



OPEN ACCESS

EDITED BY

Anas Tallou,
Institute of Agrifood Research and
Technology (IRTA), Spain

REVIEWED BY

Jitendra Rajput,
IARI, India
Hajar Saikouk,
Mohammed V University, Morocco

*CORRESPONDENCE

Hajar Hamdaoui
✉ ha.hamdaoui@ump.ac.ma

RECEIVED 28 October 2025

REVISED 16 January 2026

ACCEPTED 09 February 2026

PUBLISHED 18 March 2026

CITATION

Hamdaoui H, Zarrouk Y, Ankush MAT, Al
Kaddouri H, Alzain MN, Noman O,
Nasr FA and Kouddane N (2026)
Accurate reference evapotranspiration
estimation with limited data for
sustainable irrigation in eastern
Morocco: a machine learning approach.
Front. Sustain. Food Syst. 10:1734366.
doi: 10.3389/fsufs.2026.1734366

COPYRIGHT

© 2026 Hamdaoui, Zarrouk, Ankush,
Al Kaddouri, Alzain, Noman, Nasr and
Kouddane. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

Accurate reference evapotranspiration estimation with limited data for sustainable irrigation in eastern Morocco: a machine learning approach

Hajar Hamdaoui^{1*}, Yassine Zarrouk², Mujtaba A. T. Ankush³,
Hanae Al Kaddouri⁴, Mashail N. Alzain⁵, Omar Noman⁶,
Fahd A. Nasr⁷ and Nouredine Kouddane¹

¹Laboratory of Improvement of Agricultural Production, Biotechnology, and Environment, Mohammed First University, Oujda, Morocco, ²Laboratory of Research in Applied Sciences, Mohammed First University, Oujda, Morocco, ³Department of Fisheries & Marine Resources, College of Agriculture, Basrah, Iraq, ⁴Laboratory of Smart Information Communication and Technology, Mohammed First University, Oujda, Morocco, ⁵Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, ⁶Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia, ⁷Biology Department, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSUI), Riyadh, Saudi Arabia

Accurate estimation of daily reference evapotranspiration (ET_o) is essential for effective irrigation scheduling, improved water-use efficiency, and sustainable crop production in arid and semi-arid regions. Although the FAO-56 Penman–Monteith equation (ET_{oPM}) is widely accepted as the reference method, its reliance on complete meteorological data limits its applicability in data-scarce agricultural systems such as those in eastern Morocco. This study proposes a practical solution for daily ET_o estimation under data-limited conditions by evaluating machine learning approaches against traditional empirical models. Daily climatic data (2001–2025) were collected from airport meteorological stations, NASA POWER reanalysis, and on-farm sensors at four representative locations in Eastern Morocco (Oujda, Berkane, Taourirt, and Figuig). ET_o was computed using the ET_{oPM} equation. Three ML algorithms—Support Vector Regression (SVR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—were developed under multiple input scenarios guided by feature-importance analysis, which identified solar radiation (R_{sn}), maximum temperature (T_{max}), and relative humidity (RH) as dominant predictors. To ensure realistic evaluation and avoid information leakage, a strict chronological split was applied, using 2001–2018 for training and 2019–2025 for independent testing. The best ML configurations were then compared with widely used empirical ET_o equations (Hargreaves–Samani, Jensen–Haise, Priestley–Taylor, Makkink, and Turc). Results showed that XGBoost consistently outperformed SVR and RF, achieving the best balance between accuracy and computational efficiency. The three-input configuration (R_{sn} + T_{max} + RH) produced near-reference performance across all locations ($R^2 = 0.976$; RMSE < 0.55 mm day⁻¹; training time ≈ 4.48 s; RAM ≈ 0.13 MB), while the reduced two-input configuration (R_{sn} + T_{max}) maintained reliable performance ($R^2 = 0.936$; RMSE ≈ 0.60 mm day⁻¹; training time ≈ 0.33 s; RAM ≈ 0.13 MB). In contrast, empirical approaches exhibited poor predictive capability and weak transferability, with negative R^2 values and substantially larger errors (RMSE generally > 4 mm day⁻¹, exceeding 7 mm day⁻¹ in the weakest formulations). Overall, this machine learning–based framework provides a robust and

operational alternative for daily ETo estimation in eastern Morocco, supporting efficient irrigation scheduling, improved agricultural water management, enhanced crop productivity, and sustainable agriculture in arid and semi-arid regions.

KEYWORDS

crop productivity, evapotranspiration, irrigation, machine learning, Morocco, sustainable agriculture, water management

1 Introduction

Water scarcity and climate change are critically challenging agriculture in arid and semi-arid regions. In Eastern Morocco, prolonged droughts, erratic rainfall patterns, and rising temperatures have disrupted hydrological cycles and exacerbated water shortages (Abdullahi, 2025). Agriculture consumes about 88% of Morocco's freshwater resources, so declining precipitation (~40% drop since 2015) and groundwater over-exploitation have pushed water reserves to critical lows (dam reservoirs ~28% of capacity in 2024) (DIAEA, 2024). This water crisis threatens food security and rural livelihoods (Depeursinge et al., 2010). Eastern Morocco's farms, spanning from semi-arid plateaus to oasis zones, are predominantly rain-limited and heavily reliant on irrigation (WMO, 2023; Morocco - Water, 2023). With scarce water supplies posing serious risks to crop production, there is an urgent need for precise and judicious irrigation management. Suboptimal irrigation practices, whether over- or under-watering, not only waste water but also harm crops. Excessive watering leads to anaerobic conditions and root asphyxiation, nutrient leaching, and increased susceptibility to waterborne diseases (Kozłowski, 1984; Tasa et al., 2025). Whereas insufficient irrigation induces plant stress, impairs physiological functions, and ultimately reduces yields (Atta et al., 2023). Thus, determining the exact water requirement of crops at different growth stages is crucial for maximizing agricultural output while minimizing water wastage (Bayabil et al., 2020). To achieve this, it is essential to rely on accurate indicators of crop water demand. One of the most critical is evapotranspiration (ET). This physical process includes the combined movement of water into the atmosphere through two main processes: the evaporation of soil surface moisture and the transpiration of water occurring in plant tissues, with transpiration alone accounting for 50–70% of total ET (Allen et al., 1998; Tausif et al., 2023).

Accurate estimation of ET remains essential in developing effective irrigation schedules since it directly reflects the amount of water that needs to be replenished in the soil to meet the atmospheric demand and physiological needs in crops (Wanniarachchi and Sarukkalgige, 2022). Among the different ET forms, reference evapotranspiration (ETo) represents the rate of water loss from an ordinance-irrigated reference crop (Quej et al., 2022). ETo is considered the most convenient and commonly used method in the estimation of crop evapotranspiration (ETc), as it enables a standardized reference applicable in different crop types with the use of specific crop coefficients (Allen et al., 1998). ETo can be assessed through both direct and indirect approaches. Direct methods like lysimeters can provide accurate measurements, but they are often costly, complicated, and rarely used outside the experimental fields (Hamdaoui et al., 2024). Consequently, multiple empirical and semi-empirical models have been developed to estimate ETo based on meteorological data (Haris

et al., 2025). The Penman-Monteith FAO-56 equation (ETo_{PM}) remains the global benchmark due to its strong physical basis and high accuracy in a wide range of climatic conditions (Allen et al., 1998; Yin et al., 2008). However, the ETo_{PM} formula requires numerous meteorological parameters, often unavailable in remote or resource-limited areas (Koç and Can, 2023). This data scarcity has spurred the use of simpler empirical equations that need fewer inputs—such as Hargreaves, Blaney–Criddle, Priestley–Taylor, Turc, or Jensen–Haise—to approximate ETo (Rajput et al., 2023b). While these methods can provide rough estimates with minimal data, their simplicity often comes at the cost of substantial accuracy loss (Valipour, 2015). Many studies have confirmed that such traditional equations generally underperform compared to ETo_{PM}, especially under the variable and extreme climate conditions of arid regions (Ghiat et al., 2021; Katsoulas et al., 2019; Nagappan et al., 2025). This gap highlights a need for alternative approaches that maintain high accuracy while using limited inputs.

In recent years, artificial intelligence (AI) has introduced transformative tools to agriculture (Hamdaoui et al., 2025). Machine learning (ML), a key subset of AI, allows models to learn from data and identify complex patterns without relying on predefined equations (Khine, 2024; Al Kaddouri et al., 2025). Once trained, ML models can accurately predict and analyze agricultural processes such as crop growth, water use, and weather interactions, which are often too nonlinear for traditional empirical methods (Singla et al., 2024). Their ability to handle incomplete datasets and varying environmental conditions makes them particularly valuable for irrigation modeling in arid and semi-arid regions. A growing body of research worldwide has applied ML to ETo estimation under different climatic settings, confirming the strong capacity of algorithms such as Support Vector Machines (SVM), Random Forests (RF), Neural Networks, Long Short-Term Memory networks (LSTM), and Gradient Boosting (GB) models (Sabanci et al., 2023; Mohammadi et al., 2024). These data-driven models often surpass conventional methods in diverse climates by capturing complex nonlinear interactions among meteorological variables and adapting effectively to local climatic conditions. Recent studies in various climate conditions showed ML models providing more accurate daily ETo predictions than equations like Hargreaves or Blaney–Criddle, especially when certain climate inputs are missing. Moreover, ML models can be trained to be computationally efficient for deployment on devices, enabling integration with real-time irrigation systems (Rajput et al., 2023a; Shaloo et al., 2024; Rajput et al., 2024; Shaloo et al., 2026).

Despite these advances, very few studies have explored ML-based ETo estimation under Moroccan conditions, particularly in the harsh arid climate of Eastern Morocco. Therefore, it remains unclear whether ML can consistently outperform traditional empirical approaches while maintaining strong generalization under limited-input conditions. In addition, many existing ML studies rely on

complete meteorological datasets, which reduces their applicability in data-scarce farming systems. The novelty of the present work lies in developing and validating a minimal-input, computationally efficient ML framework for daily ETo estimation in Eastern Morocco, and systematically benchmarking it against widely used simplified empirical equations under a strict chronological evaluation strategy. Accordingly, this study aims to: (1) Develop and compare three ML models (SVR, RF, and XGBoost) for daily ETo estimation in Eastern Morocco under multiple reduced-input scenarios, using $E_{To_{PM}}$ as the reference target; (2) Compare the selected ML-minimal-input configurations against widely used simplified empirical ETo equations to evaluate whether ML provides a statistically and operationally meaningful improvement under data-scarce conditions; and (3) Deliver a practical and resource-efficient ETo prediction framework that can support precision irrigation scheduling and improve crop water management in arid and semi-arid regions, combining high accuracy with low computational cost. The proposed framework also provides a basis for future integration into IoT-enabled irrigation systems, facilitating real-time decision-making and contributing to sustainable agriculture and water security in Morocco.

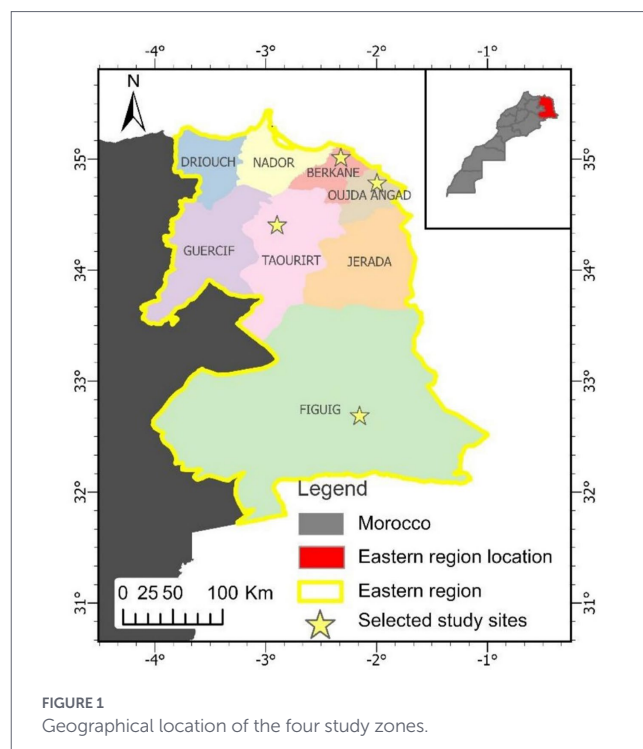
2 Materials and methods

2.1 Study zones

The present investigation was undertaken in the oriental region of Morocco. This region, encompassing approximately 82,800 km², extends from the Mediterranean coastline to the border with Algeria and includes a range of environments from relatively humid northern areas to arid desert regions in the south. It is characterized as a geographically diverse zone exhibiting a semi-arid to arid climate, marked by low annual precipitation, elevated ET rates, and recurrent drought occurrences. To adequately reflect the climatic and agronomic variability within the oriental region, four representative zones were identified for data collection and model formulation: Oujda, Berkane, Taourirt, and Figuig (Figure 1). The geographic coordinates and principal climatic features (averages for 2024) of these selected zones are presented in Table 1. This particular area was selected because of its significance in agriculture, scarcity of water resources, and substantial interannual climatic fluctuations.

2.2 Agricultural importance of the study zones

Eastern Morocco constitutes one of the country's main agricultural basins. Despite a generally arid to semi-arid climate, this region is home to diverse agricultural systems, primarily focused on arboricultural, horticultural, and cereal crops. Berkane stands out as a major center for citrus production, relying heavily on irrigation, which makes accurate ETo estimation essential for water management (Ahmed Mator et al., 2025). Oujda and Taourirt are dominated by agrosylvo-pastoral systems in which plant production is based on cereals, legumes, as well as arboricultural crops, often grown in rain-fed conditions or with supplemental irrigation. These systems are based on a tiered agriculture where phoenicultural, horticultural, and forage crops coexist (Madani and Tariq, 2006). The selected zones in this study stand out as major agricultural basins, contributing



significantly to crop production, both regionally and nationally, as well as in export sectors.

2.3 Data sources and preprocessing

2.3.1 Data sources

To build a comprehensive dataset, we gathered daily climatic records covering maximum and minimum temperatures (T_{max} and T_{min} °C), relative humidity (RH %), wind speed at 2 m (U_2 , m s⁻¹), and solar radiation (R_{sn} , MJ m⁻² day⁻¹) from 2001 to 2025 for the four study sites within the selected zones, their geographic coordinates are as in Table 2. Data collection relied on three main sources: (i) the NASA POWER platform, (ii) meteorological stations in agricultural exploitations, and (iii) airport weather stations. However, despite these efforts, the data collection process faced significant challenges. Many local stations were either inoperative or only partially equipped, or lacking certain climatic sensors, leading to gaps in measurement. Some records were incomplete, discontinuous, or difficult to access due to administrative restrictions. These obstacles highlight the lack of reliable and continuous climatic monitoring infrastructure in the Moroccan oriental region.

2.3.2 Data preprocessing and temporal splitting strategy

All data were compiled into site-specific Excel datasets, then underwent a preprocessing phase that included cleaning missing/outlier values, merging variables from different sources into unified datasets, and normalizing predictors (important for SVR). Daily ETo was then computed using the $E_{To_{PM}}$ (Allen et al., 1998), which served as the target variable for model development. Since ETo is a time-dependent variable, splitting step using a random train-test split would create information leakage, leading to overly optimistic performance estimates. Therefore,

TABLE 1 Geographic coordinates and main climatic characteristics (2024 averages) of the four study zones.

Regions	Latitude	Longitude	Avg. TMax (°C)	Avg. TMin (°C)	Avg. RH (%)	Avg. Precipitation (mm/day)	Avg. WS (m/s)	Radiation (MJ/m ² /day)
Oujda	34.63°N	-1.89°W	26.24	12.04	50.42	0.56	2.87	18.43
Berkane	34.99°N	-2.26°W	26.02	15.63	61.62	0.72	2.92	18.52
Taourirt	34.41°N	-2.88°W	29.30	14.73	49.7	0.59	2.48	18.82
Figuig	32.11°N	-1.23°W	28.92	14.22	34.01	0.21	2.94	20.28

TABLE 2 Geographic coordinates of the selected sites within the study zones.

Sites	Latitude	Longitude
Site 1 Berkane	34.9994°N	-2.2698°W
Site 2 Oujda	34.6369°N	-1.8989°W
Site 3 Figuig	32.0972°N	-1.2295°W
Site 4 Taourirt	34.4145°N	-2.8845°W

a strict chronological splitting strategy was adopted to preserve temporal causality and ensure realistic model evaluation. Specifically, data from 2001 to 2018 were used for model training, while the independent period 2019–2025 (≈20% of the dataset) was reserved for testing. This temporally consistent evaluation reflects real operational conditions, where future ETo must be predicted exclusively from past climatic observations. Accordingly, all performance metrics reported in the Results section were derived from this chronological train–test scheme.

$$ET_{oPM} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{\gamma + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

Where: ET_{oPM} denotes the daily reference evapotranspiration computed using the Penman–Monteith method (mm day⁻¹); Δ is the slope of the saturation vapor pressure curve as a function of mean daily air temperature (kPa °C⁻¹); R_n is the net radiation at the crop surface (MJ m⁻² day⁻¹); G is the soil heat flux density (MJ m⁻² day⁻¹); γ is the psychrometric constant, which varies with altitude (kPa °C⁻¹); T is the mean daily air temperature (°C); u_2 is the wind speed measured at 2 m height (m s⁻¹); e_s is the saturation vapor pressure (kPa); and e_a is the actual vapor pressure (kPa).

2.4 Machine learning framework

In this study, three ML algorithms were developed to predict daily ETo: Support Vector Regression (SVR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost).

- SVR derived from support vector machines, fits the best regression function within a predefined error margin using kernel functions, enabling it to capture complex relationships between climatic variables and ETo while performing well on relatively small datasets (Smola and Schölkopf, 2004).
- RF, an ensemble method, produces multiple decision trees from random sets of the data and features, then averages their

predictions in order to reduce variance and prevent overfitting. The method proves robust to noise, can model nonlinear interactions, and provides valuable insights into the importance of variables (Breiman, 2001).

- XGBoost, an efficient implementation of gradient boosting, constructs trees in sequence such that each fixes the mistakes of the previous one, and includes regularization in order to prevent overfitting and deals adequately with missing values, making it very suitable for structured type of datasets like climatic records (Chen et al., 2016).

Model development and evaluation were performed in Python 3.11 using the following libraries: scikit-learn (v1.5.0), XGBoost (v2.1.0), pandas (v2.2.2), NumPy (v1.26.0), Matplotlib (v3.8.0), and Seaborn (v0.13.0).

2.5 Conventional empirical ETo estimation models

Five widely used simplified empirical equations were included for comparison with the best ML model, using the ET_{oPM} method as the reference standard, to assess whether ML approaches outperform classical empirical models in estimating daily ETo:

- The Hargreaves-Samani formula (HSF) (Hargreaves and Samani, 1985):

$$ET_{oHSF} = 0.0023 R_a (T_a + 17.8) \sqrt{(T_{max} - T_{min})}$$

Where: ET_{oHSF} represents the ETo value computed using the HSF (mm day⁻¹). T_{min} , T_{max} , and T_a refer to the minimum, maximum, and mean air temperatures (°C). R_a is the extraterrestrial radiation (MJ m⁻² day⁻¹), computed following (Allen et al., 1998):

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

Here, G_{sc} is the solar constant (0.082 MJ m⁻² min⁻¹), ω_s is the sunset hour angle (rad), δ is the solar declination (rad), and d_r is the inverse relative Earth–sun distance.

b The Makkink formula (MF) (Makkink, 1957):

$$ET_{oMF} = 0.61 \left(\frac{\Delta}{\Delta + \gamma} \right) \left(\frac{R_s}{\lambda} \right) - 0.12$$

Where: ET_{oMF} is the daily ETo estimated using MF (mm day⁻¹). λ is the latent heat of vaporization (MJ kg⁻¹). R_s is the incoming shortwave solar radiation (MJ m⁻² day⁻¹). Other parameters are as defined above.

c The Priestley-Taylor formula (PTF) (Gong et al., 2021):

$$ET_{oPTF} = 1.26 \left(\frac{\Delta}{\Delta + \gamma} \right) \left(\frac{R_n - G}{\lambda} \right)$$

where ET_{oPTF} is the daily ETo estimated using the (PTF) (mm day⁻¹), and the remaining variables are as previously defined.

d Turc formula (TF) (Turc, 1961):

$$ET_{oTF} = 0.0133 \frac{T_a}{T_a + 15} (R_s + 50) \quad R_H > 50\%$$

$$ET_{oTF} = 0.0133 \frac{T_a}{T_a + 15} (R_s + 50) \left(1 + \frac{50 - R_H}{70} \right) \quad R_H < 50\%$$

Where: ET_{oTF} represents the ETo value computed using the TF (mm day⁻¹). R_H is the relative humidity (%). Other parameters are the same as the previous equations.

e The Jensen-Haise formula (JHF) (Jensen and Haise, 1963):

$$ET_{oJHF} = 0.025 \left(\frac{R_s}{\lambda} \right) (T_a - T_x)$$

Where: ET_{oJHF} represents the ETo value computed using the JHF (mm day⁻¹). 0.025 is a temperature constant. T_x is equal to -3. The other parameters are as in the previous equations.

2.6 Feature importance analysis

Feature importance was estimated for each regression algorithm to quantify the contribution of the input variables to ETo prediction. The approach depended on the nature of the algorithm. For the RF and XGBoost, importance was calculated using the mean decrease in impurity (MDI), also referred to as Gini importance. For each split in a decision tree, the reduction in squared error (MSE) is attributed to the variable used. The total

importance of a feature is the normalized sum of these reductions across all trees:

$$FI_j = \frac{\sum_{t=1}^T \sum_{s \in S_{j,t}} \Delta i(s,t)}{\sum_{t=1}^T \sum_{s \in S_t} \Delta i(s,t)}$$

where FI_j is the importance of feature j , $\Delta i(s,t)$ is the reduction of impurity at split s in tree t , $S_{j,t}$ is the set of splits using feature j , and T is the number of trees.

Since SVR does not provide direct feature importance, we used permutation importance. This method evaluates the change in model performance (measured by RMSE) when the values of one feature are randomly permuted while others remain unchanged. The importance of a feature j is then:

$$PI_j = RMSE_{perm(j)} - RMSE_{baseline}$$

where $RMSE_{perm(j)}$ is the model error after permuting feature j , and $RMSE_{baseline}$ is the original error.

Based on the feature importance analysis, five feature combination scenarios were designed to evaluate the impact of different input variables on ETo prediction performance. The details of these scenarios are presented in the Results section.

2.7 Model training and validation procedure

Following the strict chronological split described in Section 2.3, models were trained using the historical period 2001–2018 and evaluated on the independent testing period 2019–2025. Hyperparameters were optimized using a grid-search strategy, selecting the combinations that minimized the Root Mean Square Error (RMSE). Final models were retrained using the optimal hyperparameters and assessed on the unseen test period using the performance metrics defined in Section 2.8. The optimal hyperparameter settings for each algorithm are summarized in Table 3.

TABLE 3 Optimized hyperparameters for machine learning models.

Model	Hyperparameter	Tested values
XGBoost	n_estimators	[100, 300]
	learning_rate	[0.05, 0.1]
	max_depth	[3, 5]
RF	n_estimators	[100, 200]
	max_depth	[10, 20]
	min_samples_split	[2, 5]
SVR	C (Regularization)	[10]
	epsilon	[0.1]

2.8 Performance metrics

Model evaluation and comparison were conducted using multiple statistical indicators computed separately for the training and testing phases:

- a Root Mean Square Error (RMSE): RMSE is the square root of MSE and provides error values in the same unit as the predicted variable (ETo in mm/day). It offers an interpretable measure of the model's predictive performance.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- b Mean Absolute Error (MAE): MAE evaluates the average magnitude of the prediction errors without considering their direction (positive or negative). It provides a straightforward measure of the absolute discrepancies between predicted and observed ETo values.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- c Mean Absolute Percentage Error (MAPE): MAPE expresses prediction accuracy as a percentage by comparing absolute errors to actual observed values. It is particularly useful for interpreting model performance in relative terms, though it may be sensitive when actual values approach zero.

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- d Coefficient of Determination (R^2): R^2 quantifies how well the predictions explain the variance of the observed ETo values. A value close to 1 indicates strong explanatory power.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- e Execution Time (s): The time required for model training and prediction was recorded to assess computational efficiency, an important factor for large-scale or real-time applications.
- f Memory Usage (MB): Memory consumption during training and testing was also monitored, as resource efficiency is critical for future deployment on low-power or embedded devices.

The overall workflow of the proposed modeling approach is illustrated in Figure 2.

2.9 Code and data availability

All datasets used in this study were obtained from NASA POWER,¹ Moroccan airport weather stations, and private farm

¹ <https://power.larc.nasa.gov/>

networks in the Oriental region. The datasets and Python scripts used for model training and analysis are available upon reasonable request or via the corresponding author.

3 Results

3.1 Model's performance with all features

Using the complete set of input features, all ML models exhibited excellent predictive performance for ETo estimation during both training and testing phases. High coefficients of determination ($R^2 = 0.99$) were obtained for XGBoost, RF, and SVR, indicating a strong ability to reproduce the variability of reference ETo values (Table 4). The evaluation of error magnitude revealed notable differences among models: SVR achieved the lowest RMSE values (0.047 mm day⁻¹ in training and 0.053 mm day⁻¹ in testing), followed by XGBoost (0.048 and 0.103 mm day⁻¹, respectively), whereas RF showed a larger increase in testing error (RMSE \approx 0.2 mm day⁻¹), suggesting reduced generalization despite high R^2 values (Figure 3). The comparison between training and testing metrics indicates that SVR and XGBoost achieved stable performance with limited overfitting, while RF exhibited greater sensitivity to unseen data. Scatter plots of predicted versus reference ETo further corroborated these findings, with predictions from SVR and XGBoost closely clustered around the 1:1 line across the full ETo range, whereas RF displayed a slightly wider dispersion during testing (Figure 4). In addition to predictive accuracy, computational efficiency differed among models: XGBoost and SVR required shorter execution times during training, whereas RF exhibited the highest computational cost, reflecting its ensemble-based structure. Similarly, memory usage varied across models, with RF requiring substantially higher memory resources compared to XGBoost and SVR, which demonstrated more compact memory footprints (Table 4).

3.2 Feature importance

The feature importance score among the three models (RF, SVR, and XGBoost) consistently ranked Rsn first in affecting ETo estimation, with cumulative contribution in all models exceeding 80%. The second and third most influential variables were Tmax and RH, which ranged from the interval of 20 to 38% among the three models. The least contributors were the wind speed (U2), with minimum temperature (Tmin), which contributed marginally (Figure 5). From the above findings, the exploration in the subsequent section concentrated on Rsn, Tmax, and RH as the main inputs in the derivation of the reduced-feature models.

3.3 Reduced feature combinations results

To evaluate the practical impact of the dominant variables (Rsn, Tmax, and RH) and assess the models' capacity to estimate ETo with minimal inputs, each algorithm was retrained and tested on four scenario combinations of input sets (Table 4).

3.3.1 Performance with single feature (Rsn)

When solar radiation (Rsn) was used as the sole input variable, given its high relevance identified by the feature importance analysis, all three models exhibited moderate predictive performance relative to the reference ET_{oPM} during both training and testing phases. During testing, the coefficients of determination derived from the R^2 bar charts (Figures 6–8) reached 0.585 for XGBoost, 0.576 for RF, and 0.538 for SVR, indicating limited agreement between predicted and observed values when relying exclusively on Rsn. The corresponding error metrics reported in Table 4 show relatively high prediction errors for this configuration. Testing RMSE ranged from $2.138 \text{ mm day}^{-1}$ (XGBoost) to $2.314 \text{ mm day}^{-1}$ (RF), while training RMSE values exceeded 1.8 mm day^{-1} for all models. Among the evaluated algorithms, XGBoost achieved the lowest testing RMSE and slightly lower MAE and MAPE values compared to RF and SVR. From a computational perspective, all models required relatively low

resources under the single-feature scenario. XGBoost was the most efficient, achieving the shortest training time (7.79 s) and the lowest peak memory usage (0.12 MB), compared with RF (9.98 s; 19.95 MB) and SVR (8.70 s; 0.89 MB) (Table 4).

3.3.2 Performance with two features

The inclusion of a second input variable resulted in a substantial improvement in model performance compared to the single-feature configuration. In scenario 3 (Rsn + Tmax), high predictive accuracy was obtained for all models during both training and testing phases. During training, R^2 values reached 0.9956 for XGBoost, 0.9914 for RF, and 0.9473 for SVR. In the testing phase, the corresponding R^2 values remained high, attaining 0.936 for XGBoost, 0.934 for RF, and 0.931 for SVR, as illustrated in the R^2 comparison bar charts (Figures 6–8). This improvement was accompanied by a marked reduction in prediction errors. As shown in Table 4, testing RMSE values decreased to

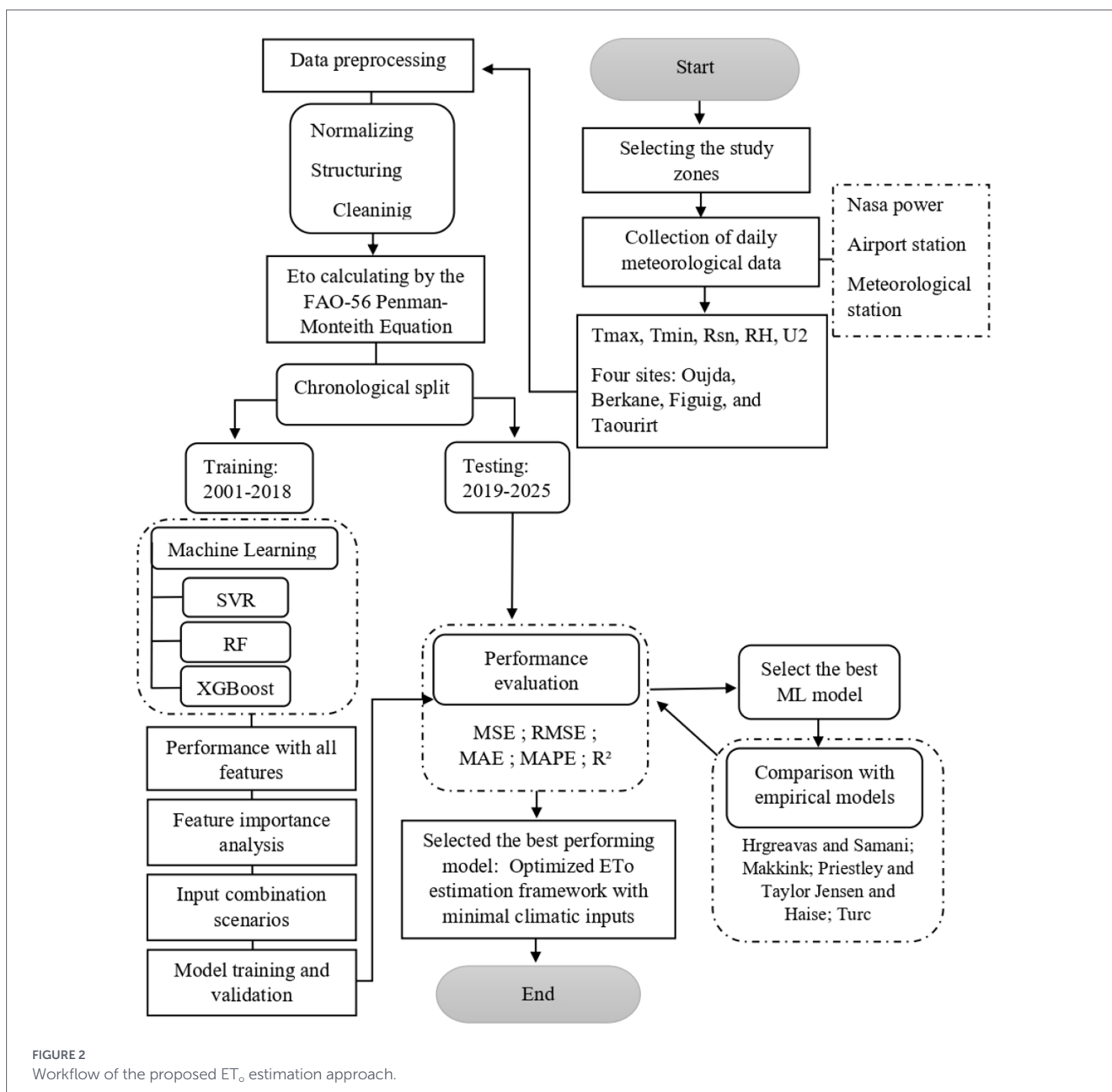
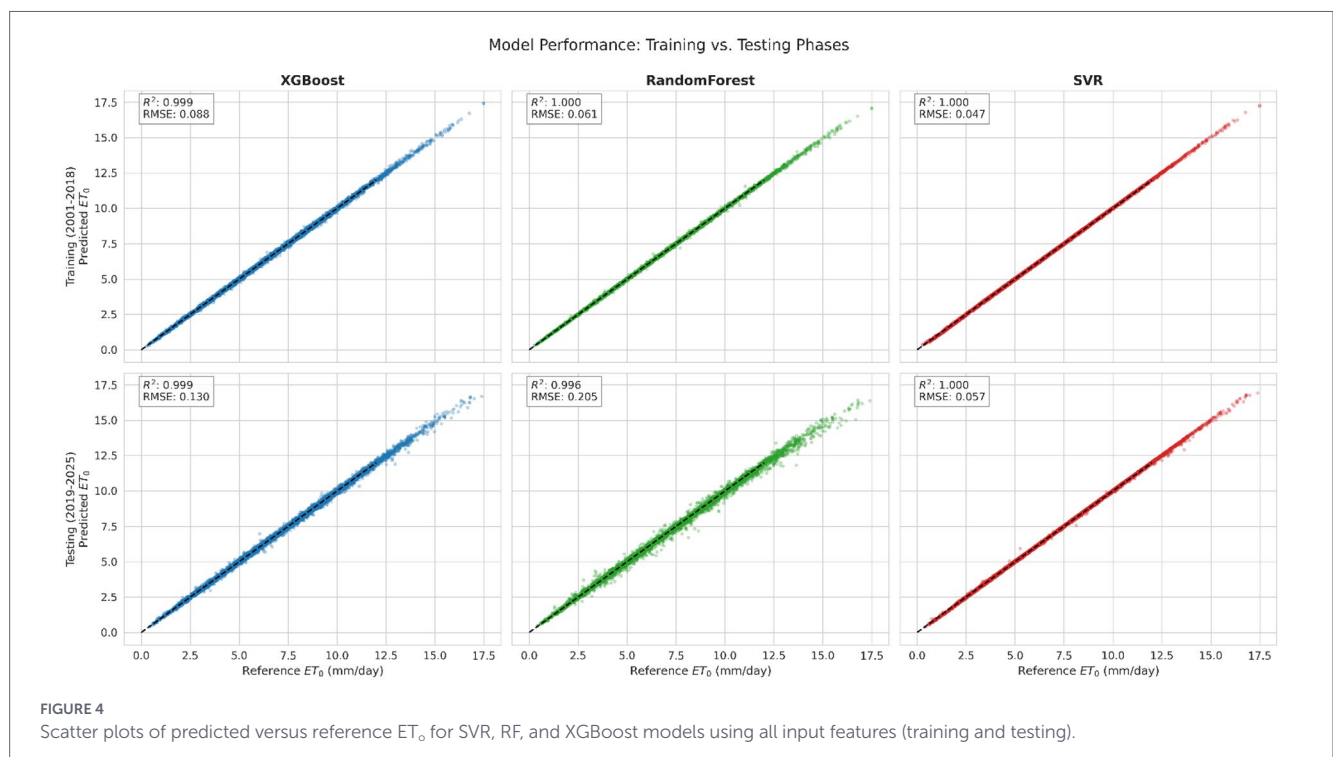


FIGURE 2 Workflow of the proposed ET_o estimation approach.

TABLE 4 Performance metrics of Random Forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost) under different input feature scenarios.

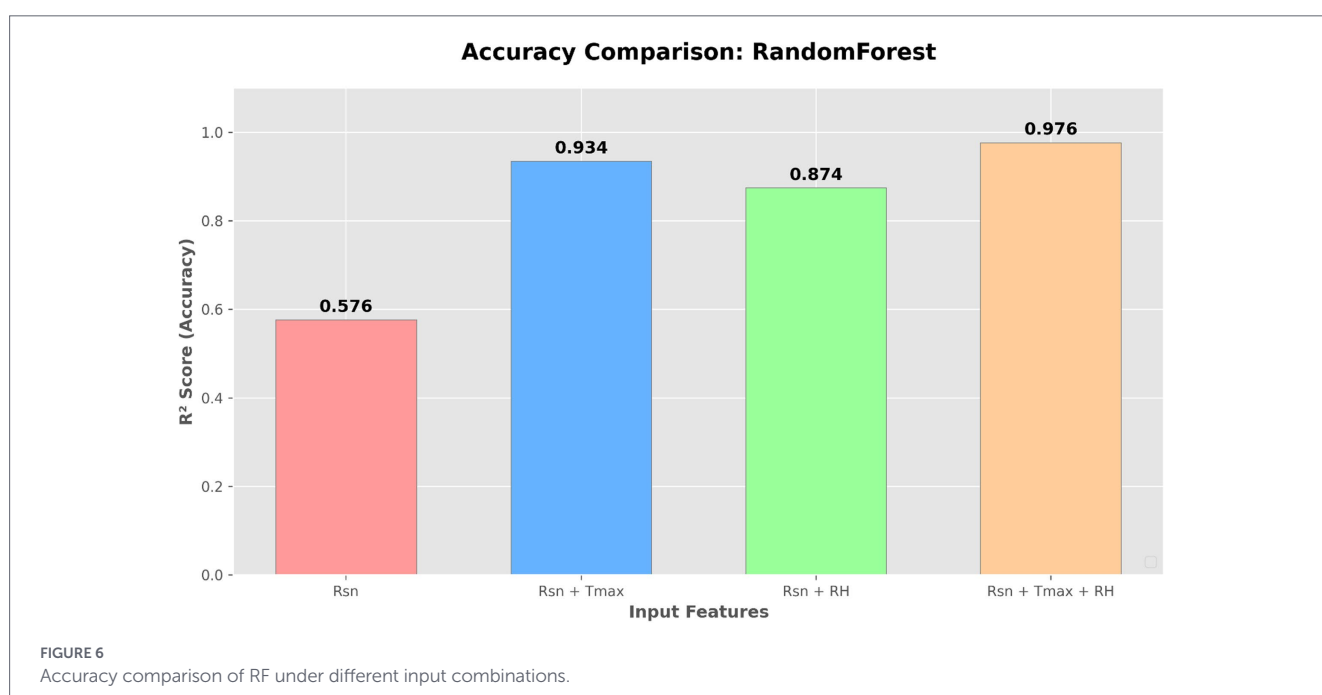
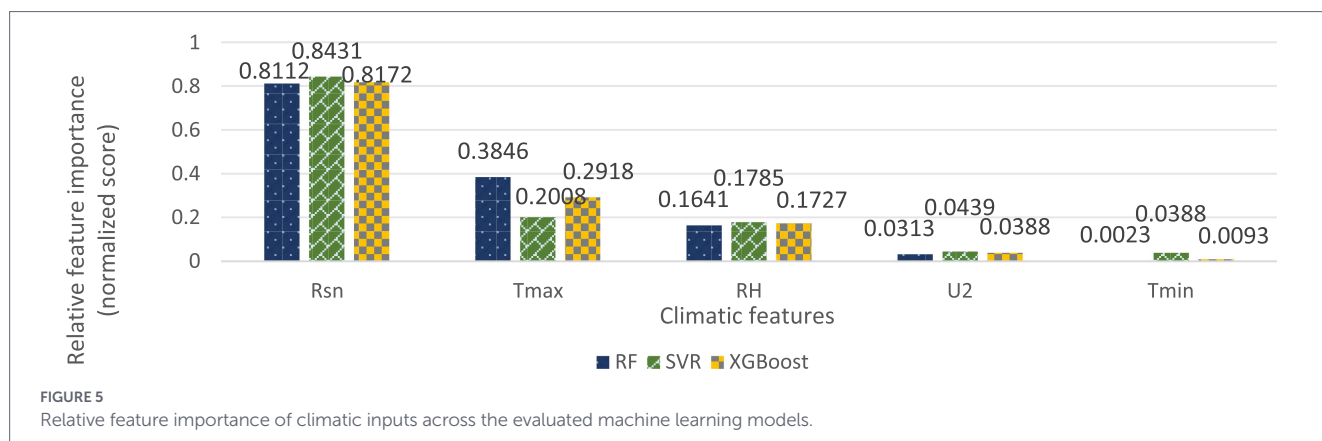
Model	Input scenarios	MSE train	MSE test	RMSE train	RMSE test	R^2 train	R^2 test	MAE train	MAE test	MAPE train	MAPE test	Execution time (s)	Memory (MB)
RF	Scenario 1: all the features	0.0037	0.0381	0.0607	0.1952	0.9996	0.9965	0.0401	0.1268	0.76	2.04	9.43	24.08
SVR		0.0023	0.0029	0.0477	0.0537	0.9998	0.9997	0.0398	0.0417	0.95	0.81	0.6	3.09
XGBoost		0.0024	0.0107	0.0488	0.1035	0.9998	0.999	0.037	0.0718	0.74	1.21	2.96	0.21
RF	Scenario 2: Rsn	3.317	5.3566	1.8213	2.3144	0.6833	0.576	1.4028	1.7841	26.19	28.46	9.98	19.95
	Scenario 3: Rsn, Tmax	0.0896	0.8113	0.2994	0.9007	0.9914	0.934	0.2053	0.6345	3.87	10.55	10.26	22.83
	Scenario 4: Rsn, RH	0.1815	1.576	0.426	1.2554	0.9827	0.874	0.3204	0.965	5.43	14.92	10.27	23.24
	Scenario 5: Rsn, RH, Tmax	0.0258	0.27	0.1608	0.5196	0.9975	0.976	0.0983	0.3284	1.83	5.35	9.32	23.69
SVR	Scenario 2: Rsn	4.1374	5.0606	2.0341	2.2496	0.605	0.5383	1.5427	1.667	26.45	26.75	8.7	0.89
	Scenario 3: Rsn, Tmax	0.5517	0.757	0.7427	0.87	0.9473	0.931	0.4921	0.5951	8.94	9.56	9.29	1.59
	Scenario 4: Rsn, RH	1.1238	1.3768	1.0601	1.1734	0.8927	0.8744	0.7946	0.8915	13.6	13.89	8.92	1.65
	Scenario 5: Rsn, RH, Tmax	0.163	0.2346	0.4038	0.4844	0.9844	0.979	0.2433	0.301	4.52	4.86	9.41	1.97
XGBoost	Scenario 2: Rsn	4.1374	5.0606	1.9822	2.1382	0.6249	0.585	1.5616	1.635	29.78	25.23	7.79	0.12
	Scenario 3: Rsn, Tmax	0.5517	0.757	0.6006	0.8782	0.9956	0.936	0.429	0.6066	8.18	10.08	0.33	0.13
	Scenario 4: Rsn, RH	1.1238	1.3768	0.8939	1.2013	0.9237	0.876	0.6873	0.9229	11.96	14.13	4.76	0.12
	Scenario 5: Rsn, RH, Tmax	0.163	0.2346	0.2689	0.5258	0.9931	0.977	0.1838	0.327	3.57	5.31	4.48	0.13



0.878 mm day⁻¹ for XGBoost, 0.90 mm day⁻¹ for RF, and 0.870 mm day⁻¹ for SVR, while training RMSE values remained below 0.75 mm day⁻¹ for all models. XGBoost maintained a favorable balance between accuracy and computational efficiency, requiring only 0.33 s of training time and 0.13 MB of memory, compared to RF (10.26 s and 22.83 MB) and SVR (9.29 s and 1.59 MB).

In the Rsn + RH scenario, intermediate predictive performance was obtained for all models during both training and testing phases. Training R^2 values ranged from 0.8927 for SVR to 0.9827 for RF, while testing R^2 values decreased to 0.876 for XGBoost, 0.874 for SVR, and

RF. Correspondingly, testing RMSE values increased to 1.173 mm day⁻¹ for SVR, 1.255 mm day⁻¹ for RF, and 1.201 mm day⁻¹ for XGBoost (Table 4). Although this configuration improved prediction accuracy relative to the single-feature case, its performance remained consistently lower than that obtained with the Rsn + Tmax combination across all three models. From a computational standpoint, XGBoost again exhibited lower training time (4.76 s) and minimal memory usage (0.12 MB), whereas RF showed substantially higher memory consumption exceeding 23 MB, and SVR required intermediate computational resources.



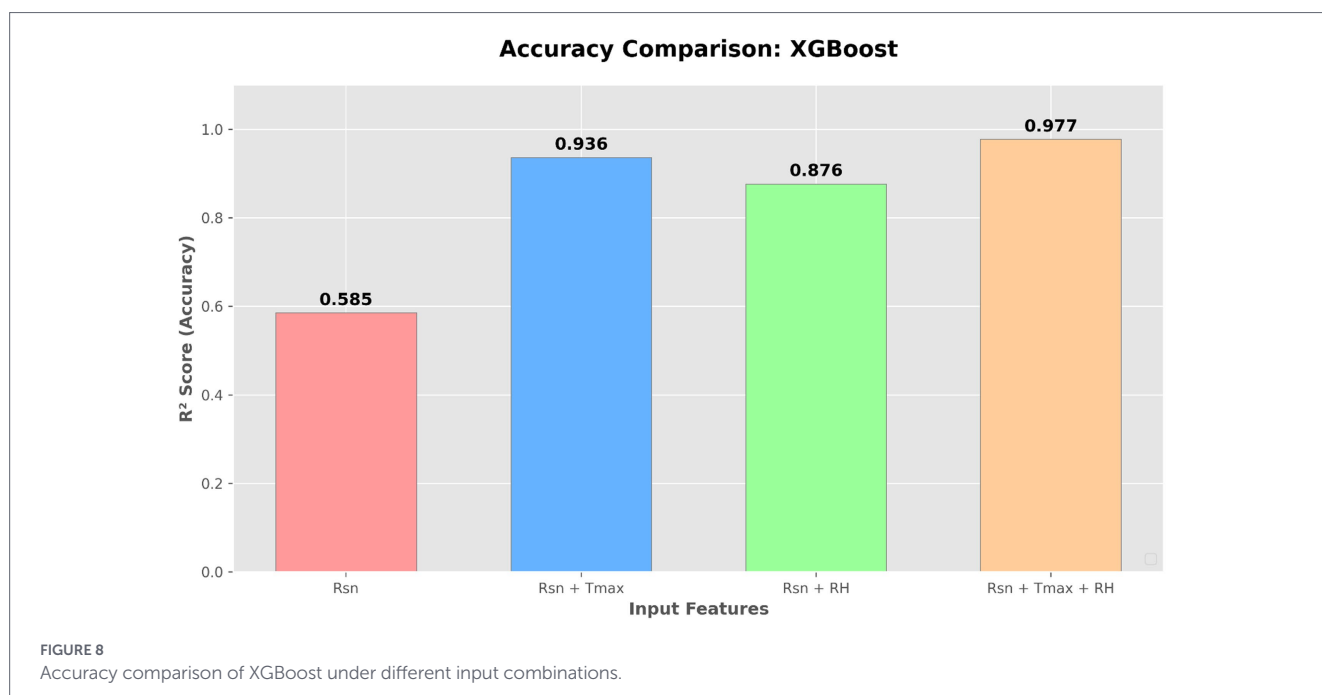
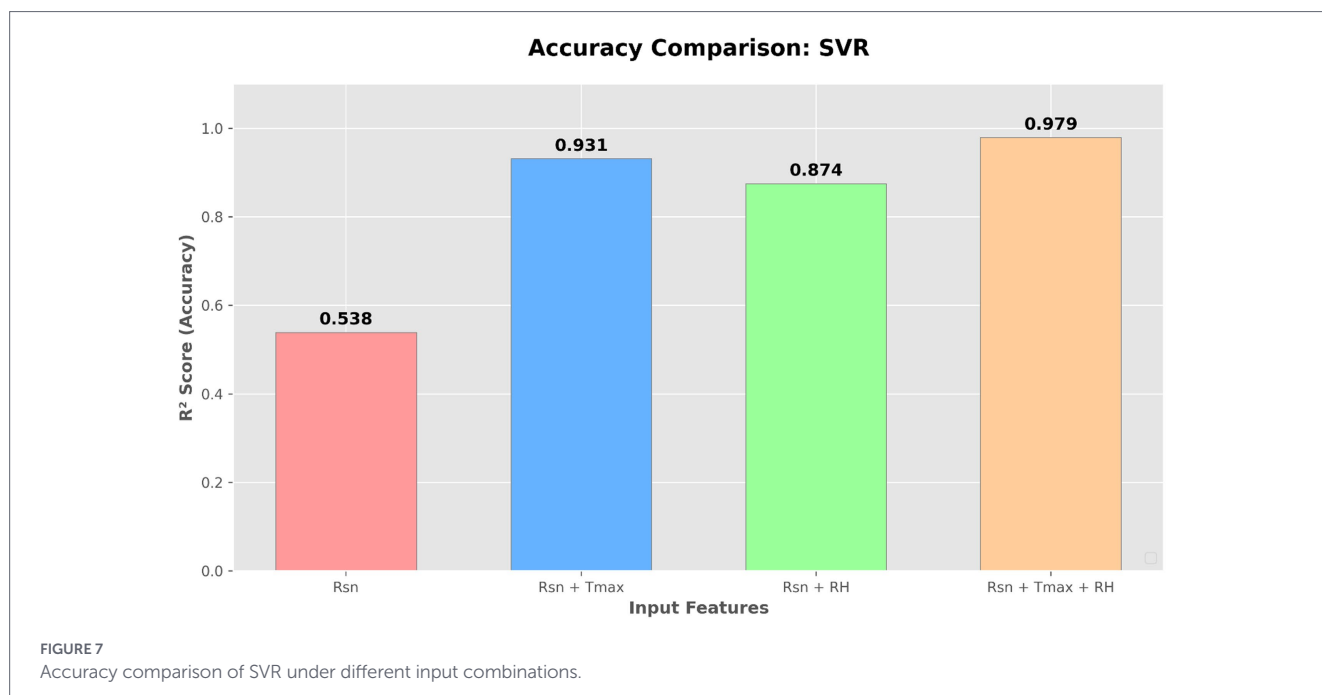
3.3.3 Performance with three-features: Rsn, Tmax, and RH

The integration of three input variables in Scenario 5 (Rsn + Tmax + RH) resulted in the highest predictive accuracy among all reduced-input configurations for the three models. During the training phase, very high coefficients of determination were obtained, with R^2 values of 0.9931 for XGBoost, 0.9975 for RF, and 0.9844 for SVR. During testing, XGBoost and RF achieved similarly high performance, with R^2 values of 0.977 and 0.976, respectively, while SVR showed a marginally higher R^2 (0.979), as illustrated in the R^2 comparison bar charts (Figures 6–8). The error metrics reported in Table 4 confirm the high predictive accuracy of this configuration. During training, RMSE values were 0.2689 mm day⁻¹ for XGBoost, 0.1608 mm day⁻¹ for RF, and 0.4038 mm day⁻¹ for SVR. In the testing phase, XGBoost achieved an RMSE of 0.5258 mm day⁻¹, slightly higher than RF (0.5196 mm day⁻¹) and SVR (0.4844 mm day⁻¹).

Testing MAE values were 0.327 mm day⁻¹ for XGBoost, 0.3284 mm day⁻¹ for RF, and 0.301 mm day⁻¹ for SVR, while testing MAPE values remained low, at 5.31, 5.35, and 4.86%, respectively. From a computational perspective (Table 4), XGBoost showed the lowest training time among the tree-based models, requiring 4.48 s with a peak memory usage of 0.13 MB. RF exhibited a longer training time of 9.32 s and substantially higher peak memory usage of 23.69 MB. SVR required 9.41 s of training time with a peak memory usage of 1.97 MB.

3.3.4 Selection of the optimal machine learning model

Based on the comparative evaluation of the three ML models under different scenarios, XGBoost showed the most reliable performance, particularly under the two- and three-input configurations. It achieved the best compromise between predictive accuracy and

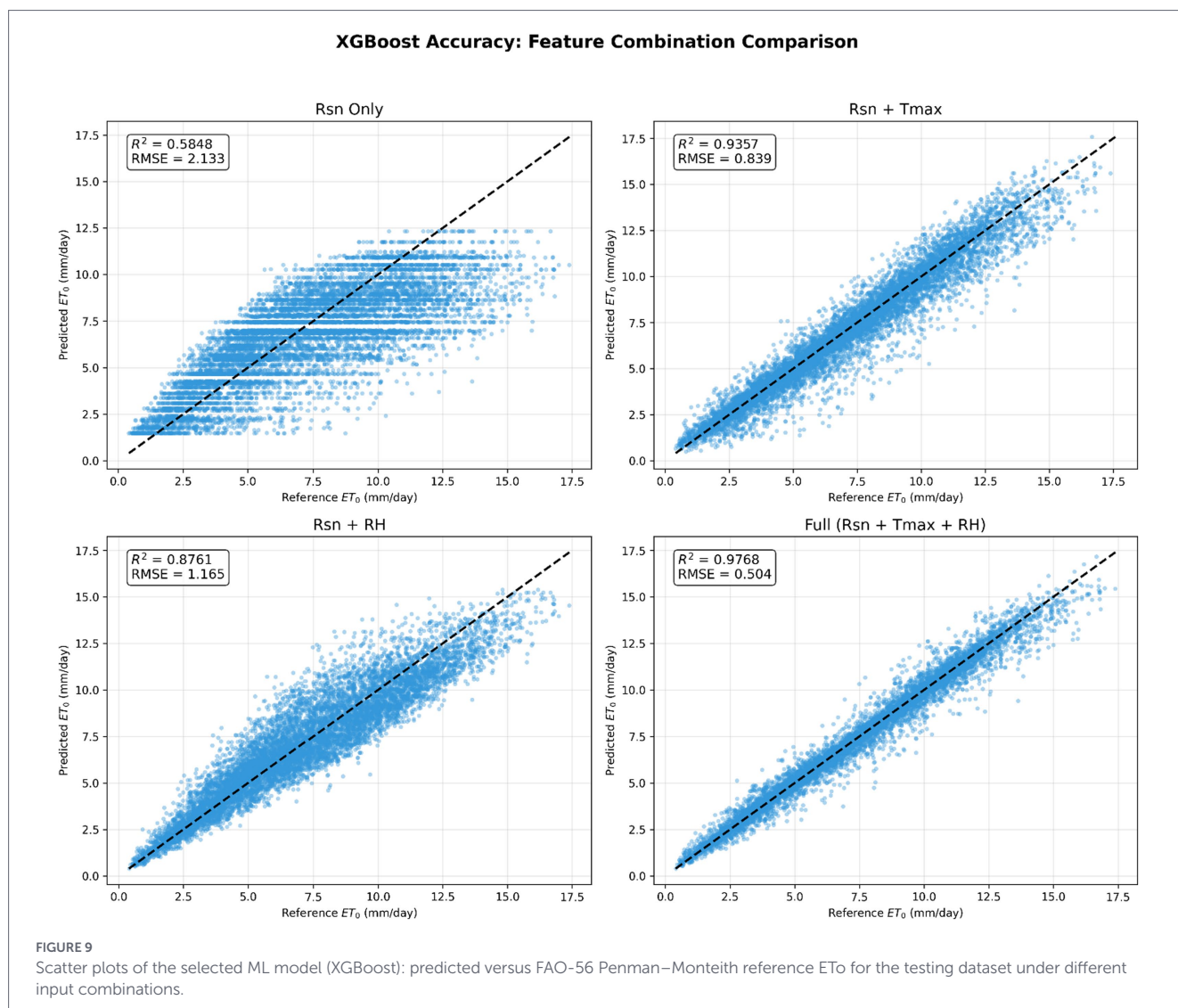


computational efficiency, including training time and memory usage, as summarized in the performance table. This result is further supported by the scatter plots obtained for the testing dataset, which show a strong agreement between XGBoost-predicted ETo values and the reference ETo_{PM} across all feature combinations, with predictions closely distributed around the 1:1 line, particularly for the scenarios using two inputs (Rsn + Tmax) and three inputs (Rsn + Tmax + RH) (Figure 9). In addition, the time-series analysis performed on the testing samples demonstrates that XGBoost accurately reproduces the temporal dynamics of ETo, capturing both short-term variability and peak values without noticeable systematic bias or temporal lag (Figure 10). These visual analyses are fully consistent with the statistical performance metrics and confirm the robustness and

generalization capability of XGBoost. Based on these results, the XGBoost model configured with two inputs (Rsn and Tmax) and three inputs (Rsn, Tmax, and RH) was selected for comparison with traditional empirical ETo estimation models in the next section.

3.4 Comparison of XGBoost models with empirical ETo methods

The performance of the proposed XGBoost models was compared with the widely used empirical ETo equations presented in Section 2.5. Based on the previous analysis of input combinations, the two XGBoost configurations that exhibited the highest predictive accuracy were retained for comparison: a reduced-input model using two



variables (Rsn + Tmax), hereafter referred to as XGBoost-2, and an extended-input model using three variables (Rsn + Tmax + RH), referred to as XGBoost-3. The comparison was conducted using the global statistical indicators described in Section 2.7, aggregated across all studied regions (Berkane, Figuig, Oujda, and Taourirt), and based on the testing period from 2019 to 2025.

3.4.1 Global performance comparison

As summarized in Table 5 and Figure 11, both XGBoost configurations (XGBoost-3 and XGBoost-2) substantially outperformed all simplified empirical equations across all evaluation metrics. XGBoost-3 achieved the highest overall accuracy, showing excellent agreement with the reference $ET_{0,PM}$, with the lowest global MAE ($0.332 \text{ mm day}^{-1}$), and a negligible bias ($-0.0206 \text{ mm day}^{-1}$), indicating the absence of systematic over- or underestimation. Although slightly less accurate, XGBoost-2 also demonstrated strong predictive capability (global $R^2 = 0.9381$) and remained clearly superior to all empirical approaches. The global RMSE comparison confirms a large performance gap, with XGBoost-3 and XGBoost-2 yielding RMSE values of 0.5077 and $0.8237 \text{ mm day}^{-1}$, respectively, whereas the empirical models exhibited substantially higher errors, with RMSE values ranging from 4.03 mm day^{-1} (Jensen–Haise)

to 7.43 mm day^{-1} (Turc). In addition, all empirical methods yielded negative R^2 values (Table 5), reflecting their limited ability to reproduce the variability of ET_0 relative to the PM reference method. Overall, the XGBoost models reduced estimation errors by approximately 80–90% compared with traditional empirical equations.

3.4.2 Regional consistency of model performance

The regional comparison of RMSE and R^2 values highlights clear differences in model performance across Berkane, Figuig, Oujda, and Taourirt. In all regions, XGBoost-3 consistently achieved the lowest RMSE values and the highest R^2 scores, followed closely by XGBoost-2, indicating stable and high predictive accuracy across the study area. Conversely, all empirical models yielded consistently negative R^2 values across all regions (Figure 12).

4 Discussion

This study employed machine learning approaches to estimate daily reference evapotranspiration with high precision using a

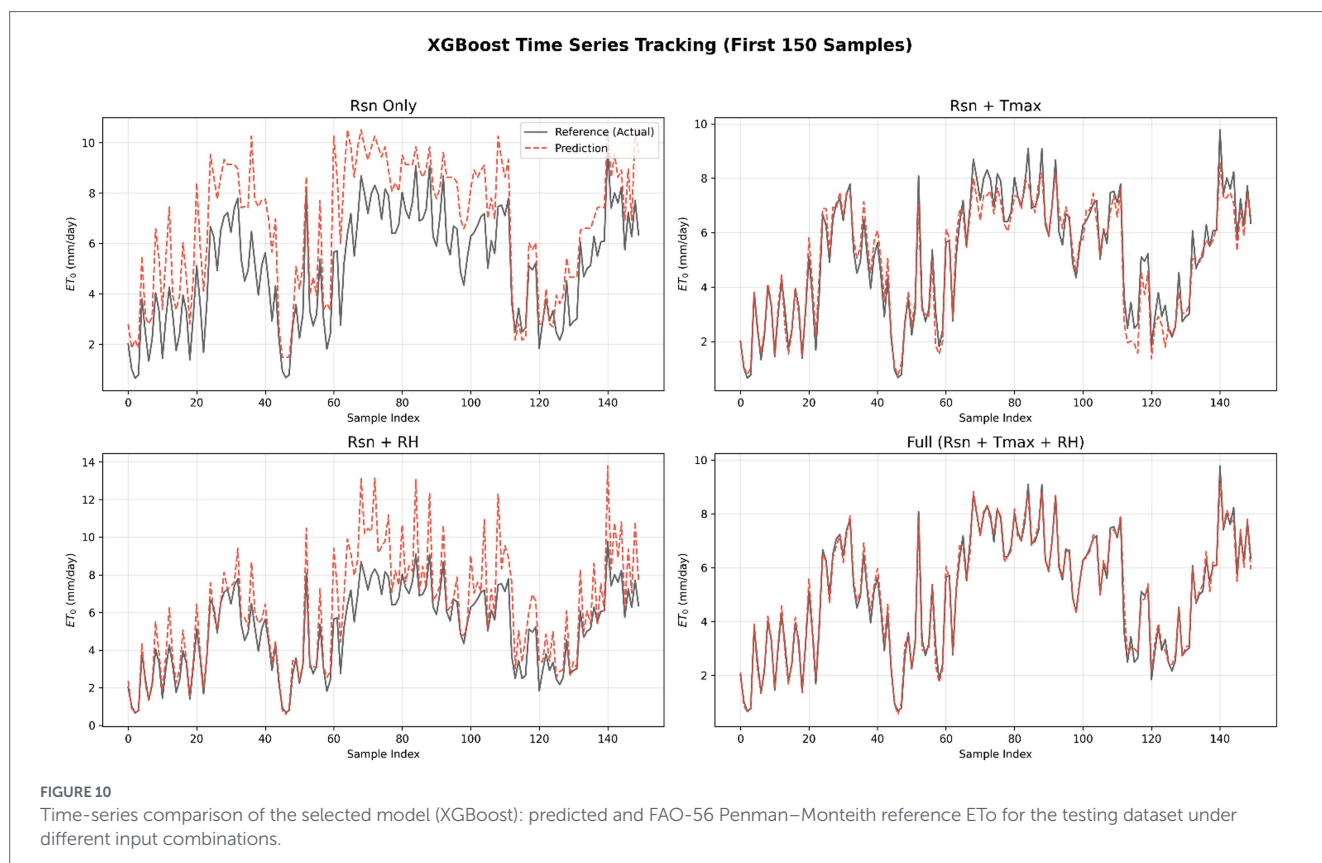


TABLE 5 Comparison of XGBoost (two- and three-input) and empirical models for ETo estimation.

Model	Global RMSE	Global MAE	Global R ²	Global bias
XGBoost 3	0.5077	0.332	0.9765	-0.0206
XGBoost 2	0.8237	0.5881	0.9381	-0.0312
Jensen and Haise (JH)	4.0314	3.2064	-0.4831	-2.2432
Hargreaves Samani (HS)	4.2483	3.392	-0.6469	2.2618
Makkink	5.2739	4.2841	-1.5381	2.1859
Priestley Taylor	5.9579	5.029	-2.2392	-0.7447
Turc	7.4341	6.6983	-4.0432	-6.6983

minimal set of climatic inputs, and systematically compared their performance against conventional empirical methods using standard statistical metrics, to enhance water-use efficiency and support sustainable irrigation practices in arid and semi-arid agricultural regions, particularly in Eastern Morocco.

The investigation commenced with an analysis of feature importance, which consistently indicated that Rsn is the primary factor influencing ETo across all employed models (RF, SVR, and XGBoost). This finding aligns with the fundamental physical principle that ET is essentially an energy-driven process (Allen et al., 1998). Comparable conclusions were reached by Liu et al. (2021) and Sh Sammen et al. (2023), who identified Rsn as the most significant input in ETo modeling. Nevertheless, the present study analysis revealed that radiation

alone (scenario 2) fails to account for the complete variability of daily ETo, as the models consistently underestimated high values while overestimating low values. This limitation was similarly recognized by Adnan et al. (2019), who highlighted the inadequate standalone efficacy of solar-radiation-based models. In Scenario 3, the incorporation of the second crucial feature, Tmax, significantly enhanced predictive accuracy, thereby emphasizing its complementary relationship with Rsn. In semi-arid and arid climates, Tmax serves as a strong indicator of atmospheric thermal demand and plays a role in vapor pressure deficit, rendering it an essential predictor (Gao et al., 2017; Yu et al., 2020). These findings are consistent with those of Abdel-Fattah et al. (2023) and Chia et al. (2020), who noted Tmax as one of the most impactful variables in ETo modeling within arid and semi-arid contexts. Research conducted in Mexico also validated Tmax as a reliable predictor in warm climates (Quej et al., 2022). In contrast, when Rsn was combined with the third most important feature, RH, in scenario 4, the models exhibited a moderate improvement compared to using Rsn alone, but the performance remained weaker than that of the Rsn + Tmax scenario. RH captures atmospheric moisture conditions and indirectly reflects vapor pressure deficit, but its influence on ETo was weaker in arid and semi-arid study sites. This finding is consistent with Pagano et al. (2023) and Salahudin et al. (2023), who showed that RH plays a more important role in humid regions than in arid ones. Nourani et al. (2019) further explained that in wetter locations, high air moisture content enhances the predictive role of RH, whereas in drier regions its contribution diminishes. Nonetheless, in data-scarce regions where Tmax is unavailable, the Rsn + RH scenario remains a practical alternative for ETo estimation. If the third and fourth scenarios gave a high performance across the three models, the combination of the most relevant three variables in the fifth scenario (Rsn, Tmax, and RH) achieved the best overall performance across all the

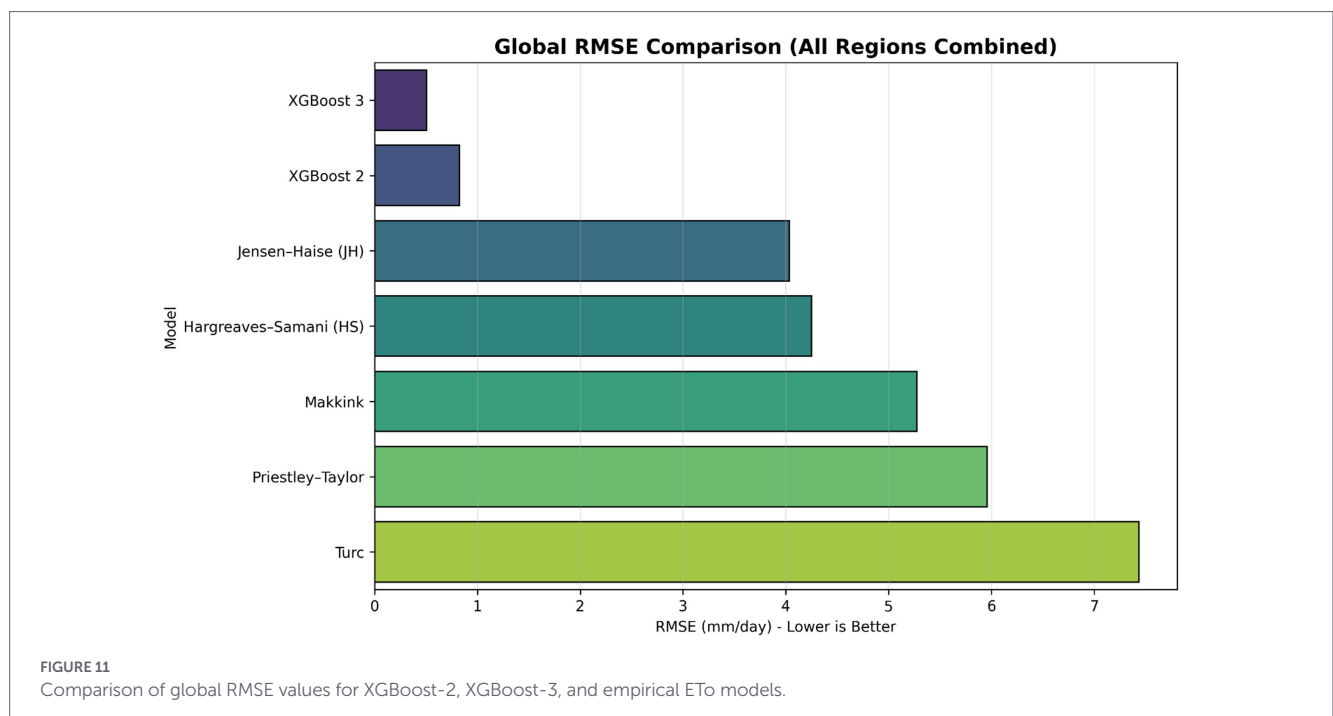


FIGURE 11 Comparison of global RMSE values for XGBoost-2, XGBoost-3, and empirical ETo models.

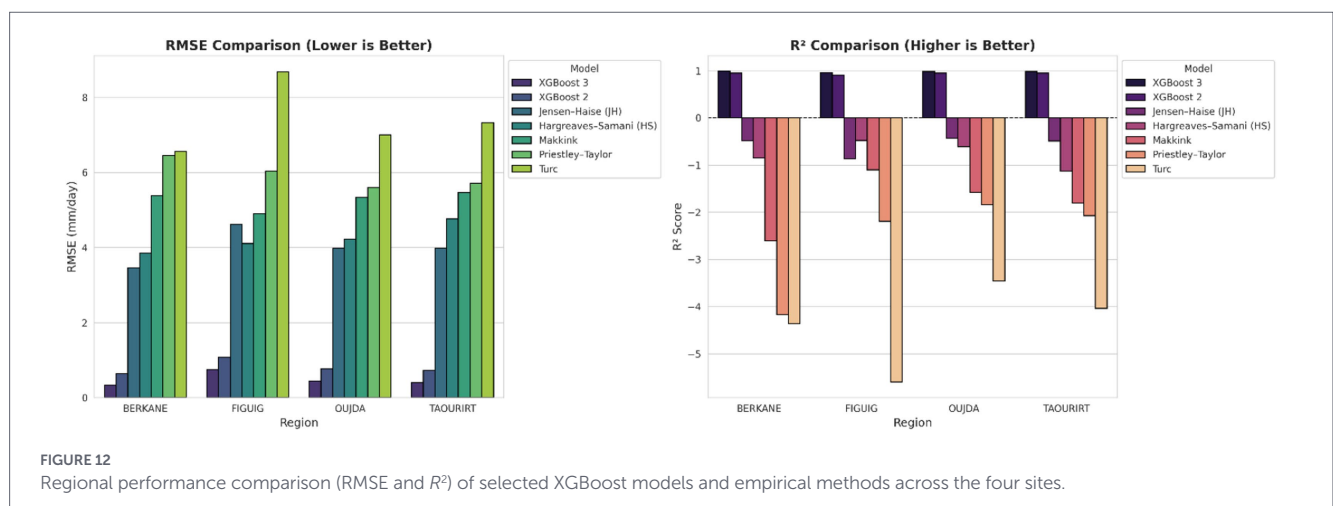


FIGURE 12 Regional performance comparison (RMSE and R²) of selected XGBoost models and empirical methods across the four sites.

models. Despite relying on fewer input variables, this level of accuracy is almost indistinguishable from the benchmark FAO-56 Penman-Monteith method and scenario 1, which incorporates all climatic inputs. These findings clearly demonstrated the critical role of Tmax in improving ETo prediction in arid and semi-arid regions. When combined with Rsn, Tmax significantly enhanced predictive accuracy. Additionally, while RH alone with Rsn provided a weaker contribution, its role became more meaningful when combined with Tmax, as the interaction between air temperature and humidity refines the representation of vapor pressure deficit as reported by Granata (2019).

When comparing the algorithms, the three ML models, XGBoost, RF, and SVR, exhibited overall similar predictive performance, with only marginal differences in accuracy across most input configurations. In several scenarios, including the single-input case (Rsn) and the two-input configurations (Rsn + Tmax and Rsn + RH), XGBoost consistently achieved the highest R² values, followed closely by RF or SVR. In the three-input configuration (Rsn + Tmax + RH), all models reached very high accuracy levels, with only slight differences in

performance, indicating that all three algorithms are capable of accurately estimating ETo using limited climatic information. Despite these relatively small differences in predictive accuracy, clear contrasts emerged when computational efficiency was considered. XGBoost consistently required substantially lower training time and memory resources than both RF and SVR. This efficiency advantage becomes particularly important in operational contexts, where models are intended for real-time use, such as in smart irrigation systems, embedded platforms, or IoT-based decision-support tools. Consequently, although all three algorithms demonstrated strong predictive capability, the combination of high accuracy, low computational cost, and reduced memory demand makes XGBoost the most suitable algorithm for practical deployment in precision irrigation applications. These findings align with previous research demonstrating the superiority of XGBoost for accurate and efficient ETo estimation (Kaissi et al., 2025; Lavarenne and Brouillet, 2025; Mohammadnezhad et al., 2025). The ability of the reduced-input ML models to achieve accuracy comparable to the FAO-56 Penman-Monteith equation represents a

key outcome of this study. Although the Penman–Monteith formulation is grounded in physical principles, its reliability depends on the availability of high-quality and complete climatic datasets. In contrast, our streamlined ML models were able to attain similar levels of accuracy by effectively capturing the complex, non-linear relationships among only three dominant variables. From an irrigation management perspective, the strong performance of the two- and three-feature configurations is particularly valuable, as it enables reliable estimation of crop water requirements while substantially reducing data dependency. This capability supports more precise irrigation scheduling, improves water-use efficiency, and contributes to sustainable irrigation practices in arid and semi-arid agricultural systems. Similar conclusions have been reported by Ge et al. (2022) and Yildirim et al. (2023).

Following its identification as the most suitable algorithm among the evaluated ML approaches, two XGBoost configurations were selected for further analysis and comparison with widely used empirical ETo estimation methods. These configurations include a reduced-input model based on solar radiation and maximum temperature (Rsn + Tmax) and an extended-input model incorporating relative humidity (Rsn + Tmax + RH). The empirical methods exhibited a limited ability to maintain stable predictive performance across the studied regions. Despite the pronounced climatic contrasts between the semi-arid Mediterranean conditions of Berkane and the arid continental climate of Figuig, the empirical formulations consistently yielded low or even negative R^2 values across all regions. This behavior indicates that such empirical equations struggle to represent the complex interactions between climatic drivers in arid and semi-arid environments, particularly when applied without local calibration. Similar limitations have been reported by Acharki et al. (2025), who showed that several commonly used empirical methods, including Hargreaves–Samani and Makkink-based formulations, performed poorly when compared to FAO-56 Penman–Monteith estimates. Among the empirical approaches evaluated in the present study, the Hargreaves–Samani and Jensen–Haise equations produced comparatively lower RMSE values than the other empirical models. This observation is consistent with the findings of Chen et al. (2005), who reported acceptable ETo estimates from Hargreaves-based formulations when temperature data are available. Similar conclusions have been reported in numerous studies (Wu et al., 2019; Chen et al., 2020; Dos Santos Farias et al., 2020; Üneş et al., 2020; Triana-Madrid et al., 2023; Hailegnaw et al., 2024; Yetik, 2025), all of which consistently demonstrate the superior performance of ML models over the Hargreaves equation, particularly under extreme or data-scarce climatic conditions. Radiation-based empirical methods, including Priestley–Taylor, Makkink, and Turc, offer intermediate model complexity and have shown competitive performance in some studies (Hailegnaw et al., 2024). However, in the present study, these methods produced the lowest R^2 values and the highest error magnitudes among all evaluated approaches. This outcome highlights the limited adaptability of fixed-parameter empirical formulations when applied across diverse climatic conditions without regional recalibration. Similar conclusions were drawn by Wu et al. (2019), who found that tree-based ML models, such as Random Forest, Gradient Boosting, and Decision Trees, consistently outperformed both temperature- and radiation-based empirical methods in local-scale applications. Overall, the superiority of XGBoost over all empirical formulations observed in this study aligns with a growing body of literature demonstrating the

advantages of ML for ETo estimation. Previous studies have reported that ML approaches typically achieve coefficients of determination exceeding 0.95 and reduce RMSE by 10–50% compared to traditional empirical methods (Raza et al., 2020; Ferreira et al., 2021; Shaloo et al., 2024; Sharafi and Mohammadi Ghalehi, 2024).

Importantly, the present study advances existing knowledge by demonstrating that near-reference daily ETo accuracy can be achieved using only two or three carefully selected climatic variables, without reliance on the full set of FAO-56 Penman–Monteith inputs. These results represent a direct and practical solution for improving water-use efficiency. Accurate ETo estimation plays a critical role in irrigation scheduling, as the use of inappropriate models can lead to substantial over-irrigation, reported in some studies to reach up to 54 mm per crop season (Hailegnaw et al., 2024). Consequently, the proposed XGBoost models based on (Rsn + Tmax) and (Rsn + Tmax + RH) provide reliable ETo estimates that translate into tangible benefits, including reduced water losses, improved crop water productivity, and enhanced sustainability of irrigation practices. These advantages are particularly crucial in water-scarce regions facing increasing pressures from climate change, population growth, and competing water demands.

5 Conclusion

This study demonstrated the strong potential of machine learning models for accurate daily reference evapotranspiration (ETo) estimation under data-limited conditions in Eastern Morocco. Using a strict chronological evaluation strategy (training: 2001–2018; testing: 2019–2025) to preserve the time-series structure and avoid information leakage, machine learning approaches clearly outperformed conventional empirical equations, which showed weak transferability and poor predictive accuracy across the studied regions. Among the tested algorithms, XGBoost provided the most robust and operational solution, achieving near-reference performance with only three inputs (Rsn, Tmax, RH) ($R^2 = 0.976$; $RMSE < 0.55 \text{ mm day}^{-1}$), while the two-input configuration (Rsn + Tmax) maintained reliable accuracy ($R^2 = 0.936$; $RMSE \approx 0.60 \text{ mm day}^{-1}$) with minimal computational cost. These findings confirm that reliable ETo estimation can be achieved with minimal climatic information, supporting practical irrigation scheduling and improved water-use efficiency in arid and semi-arid environments. Despite these promising outcomes, several limitations should be acknowledged. First, model development and validation were based on four locations within Eastern Morocco, which may limit direct transferability to other climatic zones without further testing. Second, FAO-56 Penman–Monteith ETo was used as the reference target rather than lysimeter-based measurements, which may introduce uncertainty in absolute accuracy. Third, although the minimal-input design enhances operational applicability, extreme events and local microclimatic variability may require additional predictors to further improve robustness. Future work should therefore focus on expanding validation across broader agro-climatic regions, integrating field-measured ETo observations when available, and evaluating long-term robustness under climate variability. In addition, integrating the proposed XGBoost models into decision-support systems and IoT-enabled irrigation platforms could enable real-time ETo estimation and fully automated irrigation scheduling, contributing to sustainable agricultural production and water security in Morocco.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://power.larc.nasa.gov/>.

Author contributions

HH: Writing – review & editing, Writing – original draft, Software, Conceptualization, Methodology, Formal analysis. YZ: Investigation, Conceptualization, Resources, Methodology, Writing – review & editing. MuA: Validation, Writing – review & editing, Investigation, Formal analysis. HA: Investigation, Visualization, Writing – review & editing, Data curation. MaA: Conceptualization, Methodology, Funding acquisition, Writing – review & editing. ON: Data curation, Writing – review & editing, Formal analysis, Visualization. FN: Investigation, Writing – review & editing, Data curation, Methodology, Resources, Formal analysis. NK: Investigation, Validation, Formal analysis, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was funded by the Princess Nourah bint Abdulrahman University, Researchers Supporting Project Number PNURSP2026R103, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

- Abdel-Fattah, M. K., Kotb Abd-Elmabod, S., Zhang, Z., and Merwad, A. R. M. A. (2023). Exploring the applicability of regression models and artificial neural networks for calculating reference evapotranspiration in arid regions. *Sustainability (Switzerland)* 15:15494. doi: 10.3390/su152115494
- Abdullahi, A. A. (2025). Agriculture and sustainability in Somalia: challenges and pathways to sustainable development. *Selcuk J. Agric. Food Sci.* 39, 310–319. doi: 10.15316/SELCKUJAFSCL1344646
- Acharki, S., Raza, A., Vishwakarma, D. K., Amharref, M., Bernoussi, A. S., Singh, S. K., et al. (2025). Comparative assessment of empirical and hybrid machine learning models for estimating daily reference evapotranspiration in sub-humid and semi-arid climates. *Sci. Rep.* 15:2542. doi: 10.1038/s41598-024-83859-6
- Adnan, R. M., Malik, A., Kumar, A., Parmar, K. S., and Kisi, O. (2019). Pan evaporation modeling by three different neuro-fuzzy intelligent systems using climatic inputs. *Arab. J. Geosci.* 12, 1–14. doi: 10.1007/S12517-019-4781-6
- Ahmed Matoir, M., Abdelmajid, B., and Belati, F. (2025). Adding value to the Berkane Clementine. doi: 10.20944/PREPRINTS202503.2019.V1
- Al Kaddouri, H., Blaacha, J., Hamdaoui, H., El Mehdi, A., Douzi, Y., Messbah, H., et al. (2025). Optimization of hyperparameters for SVM classification of Citrus diseases using grid search and cross-validation. *Lect. Notes Electric. Eng.* 1306, 489–497. doi: 10.1007/978-981-96-0644-3_44
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper. Rome, Italy: Food and Agriculture Organization of the United Nations (FAO), 56.
- Atta, K., Mondal, S., Gorai, S., Singh, A. P., Kumari, A., Ghosh, T., et al. (2023). Impacts of salinity stress on crop plants: improving salt tolerance through genetic and molecular dissection. *Front. Plant Sci.* 14:1241736. doi: 10.3389/fpls.2023.1241736
- Bayabil, H. K., Migliaccio, K. W., Dukes, M. D., and Vasquez, L. (2020). Basic tips for designing efficient irrigation systems. *EDIS* 2020. doi: 10.32473/EDIS-AE539-2020
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Chen, T., and Guestrin, C. (2016). “XGBoost: a scalable tree boosting system” in Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, New York, NY, USA: Association for Computing Machinery (ACM) 13-17-August-2016, 785–794.
- Chen, J.-F., Yeh, H.-F., Lee, C.-H., and Lo, W.-C. (2005). “Comparison of empirical equations for estimating potential evapotranspiration in Taiwan” in Proceedings of the 31st IAHR world congress.
- Chen, Z., Zhu, Z., Jiang, H., and Sun, S. (2020). Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J. Hydrol.* 591:125286. doi: 10.1016/J.JHYDROL.2020.125286
- Chia, M. Y., Huang, Y. F., and Koo, C. H. (2020). Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. *Comput. Electron. Agric.* 175:105577. doi: 10.1016/J.COMPAE.2020.105577
- Depeursinge, A., Racoceanu, D., Iavindrasana, J., Cohen, G., Platon, A., Poletti, P.-A., et al. (2010). Fusing visual and clinical information for lung tissue classification in HRCT data. *Artif. Intell. Med.:ARTMED*1118. doi: 10.1016/j.artmed.2010.04.006
- DIAEA (2024). Directeur de l'irrigation et de l'aménagement de l'espace agricole du département de l'agriculture | Ministère de l'agriculture. Available online at: <https://www.agriculture.gov.ma/fr/node/15308> (Accessed February 16, 2026).
- Dos Santos Farias, D. B., Althoff, D., Rodrigues, L. N., and Filgueiras, R. (2020). Performance evaluation of numerical and machine learning methods in estimating reference evapotranspiration in a Brazilian agricultural frontier. *Theor. Appl. Climatol.* 142, 1481–1492. doi: 10.1007/S00704-020-03380-4
- Ferreira, L. B., Da Cunha, F. F., Da Silva, G. H., Campos, F. B., Dias, S. H. B., and Santos, J. E. O. (2021). Generalizability of machine learning models and empirical equations for the estimation of reference evapotranspiration from temperature in a semiarid region. *An. Acad. Bras. Cienc.* 93:e20200304. doi: 10.1590/0001-3765202120200304

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. The authors used ChatGPT (GPT-5, OpenAI, 2025) to assist in language polishing and consistency checking. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final version of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gao, Z., He, J., Dong, K., and Li, X. (2017). Trends in reference evapotranspiration and their causative factors in the West Liao River basin, China. *Agric. For. Meteorol.* 232, 106–117. doi: 10.1016/j.agrformet.2016.08.006
- Ge, J., Zhao, L., Yu, Z., Liu, H., Zhang, L., Gong, X., et al. (2022). Prediction of greenhouse tomato crop evapotranspiration using XGBoost machine learning model. *Plants* 11, 1923–1923. doi: 10.3390/PLANTS11151923
- Ghiat, I., Mackey, H. R., Al-Ansari, T., Ghiat, I., Mackey, H. R., and Al-Ansari, T. (2021). A Review of Evapotranspiration Measurement Models, Techniques and Methods for Open and Closed Agricultural Field Applications. *Water* 13:13. doi: 10.3390/W13182523
- Gong, X., Qiu, R., Ge, J., Bo, G., Ping, Y., Xin, Q., et al. (2021). Evapotranspiration partitioning of greenhouse grown tomato using a modified Priestley–Taylor model. *Agric. Water Manag.* 247:106709. doi: 10.1016/j.agwat.2020.106709
- Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—a comparative study. *Agric. Water Manag.* 217, 303–315. doi: 10.1016/j.agwat.2019.03.015
- Hailegnaw, N. S., Bayabil, H. K., Berihun, M. L., Teshome, F. T., Shelia, V., and Getachew, F. (2024). Integrating machine learning and empirical evapotranspiration modeling with DSSAT: implications for agricultural water management. *Sci. Total Environ.* 912. doi: 10.1016/j.scitotenv.2023.169403
- Hamdaoui, H., Hsana, Y., Hamdi, I., Al Kaddouri, H., and Kouddane, N. E. (2024). Revolutionizing agriculture: a comprehensive review of AI-enabled precision irrigation and water quality forecasting. *Basrah J. Agric. Sci.* 37, 354–380. doi: 10.30770/25200860.2024.37.2.26
- Hamdaoui, H., Zarrouk, Y., Kouddane, N. E., Hsana, Y., Al Kaddouri, H., and Chkird, F. (2025). A deep learning-based model for efficient olive leaf disease classification. *Lect. Notes Electric. Eng.* 1306, 477–487. doi: 10.1007/978-981-96-0644-3_43
- Hargreaves, G. H., and Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.* 1, 96–99. doi: 10.13031/2013.26773
- Haris, A., Marimin, Wahjuni, S., and Setiawan, B. I. (2025). The use of artificial neural networks to estimate reference evapotranspiration. *Agromet* 39, 1–7. doi: 10.29244/J.AGROMET.39.1.1-7
- Jensen, M. E., and Haise, H. R. (1963). Estimating evapotranspiration from solar radiation. *J. Irrig. Drain. Div.* 89, 15–41. doi: 10.1061/JRCEA4.0000287
- Kaissi, O., Er-Raki, S., Belaqqiz, S., Kharrrou, M. H., Amazirh, A., Bouras, E. H., et al. (2025). Enhancing water management in Morocco's arid regions: using advanced models for accurate ETO estimation. *Acta Hort.* 1, 267–274. doi: 10.17660/ACTAHORTIC.2025.1422.33
- Katsoulas, N., Stanghellini, C., Katsoulas, N., and Stanghellini, C. (2019). Modelling Crop Transpiration in Greenhouses: Different Models for Different Applications. *Agronomy* 9:9. doi: 10.3390/AGRONOMY9070392
- Khine, M. S. (2024). "Machine learning in education" in *Artificial intelligence in education*, Hershey, PA, USA: IGI Global. 571–627. doi: 10.1007/978-981-97-9350-1_5
- Koç, D. L., and Can, M. E. (2023). Reference evapotranspiration estimate with missing climatic data and multiple linear regression models. *PeerJ* 11:e15252. doi: 10.7717/PEERJ.15252/SUPP-1
- Kozłowski, T. T. (1984). Plant responses to flooding of soil. *Bioscience* 34, 162–167. doi: 10.2307/1309751
- Lavarenne, J., and Brouillet, A. (2025). Training local models from reanalysis data to estimate reference evapotranspiration with fewer onsite sensors, an evaluation in West Africa. Charlottesville, VA, USA: Center for Open Science (OSF).
- Liu, X., Wu, L., Zhang, F., Huang, G., Yan, F., and Bai, W. (2021). Splitting and length of years for improving tree-based models to predict reference crop evapotranspiration in the humid regions of China. *Water (Switz.)* 13:3478. doi: 10.3390/w13233478
- Madani, and Tariq (2006). Le partage de l'eau dans l'oasis de Figuig (Maroc oriental). Approche historique et archéologique. *Mélanges de la Casa de Velázquez. Nouvelle série* 36–2, 61–81. doi: 10.4000/MCV.2016
- Makkink, F. G. (1957). Testing the penman formula by means of lysimeters. *J. Inst. Water Eng.* 11, 277–288.
- Mohammadi, B., Chen, M., Reza Nikoo, M., Cheraghalizadeh, M., Yu, Y., Zhang, H., et al. (2024). An explainable hybrid framework for estimating daily reference evapotranspiration: combining extreme gradient boosting with Nelder-Mead method. *J. Hydrol.* 644:132130. doi: 10.1016/j.jhydrol.2024.132130
- Mohammadnezhad, M., Davary, K., Shirazi, P., Rezvanpour, M. J., and Hashemini, S. M. (2025). A novel hybrid model for actual evapotranspiration estimation in data-scarce arid regions: integrating modified Budyko and machine learning models using deep learning. *Sci. Total Environ.* 1001:180438. doi: 10.1016/j.scitotenv.2025.180438
- Morocco - Water. (2023). Available online at: <https://www.trade.gov/country-commercial-guides/morocco-water> (Accessed February 16, 2026).
- Nagappan, M., Kudiyarasumani, A., and Jayamurugan, S. (2025). Assessment of Reference Evapotranspiration: A Case Study Using Hargreaves and Penman-Monteith. *International Journal of Innovative Science and Research Technology*, 2354–2360. doi: 10.38124/IJISRT/25MAR129
- Nourani, V., Elkiran, G., and Abdullahi, J. (2019). Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J. Hydrol.* 577:123958. doi: 10.1016/j.jhydrol.2019.123958
- Pagano, A., Amato, F., Ippolito, M., De Caro, D., Croce, D., Motisi, A., et al. (2023). Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation. *Ecol. Inform.* 76. doi: 10.1016/j.ecoinf.2023.102133
- Quej, V. H., Castillo, C. D. L. C., Almorox, J., and Rivera-Hernandez, B. (2022). Evaluation of artificial intelligence models for daily prediction of reference evapotranspiration using temperature, rainfall and relative humidity in a warm sub-humid environment. *Ital. J. Agrometeorol.* 2022, 49–63. doi: 10.36253/IJAM-1373
- Rajput, J., Singh, M., Lal, K., Khanna, M., Sarangi, A., Mukherjee, J., et al. (2023a). Assessment of data intelligence algorithms in modeling daily reference evapotranspiration under input data limitation scenarios in semi-arid climatic condition. *Water Sci. Technol.* 87, 2504–2528. doi: 10.2166/WST.2023.137
- Rajput, J., Singh, M., Lal, K., Khanna, M., Sarangi, A., Mukherjee, J., et al. (2023b). Selection of alternate reference evapotranspiration models based on multi-criteria decision ranking for semiarid climate. *Environ. Dev. Sustain.* 26, 11171–11216. doi: 10.1007/S10668-023-03234-9
- Rajput, J., Singh, M., Lal, K., Khanna, M., Sarangi, A., Mukherjee, J., et al. (2024). Development of single and dual crop coefficients for drip-irrigated broccoli using weighing type field lysimeters in semi-arid environment. *Environ. Dev. Sustain.* doi: 10.1007/S10668-024-05416-5
- Raza, A., Shoaib, M., Faiz, M. A., Baig, F., Khan, M. M., Ullah, M. K., et al. (2020). Comparative assessment of reference evapotranspiration estimation using conventional method and machine learning algorithms in four climatic regions. *Pure Appl. Geophys.* 177, 4479–4508. doi: 10.1007/S00024-020-02473-5
- Sabancı, D., Yurekli, K., Comert, M. M., Kilicarslan, S., and Erdogan, M. (2023). Predicting reference evapotranspiration based on hydro-climatic variables: comparison of different machine learning models. *Hydrol. Sci. J.* 68, 1050–1063. doi: 10.1080/02626667.2023.2203824
- Salahudin, H., Shoaib, M., Albano, R., Inam Baig, M. A., Hammad, M., Raza, A., et al. (2023). Using ensembles of machine learning techniques to predict reference evapotranspiration (ETO) using limited meteorological data. *Hydrology* 10. doi: 10.3390/HYDROLOGY10080169
- Sh Sammen, S., Kisi, O., Mohammed Sami Al-Janabi, A., and Elbeltagi, A. (2023). Predicting reference evapotranspiration in semi-arid-region by regression-based machine learning methods using limited climatic inputs. *Res. Sq.* doi: 10.21203/RS.3.RS-2600302/V1
- Shaloo, Bisht, H., Kumar, B., Rajput, J., and Brahmanand, P. S. (2026). Modeling daily reference evapotranspiration and evaluating uncertainty analysis in machine learning under limited meteorological data conditions for northern India. *J. Atmos. Solar-Terr. Phys.* 278:106696. doi: 10.1016/j.jastp.2025.106696
- Shaloo, Kumar, B., Bisht, H., Rajput, J., Mishra, A. K., Kiran Kumara, T. M., et al. (2024). Reference evapotranspiration prediction using machine learning models: an empirical study from minimal climate data. *Agron. J.* 116, 956–972. doi: 10.1002/AGJ2.21504
- Sharafi, S., and Mohammadi Ghaleni, M. (2024). Revealing accuracy in climate dynamics: enhancing evapotranspiration estimation using advanced quantile regression and machine learning models. *Appl. Water Sci.* 14:162. doi: 10.1007/S13201-024-02211-5
- Singla, A., Nehra, A., Joshi, K., Kumar, A., Tuteja, N., Varshney, R. K., et al. (2024). Exploration of machine learning approaches for automated crop disease detection. *Curr. Plant Biol.* 40:100382. doi: 10.1016/j.cpb.2024.100382
- Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Stat. Comput.* 14, 199–222. doi: 10.1023/B:STCO.0000035301.49549.88
- Tasa, M., González-Guzmán, M., Arbona, V., and Pérez-Pérez, J. G. (2025). High boron in irrigation water mitigates drought stress effects on "Carrizo" citrange seedlings. *Environ. Exp. Bot.* 238:106228. doi: 10.1016/j.envexpbot.2025.106228
- Tausif, M., Dilshad, S., Umer, Q., Iqbal, M. W., Latif, Z., Lee, C., et al. (2023). Ensemble learning-based estimation of reference evapotranspiration (ET_o). *Internet Things* 24:100973. doi: 10.1016/j.iot.2023.100973
- Triana-Madrid, J. C., Ocampo-Marulanda, C., Carvajal-Escobar, Y., Torres-López, W. A., Triana, J., and Canchala, T. (2023). Estimation of monthly reference evapotranspiration with scarce information using machine learning in southwestern Colombia. *Meteorologica* 48:024. doi: 10.24215/1850-468XE024
- Turc, L. (1961). Evaluation des besoins en eau d'irrigation, evapotranspiration potentielle. *Ann. Agron.* 12, 13–49.
- Üneş, F., Kaya, Y. Z., and Mamak, M. (2020). Daily reference evapotranspiration prediction based on climatic conditions applying different data mining techniques and empirical equations. *Theor. Appl. Climatol.* 141, 763–773. doi: 10.1007/S00704-020-03225-0
- Valipour, M. (2015). Temperature analysis of reference evapotranspiration models. *Meteorol. Appl.* 22, 385–394. doi: 10.1002/met.1465

- Wanniarachchi, S., and Sarukkalige, R. (2022). A review on evapotranspiration estimation in agricultural water management: past, present, and future. *Hydrology* 9:123. doi: 10.3390/HYDROLOGY9070123
- WMO (2023). World Meteorological Organization. Available online at: <https://wmo.int/> (Accessed February 16, 2026).
- Wu, L., Peng, Y., Fan, J., and Wang, Y. (2019). Machine learning models for the estimation of monthly mean daily reference evapotranspiration based on crossstation and synthetic data. *Hydrol. Res.* 50, 1730–1750. doi: 10.2166/NH.2019.060
- Yetik, A. K. (2025). Machine learning-based estimation of daily ETo under limited meteorological data. *J. Agric. Faculty Gaziosmanpasa Univ.* 42, 250–261. doi: 10.55507/GOPZFD.1709027
- Yildirim, D., Küçüktopcu, E., Cemek, B., and Simsek, H. (2023). Comparison of machine learning techniques and spatial distribution of daily reference evapotranspiration in Türkiye. *Appl. Water Sci.* 13. doi: 10.1007/S13201-023-01912-7
- Yin, Y., Wu, S., Zheng, D., and Yang, Q. (2008). Radiation calibration of FAO56 penman-monteith model to estimate reference crop evapotranspiration in China. *Agric. Water Manag.* 95, 77–84. doi: 10.1016/J.AGWAT.2007.09.002
- Yu, J., Zheng, W., Xu, L., Zhangzhong, L., Zhang, G., and Shan, F. (2020). A PSO-XGBoost model for estimating daily reference evapotranspiration in the solar greenhouse. *Intell. Autom. Soft Comput.* 26, 989–1003. doi: 10.32604/iasc.2020.010130