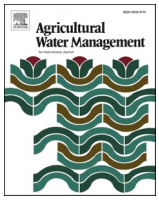
Agricultural Water Management 295 (2024) 108779

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat



SolarET: A generalizable machine learning approach to estimate reference evapotranspiration from solar radiation

Arman Ahmadi [a], Mohammad Hossein Kazemi [b], Andre Daccache [a,*], Richard L. Snyder [c]

[a] Department of Biological and Agricultural Engineering, University of California, Davis, CA 95616, USA
[b] Department of Civil Engineering, Sharif University of Technology, Tehran, Iran
[c] Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

ARTICLE INFO

Handling Editor: J.E. Fernández

Keywords:
Irrigation scheduling
Weather station
Reference grass surface
Pyranometer
CatBoost
California

ABSTRACT

Irrigation is the most significant consumer of freshwater worldwide. Deciding on the right amount of irrigation is crucial for sustainable water management and food production. The Penman-Monteith (P-M) reference crop evapotranspiration (ET_O) is the gold standard in irrigation management and scheduling; however, its calculation requires measurements from multiple sensors over an extended reference grass surface. The cost of land, sensors, maintenance, and water to keep the grass surface green impedes having a dense network of ET_O stations. To solve this challenge, this research aims to develop an input-limited ET_O estimation approach based on historical weather data and machine learning (ML) algorithms to relax the need for a reference grass surface. This approach, called "SolarET," takes solar radiation (R_S) data as its sole input. R_S is the only meteorological driving factor of ET_O that does not rely on the measuring surface. To test the generalizability of SolarET, we test its performance over unseen arbitrary locations across California. California is chosen as the case study since it is one of the world's most hydrologically altered and agriculturally productive regions. In total, 19,088,736 hourly data samples from 131 automated weather stations have been used in this study. The ML models have been trained over 114 stations and tested over 17 unseen stations, each representing a California climatic zone. Our findings point to the superiority of decision tree-based algorithms versus neural networks. SolarET works best in irrigation-oriented regions of California (e.g., Central Valley) and is less accurate in coastal and desert zones. Our results demonstrate the higher accuracy of SolarET using hourly (RMSE = 0.93 mm/day) and daily (RMSE = 0.97 mm/day) R_S data in comparison to well-known empirical alternatives like Priestley-Taylor (PT) (RMSE = 1.35 mm/day) and Hargreaves-Samani (HS) (RMSE = 2.69 mm/day).

1. Introduction

Irrigation is considered the most significant anthropogenic alternation to the natural hydrological cycle, accounting for about 70% of the global freshwater withdrawal (Siebert et al., 2010; Foley et al., 2011; Zhang et al., 2022). Population growth and economic development require increasing food production in the future, and expanding sustainable irrigation is essential to satisfy this burgeoning demand (Schmitt et al., 2022; Karimzadeh et al., 2024). Reference evapotranspiration (ET_O) is a decisive factor for water resources management in weekly and monthly resolutions and for irrigation scheduling in daily intervals. A common approach for irrigation management is to adjust the daily ET_O with crop- and growth stage-specific coefficients to determine the potential crop evapotranspiration (ET_C) (Ji et al., 2017; Fernández, 2023). Under the hypothesis that the crop is fully irrigated and without any growth-limiting factors (biotic and abiotic factors), the volume of water lost by the crop should be topped up with irrigation water (Haghverdi et al., 2021). Anything more than this "correct amount" (non-consumptive water use) is considered unproductive as it will be lost by evaporation, runoff, or deep percolation, while anything less than this amount will develop water stress and potentially yield loss. Therefore, a realistic estimate of daily ET_O is essential for sustainable water management and food production.

The FAO 56 Penman-Monteith (P-M) equation is the gold standard of ET_O calculation (Penman, 1948; Monteith, 1965; Allen et al., 1998). Compared to other empirical or semi-empirical equations, P-M is a physical-based equation that combines energy balance and mass transfer method and considers aerodynamic and surface resistance factors to

* Corresponding author.
E-mail address: adaccache@ucdavis.edu (A. Daccache).

https://doi.org/10.1016/j.agwat.2024.108779
Received 13 January 2024; Received in revised form 11 March 2024; Accepted 12 March 2024
Available online 15 March 2024

calculate the evapotranspiration rate over a standardized cropped surface (Allen et al., 1998). Therefore, P-M requires a complete weather station (several sensors for measuring multiple parameters) to be installed over an extended reference grass surface to achieve reliable results. This surface should be homogeneously extended towards different directions (enough fetch) to satisfy the heat and vapor advection assumptions. Moreover, the grass should be well-watered to avoid any water stress. Meeting these conditions is troubling, especially where labor for maintenance is scarce or expensive, and land and water resources are limited, which is very common in agriculturally productive regions. Utilizing fertile land and scarce water and labor resources for weather stations is not appealing to growers and stakeholders. The other problem is the number of stations required for irrigation management. Due to the diversity of micro-climates in croplands, only a dense network of stations is practically helpful for irrigation management.

To address the abovementioned challenges, this research aims to develop an ET_O estimation approach that relaxes the need for a reference surface. ET_O has four meteorological driving factors: solar radiation, air temperature, air humidity, and wind speed (Ahmadi et al., 2022). From these factors, solar radiation (R_s) is the only one that does not rely on surface characteristics, as it is a function of incoming solar energy only. Therefore, our approach called "SolarET" uses R_S, which is measured with only one sensor (i.e., pyranometer) placed on any arbitrary surface as the sole input parameter.

There are numerous input-limited alternatives for ET_O estimation in the literature, ranging from empirical models (e.g., Hargreaves and Samani, 1985) to machine learning (ML) applications (e.g., Chen et al., 2020). Almost all well-known empirical models were developed decades ago over restricted training data samples, sampling locations, and computational resources (Priestley and Taylor, 1972; Hargreaves and Samani, 1985). These methods tend to aggregate all un-measured inputs and uncertainties deterministically into one empirical coefficient, which might result in inaccurate results. Also, none of these methods rely only on solar radiation data. On the contrary, these methods need information on air temperature and humidity. Since this information is a function of the surface, the empirical equations cannot relax the need for the reference grass surface. Similarly, ML alternatives available in the literature take air temperature and humidity as their inputs and are not primarily focused on bypassing the reference surface (Chen et al., 2020; Dong et al., 2022; Kushwaha et al., 2022).

Our proposed approach uses decision tree- and neural network-based ML regression models to estimate daily ET_O with hourly and daily R_S data. The ML models employed in this study have shown superior performance for similar regression problems based on the available literature and ML competitions (Rodriguez-Galiano et al., 2015; Zhang et al., 2020; Hancock and Khoshgoftaar, 2020; Zhangzhong et al., 2023). Contrary to the available ML methods for daily ET_O forecasting, which use only daily meteorological inputs, this research leverages hourly R_S inputs to take advantage of the information gained from this finer temporal resolution data. Another essential difference between SolarET and available ML-based alternatives for ET_O estimation lies in its generalizability. Our test set consists of unseen weather stations to ensure SolarET's generalizability and applicability to new arbitrary locations in California. In other words, we train our models on data collected from some stations/locations and then test them using a previously unseen dataset from new stations.

We use California as the case study to evaluate the applicability of SolarET. California is one of the most hydrologically altered regions in the world (Zimmerman et al., 2018). California is also one of the most agriculturally productive regions on the planet, with heavy reliance on irrigation (Lobell and Bonfils, 2008; Tindula et al., 2013). Another reason California is a suitable case study for this research is the reliable, long-term data available from several weather stations in California (Ahmadi et al., 2023).

To examine the performance of SolarET, we test it against Hargreaves-Samani (HS), Priestley-Taylor (PT), and Romanenko empirical models. HS and PT models are developed using California data and are widely used and highly trusted for ET_O estimation (Tabari et al., 2013). Leveraging high temporal resolution data, cutting-edge regression algorithms, and big training data availability, we hypothesize that SolarET can outperform empirical alternatives without requiring surface-dependent inputs. In other words, we hypothesize that SolarET can surpass comparable methods in terms of prediction accuracy, cost-effectiveness, and generalizability.

2. Study area and dataset

California was chosen as the case study of this research. It is one of the world's most prominent agriculturally productive regions, heavily dependent on irrigation. According to the California Department of Water Resources (DWR), in an average year, approximately 9.6 million acres are irrigated with roughly 42 billion cubic meters of water. To assist irrigators in managing this immense water demand, DWR and the University of California, Davis (UC Davis) developed a program called the *California Irrigation Management Information System* (CIMIS) in 1982. CIMIS consists of over 145 automated weather stations that measure meteorological inputs of the Penman-Monteith equation in a standardized condition. Daily and hourly reports of ET_O and its meteorological driving factors are publicly available on the CIMIS website (https://cimis.water.ca.gov/). More information about the CIMIS program can be found on the CIMIS website.

For this study, daily and hourly data on required variables from 131 active CIMIS stations are acquired. While some of those stations are located in research facilities, most are installed on private lands (often growers) where proper maintenance (mainly grass and fetch requirements) is a limiting factor. For this work, stations are chosen according to their condition and maintenance records, and problematic stations (e.g., stations with poor maintenance, inadequate or infrequent irrigation, and non-grass reference surface) are eliminated to ensure the quality of the input data. The train set consists of 114 stations. The train set data is acquired from January 1, 1995, to December 31, 2022 (i.e., 28 years of data). Although all the stations in the train set have data until the end of 2022 (i.e., they are active stations), their installation date varies, and some of them have been installed after 1995 and, therefore, have less than 28 years of data.

CIMIS divides California into 18 homogeneous climatic zones with similar meteorological and evapotranspiration characteristics. To test the applicability of the proposed methodology in estimating ET_O at climatically diverse and unseen arbitrary locations, 17 stations, each representing a climatic zone, were chosen as the test set. Zone 11 is eliminated, as it does not have any reliable station. Names and characteristics of these climatic zones can be found in Table A1 in the appendix. Readers are referred to Ahmadi et al. (2022) for more information about these zones.

To evaluate the generalizability of the proposed ML models, they have never been exposed to the test set, even for validation purposes. All the results reported in this paper represent the models' performance on this unseen test set. To eliminate biases due to data length and climatic variability in longer time spans (e.g., wet or dry periods), the test set consists of data from January 1, 2018, to December 31, 2022 (five years) for all stations. In total, 19,088,736 hourly data records have been used in this study, with 18,447,984 records used as the training set and the remaining 640,752 records employed as the test set. Fig. 1 shows the distribution of training and testing stations across California. The zoning of Fig. 1 refers to the ET_O homogeneous zones.

3. Methodology

3.1. Reference evapotranspiration

The Penman-Monteith (P-M) equation is the gold standard for calculating ET_O (Penman, 1948; Monteith, 1965). CIMIS calculates ET_O

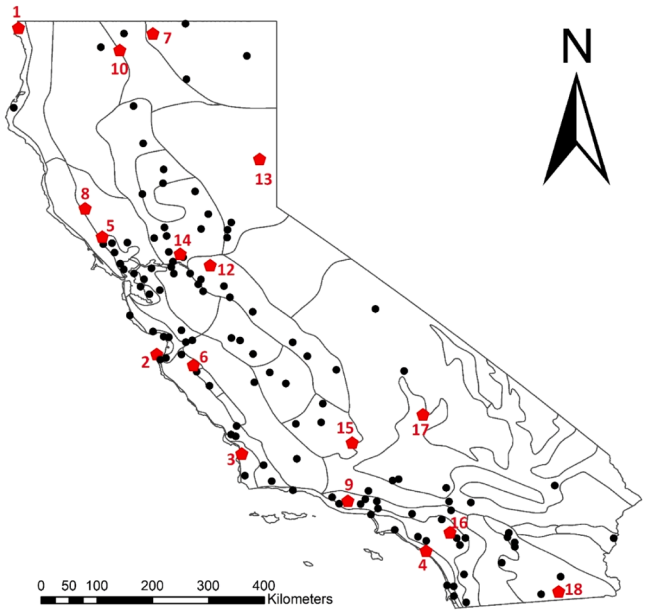


Fig. 1. Location map of CIMIS stations used in this study. Black circles represent the training stations, and red pentagons depict the stations used as the test set. The values adjacent to the test stations are the number of the climatic zone they represent.

using the P-M equation version described in the American Society of Civil Engineering-Environmental Water Resources Institute paper (Allen et al., 2005). CIMIS does not use the daily meteorological variables and daily P-M equation to calculate daily ET_O. It reports the summation of 24 hourly ET_O values from midnight to midnight as the daily ET_O. Ahmadi et al. (2022) showed that the summation of hourly ET_O values results in similar outputs as using the daily P-M equation. CIMIS uses Eq. 1 to calculate hourly ET_O (Allen et al., 1998, 2005; Walter et al., 2000; Pereira et al., 2015).

$$ET_O = \frac{\Delta(R_n - G)}{\lambda[\Delta + \gamma(1 + C_d u_2)]} + \frac{\gamma \frac{37}{T_a+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + C_d u_2)} \quad (1)$$

Where C_d (the ratio of bulk surface to aerodynamic resistance) is equal to 0.24 and 0.96 during daytime and nighttime, respectively.

ET_O: standardized reference evapotranspiration (mm h^{-1}), which approximates grass ET

Δ: the slope of saturation vapor pressure curve (kPa $°C^{-1}$) at mean air temperature (T)

R_n: net radiation (MJ m^{-2} h^{-1}) estimated from solar radiation and other input variables

G: soil heat flux density (MJ m^{-2} h^{-1}): $G=0.1 \cdot R_n$ daytime and $G=0.5 \cdot R_n$ nighttime

γ: psychrometric constant (kPa°C^{-1})

T_a: mean hourly air temperature (°C)

u_2: mean hourly wind speed at 2 m height (m s^{-1})

e_s: saturation vapor pressure (kPa) at T_a

e_a: mean actual vapor pressure (kPa) from dew point temperature (°C), or relative humidity (%) and temperature (°C)

λ: latent heat of vaporization (2.45 MJ kg^{-1} for water temperature of 20 °C)

This equation's first and second fractions are the radiation and aerodynamic terms of the P-M equation. This research uses the daily ET_O value reported by CIMIS (i.e., the summation of hourly values) as the target value or predictand of ML models.

3.2. Data preprocessing and feature selection

To relax the need of reference surface to estimate ET_O, this research uses only solar radiation data measured by a pyranometer as the input of ML models. This study uses hourly R_S as input to leverage more detailed measurements and capture the variations in solar radiation during the day. Therefore, each sample consists of an input vector of size 24 and a target value, which is the daily ET_O.

To ensure the reliability of input data, all samples with zero R_S values for daytime (i.e., from 8 AM to 5 PM) have been removed from the dataset. Therefore, of the 795,364 daily samples used in this study, 1224 samples were eliminated. As only 69 of these eliminated samples are from the test set, we hypothesized that this process would not bias the distribution of samples in the test set and the study results.

To ensure the robustness of ML models and improve their accuracy, three data preprocessing steps were taken: 1) eliminating records with null values: Null values are removed from the dataset to ensure accuracy, eliminate inconsistencies, and transform the data into a suitable format for inputting into the models. This filtering removed 6752 records from the training set of ML models and 4830 records from the test set. 2) Eliminating records with ET_O values higher than 15 mm/day: we hypothesize these values are due to sensor errors or poor maintenance conditions. This filtering resulted in the elimination of 10 records from the training set. 3) Using robust scaler: While traditional scaling techniques such as min-max scaling or standard scaling rely on the mean and standard deviation of the data, robust scaling uses different statistics that are less sensitive to outliers. Specifically, it leverages the median and interquartile range (IQR) to scale the input data. By reducing the potential impact of extreme values on the performance of ML models, robust scaling helps to mitigate the influence of outliers in the input variables. Eq. 2 shows how scaled data records are calculated using the robust scaling method:

$$x_{scaled} = \frac{x_i - x_{median}}{Q_3 - Q_1} \quad (2)$$

Where, as the names suggest, x_i is the value before scaling, and x_{median} is the median of the data distribution; x_{scaled} is the value after scaling. Q_3 and Q_1 are the 75th quantile and 25th quantile of the data distribution, respectively.

Two strategies are employed for selecting the inputs of ML models. In the first strategy, all zero R_S values (i.e., nighttime) have been substituted with the daily mean R_S value. In this strategy, which is called "mean substitution," each input record for the ML models has 24 features. To implicitly capture the daylight period, the mean daily R_S is calculated as the sum of all hourly R_S values divided by the number of non-zero R_S values. This procedure differs from calculating daily R_S, where 24 is the denominator. As the second strategy, we used feature selection methods to assign quantitative feature importance measures to each hour (i.e., each feature). In this strategy, only the most essential features are used as the input to train the models. To analyze the model performance under different feature selection strategies, we used correlation-based (Pearson and Spearman), information-based (Mutual Information), and intrinsic (CatBoost) feature selection strategies.

To compare the use of hourly R_S values against coarser temporal resolution inputs, we also tested using daily R_S as input. In the case of daily R_s, we used a sine function of Julian date as the second feature to capture the seasonality. Using a sine function of the Julian date has merits over using the raw Julian date, as it can simulate the cyclic nature of input records (Eq. 3).

$$J_C = \sin(\frac{2\pi}{365} J) \quad (3)$$

Where J is the Julian date and J_C is the cyclic Julian date value used as the second feature along with daily R_S for the daily ML models.

3.2.1. Pearson correlation

The Pearson correlation coefficient or Pearson's "*r*" measures the linear correlation between two data sets or two random variables. It is a normalized measure of covariance, and therefore, its value is always between −1 and 1, while 1 shows the perfect positive linear correlation and −1 indicates the perfect negative linear correlation. When there is no linear correlation between the two variables r equals zero. The Pearson correlation of two random variables X and Y ($\rho_{X,Y}$) is defined as:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{4}$$

Where $\mathrm{cov}(X,Y)$ is the covariance of two variables, and σ_X and σ_Y are the standard deviations of X and Y, respectively. We used the *Pandas* library in Python to calculate the Pearson correlation.

3.2.2. Spearman correlation

Spearman's rank correlation (r_s) coefficient is a nonparametric (distribution-free) rank statistic that measures the strength of correlation between the ranking of two variables (Hauke and Kossowski, 2011). Although the Spearman correlation between two variables X and Y is equal to the Pearson correlation between the rank values of those two variables, i.e., R(X) and R(Y), the Spearman correlation is not a measure of the linear relationship. Spearman's coefficient assesses how well an arbitrary monotonic function (linear or not) can describe the relationship between two variables. Eq. 5 shows how r_s is calculated. Like the Pearson correlation, r_s values are between −1 and 1. We used the *Pandas* library in Python to calculate the Spearman correlation.

$$r_s = \rho_{R(X),R(Y)} = \frac{\mathrm{cov}(R(X),R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{5}$$

3.2.3. Mutual information

Mutual information (MI) measures how much information about a random variable we can acquire by knowing another random variable. MI is a non-negative dimensionless quantity with units of information (bits) that quantifies the dependency between two variables. MI measures the amount of information gain or uncertainty reduction about one variable given knowledge of another. MI is a non-negative measure between 0 and 1. It is equal to zero if and only if two random variables are strictly independent, and higher values mean higher dependency (Kraskov et al., 2004).

In contrast to the Pearson correlation coefficient or other linear correlation measures, MI is also sensitive to dependencies that do not demonstrate themselves in covariance; therefore, it is more generalizable and reliable for feature importance evaluation, especially in nonlinear contexts like evapotranspiration equations. Detailed information about MI can be found in the references (Kozachenko et al., 1987; Kraskov et al., 2004; Ross, 2014). We used the *Scikit-learn* software library in Python to calculate MI.

3.2.4. CatBoost feature importance

The CatBoost regression model performs automatic feature importance analysis on the input data during training. After training the model, the relative importance of each input feature can be acquired. This measure shows how much, on average, the model prediction changes if the feature value changes. Higher feature importance values indicate more significant changes to the prediction if the feature value changes. CatBoost feature importance is a non-negative value. Feature importance values are normalized so that the sum of the importance values of all features is equal to 100. This normalization enables a relative assessment of the importance values across various features.

For feature importance analysis, we chose the hyperparameters of the CatBoost model according to outperforming CatBoost models trained on similar datasets on Kaggle, with no further hyperparameter tuning. 70% of the data was used to train the model, and 30% as the test set for model evaluation. More information about the CatBoost model is provided in the following sections. We used the *catboost* library in Python to calculate the CatBoost feature importance.

3.3. Machine learning

Five ML algorithms were chosen to test if only R_S data can predict ET_O over California. We selected highly used and well-known regression algorithms. These algorithms have demonstrated superior performance in literature and ML competitions with datasets similar to the one used in our study (Zhang et al., 2020). We also used an automatic hyperparameter tuning framework to optimize the hyperparameters of the CatBoost algorithm. We used this automatic hyperparameter optimization only for CatBoost since CatBoost was the best model in our pilot model training, utilizing only a portion of input data. Also, according to the literature and ML competition results, CatBoost has demonstrated superiority in similar regression problems. Specifically, Zhang et al. (2020) showed that the CatBoost model is the most accurate compared to similar ML models for daily ET_O estimation.

3.3.1. CatBoost

CatBoost is an open-source gradient-boosted decision tree with widespread applications for classification and regression tasks to many data types (Prokhorenkova et al., 2018; Dorogush et al., 2018). CatBoost aims to mitigate the challenges of overfitting and target leakage through its innovative techniques (Asghari et al., 2023). The algorithm introduces two main innovations compared to other gradient-boosted decision trees: *ordered target statistics* and *ordered boosting* (Hancock and Khoshgoftaar, 2020). The term "ordered target statistics" refers to the technique CatBoost uses to encode categorical variables. "Ordered boosting" is a refinement of gradient boosting (Prokhorenkova et al., 2018). Through this technique, a decision tree model is recursively trained on the prediction residuals of data points. This approach allows the model to obtain unshifted residuals by applying the current model to new training examples at each step, leading to reduced overfitting.

The main difference between CatBoost and other gradient boosting algorithms like LightGBM and XGBoost is that CatBoost uses symmetric or balanced trees. In other words, in CatBoost, the splitting condition is consistent across all nodes at the same tree depth. Symmetric trees are faster to train and less prone to overfitting. Nevertheless, symmetric trees are weaker learners than asymmetric trees, and therefore, they generally make worse predictions. However, as the main idea of gradient boosting is to combine numerous weak learners to make predictions, the CatBoost algorithm tends to make more accurate predictions than LightGBM and XGBoost in several cases. Detailed information about the CatBoost algorithm can be found in Prokhorenkova et al. (2018). We used the *CatBoost* package in Python to develop the CatBoost model.

We employed Optuna, a hyperparameter optimization framework for automatic hyperparameter tuning of the CatBoost model (Akiba et al., 2019). Optuna allows users to dynamically construct the hyperparameter search space by offering a define-by-run API. Optuna has a user-friendly setup, and it provides efficient sampling and pruning algorithms for customization. Detailed information about Optuna and its algorithm can be found in Akiba et al. (2019).

We did not use the test set in automatic hyperparameter tuning to ensure our ML models' generalizability and avoid data leakage in our study. Instead, we utilized a cross-validation set consisting of 16 stations from the training data, where each represents a distinct climatic zone. This cross-validation set mimics the unseen test set. With Optuna, we conducted 100 trials to optimize the CatBoost model hyperparameters.

3.3.2. Deep neural network

The feed-forward neural network is the ancestor of all deep learning architectures. Deep neural networks (DNNs) are well-known and well-studied architectures with innumerable applications as predictive

regression models and are commonplace in ET prediction and forecasting (Kumar et al., 2002; Kim and Kim, 2008). Although gradient-boosted algorithms, such as XGBoost, CatBoost, and LightGBM, are generally considered more suitable for tabular datasets (Shwartz-Ziv and Armon, 2022), such as the dataset used in our study, we also included a deep neural network (DNN) algorithm in our models to assess its performance relative to other methods. We used the *TensorFlow* library in Python to build our DNN model. Our model has three hidden layers with 16, 32, and 16 units, respectively. The activation function of input and hidden layers is the Rectified Linear Unit (ReLU). Mean Squared Error (MSE) is the loss function of the network. Ridge Regression (L2) and Lasso Regression (L1) are used as kernel and activity regularizers, respectively, with a regularization parameter of 0.01. The model uses Adam optimizer and is trained over 30 epochs with a batch size of 64.

3.3.3. LightGBM

LightGBM (Light Gradient Boosting Machine), originally proposed by Ke et al. (2017) and developed by Microsoft as a free and open-source framework, efficiently implements the gradient boosting algorithm and reduces memory usage. Gradient boosting is an ensemble method where ensembles are constructed from decision trees. Models are fit through a gradient descent optimization algorithm, where the loss gradient is minimized as the model is tuned (Brownlee, 2020). We used the LightGBM package in Python to develop the LightGBM model. We used a maximum of 30 tree leaves and a maximum of 7 tree depths for base learners. We used 2000 boosted trees to fit. More information about the LightGBM method can be found in Ke et al. (2017) and Al Daoud (2019).

3.3.4. Random forest

Random forest (RF) is a supervised learning algorithm used to solve classification and regression problems. RF is a bagging technique that trains many decision trees in parallel, and its results are based on an ensemble of the trained trees. When used as a regression model, RF returns the mean or average prediction of the individual trees as the model's output. Detailed information about the RF algorithm can be found in the references (Ho, 1995; Breiman, 2001). We used the *Scikit-learn* software library in Python to build the RF model. We used 1000 trees in the forest, with a maximum depth of 9.

3.3.5. XGBoost

XGBoost (eXtreme Gradient Boosting) is an open-source gradient-boosting software library. XGBoost provides a parallel tree boosting (also known as gradient-boosted decision trees) that efficiently solves classification and regression problems. More information about the XGBoost model can be found in the references (Chen et al., 2015; Chen and Guestrin, 2016). We used the *XGBoost* library in Python to build the XGBoost model. We used 1000 trees in the ensemble, with minimum depth and maximum leaves of 7 and 9, respectively. The learning rate was set to 0.3.

3.4. Empirical models

We evaluated the performance of our input-limited data-driven approach against available input-limited alternatives for ET_O estimation. Three well-known and reliable methods were chosen for this aim: a radiation-based method (i.e., Priestley-Taylor), a temperature-based method (i.e., Hargreaves-Samani), and a mass transfer-based method (i.e., Romanenko). Although none of these methods eliminate the need for a standard reference surface and their meteorological inputs rely on the measurement conditions, the comparison between SolarET and the methods is beneficial as they are well-established, widely-used procedures for ET_O estimation.

3.4.1. Priestley-Taylor

As shown in Eq. 1, the P-M equation consists of a radiation term and an aerodynamic term, where the latter groups the elements that represent the surface-atmosphere interactions (Pereira, 2004). By analyzing data collected over various surfaces, Priestley and Taylor (1972) proposed an empirical adjustment to the P-M equation (Mallick et al., 2014). The Priestley-Taylor (PT) equation neglects the aerodynamic term and corrects the estimated ET by a dimensionless coefficient α (the Priestley-Taylor parameter). This coefficient accounts for the atmospheric vapor pressure deficit and air and surface resistances to estimate the ET from a short, well-watered grass surface. Experimental results from several sites worldwide resulted in an average value of α=1.26. We use this value in this research. Pereira (2004) states that 1.26 is a good estimate for α in Davis, California. Eq. 6 depicts how the PT method calculates ET_O:

$$ET_O = 1.26\frac{\Delta}{\Delta+\gamma}\frac{R_n - G}{\lambda} \tag{6}$$

Where Δ is the slope of the saturation vapor pressure curve (kPa/℃), γ is the psychrometric constant (kPa/℃), λ is the latent heat of vaporization (2.45 MJ/kg), R_n is the net radiation ($(MJ/m^2)/day$), and G is the soil heat flux ($(MJ/m^2)/day$). The Δ over Δ+γ represents the fraction of diabatic energy (R_n - G) contributing to ET from the grass surface, and the adiabatic contribution to ET is approximately 26% of the diabatic contribution. The daily ground heat flux is assumed to be G=0.

3.4.2. Hargreaves-Samani

Hargreaves-Samani (HS) is an input-limited empirical equation to estimate ET_O using air temperature and extraterrestrial radiation (Hargreaves and Samani, 1985). This equation is based on the Hargreaves original equation, approximating solar radiation with extraterrestrial radiation and maximum minus minimum air temperature. Since the HS equation is developed with daily lysimeter data from Davis, California, it is well-suited for and highly used in California, which makes it a good benchmark for validating our approach. The HS equation is as follows:

$$ET_O = 0.0023\, R_a (T_{mean} + 17.8)(T_{\max} - T_{\min})^{0.5} \tag{7}$$

Where ET_O is reference evapotranspiration (mm/day), R_a is the extraterrestrial radiation (mm/day, for equivalent evaporated water depth), $T_{mean} = (\frac{T_{\max}+T_{\min}}{2})$, $T_{\max}$, and $T_{\min}$ are the mean, maximum, and minimum daily air temperature (ºC), respectively. Since data from Davis, California, is used to calibrate the constant value of 0.0023, we use the same value in this research.

3.4.3. Romanenko

Romanenko equation is a mass transfer-based empirical equation that uses air temperature and relative humidity to estimate ET_O. The mass transfer-based methods are essentially based on Dalton's gas law and employ the concept of eddy transfer of water vapor from an evaporating surface to the atmosphere (Mehdizadeh et al., 2017). Romanenko equation estimates daily ET_O as follows:

$$ET_O = 0.00006\left(25 + T_{avg}\right)^2 (100 - RH) \tag{8}$$

Where RH is the relative humidity (%).

3.5. Performance measures

In this study, two deterministic performance measures are employed to investigate the accuracy of ET_O estimation models: root mean square error (RMSE) and mean absolute error (MAE) (Moriasi et al., 2015):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \tag{9}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(|O_i - P_i|) \quad (10)$$

Where N is the number of samples, and O_i and P_i are observed and predicted ET_O values at i^{th} sample, respectively. Lower RMSE and MAE values indicate higher accuracies and better performance. Although both RMSE and MAE indicate average model prediction error in units of the variable of interest, they have fundamental differences. Since RMSE squares the errors before averaging them, it gives higher weights to large errors. Consequently, RMSE tends to penalize large errors more severely than MAE. According to Chai and Draxler (2014), each of these metrics has its own merits in explaining the model performance; therefore, in this research, we use them in tandem.

4. Results and discussion

4.1. Data investigation

Fig. 2 shows the number of data samples and the training data distribution for each climatic zone. This figure shows that zones 14, 6, and 12 have the highest number of stations and input data samples in the training set. On the other hand, there is no data from zone 4 in the training set, as all the samples from this zone have been eliminated in the data preprocessing. However, zone 4 is present in the test set. This can further evaluate the generalizability of the presented methodology for ET_O estimation in unseen locations in California. Zones 12 and 14 cover a vast area of the Central Valley. The Central Valley is one of the most productive agricultural regions in the world, with ideal soil and climate for a variety of crops but limited water resources. With limited to no rainfall in the summer, the Central Valley's agriculture relies on irrigation. About one-sixth of the US irrigated land lies in the Central Valley (Reilly et al., 2008). This heavy irrigation dependence explains this region's high density of CIMIS stations.

Fig. 2 suggests a relatively uniform distribution of solar radiation across California, with lower values in coastal zones (e.g., zones 1 and 2) and higher values in deserts (e.g., zones 17 and 18). The lower R_S values on the coast can be attributed to dense fog.

The ET_O distribution in California reveals more variability across different zones. The most anomalous zones in terms of ET_O distribution are coastal zones with very low ET_O and desert zones with very high ET_O. Also, the box plots of Fig. 2 illustrate a broader ET_O distribution for desert zones. Being abnormal regarding R_S and ET_O distribution and having fewer input data, we hypothesize that our data-driven models struggle with accurate ET_O prediction in coastal and desert zones.

4.2. Feature selection

The results of the feature selection methods are presented in Fig. 3, which suggests no sharp difference between the features' importance, especially from Pearson and Spearman correlations. As expected, all feature selection methods find the hourly R_S values from morning and

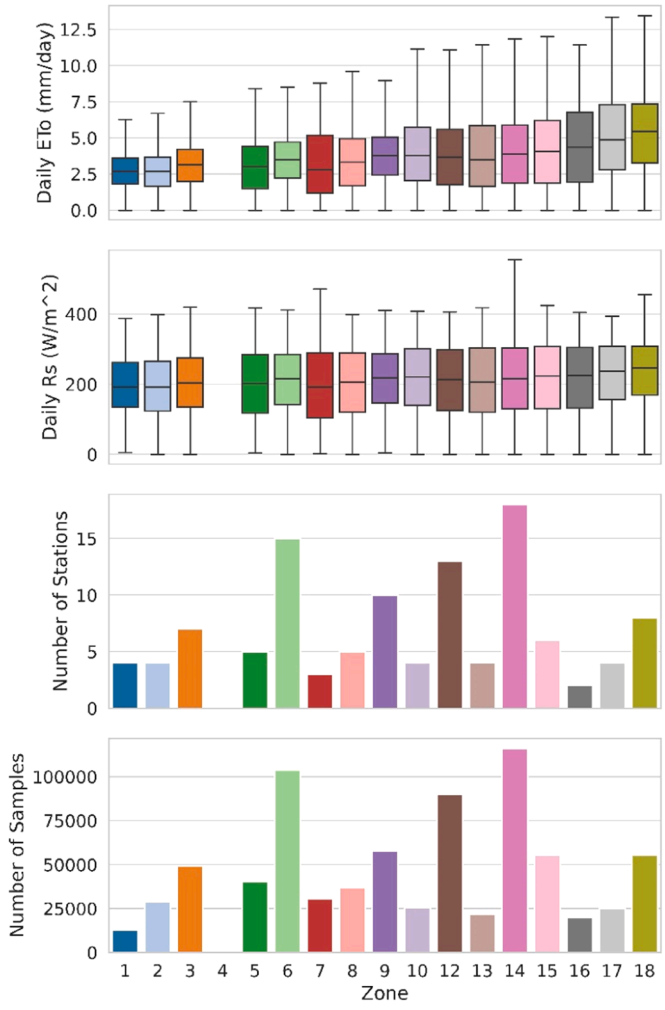


Fig. 2. Number of samples, stations, and distribution of the training data for each climatic zone.

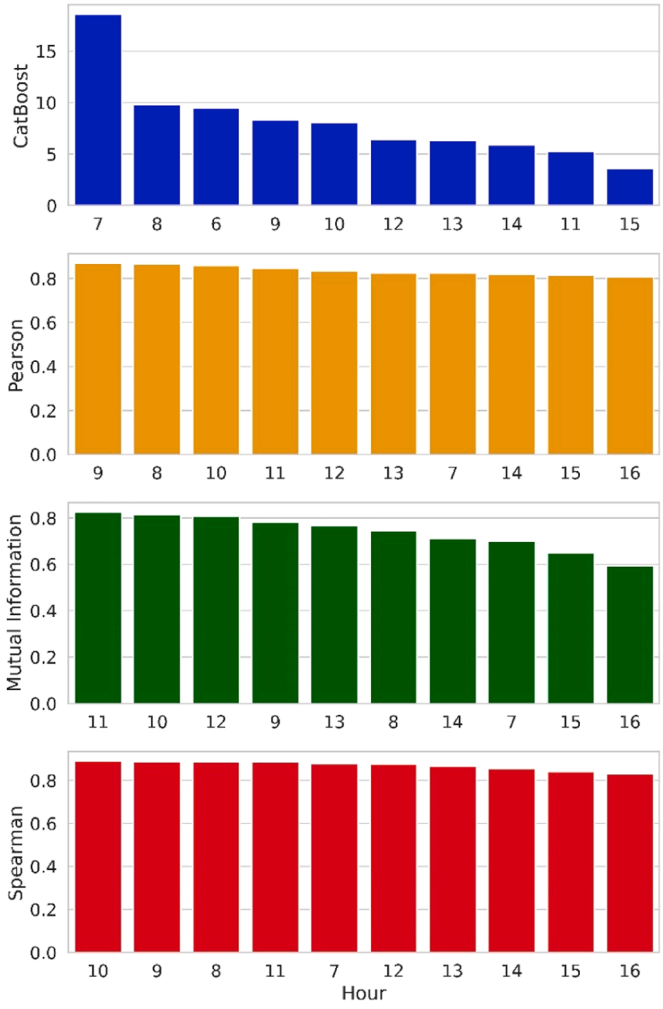


Fig. 3. The most important features (i.e., solar radiation from different hours of the day) selected by each feature selection method and their corresponding score.

around noon to be the most important features. Based on these results, we decided to use the six most essential features/hours as the input of the ML models. For example, the six most important hourly Rs values selected by the CatBoost feature importance method are 7, 8, 6, 9, 10, and 12, respectively (Fig. 3).

It should be noted that both Pearson and MI methods result in hours 8–13 of Rs as the six most important features, only with various orders. Therefore, the results for Pearson and MI methods are the same and combined hereafter. Because different methods suggest a uniform feature importance, we hypothesize that the mean substitution procedure (i.e., using all hourly R_S values) might be superior to using only selected features as input.

4.3. Machine learning models

The overall performance of the ML models and feature selection methods is depicted in Fig. 4 and Figure A1 in the appendix. As these figures show, CatBoost and RF have very similar performances in terms of prediction accuracy. It should be noted that while Optuna has automatically optimized the CatBoost model's hyperparameters, RF hyperparameters are set manually. This would suggest that automatic hyperparameter tuning for this type of regression problem does not guarantee higher accuracy, and a manual hyperparameter tuning strategy can result in equally accurate models. Also, most of the available automatic hyperparameter tuning frameworks, including Optuna, are specifically designed to optimize the hyperparameters of deep learning architectures, where the high dimensionality of the hyperparameter space interferes with manual hyperparameter tuning strategies like grid or random search and might result in sub-efficient models (Akiba et al., 2019).

Fig. 4 and A1 illustrate the overall superiority of using all the hourly R_S values with the mean substitution method over selecting the most important hours. Among feature selection methods, ML models using CatBoost intrinsic feature importance and Spearman correlation perform better than MI and Pearson correlation. The quality and relevance of the selected features can influence the performance of ML models. In our case, the difference in performance between the feature selection methods can be attributed to 1) the ability to handle non-linear relationships: ML models can effectively capture non-linear relationships between features and the target variable. In this case, CatBoost's intrinsic feature importance considers the inherent non-linear relationship between the input and output of the model and selects features that have the highest impact on the model's performance. 2) Robustness to outliers and noise: These observations can also be attributed to CatBoost intrinsic feature importance and Spearman correlation's more effective robustness to outliers than Pearson correlation and Mutual Information correlation. Considering the rank-based correlation in Spearman's correlation or the model-specific feature importance in CatBoost, these methods may have selected more robust features that lead to better model performance.

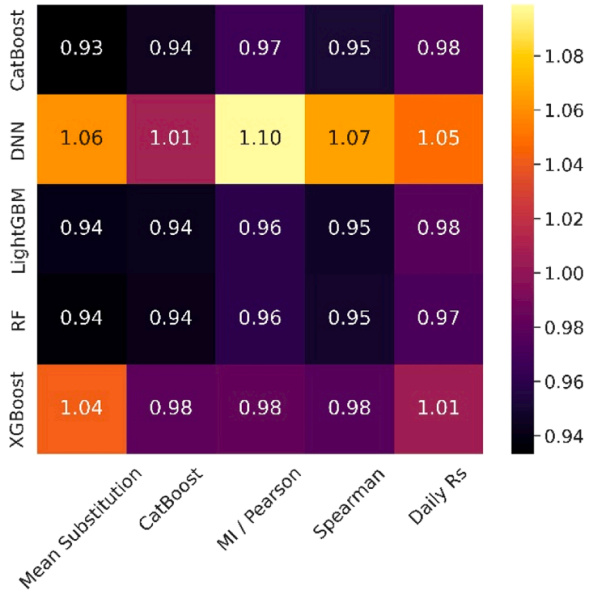


Fig. 4. Root Mean Square Error (RMSE) values (mm/day) for machine learning models and feature selection methods over the whole testing set.

Using feature selection or mean substitution, the CatBoost model with the mean substitution method stands out as the most accurate hourly model. Additionally, the RF excels as the preferred choice between daily models. We compare these models as the best-performing frameworks with empirical ET_O estimation models.

Fig. 4 and A1 clearly show that the DNN model falls behind all ensemble decision tree-based algorithms (i.e., RF and gradient boosting-based frameworks). This finding agrees with previous studies where tree-based algorithms outperformed vanilla neural networks in predictive regression problems (Rodriguez-Galiano et al., 2015; Pham et al., 2020; Jun, 2021).

To better understand how SolarET performs in different climatic zones of California, we included one CIMIS station with around five years of data from each zone. These stations have not been used to train the model; therefore, SolarET's performance in these locations can be claimed as its performance in any arbitrary location in those climatic zones. Fig. 5 and A2 depict the models' accuracy for each zone in the test set. These results clearly show that the performance of various ML models and feature selection methods is similar in each climatic zone. In other words, we cannot claim that a specific combination of the ML model and feature selection technique is more suitable for a geographical location or particular climatic characteristics. All successful prediction frameworks (i.e., ensemble decision tree-based algorithms) have lower accuracies in estimating ET_O in coastal (zones 1 and 2) and desert (zones 17 and 18) areas (Fig. 5 and A2).

From an ML perspective, three reasons might contribute to the lower accuracy of SolarET in the coastal and desert regions: 1) the low number of training data from these zones: Fig. 2 shows limited training data in these zones- leading to their insufficient representation of data patterns. Other zones also have low numbers of training samples but high predictive accuracy (e.g., zones 16, 13, and 10). Although, these zones are more similar to the majority of the training set in terms of ET_O distribution (Fig. 2). Therefore, we hypothesize that the first reason for the inferior performance of SolarET in the coast and desert data is the low number of training samples resembling the statistical characteristics of the data from these zones. 2) Higher level of outlier values in both training and test sets: As depicted in Fig. 6, there is a higher percentage of outlier values in stations 17 and 18 compared to other stations, which could impact the performance of the ML models in these specific zones. These outliers introduce noise and may not follow the patterns that the ML models have learned from the training data, leading to less accurate predictions. 3) Lack of relevant features: According to Fig. 7, we hypothesize that one of the other reasons for less accurate predictions in zones 17 and 18 is the lower correlation between the daily R_S and ET_O in these regions. Various climatic variables, such as temperature, humidity, wind speed, and solar radiation influence ETO. In zones 17 and 18, hourly or daily solar radiation may not accurately capture all the required information to estimate ET_O. All in all, it is worth mentioning that the combination of these reasons, rather than a single factor, may have led to inferior performance of the ML models in these zones. It is also expected that advection is a more considerable contribution to ET_O in desert regions than in the Central Valley or along the coast, and the lack of information on temperature, humidity, and wind speed in the ML models should lead to more considerable errors in regions having large fluctuations in aerodynamic contributions to ET_O (Berengena and Gavilán, 2005).

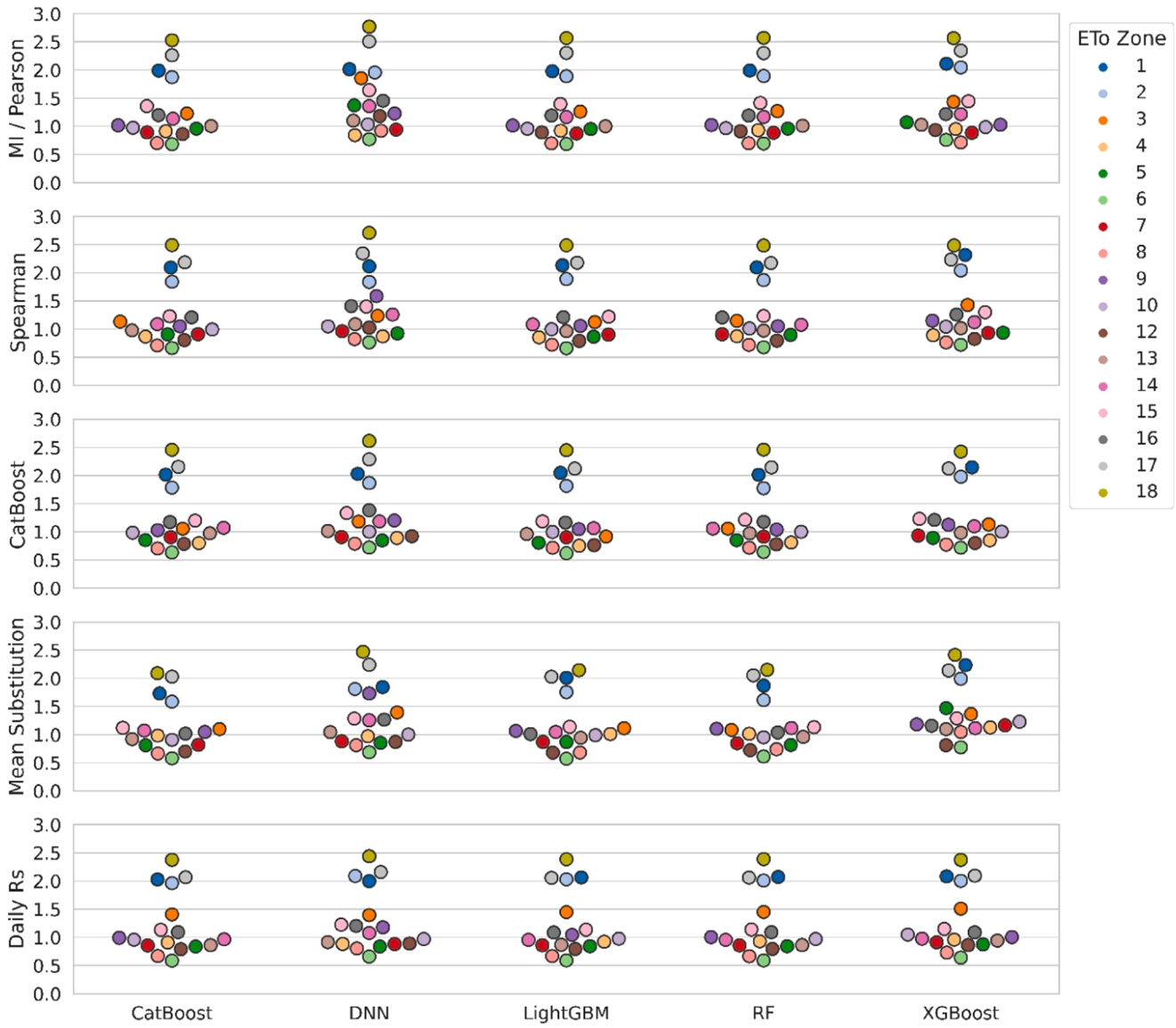


Fig. 5. Root Mean Square Error (RMSE) values (mm/day) for machine learning models and feature selection methods over climatic zones of California.

Other possible physical and climatological causes of this lower performance are discussed in Section 4.4.

4.4. *Reference evapotranspiration prediction*

To evaluate our approach of using ML models with only pyranometer data for ET_O estimation, we compared its performance against three empirical limited-input ET_O estimation models. It should be noted that both approaches, ML and empirical models, were evaluated using the same test set but with different features. Fig. 8 and A3 depict the findings of this comparison over 17 climatic zones of the test set. In these figures, we displayed the results of only the best-performing data-driven models. In the case of hourly inputs, the CatBoost ML model with mean substitution is the best, and in the case of daily inputs, the random forest model gives the highest accuracy (Fig. 4 and A1).

As seen in Fig. 8 and A3, our hourly model generally works as well or even better than the most accurate empirical model (i.e., Priestley-Taylor). This is especially important as our model takes only solar radiation data as input, while PT takes information from minimum and maximum air temperature and elevation in addition to net radiation. Also, our hourly and daily models clearly beat Hargreaves-Samani and Romanenko in terms of prediction accuracy. Our findings suggest that SolarET works very well unless there are extremely high or low contributions of aerodynamics to the ET_O (e.g., desert and coastal climates). This is not surprising, as SolarET can only mimic the radiation term of the P-M equation, as it takes no information about the aerodynamics term (Eq. 1). Still, Fig. 8 and A3 illustrate that SolarET works better than PT as a radiation-based ET estimation model in desert climates. The inferior performance of SolarET in comparison to PT in coastal regions might be due to the high humidity fluctuation in these regions due to variations in wind direction, e.g., from the land or from the ocean, which lead to significant variations in temperature and humidity. Moreover, the presence of intermittent clouds and fog in these regions might be another reason for the lower accuracy of our model as opposed to PT. The low accuracy of empirical models in desert climate might be due to two reasons: 1) these models are not calibrated for desert climate (Priestley and Taylor, 1972; Hargreaves and Samani, 1985; Tabari et al., 2013); and 2) these models do not adequately account for heat advection in hot desert climates (Tolk et al., 2006; Wang et al., 2023).

We hypothesize that another reason for the inferior performance of SolarET in deserts and coast is not accounting for aridity. This pattern is the same for PT and HS in desert regions. This goes along with not

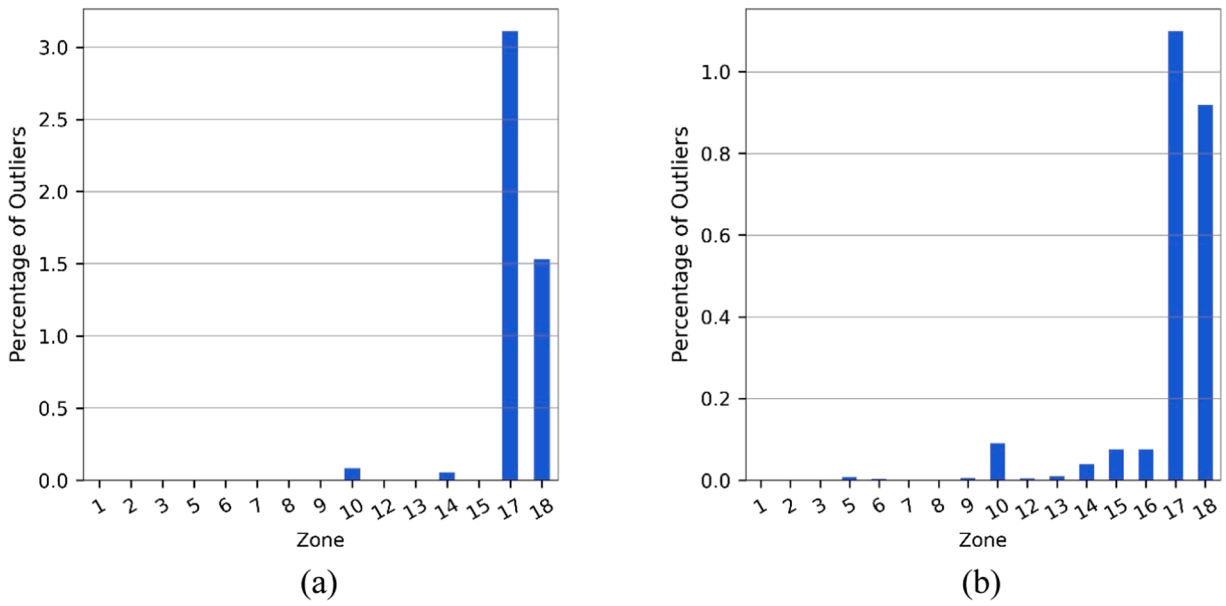


Fig. 6. Outlier percentage for each climatic zone of California in a) test set, and b) training set.

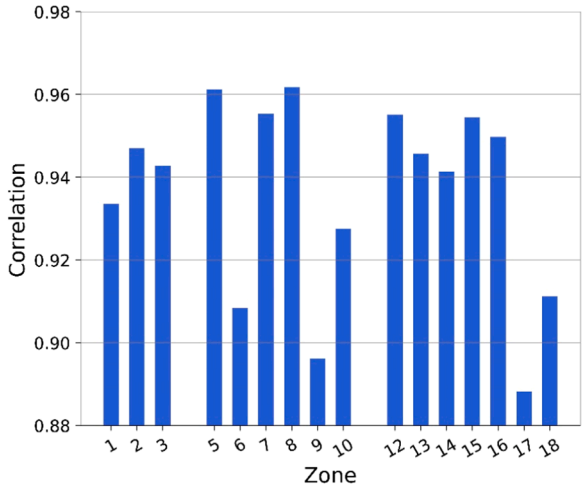


Fig. 7. Spearman correlation between daily solar radiation (R_S) and reference evapotranspiration (ET_O) over climatic zones of California.

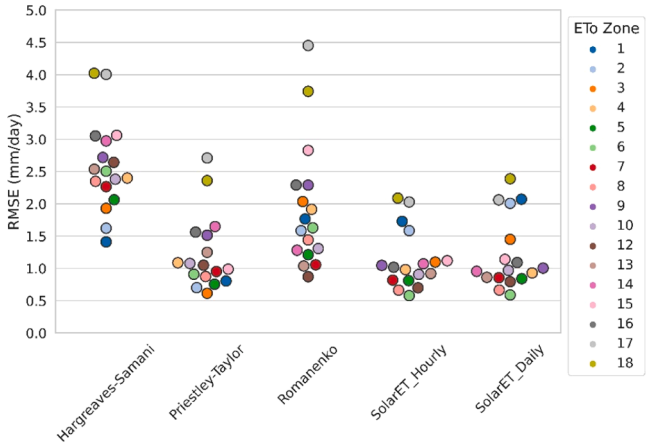


Fig. 8. Root Mean Square Error (RMSE) values for empirical models and SolarET over climatic zones of California.

considering the aerodynamics contribution to the daily ET_O. We hypothesize that adding information about vapor pressure can solve this problem.

Fig. 9 analyzes the performance of SolarET against available empirical alternatives over the entire test set in terms of prediction accuracy. RMSE and MAE measure the dissimilarity between input-limited alternatives and P-M equation prediction. We use these performance measures to assess which input-limited ET_O estimation alternative can simulate the daily P-M equation more realistically. As Fig. 9 demonstrates, our hourly and daily models beat all empirical alternatives in terms of similarity to the P-M daily ET_O. Our hourly model is the winner of this comparison. Although the overall performance of the hourly model is close to the daily model according to Fig. 9, as Fig. 8 depicts, the hourly model results in identical or more accurate predictions for all climatic zones. Interestingly, the hourly model works much better than the daily model in coastal regions (zones 1, 2, and 3 in Fig. 8). This finding suggests that the information gained from the higher temporal resolution input (i.e., hourly R_s) is beneficial in simulating climatically diverse samples and can boost generalizability.

Fig. 9 shows that the PT equation, which is a radiation-based model, is the most accurate empirical ET_O estimation model. The superiority of SolarET and the high accuracy of the PT model signifies the suitability of radiation data for input-limited ET_O estimation in California. This finding agrees with Ahmadi et al. (2022) findings on the high correlation between radiation and ET_O over California. Fig. 9 illustrates that HS and

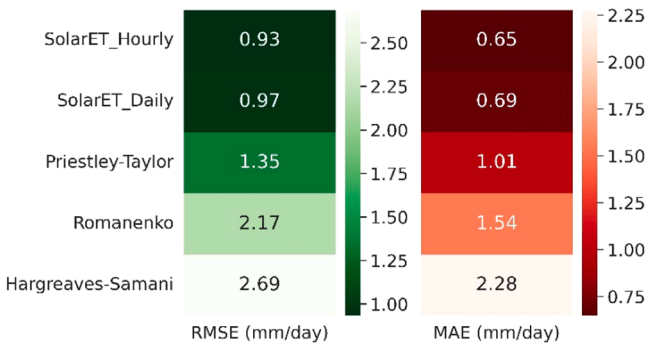


Fig. 9. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values for empirical models and SolarET over the entire test set.

Romanenko models are the least realistic representations of P-M ET_O over California. However, it should be noted that since we aim to compare not only the accuracy but also the generalizability of models, the HS model was not calibrated in this study. We hypothesize that calibrating HS will improve its accuracy in ET_O estimation.

Analyzing the errors of the ML and empirical models can provide valuable insights into their performance and help identify areas for improvement. In this vein, we divided the ET_O distribution into distinct ranges and compared the best-performing ML model, CatBoost, using hourly R_s, with the PT empirical model. Fig. 10 illustrates the percentage of values within each specific ET_O range where the value of $\frac{|ETo_{actual}-ETo_{predicted}|}{ETo_{actual}}$ exceeds 0.5. In this equation, ETo_{actual} Refers to P-M value and $ETo_{predicted}$ is what the model predicts. This figure indicates the more accurate predictions of CatBoost compared to PT across various ranges of ET_O values. Overall, CatBoost maintains its superiority not only in the whole distribution of ET_O over the entire test set but also in cases of low or high values of ET_O. Moreover, as Figure A4 in the appendix shows, the distribution of SolarET error is similar to a Gaussian distribution with a mean value close to zero. This distribution suggests the error is random, not a systematic over- or under-estimation of ET_O. It is also worth mentioning that SolarET accuracy is comparable to the best-performing ML-based ET_O estimation models available in the literature, while it does not use surface-related inputs such as air temperature (Chen et al., 2020).

5. Summary and conclusion

This paper develops a generalizable data-driven approach to estimate ET_O using only solar radiation data. Since solar radiation is the only driving force of ET_O, whose measurement does not require a reference crop surface, the developed approach (i.e., SolarET) can cut the cost of land, irrigation, maintenance, and extra sensors while assisting irrigators and water managers with their farm-level and regional decisions. As incoming solar radiation is not a function of the surface, a simple pyranometer placed on any location can measure the inputs of SolarET. Using a test set from unseen arbitrary locations, we showed the generalizability of SolarET over California. Our findings reveal that SolarET beats the Hargraves-Samani and Romanenko models in terms of accuracy. SolarET works better than Priestley-Taylor, a well-known radiation-based ET_O estimation method, without using any information about air temperature. Although SolarET works accurately in heavily irrigated regions of California and beats all empirical models in desert climates, it is less accurate in coastal regions. Future studies can include publicly available air temperature data as another input, along with measured solar radiation. Adding publicly available air temperature data can result in better performance in coastal and desert regions, where the contribution of the aerodynamics term of the Penman-Monteith equation is different from other inland regions of California. Moreover, developing models tailored to a specific region or climate might result in higher accuracies but at the cost of lower generalizability. The accuracy-generalizability trade-off of this data-driven approach is another area for future research.

CRediT authorship contribution statement

Mohammad Hossein Kazemi: Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andre Daccache:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Arman Ahmadi:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Richard Snyder:** Writing – review & editing, Conceptualization.

Declaration of Competing Interest

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.

Data availability

Data will be made available on request.

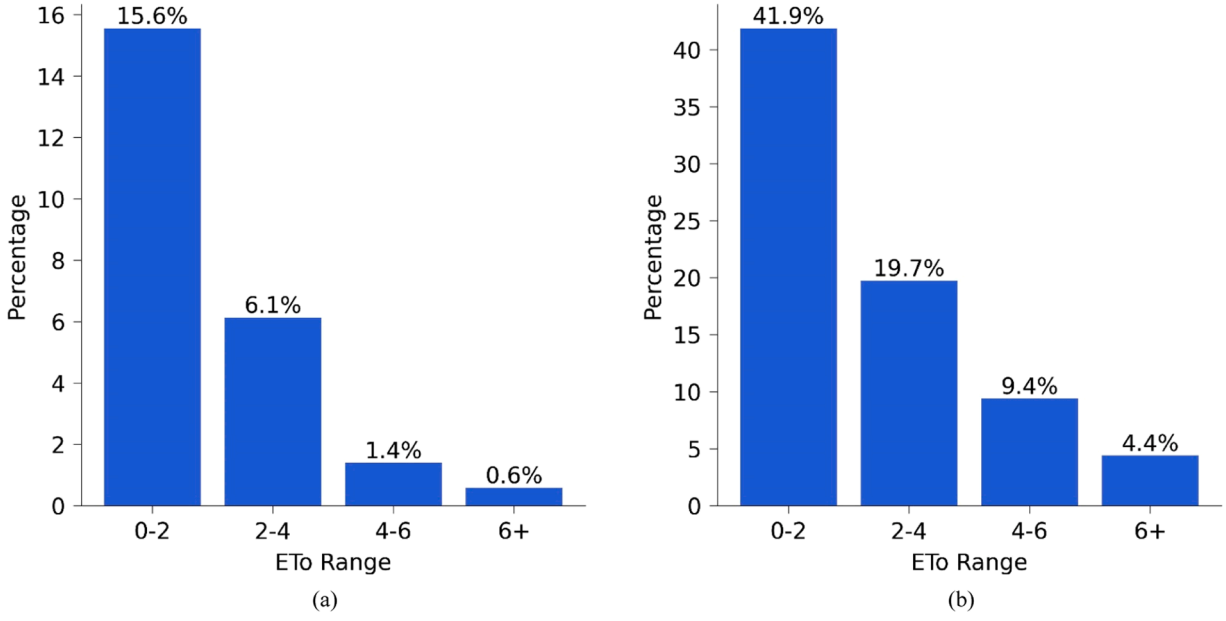


Fig. 10. Percentage of values with $\frac{|ETo_{actual}-ETo_{predicted}|}{ETo_{actual}} > 0.5$ for a) CatBoost hourly model and b) Priestley-Taylor in each specific reference evapotranspiration (ETo) range.

Appendix

Table A1
Number, name, and characteristics of reference evapotranspiration zones according to the CIMIS website

Zone #	Name	Characteristics
1	Coastal plains heavy fog belt	lowest ET_O in California, characterized by dense fog
2	Coastal mixed fog area	less fog and higher ET_O than zone 1
3	Coastal valleys & plains & north coast mountains	more sunlight than zone 2
4	South coast inland plains & mountains north of San Francisco	more sunlight and higher summer ET_O than zone 3
5	Northern inland valleys	valleys north of San Francisco
6	Upland central coast & Los Angeles basin	higher elevation coastal areas
7	Northeastern plains	
8	Inland San Francisco Bay area	inland area near San Francisco with some marine influence
9	South coast marine to desert transition	inland area between marine & desert climates
10	North central plateau & central coast range	cool, high elevation areas with strong summer sunlight
11	Central Sierra Nevada	mountain valleys east of Sacramento with some influence from delta breeze in summer
12	East side Sacramento-San Joaquin valley	low winter & high summer ET_O with slightly lower ET_O than zone 14
13	Northern Sierra Nevada	northern Sierra Nevada Mountain valleys with less marine influence than zone 11
14	Mid-central valley, southern Sierra Nevada, Tehachapi & high desert mountains	high summer sunshine and wind in some locations
15	Northern and southern San Joaquin valley	slightly lower winter ET_O due to fog and slightly higher summer ET_O than zones 12 & 14
16	Westside San Juaquin valley & mountains east & west of Imperial valley	
17	High desert valleys	valleys in the high desert near Nevada and Arizona
18	Imperial Valley, Death Valley & Palo Verde	low desert areas with high sunlight & considerable heat advection

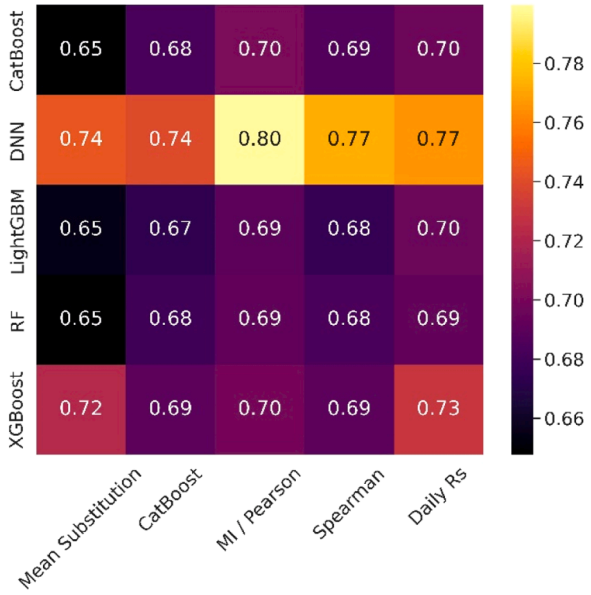


Figure A1. Mean Absolute Error (MAE) values (mm/day) for machine learning models and feature selection methods over the whole testing set.

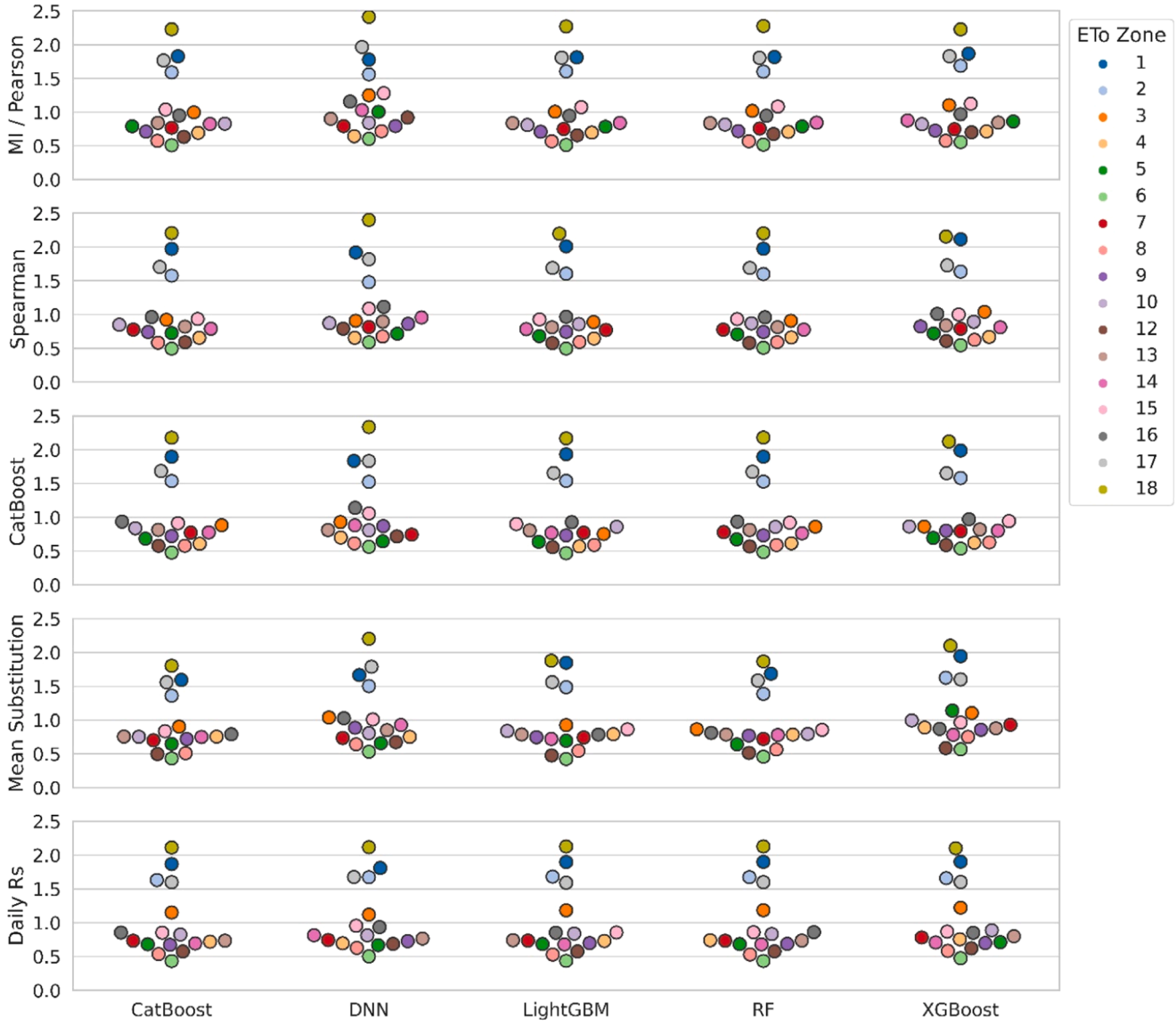


Figure A2. Mean Absolute Error (MAE) values (mm/day) for machine learning models and feature selection methods over climatic zones of California.

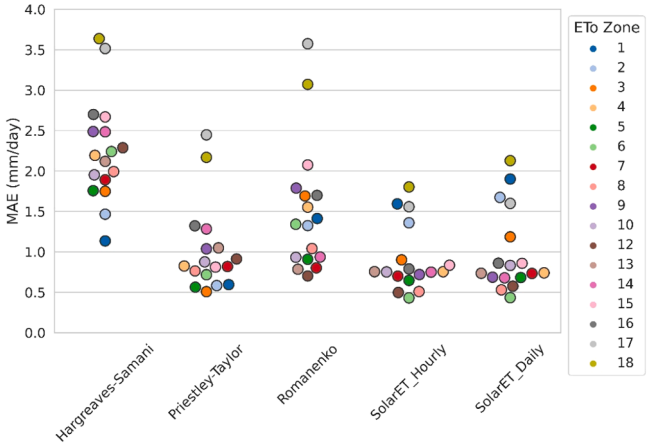


Figure A3. Mean Absolute Error (MAE) values for empirical models and SolarET over climatic zones of California.

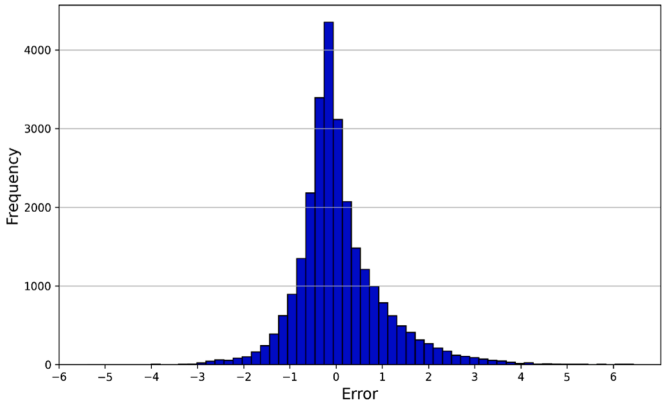


Figure A4. Distribution of error (the difference between predicted and observed ET_O) for SolarET.

References

Ahmadi, A., Daccache, A., Snyder, R.L., Suvočarev, K., 2022. Meteorological driving forces of reference evapotranspiration and their trends in California. Sci. Total Environ. *849*, 157823.

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. Comput. Electron. Agric. *215*, 108424.

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., 2019, July. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2623-2631).

Al Daoud, E., 2019. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. Int. J. Comput. Inf. Eng. *13* (1), 6–10.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage, *300*. Fao, *Rome*, p. D05109.

Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., Snyder, R. L., 2005. The ASCE Standardized Reference Evapotranspiration Equation. Amer. Soc. Of Civil Eng., Reston, Virginia 192p.

Asghari, V., Kazemi, M.H., Duan, H.F., Hsu, S.C., Keramat, A., 2023. Machine learning modeling for spectral transient-based leak detection. Autom. Constr. *146*, 104686.

Berengena, J., Gavilán, P., 2005. Reference evapotranspiration estimation in a highly advective semiarid environment. J. Irrig. Drain. Eng. *131* (2), 147–163.

Breiman, L., 2001. Random forests. Mach. Learn. *45*, 5–32.

Brownlee, J., 2020. Gradient Boosting With Scikit-Learn, Xgboost, Lightgbm, and Catboost. *Machine Learning Mastery*.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geosci. Model Dev. *7* (3), 1247–1250.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system (August). Proc. 22nd acm sigkdd Int. Conf. Knowl. Discov. data Min. 785–794.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I. and Zhou, T., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), pp.1-4.

Chen, Z., Zhu, Z., Jiang, H., Sun, S., 2020. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. J. Hydrol. *591*, 125286.

Dong, J., Zhu, Y., Jia, X., 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. J. Hydrol. *604*, 127207.

Dorogush, A.V., Ershov, V., Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv Prepr. arXiv *1810*, *11363*.

Fernández, E., 2023. Editorial note on terms for crop evapotranspiration, water use efficiency and water productivity. Agric. Water Manag. *289*, 108548.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., 2011. Solutions for a cultivated planet. Nature *478* (7369), 337–342.

Haghverdi, A., Singh, A., Sapkota, A., Reiter, M., Ghodsi, S., 2021. Developing irrigation water conservation strategies for hybrid bermudagrass using an evapotranspiration-based smart irrigation controller in inland southern California. Agric. Water Manag. *245*, 106586.

Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. J. big data *7* (1), 1–45.

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Appl. Eng. Agric. *1* (2), 96–99.

Hauke, J., Kossowski, T., 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest. Geogr. *30* (2), 87–93.

Ho, T.K., 1995, August. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

Ji, X.B., Chen, J.M., Zhao, W.Z., Kang, E.S., Jin, B.W., Xu, S.Q., 2017. Comparison of hourly and daily Penman-Monteith grass-and alfalfa-reference evapotranspiration equations and crop coefficients for maize under arid climatic conditions. Agric. Water Manag. *192*, 1–11.

Karimzadeh, S., Hartman, S., Chiarelli, D.D., Rulli, M.C., D'Odorico, P., 2024. The trade-off between water savings and salinization prevention in dryland irrigation. Adv. Water Resour. *183*, 104604.

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. Adv. Neural Inf. Process. Syst. *30*.

Kim, S., Kim, H.S., 2008. Neural networks and genetic algorithm approach for non-linear evaporation and evapotranspiration modeling. J. Hydrol. *351* (3-4), 299–317.

Kozachenko, L.F., Leonenko, N.N., 1987. Sample estimate of the entropy of a random vector. Probl. Peredachi Inf. *23* (2), 9–16.

Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Phys. Rev. E *69* (6), 066138.

Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W., Pruitt, W.O., 2002. Estimating evapotranspiration using artificial neural network. J. Irrig. Drain. Eng. *128* (4), 224–233.

Kushwaha, N.L., Rajput, J., Sena, D.R., Elbeltagi, A., Singh, D.K., Mani, I., 2022. Evaluation of data-driven hybrid machine learning algorithms for modelling daily reference evapotranspiration. Atmosphere-Ocean *60* (5), 519–540.

Lobell, D.B., Bonfils, C., 2008. The effect of irrigation on regional temperatures: a spatial and temporal analysis of trends in California, 1934–2002. J. Clim. *21* (10), 2063–2071.

Mallick, K., Jarvis, A.J., Boegh, E., Fisher, J.B., Drewry, D.T., Tu, K.P., Hook, S.J., Hulley, G., Ardö, J., Beringer, J., Arain, A., 2014. A surface temperature initiated closure (STIC) for surface energy balance fluxes. Remote Sens. Environ. *141*, 243–261.

Mehdizadeh, S., Behmanesh, J., Khalili, K., 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. Comput. Electron. Agric. *139*, 103–114.

Monteith, J.L., 1965. Evaporation and environment. In. In: Symposia of the society for experimental biology, Vol. 19. Cambridge University Press (CUP), Cambridge, pp. 205–234.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Trans. ASABE *58* (6), 1763–1785.

Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci. *193* (1032), 120–145.

Pereira, A.R., 2004. The Priestley–Taylor parameter and the decoupling factor for estimating reference evapotranspiration. Agric. For. Meteorol. *125* (3-4), 305–313.

Pereira, L.S., Allen, R.G., Smith, M., Raes, D., 2015. Crop evapotranspiration estimation with FAO56: past and future. Agric. Water Manag. *147*, 4–20.

Pham, T.A., Ly, H.B., Tran, V.Q., Giap, L.V., Vu, H.L.T., Duong, H.A.T., 2020. Prediction of pile axial bearing capacity using artificial neural network and random forest. Appl. Sci. *10* (5), 1871.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. *100* (2), 81–92.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. *31*.

Reilly, T.E., Dennehy, K.F., Alley, W.M. and Cunningham, W.L., 2008. Ground-water availability in the United States (No. 1323). Geological Survey (US).

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818.

Ross, B.C., 2014. Mutual information between discrete and continuous data sets. PloS One *9* (2) p.e87357.

Schmitt, R.J., Rosa, L., Daily, G.C., 2022. Global expansion of sustainable irrigation limited by water storage. Proc. Natl. Acad. Sci. *119* (47), e2214291119.
Shwartz-Ziv, R., Armon, A., 2022. Tabular data: deep learning is not all you need. Inf. Fusion *81*, 84–90.
Siebert, S., Burke, J., Faures, J.M., Frenken, K., Hoogeveen, J., Döll, P., Portmann, F.T., 2010. Groundwater use for irrigation–a global inventory. Hydrol. earth Syst. Sci. *14* (10), 1863–1880.
Tabari, H., Grismer, M.E., Trajkovic, S., 2013. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. Irrig. Sci. *31*, 107–117.
Tindula, G.N., Orang, M.N., Snyder, R.L., 2013. Survey of irrigation methods in California in 2010. J. Irrig. Drain. Eng. *139* (3), 233–238.
Tolk, J.A., Evett, S.R., Howell, T.A., 2006. Advection influences on evapotranspiration of alfalfa in a semiarid climate. Agron. J. *98* (6), 1646–1654.
Walter, I.A., Allen, R.G., Elliott, R., Jensen, M.E., Itenfisu, D., Mecham, B., Howell, T.A., Snyder, R., Brown, P., Eching, S., Spofford, T., Hattendorf, M., Cuenca, R.H., Wright, J.L., Martin, D., 2000. ASCE's standardized reference evapotranspiration equation. Proc. of the Watershed Management 2000 Conference, June 2000, Ft. Collins, CO. American Society of Civil Engineers, St. Joseph, MI.
Wang, T., Verfaillie, J., Szutu, D., Baldocchi, D., 2023. Handily measuring sensible and latent heat exchanges at a bargain: a test of the variance-Bowen ratio approach. Agric. For. Meteorol. *333*, 109399.
Zhang, K., Li, X., Zheng, D., Zhang, L., Zhu, G., 2022. Estimation of global irrigation water use by the integration of multiple satellite observations. Water Resour. Res. *58* (3), e2021WR030031.
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. J. Hydrol. *588*, 125087.
Zhangzhong, L., Gao, H., Zheng, W., Wu, J., Li, J., Wang, D., 2023. Development of an evapotranspiration estimation method for lettuce via mobile phones using machine vision: proof of concept. Agric. Water Manag. *275*, 108003.
Zimmerman, J.K., Carlisle, D.M., May, J.T., Klausmeyer, K.R., Grantham, T.E., Brown, L. R., Howard, J.K., 2018. Patterns and magnitude of flow alteration in California, USA. Freshw. Biol. *63* (8), 859–873.