrol.2025.133045
19. Elzain, H. E., Abdalla, O. A., Abdallah, M., Al-Maktoumi, A., Eltayeb, M., & Abba, S. I. (2024). Innovative approach for predicting daily reference evapotranspiration using improved shallow and deep learning models in a coastal region: A comparative study. Journal of Environmental Management, 354, 120246. https://doi.org/10.1016/j.jenvman.2024.120246
20. Fu, T., & Li, X. (2022). Hybrid the long short-term memory with whale optimization algorithm and variational mode decomposition for monthly evapotranspiration estimation. Scientific Reports, 12, 20717. https://doi.org/10.1038/s41598-022-25208-z
21. Çırağ, B., Acar, R., Katipoğlu, O. M., Yağanoğlu, M., & Şengül, S. (2025). Reconstructing missing streamflow data in mountainous, snow-dominated regions using optimized hybrid machine learning. Water Resources Management, 39(12), 6165–6187. https://doi.org/10.1007/s11269-025-04245-z
22. Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration: Guidelines for computing crop water requirements (FAO Irrigation and Drainage Paper No. 56). Food and Agriculture Organization of the United Nations.
23. Allen, R. G., Jensen, M. E., Wright, J. L., & Burman, R. D. (1989). Operational estimates of reference evapotranspiration. Agronomy Journal, 81(4), 650–662. https://doi.org/10.2134/agronj1989.00021962008100040019x
24. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154. https://doi.org/10.5555/3294996.3295074
25. Xu, K., Han, Z., Xu, H., & Bin, L. (2023). Rapid prediction model for urban floods based on a light gradient boosting machine approach and hydrological-hydraulic model. International Journal of Disaster Risk Science, 14(1), 79–97. https://doi.org/10.1007/s13753-023-00465-2
26. Chen, C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometrics and Intelligent Laboratory Systems, 191, 54–64. https://doi.org/10.1016/j.chemolab.2019.06.003
27. Lin, K., Sheng, S., Chen, H., Zhou, Y., Luo, Y., Xiong, L., & Xu, C.-Y. (2023). Exploring an intelligent adaptation method of hydrological model parameters for flood simulations based on the light gradient-boosting machine. Journal of Hydrology, 626, 130340. https://doi.org/10.1016/j.jhydrol.2023.130340

28. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31, 6639–6649. https://doi.org/10.5555/3327757.3327770
29. Zhang, K., Schölkopf, B., Muandet, K., & Wang, Z. (2013). Domain adaptation under target and conditional shift. In Proceedings of the 30th International Conference on Machine Learning (pp. III-819–III-827). PMLR. https://doi.org/10.5555/3042817.3043028
30. Friedman, J. H. (2002). Stochastic gradient boosting. Computational Statistics & Data Analysis, 38(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2
31. Zhang, Y., Zhao, Z., & Zheng, J. (2020). CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. Journal of Hydrology, 588, 125087. https://doi.org/10.1016/j.jhydrol.2020.125087
32. Vishwakarma, D. K., Kumar, P., Yadav, K. K., Ali, R., Markuna, S., Chauhan, S., Heddam, S., Kuriqi, A., Srivastava, A., Alam, M., & Vinayak, V. (2024). Evaluation of CatBoost method for predicting weekly pan evaporation in subtropical and sub-humid regions. Pure and Applied Geophysics, 181(2), 719–747. https://doi.org/10.1007/s00024-023-03426-4
33. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139. https://doi.org/10.1006/jcss.1997.1504
34. Busari, G. A., & Lim, D. H. (2021). Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. Computers and Chemical Engineering, 155, 107513. https://doi.org/10.1016/j.compchemeng.2021.107513
35. Yin, G., Alazzawi, F. J. I., Bokov, D., Marhoon, H. A., El-Shafay, A. S., Rahman, M. L., Su, C.-H., Lu, Y.-Z., & Nguyen, H. C. (2022). Multiple machine learning models for prediction of CO2 solubility in potassium and sodium based amino acid salt solutions. Arabian Journal of Chemistry, 15(3), 103608. https://doi.org/10.1016/j.arabjc.2021.103608
36. Çelik, A., & Kunt, M. A. (2026). A study on predicting engine performance outputs by machine learning algorithms in a single cylinder HCCI engine. Engineering Perspective, 6(1), 57–68. https://doi.org/10.64808/engineeringperspective.1782349
37. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
38. Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011
39. Kriegeskorte, N., & Golan, T. (2019). Neural network models and deep learning. Current Biology, 29(7), R231–R236. https://doi.org/10.1016/j.cub.2019.02.034
40. Elbeltagi, A., Srivastava, A., Li, P., Jiang, J., Jinsong, D., Rajput, J., Khadke, L., & Awad, A. (2023). Forecasting actual evapotranspiration without climate data based on stacked integration of DNN and meta-heuristic models across China from 1958 to 2021. Journal of Environmental Management, 345, 118697. https://doi.org/10.1016/j.jenvman.2023.118697
41. Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A review of deep learning models for time series prediction. IEEE Sensors Journal, 21(6), 7833–7848. https://doi.org/10.1109/JSEN.2019.2923982
42. Zhu, X., Li, J., Zhu, M., Jiang, Z., & Li, Y. (2018). An evaporation duct height prediction method based on deep learning. IEEE Geoscience and Remote Sensing Letters, 15(9), 1307–1311. https://doi.org/10.1109/LGRS.2018.2842235
43. Huang, X., Gao, L., Crosbie, R. S., Zhang, N., Fu, G., & Doble, R. (2019). Groundwater recharge prediction using linear regression, multi-layer perception network, and deep learning. Water, 11(9), 1879. https://doi.org/10.3390/w11091879
44. Sowmya, M. R., Kumar, M. S., & Ambat, S. K. (2020). Comparison of deep neural networks for reference evapotranspiration prediction using minimal meteorological data. In 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA) (pp. 27–33). IEEE. https://doi.org/10.1109/ACCTHPA49271.2020.9213201
45. Saroughi, M., Mirzania, E., Achite, M., Katipoğlu, O. M., Al-Ansari, N., Vishwakarma, D. K., Chung, I.-M., Alreshidi, M. A., & Yadav, K. K. (2024). Evaluate effect of 126 pre-processing methods on various artificial intelligence models accuracy versus normal mode to predict groundwater level (case study: Hamedan-Bahar Plain, Iran). Heliyon, 10(7), e29006. https://doi.org/10.1016/j.heliyon.2024.e29006
46. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 454(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193
47. Sankaran, A., Plocoste, T., Nourani, V., Vahab, S., & Salim, A. (2023). Assessment of multifractal fingerprints of reference evapotranspiration based on multivariate empirical mode decomposition. Atmosphere, 14(8), 1219. https://doi.org/10.3390/atmos14081219
48. Rao, A. R., & Hsu, E. C. (2008). Hilbert-Huang transform analysis of hydrological and environmental time series. Springer.
49. Johny, K., Pai, M. L., & Adarsh, S. (2019). Empirical forecasting and Indian Ocean Dipole teleconnections of south-west monsoon rainfall in Kerala. Meteorology and Atmospheric Physics, 131(4), 1055–1065. https://doi.org/10.1007/s00703-018-0620-7
50. Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. IEEE Transactions on Signal Processing, 62(3), 531–544. https://doi.org/10.1109/TSP.2013.2288675
51. Ahmadi, F., Tohidi, M., & Sadrianzade, M. (2023). Streamflow prediction using a hybrid methodology based on variational mode decomposition (VMD) and machine learning approaches. Applied Water Science, 13, 135. https://doi.org/10.1007/s13201-023-01943-0
52. Ali, M., Khan, A., & Rehman, N. U. (2018). Hybrid multiscale wind speed forecasting based on variational mode decomposition. International Transactions on Electrical Energy Systems, 28(1), e2466. https://doi.org/10.1002/etep.2466
53. Li, F., Ma, G., Chen, S., & Huang, W. (2021). An ensemble modeling approach to forecast daily reservoir inflow using bidirectional long- and short-term memory (Bi-LSTM), variational mode decomposition (VMD), and energy entropy method. Water Resources Management, 35(9), 2941–2963. https://doi.org/10.1007/s11269-021-02879-3
54. Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

55. Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
56. Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. Transactions of the ASABE, 58(6), 1763–1785. https://doi.org/10.13031/trans.58.10715
57. Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development, 15(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022
58. Katipoğlu, O. M., Yeşilyurt, S. N., Dalkılıç, H. Y., & Akar, F. (2023). Application of empirical mode decomposition, particle swarm optimization, and support vector machine methods to predict stream flows. Environmental Monitoring and Assessment, 195, 1108. https://doi.org/10.1007/s10661-023-11700-0
59. Perera, K. C., Western, A. W., Nawarathna, B., & George, B. (2014). Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. Agricultural and Forest Meteorology, 194, 50–63. https://doi.org/10.1016/j.agrformet.2014.03.014
60. Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., Xiang, Y., & Siddique, K. H. M. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agricultural and Forest Meteorology, 263, 225–241. https://doi.org/10.1016/j.agrformet.2018.08.019
61. Farag, A. A. (2025). Machine learning approaches for enhanced estimation of reference evapotranspiration (ETo): A comparative evaluation. Scientific Reports, 15, 38485. https://doi.org/10.1038/s41598-025-23166-w
62. Güzel, H., Üneş, F., Erginer, M., Kaya, Y. Z., Taşar, B., Erginer, İ., & Demirci, M. (2023). A comparative study on daily evapotranspiration estimation by using various artificial intelligence techniques and traditional regression calculations. Mathematical Biosciences and Engineering, 20(6), 11328–11352. https://doi.org/10.3934/mbe.2023502
63. Üneş, F., Doğan, S., Taşar, B., Kaya, Y., & Demirci, M. (2018). The evaluation and comparison of daily reference evapotranspiration with ANN and empirical methods. Natural and Engineering Sciences, 3(3), 54–64.
64. 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. Computers and Electronics in Agriculture, 173, 105430. https://doi.org/10.1016/j.compag.2020.105430
65. Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., & Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. Journal of Hydrology, 574, 1029–1041. https://doi.org/10.1016/j.jhydrol.2019.04.085
66. Liang, Y., Feng, D., Sun, Z., & Zhu, Y. (2023). Evaluation of empirical equations and machine learning models for daily reference evapotranspiration prediction using public weather forecasts. Water, 15(22), 3954. https://doi.org/10.3390/w15223954
67. Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., & Zeng, W. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agricultural Water Management, 225, 105758. https://doi.org/10.1016/j.agwat.2019.105758
68. Zhou, Z., Zhao, L., Lin, A., Qin, W., Lu, Y., Li, J., & He, L. (2020). Exploring the potential of deep factorization machine and various gradient boosting models in modeling daily reference evapotranspiration in China. Arabian Journal of Geosciences, 13, 1287. https://doi.org/10.1007/s12517-020-06293-8
69. Ahmadi, A., Daccache, A., Sadegh, M., & Snyder, R. L. (2023). Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. Computers and Electronics in Agriculture, 215, 108424. https://doi.org/10.1016/j.compag.2023.108424
70. Wu, T., Zhang, W., Jiao, X., Guo, W., & Hamoud, Y. A. (2020). Comparison of five boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. PLOS ONE, 15(6), e0235324. https://doi.org/10.1371/journal.pone.0235324