

Article

Spatiotemporal Modeling and Uncertainty Quantification of Reference Evapotranspiration Using Machine Learning and Bayesian Model Averaging in Benin

Bienvenue Christela Finounou Mizele ^{1,*}, Modeste Meliho ^{2,*}, Vinasetan Ratheil Houndji ³ ,
Semevo Arnaud R. M. Ahouandjinou ³ and Collins A. Orlando ⁴

¹ École Doctorale des Sciences de l'Ingénieur, Université d'Abomey-Calavi, Abomey-Calavi BP 526, Benin

² Centre d'Enseignement et de Recherche en Foresterie (CERFO), 2440 Ch Ste-Foy, Québec, QC G1V 1T2, Canada

³ Institut de Formation et de Recherche en Informatique, Université d'Abomey-Calavi, Abomey-Calavi 01 BP 526, Benin; ratheil.houndji@uac.bj (V.R.H.); arnaud.ahouandjinou@imsp-uac.org (S.A.R.M.A.)

⁴ Université du Québec à Rimouski, 300 All. des Ursulines, Rimouski, QC G5L 3A1, Canada; collins.orlando@uqar.ca

* Correspondence: bienvenuechristela@gmail.com (B.C.F.M.); mmeliho@cerfo.qc.ca (M.M.)

Abstract

Reference evapotranspiration (ET_0) represents the atmospheric demand for water from a well-watered vegetated surface and is a key component of the hydrological cycle and agricultural water management. This study evaluated the performance of seven machine learning (ML) models: linear regression (LR), Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), Decision Trees (DT), and Cubist, for predicting monthly FAO-56 Penman–Monteith ET_0 in Benin. The target variable was calculated from data collected at six synoptic stations over the 2017–2021 period. Ten remote-sensing and topographic predictors were used: MODIS Land Surface Temperature (LST), six Sentinel-2 optical vegetation indices (NDVI, EVI, NDMI, NDWI, MSI, NDRE), elevation, and cyclic month encoding. Models were trained on the 2017–2019 period and evaluated on an independent temporal test set (2020–2021). All models showed positive predictive performance, with the BMA ensemble achieving the highest accuracy (RMSE = 7.0% of mean ET_0 , $R^2 = 0.802$), followed by Cubist (RMSE = 7.3%, $R^2 = 0.787$) and DT (RMSE = 7.5%, $R^2 = 0.776$). The seven models were combined via Bayesian Model Averaging (BMA) with posterior weights estimated by the EM algorithm to produce 1 km monthly ET_0 maps for Benin for 2025. BMA-derived inter-model standard deviation provided spatially explicit uncertainty estimates, revealing that prediction uncertainty is greatest in the northern Sudanian zone during the dry season. The ET_0 target variable was constructed as a hybrid product combining station temperature observations with solar radiation, wind speed, and vapor pressure deficit extracted from the TerraClimate gridded reanalysis dataset; this methodological choice is discussed as a study limitation.

Keywords: reference evapotranspiration; remote sensing; machine learning; Bayesian Model Averaging; uncertainty quantification; Sudanian zone; Benin



Academic Editor: Alessandro Vitale

Received: 24 May 2026

Revised: 23 June 2026

Accepted: 28 June 2026

Published: 2 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Water is a fundamental component of life on Earth. It is an extremely important factor in maintaining the sustainable development of human society and socioeconomic systems. In recent decades, the rapid increase in demand for water due in part to high human population growth has led to the depletion of water supplies in several regions [1–3]. This growing problem is increasingly accentuated by the effects of global climate change [4]. As a result, the monitoring of the water cycle and allocation of water resources, particularly in regions already suffering from water shortages, has become a major concern for scientists and policymakers.

A major component of the water cycle and of particular interest for agricultural water balance is evapotranspiration (ET). ET relates to the emission of water vapor into the atmosphere through the evaporation of water from the soil and aquatic environments, and through the transpiration of plants. It represents the largest outflow of water from the land surface, with approximately 60–80% of precipitation at the land surface returning to the atmosphere where it becomes the source of future precipitation [5]. One of the main types of ET explored is reference evapotranspiration (ET_0), a concept introduced in the 56th Irrigation Report of the Food and Agriculture Organization of the United Nations (FAO) in the 1990s. It defines the evaporation potential of a standard canopy abundantly supplied with water [6–8]. ET_0 provides a reasonably reliable and predictive estimate of plant water requirements during active growth. Hence, data on this phenomenon can be used to help farmers schedule irrigation more accurately to maximize yields and improve water productivity.

In line with other developing countries, agriculture remains a major sector contributing to food security and significantly to the national economy in Benin. It is a leading contributor to Benin's economy, accounting for about 29% of the GDP and employing nearly 70% of the working population [9]. However, the country's agricultural sector is subject to major challenges, particularly in relation to water availability. Benin is located in a region where rainfall is erratic and limited over time, making water availability a critical factor for crop production [10–12]. The Beninese government has implemented several programs to improve access to water for agriculture, including the construction of dams, irrigation canals and the promotion of drip irrigation. However, despite these efforts, many farmers still struggle to access water, thereby limiting agricultural production and productivity. According to regional climate projections for West Africa, rainfall variability is expected to intensify, with potentially negative consequences for rainfed agricultural production [13,14]. Overall, the need for water, combined with the decrease in water resources, especially due to overexploitation, remains a crucial issue for the economic and social development of the country. Therefore, it is necessary to conduct research on the components of the hydrological cycle, including the trends of evapotranspiration over space and time.

Although ET_0 is one of the most difficult components of the hydrologic cycle to quantify, primarily due to its complexity in the soil-land-plant system, various hydrological, micrometeorological, physiological, analytical and empirical approaches have been used to estimate it. These include: the Penman-Monteith (FAO-56 PM) method [6] which employs meteorological data such as air temperature, wind speed, relative humidity and solar radiation; the Hargreaves method [15] which uses maximum and minimum air temperatures; the Jensen-Haise method [16] where solar radiation and air temperature are used; the Thornthwaite method [17] where estimation is based on air temperature and latitude. In addition to these, remotely sensed data can be leveraged for use in semi-empirical models that calculate fluxes at the time of measurement and/or by extrapolation to periodic scales [18–22].

An increasingly popular approach to address challenges in ET_0 spatiotemporal estimation is to adopt machine learning (ML) techniques. Indeed, in recent years, ML has been successfully used to estimate ET_0 rates in regions often associated with sparse and/or unavailable weather data. A diverse range of models has been investigated, including DT [23], RF [23–25], XGBoost [26], Artificial Neural Network (ANN: [27,28]), SVM [29,30], adaptive neuro-fuzzy inference system (ANFIS: [31,32]), among others. In general, ML models have demonstrated better performance than conventional methods for the same data requirements [33], further emphasizing their potential use in hydrology as well as providing powerful tools for building a knowledge base for water management [34,35].

Beyond individual ML model comparison, ensemble model combination offers a complementary strategy to improve predictive robustness and to quantify prediction uncertainty. Bayesian Model Averaging (BMA) provides a principled ensemble framework that weights each constituent model in proportion to its posterior probability given the observations, typically estimated via the Expectation–Maximization algorithm [36]. In hydrology and related disciplines, BMA has been applied to combine multiple evapotranspiration estimates, consistently demonstrating improved accuracy and broader reliability compared to any single model [37,38]. A key advantage of BMA over simple arithmetic ensembles is that it yields spatially explicit inter-model uncertainty maps, quantified as the inter-model standard deviation, which are essential for assessing prediction reliability in data-sparse regions where ground-truth validation is inherently limited. Despite these advantages, BMA-based ensemble approaches remain largely unexplored for ET_0 estimation in sub-Saharan West Africa.

Despite its demonstrated utility in ecological and hydrological ensemble modelling [36–38], BMA combined with an EM algorithm has not previously been applied to ET_0 prediction in West Africa. Regional studies in this zone have relied either on simple model averaging or on single best-model selection, overlooking the uncertainty quantification capabilities that BMA uniquely provides. This gap is particularly consequential given the high spatial heterogeneity of Benin's three climatic zones (Guinean, Sudano-Guinean, and Sudanian), which challenges the generalizability of any single ML model across the national territory.

Although ML has shown strong potential for ET_0 estimation, its application in Benin and West Africa remains limited in scope. In particular, the current body of work still lacks several key elements needed to support reliable monthly ET_0 prediction and mapping in the region. There is currently no high-resolution monthly ET_0 product specifically calibrated for Benin. Long-term ET_0 datasets integrating station observations with remote-sensing predictors also remain scarce. In addition, comparative evaluations of multiple machine-learning models for ET_0 estimation are still lacking in the region. Finally, the use of MODIS land surface temperature as a continuous predictor of ET_0 remains limited in West Africa.

Against this background, the principal scientific novelties of this study are threefold: (i) it presents the first application of the EM-based BMA framework to monthly ET_0 prediction and mapping in West Africa, advancing beyond simple model selection by providing calibrated posterior model weights and spatially explicit uncertainty maps; (ii) it demonstrates that a satellite-only predictor set, excluding all direct meteorological inputs to avoid circular dependency with the FAO-56 PM equation, achieves competitive accuracy while remaining fully operational in the absence of ground-based weather station networks; and (iii) it produces the first 1 km resolution monthly ET_0 map for Benin at the national scale, with an associated inter-model uncertainty layer that can guide water management decisions.

To address some of these gaps, this study aimed to: (i) calculate monthly FAO-56 Penman-Monteith ET_0 at six synoptic stations in Benin over the 2017–2021 period; (ii) compare the predictive performance of seven models using MODIS remote-sensing

predictors through an independent temporal test set evaluation (2020–2021); (iii) generate 1 km monthly ET_0 maps for Benin for the year 2025 via Bayesian Model Averaging (BMA) of the seven models; and (iv) quantify prediction uncertainty using the BMA inter-model standard deviation.

2. Materials and Methods

2.1. Study Area

The study area (Figure 1) covers the whole of Benin, which is a West African country located between latitudes $6^{\circ}30'$ and $12^{\circ}30'$ North and longitudes 1° and $3^{\circ}40'$ East. The country spans an area of 114,763 km² and has a population of approximately 13 million inhabitants [39]. It is bounded to the north by Niger, to the northwest by Burkina Faso, to the west by Togo, to the east by Nigeria and to the south by the Atlantic Ocean. Benin is subject to two distinct climates: (i) in the south, an equatorial climate marked by high humidity that features alternating dry seasons (November to March and mid-July to mid-September) and rainy seasons (April to mid-July and mid-September to October); and (ii) in the central and northern regions, a tropical climate characterized by a dry season from November to April and a rainy season from June to September. From a pedological point of view, Benin has five groups of soils including ferruginous soils, weakly evolved soils, ferralitic soils, hydromorphic soils and vertisols, which form on sedimentary rocks in the south, crystalline rocks in the center and north, and on alluvial or marine deposits in intra-zonal environments.

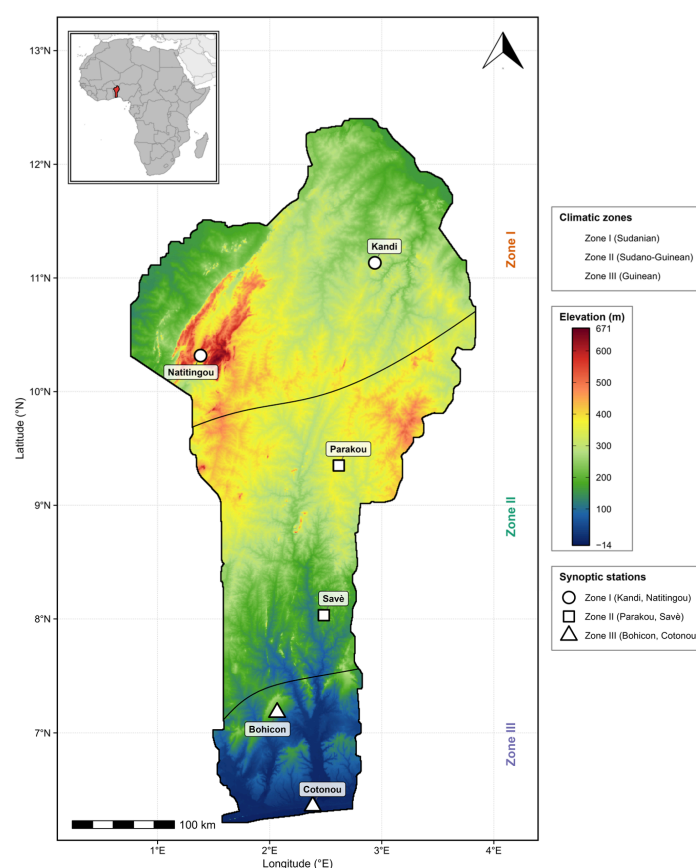


Figure 1. Geographic location of the study area showing the three distinct climatic zones of Benin (Zone I: Sudanian; Zone II: Sudano-Guinean; Zone III: Guinean), delimited by black boundary lines. The elevation model highlights the relief gradient across the country, with the highest elevations (up to 671 m) concentrated in the northwestern region. Synoptic stations used in the study are indicated by symbols according to their climatic zone. The inset map shows the location of Benin (red) within Africa.

2.2. Data Acquisition

2.2.1. Remote Sensing Data

The satellite images used in this work were collected from monthly time series over the 2017–2021 period (corresponding to the model training and test periods). To obtain the predictor variables for ET_0 modeling, spatial data from the Moderate Resolution Imaging Spectroradiometer were used using Google Earth Engine (GEE). MODIS is a sensor aboard NASA's Terra and Aqua satellites that records multiple images of the globe every 1–2 days in 36 spectral bands ranging from 0.4 to 14.4 μm with a spatial resolution of 250 m to 1 km. In this study, MODIS daytime land surface temperature (LST) was extracted from the MOD11A2 product (Terra, 8-day composites, aggregated to monthly means). This LST variable represents thermal emittance directly linked to the surface energy balance controlling evaporative demand. In addition, TerraClimate monthly climate data [40] were accessed through GEE [41] exclusively to provide the meteorological variables required for the calculation of the FAO-56 PM ET_0 target variable (solar radiation, wind speed, and vapor pressure deficit). TerraClimate data were not used as ML predictor variables, so as to avoid the circular dependency that would arise from using variables that are components of the target variable also as model inputs.

The predictor variables used as inputs to the ML models are listed in Table A1. They comprise 10 remote-sensing and geographic/temporal variables: (1) MODIS Land Surface Temperature (LST, MOD11A2 product), representing the thermal surface state; (2) six Sentinel-2 optical spectral indices (NDVI, EVI, NDMI, NDWI, MSI, and NDRE) capturing vegetation state and surface moisture conditions; (3) elevation from SRTM as a static topographic proxy; and (4) cyclic month encoding (month_sin, month_cos) capturing intra-annual seasonality. To ensure temporal consistency across the full 2017–2021 modeling period, only satellite predictor sources with uninterrupted global coverage were retained. Optical indices were extracted from Google Earth Engine monthly composites at each station location. Predictor variables were extracted at 1 km resolution across Benin for spatial mapping. All data processing was performed in an R 4.5.2 environment; data extraction was performed in a Google Earth Engine JavaScript environment.

Pearson correlations among the predictor variables are presented in Table A4. NDVI and EVI were strongly correlated ($r = 0.97$), as expected given their shared spectral basis, indicating slight redundancy in the vegetation canopy signal. LAI and FPAR were likewise highly correlated ($r = 0.96$). The cyclic month encodings (sin, cos) were orthogonal by construction ($r = 0.00$). Among all predictors, sin (month) showed the strongest correlation with ET_0 ($r = 0.69$), confirming the dominant role of intra-annual seasonality, followed by LST ($r = 0.38$).

2.2.2. Climate Data

The meteorological data collected from the six synoptic stations in Benin (Figure 1; Table A2) consisted of maximum (T_{max}), minimum (T_{min}) and mean (T_{mean}) air temperatures. The corresponding datasets were downloaded from the National Center for Environmental Information (NCEI: <https://www.ncei.noaa.gov/>, accessed on 4 March 2023) website. Because the NCEI records did not provide all the variables required by the FAO-56 Penman-Monteith equation, the remaining inputs (solar radiation, wind speed, and vapor pressure deficit) were obtained from TerraClimate. Thus, station-observed temperature variables were combined with TerraClimate-derived monthly variables to calculate FAO-56 PM ET_0 at each station. Accordingly, the target variable used in this study was not directly measured ET_0 , but reference ET_0 estimated from station meteorological observations complemented by gridded climate data. Station-based temperature observations (T_{max} , T_{min} , T_{mean}) were used exclusively for the construction of the target variable and were not

included as ML predictors. The variation of T_{mean} , T_{max} and T_{min} during the period of observation is presented in Figure 2.

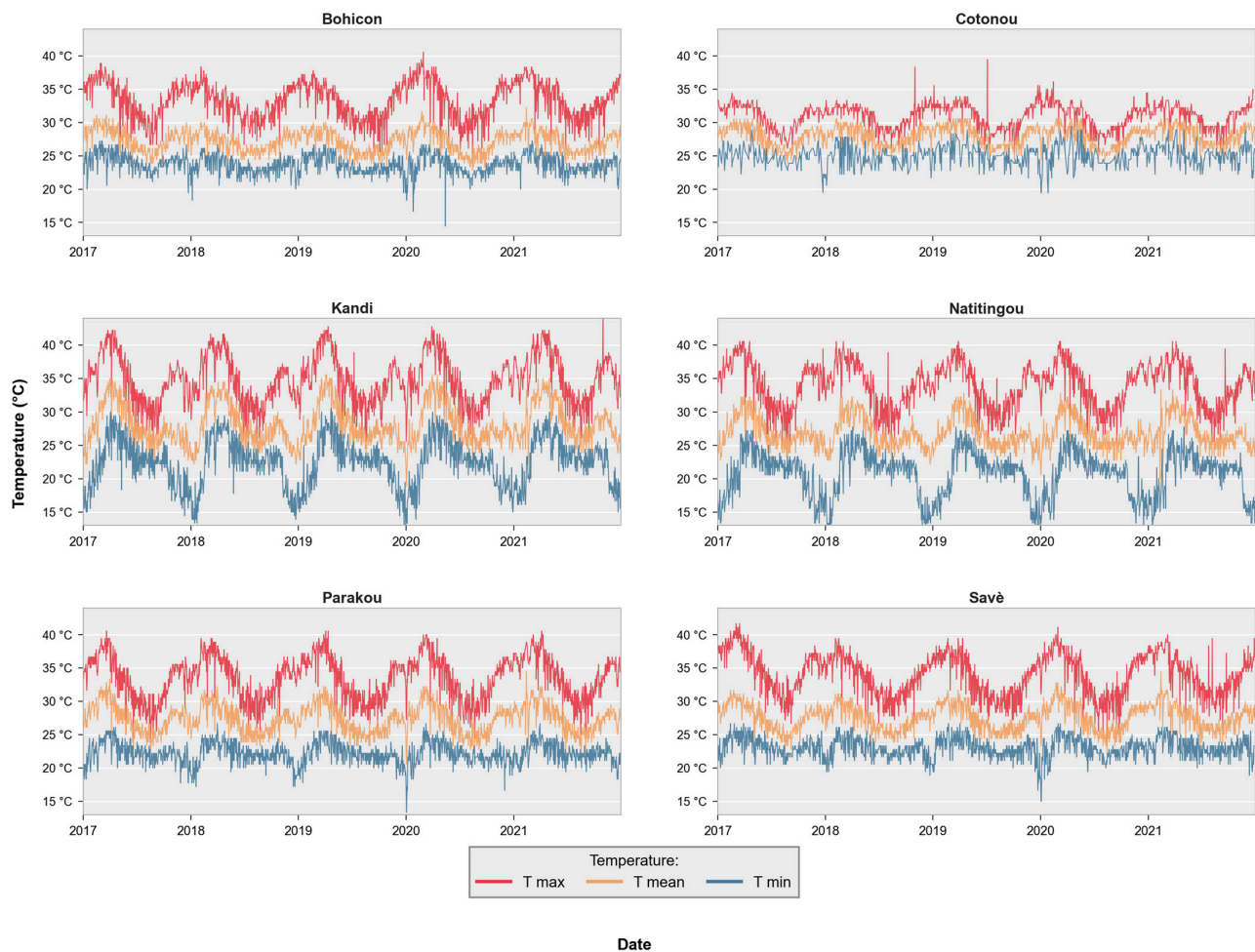


Figure 2. Variation of observed temperatures (T_{mean} , T_{max} , T_{min}) in the six stations over the 2017–2021 period.

It should be noted that solar radiation, wind speed, and vapor pressure deficit—variables that are mandatory for FAO-56 PM ET_0 computation but were not systematically measured at the six synoptic stations, were extracted from the TerraClimate gridded reanalysis dataset [40]. Consequently, the ET_0 target variable constitutes a hybrid product in which station temperature observations are combined with reanalysis-derived forcing variables. While this approach is widely adopted in data-sparse environments [34,35] and TerraClimate has been validated across West Africa [40], it introduces a systematic reanalysis bias into the training target. This bias is bounded by the TerraClimate uncertainty envelope (RMSE < 15 W m^{−2} for solar radiation; <0.5 m s^{−1} for wind speed) and affects all six stations equally, so that model rankings are internally consistent. Nevertheless, a proportion of the residual prediction error reported in Section 3.3 is attributable to this reanalysis uncertainty rather than to genuine model deficiencies.

2.3. Methodological Approach

2.3.1. Estimation of Observed ET_0 Using the FAO-56 Penman–Monteith Method

The estimation of the observed ET_0 at the six stations over the 2017–2021 period was based on the standard FAO Penman–Monteith (FAO-56 PM) method. The FAO-56 PM equation is a simple and close representation of the physical and physiological factors

governing the evapotranspiration process, which has been shown to provide consistent estimates of ET_0 in many regions and climates [6]. It is based on the formula in Equation (1):

$$ET_0 = [0.408\Delta(R_n - G) + \gamma(900/(T + 273.15))u_2(e_s - e_a)]/[\Delta + \gamma(1 + 0.34u_2)] \quad (1)$$

where R_n is the net radiation at the reference surface ($MJ\ m^{-2}\ day^{-1}$); G is the soil heat flux density ($MJ\ m^{-2}\ day^{-1}$); T is the near surface daily temperature at 2 m ($^{\circ}C$); u_2 is the wind speed at 2 m ($m\ s^{-1}$); e_s is the saturation vapor pressure (kPa); e_a is the actual vapor pressure (kPa); $e_s - e_a$ represents the saturation vapor pressure deficit (kPa); Δ is the slope of the saturation vapor pressure curve ($kPa\ ^{\circ}C^{-1}$); and γ is the psychrometric constant ($kPa\ ^{\circ}C^{-1}$). For monthly time steps, the soil heat flux density G was set to zero in accordance with standard FAO-56 guidelines [6], as monthly G is negligible relative to net radiation.

2.3.2. Machine Learning Modeling of ET_0 FAO-56 PM

Overview of the Regression Algorithms

Linear Regression (LR) is a simple but powerful machine learning algorithm which works by fitting a linear equation to the training data, where the goal is to minimize the difference between the predicted values and the actual values. The equation takes the form $y = mx + b$, where y is the value of the target variable, x is the input variable, m is the slope and b is the y-intercept.

Decision Tree (DT) works by constructing a tree-like model of decisions and their possible consequences, where each node represents a decision based on an input variable and each branch represents the outcome of the decision. For regression, the model minimizes the residual sum of squares (RSS) at each split to identify the most informative partition of the predictor space.

Random Forest (RF) is a popular supervised ML algorithm that consists of a set of decision trees used to predict a quantitative or qualitative variable. Developed by Breiman [42], it is based on a set of randomly computed decision trees, with the final prediction being the average of all predictions in the case of regression.

K-Nearest Neighbors (KNN) is a simple supervised ML algorithm first developed by Fix and Hodges [43], which predicts the value of a new input by finding its k nearest neighbors in the training set (using Euclidean distance) and computing their average value.

Support Vector Machines (SVM) are a supervised learning algorithm used in both regression and classification problems, developed by Cortes and Vapnik [44]. For regression, SVM finds the function that deviates from the observed values by a value no greater than ϵ , while simultaneously being as flat as possible.

Cubist is a hybrid algorithm that combines linear regression and decision trees into one powerful model. It is a rule-based model that is an extension of Quinlan's M5 model tree [45]. Terminal leaves contain linear regression models, and a prediction is smoothed by considering the linear model at the previous node.

Extreme Gradient Boosting (XGBoost) is a scalable and distributed gradient-boosted decision tree developed by Chen and Guestrin [46]. It combines the results of a set of simpler decision trees using boosting, where each tree corrects the residuals of the previous one, resulting in strong predictive performance.

Data Preparation and Processing

The preparation of the datasets included (i) spatial data transformation, (ii) extraction of predictor variables, and (iii) handling of missing values prior to ET_0 modeling. Two types of missing data were identified in the station observations. The first corresponded to entirely missing days, whose proportions remained low and relatively homogeneous across stations, with Cotonou recording the fewest missing days. The second was related to

missing temperature values for $T_{\max}$ and $T_{\min}$. Overall, missing data proportions remained low to moderate for most stations. No missing values were observed for mean temperature.

To standardize the temporal structure of the station data, a complete daily grid was first constructed. Missing station values were then imputed using the Multiple Imputation by Chained Equations (MICE) approach [47], performed independently for each station using predictive mean matching (PMM) with $m = 5$ imputations. The assumption of missing at random (MAR) was considered plausible because the missingness was mainly associated with instrumental failures. To avoid information leakage, this imputation model was calibrated exclusively on the training set and then applied to the test data.

Model Implementation

The dataset was partitioned using a strict temporal criterion into a training subset covering 2017–2019 (216 observations, 60%) and an independent test subset covering 2020–2021 (144 observations, 40%). This temporal split ensures that model performance is evaluated on data strictly posterior to the training period, providing an honest assessment of predictive ability on unseen temporal data [48]. Model implementation was carried out using the Caret package in R [49]. Hyperparameter tuning was performed using repeated cross-validation with 10 folds and 10 repetitions for all models. The hyperparameters of the seven models are presented in Table A3.

To assess the spatial generalizability of the ML models, that is, their ability to predict ET_0 at locations not represented during training, a spatial leave-one-station-out (LOSO) cross-validation was performed for all seven individual models and the BMA ensemble. In each of six folds, one synoptic station was withheld entirely (all years, 2017–2021), and the remaining five stations were used to train the model, which was then evaluated on the held-out station. This procedure was repeated six times (one per station), producing station-level RMSE, RMSE%, R^2 , and Bias estimates.

Assessment of Model Performance

Model performance was evaluated exclusively on the independent temporal test set (2020–2021). Two metrics were used: (i) the root mean square error expressed as a percentage of the mean observed ET_0 (RMSE%), which provides a scale-invariant measure of prediction error, and (ii) the coefficient of determination (R^2), which measures the proportion of variance in the observed ET_0 explained by the model. These metrics are defined as:

$$RMSE (\%) = [\sqrt{(1/N \times \sum_i (y_i - \hat{y}_i)^2) / \bar{y}}] \times 100 \quad (2)$$

$$R^2 = 1 - [\sum_i (y_i - \hat{y}_i)^2] / [\sum_i (y_i - \bar{y})^2] \quad (3)$$

where y_i is the observed ET_0 , $\hat{y}_i$ is the predicted ET_0 , $\bar{y}$ is the mean observed ET_0 over the test set, and N is the number of test observations. Systematic bias expressed as a percentage of mean observed ET_0 (Bias%) was additionally computed as:

$$Bias (\%) = [(1/N) \times \sum_i (\hat{y}_i - y_i) / \bar{y}] \times 100 \quad (4)$$

where positive values indicate systematic overestimation.

To assess whether differences in predictive performance among models are statistically significant, pairwise Wilcoxon signed-rank tests were applied to squared prediction errors on the 144 test observations (2020–2021). The non-parametric Wilcoxon test was preferred because squared errors are right-skewed and sample sizes are small. Significance was evaluated at $\alpha = 0.05$. The null hypothesis in each test is that the two models have equal median squared prediction error.

2.3.3. Bayesian Model Averaging and Uncertainty Quantification

To produce the final 1 km monthly ET_0 maps and quantify prediction uncertainty, Bayesian Model Averaging (BMA) was applied to all seven models following Picard et al. [36]. BMA estimates the posterior probability w_k of each model k as a latent variable via the Expectation–Maximization (EM) algorithm. Weights are initialised equally ($w_k = 1/K$). Each EM iteration alternates between an E-step that computes the responsibility of model k for observation i :

$$z_{ik} = w_k \varphi(y_i | \hat{y}_{ik}, \sigma_k^2) / \sum_j w_j \varphi(y_i | \hat{y}_{ij}, \sigma_j^2) \quad (5)$$

where $\varphi(\cdot)$ is the Gaussian density with within-model variance σ_k^2 ; and an M-step that updates the weights and variances:

$$w_k = (1/N) \sum_i z_{ik} \quad (6)$$

$$\text{with } \sigma_k^2 = [\sum_i z_{ik} (y_i - \hat{y}_{ik})^2] / \sum_i z_{ik}$$

The algorithm converged in 321 iterations. The resulting BMA posterior weights (with the corresponding within-model residual standard deviation σ_k in mm month⁻¹ in parentheses) were: DT = 41.8% ($\sigma = 7.9$), KNN = 25.4% ($\sigma = 4.2$), SVM = 21.7% ($\sigma = 3.8$), Cubist = 8.7% ($\sigma = 10.5$), XGBoost = 2.4% ($\sigma = 2.9$), RF $\approx 0\%$ ($\sigma = 11.6$), LR $\approx 0\%$ ($\sigma = 11.0$). The EM algorithm assigns higher weights to models with lower within-model residual variance σ_k^2 , rather than simply rewarding low test RMSE. The BMA ensemble prediction at each pixel was then computed as the weighted average:

$$ET_{0p}^B = \sum_k w_k \times ET_{0k} \quad (7)$$

Prediction uncertainty was quantified using the BMA inter-model standard deviation (σ_{inter}), which measures the spread of individual model predictions around the BMA ensemble mean:

$$\sigma_{\text{inter}}^2 = \sum_k w_k \times (ET_{0k} - ET_{0p}^B)^2 \quad (8)$$

$$\sigma_{\text{inter}} = \sqrt{(\sigma_{\text{inter}}^2)} \quad (9)$$

A high σ_{inter} indicates large disagreement among models and thus higher prediction uncertainty, whereas a low σ_{inter} indicates strong model consensus. The spatial coefficient of variation (CV_{inter} , %) was also computed as:

$$CV_{\text{inter}} = (\sigma_{\text{inter}} / ET_{0p}^B) \times 100 \quad (10)$$

This provides a relative measure of inter-model uncertainty that is comparable across regions with different ET_0 magnitudes.

3. Results

3.1. Estimation of Observed ET_0 FAO-56 PM

The monthly FAO-56 PM ET_0 values calculated for the six stations over the 2017–2021 period are presented in Figure 3 and summarized in Table 1. Monthly ET_0 across the six stations ranged between 73.4 and 157.8 mm month⁻¹, with station-level means between 106.0 mm month⁻¹ (Cotonou) and 118.5 mm month⁻¹ (Kandi), consistent with the north-south climatic gradient of Benin. The largest seasonal range was observed at Kandi (94.4–157.8 mm month⁻¹), reflecting the marked contrast between dry and rainy conditions at this northern station. Cotonou, the southernmost coastal station, showed the smallest

range (78.4–128.0 mm month^{−1}). The grand mean ET₀ across all stations and months was 110.5 mm month^{−1}, which was used as the denominator for computing RMSE (%).

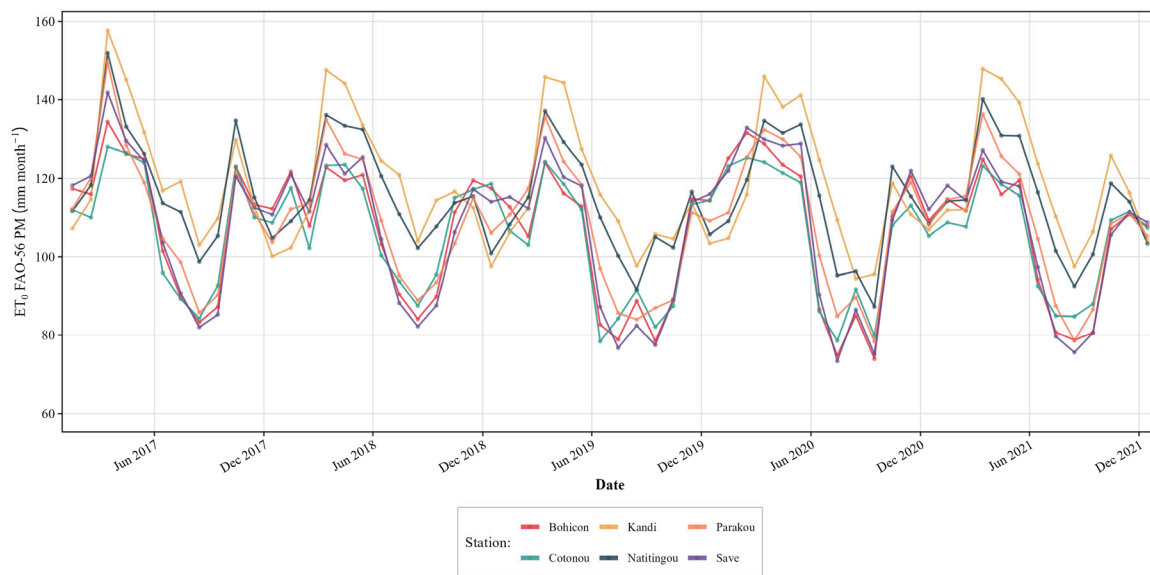


Figure 3. Monthly FAO-56 PM ET₀ variation at the six synoptic stations (2017–2021).

Table 1. Monthly FAO-56 PM ET₀ statistics by station (mm month^{−1}) for the 2017–2021 period ($n = 60$ months per station).

Station	<i>n</i>	Min	Mean	Max	Std Dev
Bohicon	60	74.0	106.5	134.4	17.0
Cotonou	60	78.4	106.0	128.0	14.9
Kandi	60	94.4	118.5	157.8	16.1
Natitingou	60	87.2	115.3	152.0	13.7
Parakou	60	78.3	109.1	149.6	16.3
Savè	60	73.4	107.6	141.8	18.3

3.2. Variable Importance Analysis

The variable importance analysis revealed a clear and consistent pattern across all seven models (Figure 4). Among the 10 predictors, MODIS Land Surface Temperature (LST) emerged as the overwhelmingly dominant predictor across all seven models, achieving maximum relative importance (100%) across all models. This result is physically grounded: LST integrates net radiation, soil moisture, and aerodynamic resistance—the primary controls on Penman–Monteith evaporative demand. Crucially, LST is a spatially variable predictor, ensuring that the 1 km ET₀ maps carry genuine spatial information. Cyclic month encoding (month_sin) ranked second (mean: 54%), encoding intra-annual solar-geometry seasonality. The cosine component (month_cos) provided complementary phase information (mean: 22%). Among Sentinel-2 optical indices, MSI ranked highest (mean: 33%), followed by NDVI (24%), NDMI (21%), NDRE (20%), NDWI (20%), and EVI (18%). Elevation had a mean importance of 14%. Overall, the predictor set confirms that ET₀ in Benin is primarily controlled by the thermal surface state (LST) and seasonal timing, with optical vegetation indices providing secondary information.

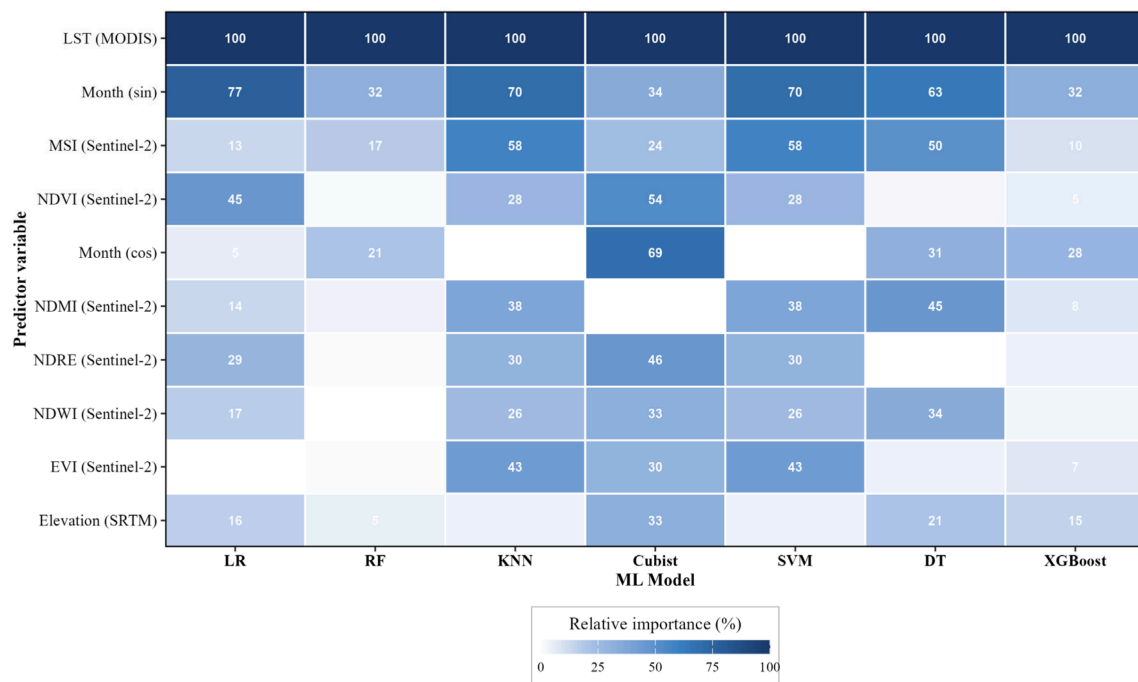


Figure 4. Relative importance of the 10 predictor variables across the seven models. Values represent the percentage of maximum relative importance (0–100%), normalised so that the most influential predictor for each model scores 100%.

3.3. Model Performance Evaluation

The performance of the seven models on the independent temporal test set (2020–2021) is presented in Table 2 and Figure 5. All models achieved positive predictive performance ($R^2 > 0.66$), confirming that the 10 predictors capture a substantial portion of ET_0 variability. All models showed a small positive systematic bias (Bias% = +1.6–3.1%), indicating a tendency to overestimate ET_0 .

Table 2. Performance metrics of the seven models and BMA ensemble on the independent test set (2020–2021). RMSE% and Bias% are expressed as percentages of the grand mean observed ET_0 (110.5 mm month^{−1}). Positive bias indicates systematic overestimation. BMA posterior weights estimated by EM algorithm: DT = 41.8%, KNN = 25.4%, SVM = 21.7%, Cubist = 8.7%, XGBoost = 2.4%, RF ≈ 0%, LR ≈ 0%.

Model	RMSE (mm month ^{−1})	RMSE (%)	R ²	Bias (mm month ^{−1})	Bias (%)
Cubist	8.06	7.3	0.787	2.79	2.5
BMA	7.76	7.0	0.802	2.57	2.3
RF	8.82	8.0	0.745	3.40	3.1
XGBoost	8.55	7.7	0.760	3.02	2.7
KNN	8.79	8.0	0.746	2.83	2.6
SVM	8.47	7.7	0.765	2.48	2.2
DT	8.26	7.5	0.776	2.39	2.2
LR	10.15	9.2	0.662	1.75	1.6

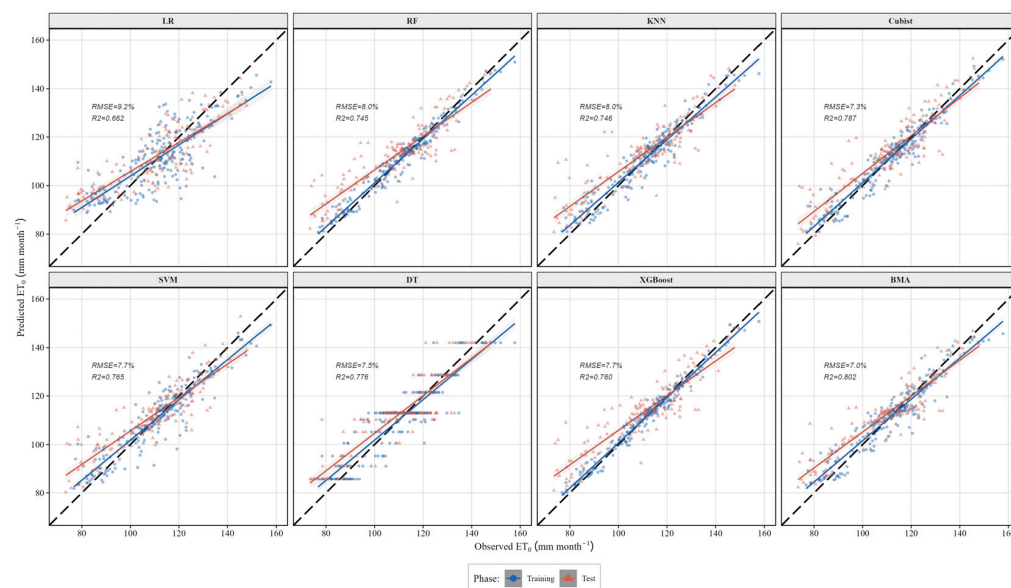


Figure 5. Performance comparison of the seven models and the BMA ensemble on the test set (2020–2021). Dashed line: 1:1 reference; metrics shown for test phase. Panels show observed vs. predicted ET_0 for each model, with RMSE% and R^2 annotated. The dashed line is the 1:1 reference; the solid line is the ordinary least-squares linear fit (intercept unconstrained).

The BMA ensemble achieved the highest accuracy (RMSE = 7.0%, $R^2 = 0.802$), outperforming all individual models. Cubist was the best individual model (RMSE = 7.3%, $R^2 = 0.787$), followed by DT (RMSE = 7.5%, $R^2 = 0.776$) and SVM (RMSE = 7.7%, $R^2 = 0.765$). XGBoost (RMSE = 7.7%, $R^2 = 0.760$), KNN (RMSE = 8.0%, $R^2 = 0.746$), and RF (RMSE = 8.0%, $R^2 = 0.745$) showed moderate performance, while LR (RMSE = 9.2%, $R^2 = 0.662$) was the weakest. These metrics are expressed as a percentage of the grand mean observed ET_0 ($110.5 \text{ mm month}^{-1}$).

Wilcoxon signed-rank tests on the 144 test-set squared errors confirmed that BMA significantly outperformed LR ($W = 1372$, $p < 0.001$), KNN ($W = 2990$, $p < 0.001$), DT ($W = 2529$, $p < 0.001$), and SVM ($W = 3798$, $p = 0.005$). Differences between BMA and RF ($p = 0.542$) and between BMA and Cubist ($p = 0.065$) were not statistically significant. XGBoost achieved a significantly lower squared error than BMA on the temporal test set ($p < 0.001$); however, the LOSO cross-validation (Table 3) reveals that this advantage does not hold spatially.

Table 3. Spatial Leave-One-Station-Out (LOSO) cross-validation performance for all individual models and the BMA ensemble. RMSE% is expressed as a percentage of mean observed ET_0 ($110.5 \text{ mm month}^{-1}$). Values are averages across six station-level folds; standard deviations across folds ($\pm$ sd) are shown in parentheses for RMSE. Models are ranked by ascending LOSO RMSE.

Model	LOSO RMSE (mm month^{-1})	LOSO RMSE (%)	LOSO R^2	LOSO Bias (mm month^{-1})
BMA	8.21 (± 1.20)	7.4	0.722	−0.67
RF	8.49 (± 1.67)	7.7	0.698	−0.96
XGBoost	8.84 (± 2.29)	8.0	0.666	−1.72
SVM	9.44 (± 1.50)	8.5	0.624	−1.81
KNN	9.46 (± 0.88)	8.6	0.633	−0.88
DT	10.67 (± 2.22)	9.7	0.517	−1.80
Cubist	11.08 (± 3.16)	10.0	0.440	−3.13
LR	11.83 (± 1.90)	10.7	0.408	−0.47

Table 3 presents the LOSO spatial cross-validation results for the models and the BMA ensemble. Under spatial leave-one-station-out evaluation, BMA remained the most

accurate model (LOSO RMSE = 8.21 ± 1.20 mm month⁻¹, RMSE% = 7.4%, $R^2 = 0.722$), followed by RF (8.49 mm, 7.7%, $R^2 = 0.698$) and XGBoost (8.84 mm, 8.0%, $R^2 = 0.666$). SVM and KNN performed similarly (RMSE ≈ 9.4 – 9.5 mm, $R^2 \approx 0.62$ – 0.63), while DT (10.67 mm, $R^2 = 0.517$), Cubist (11.08 mm, $R^2 = 0.440$), and LR (11.83 mm, $R^2 = 0.408$) showed the weakest spatial generalization. Critically, the LOSO ranking diverges sharply from the temporal test ranking: Cubist, which achieved the best individual-model RMSE on the temporal test (7.3%), degraded to the worst LOSO performance ($R^2 = 0.440$), indicating overfitting to the training-station pattern. Conversely, the BMA ensemble, the best model in both evaluations, confirmed the robustness of EM-based ensemble averaging under spatial data scarcity.

The temporal evolution of predicted ET₀ compared to the observed FAO-56 PM series at the six stations is illustrated in Figure 6. Across all stations, the BMA ensemble (solid dark blue) consistently tracked observations more closely than individual models, confirming the added value of ensemble combination. At Bohicon, Parakou, and Savè, all seven models reproduced the seasonal cycle with reasonable accuracy, including the wet-season trough (July–August). At Kandi and Natitingou, characterized by the highest ET₀ values in the northern Sudanian zone, a systematic underestimation was observed during the dry-season peak (January–March), when values approach or exceed 140 mm month⁻¹. This residual bias reflects the limited capacity of the satellite-derived predictors to fully resolve the high evaporative demand driven by wind speed and vapour pressure deficit in the Sudanian dry season. At Cotonou, most models produced a modest overestimation during the transition months (March–May), consistent with the positive systematic bias documented in Table 2, while the wet-season minimum (July–August) was generally well captured.

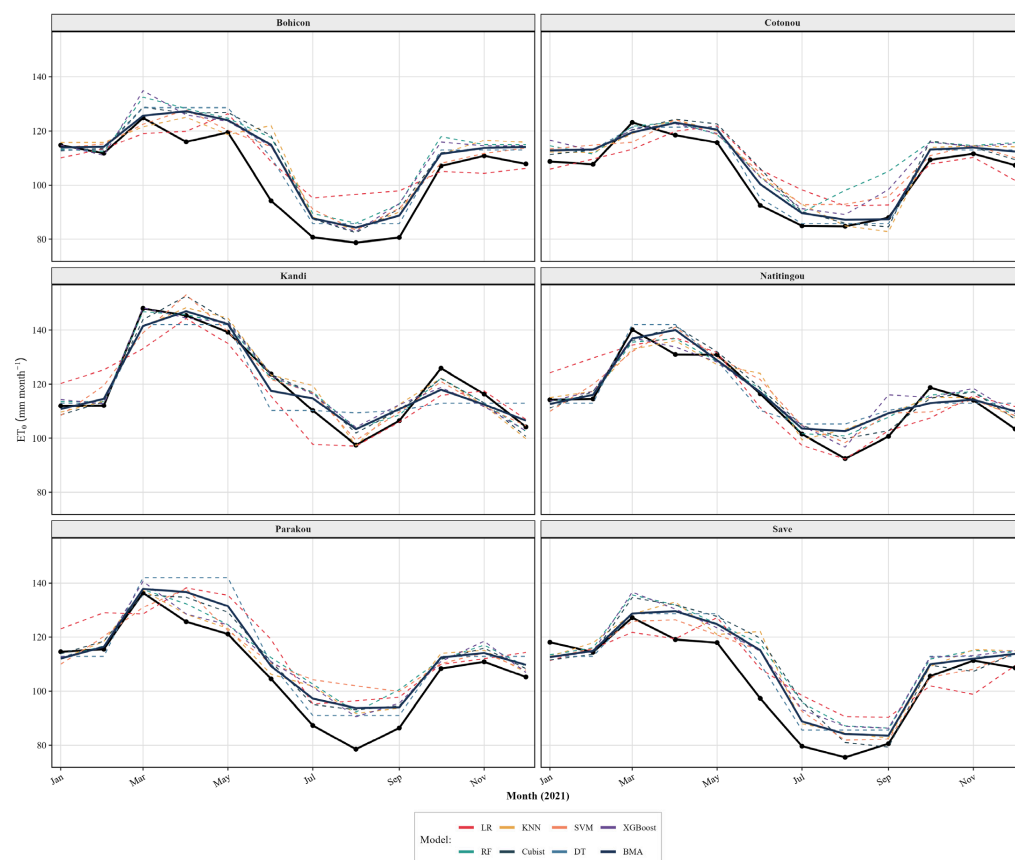


Figure 6. Monthly ET₀ predictions vs. FAO-56 PM observations at the six stations for the year 2021 (final year of the 2020–2021 independent test period; 2020 omitted for visual clarity). Black line: FAO-56 PM observations; coloured lines: seven ML models and BMA ensemble.

3.4. ET_0 Spatial Prediction and Mapping

Following the temporal validation on the 2020–2021 test set, all seven trained ML models were applied to spatially continuous 1 km predictor grids to generate monthly ET_0 maps covering the entire territory of Benin for 2025. The predictor dataset assembled for 2025 comprised the same 10 variables used during training: monthly MODIS daytime LST composites, six Sentinel-2 optical vegetation indices (NDVI, EVI, NDMI, NDWI, MSI, NDRE), the static SRTM elevation grid, and the month-specific cyclic encoding, all retrieved from Google Earth Engine for each of the 12 months. The year 2025 was selected as a prospective application year to demonstrate the temporal generalizability of the satellite-only framework beyond the training and test periods. This prospective application implicitly relies on an assumption of distributional stationarity: that satellite predictor values in 2025, particularly MODIS LST and Sentinel-2 optical indices, remain within the range encountered during the 2017–2021 training period. Changes in land cover, vegetation phenology, or regional surface conditions between 2021 and 2025 could introduce predictor distributions outside the training domain, leading to extrapolation rather than interpolation. Practitioners wishing to apply this framework to future years are encouraged to verify that predictor values remain within the training range before operational use of the resulting ET_0 estimates. The resulting BMA ensemble monthly ET_0 maps are presented in Figure 7.

The BMA ensemble monthly ET_0 maps for 2025 (Figure 7) exhibit a clear north–south gradient, with the highest ET_0 rates in the central and northern Sudanian zone and the lowest values in the southern coastal region around Cotonou. The northern zone, subject to the Harmattan wind and longer dry episodes, consistently shows ET_0 values exceeding $140 \text{ mm month}^{-1}$ during the dry-season peak (November–March), while the more humid equatorial climate of the south produces the lowest values during the wet season (June–September). Monthly BMA-estimated ET_0 ranged from approximately 70 mm month^{-1} in the southern zone during the wet season to over $150 \text{ mm month}^{-1}$ in the northern Sudanian zone during the dry season, consistent with station-based observations. Individual model maps reproduced the same dominant spatial patterns, with minor inter-model differences in the amplitude of the north–south gradient reflecting each algorithm’s structural properties—differences that motivate the use of BMA to produce a consensus ensemble prediction. The BMA posterior weights estimated by the EM algorithm were: DT = 41.8%, KNN = 25.4%, SVM = 21.7%, Cubist = 8.7%, XGBoost = 2.4%, RF $\approx 0\%$, LR $\approx 0\%$. The resulting ensemble effectively smoothed local inter-model inconsistencies while preserving the dominant spatial and seasonal ET_0 patterns.

The spatial distribution of prediction uncertainty, expressed as the BMA inter-model standard deviation (σ_{inter} , Figure 8), revealed that model disagreement was systematically highest in the northern Sudanian zone and during the dry season months (November–March). In this zone, σ_{inter} reached values of $5\text{--}12 \text{ mm month}^{-1}$ during dry season peaks, corresponding to a coefficient of variation (CV_{inter}) of approximately 6–12% (Figure A1). By contrast, the southern sub-equatorial zone showed much lower σ_{inter} values ($1\text{--}4 \text{ mm month}^{-1}$) throughout the year, indicating stronger model consensus in areas of more moderate and stable ET_0 . During the wet season (June–September), σ_{inter} was uniformly low across all three climatic zones, consistent with the convergence of model predictions under conditions of reduced evaporative demand and higher atmospheric humidity.

The spatial pattern of CV_{inter} (Figure A1) further confirms that relative prediction uncertainty is inversely related to ET_0 magnitude: the highest relative uncertainty ($CV_{\text{inter}} > 10\%$) was concentrated in transitional areas between climatic zones and in months where the predictor variables, particularly LST and optical indices, showed the largest spatial gradients across Benin’s topographic and climatic landscape. These findings indicate that future model improvements

should prioritize the northern Sudanian zone and the dry season months, where both absolute and relative prediction uncertainty are highest.

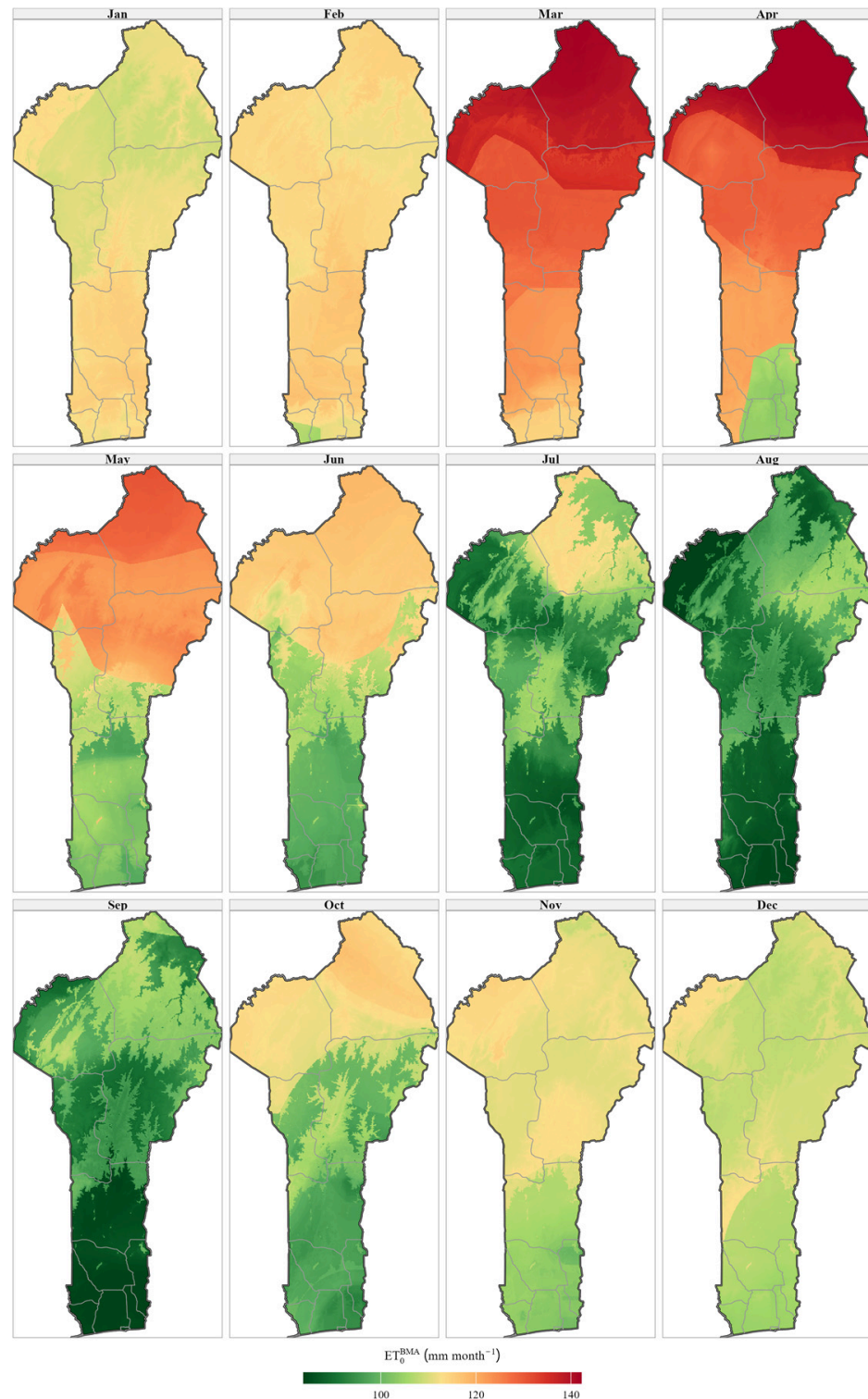


Figure 7. Spatially distributed BMA ensemble monthly ET_0 maps for Benin, 2025 (1 km resolution; 10 remote-sensing predictors). BMA posterior weights estimated by EM algorithm: DT = 41.8%, KNN = 25.4%, SVM = 21.7%, Cubist = 8.7%, XGBoost = 2.4%, RF $\approx$ 0%, LR $\approx$ 0%. The 12 panels correspond to January–December. Departmental boundaries are shown.

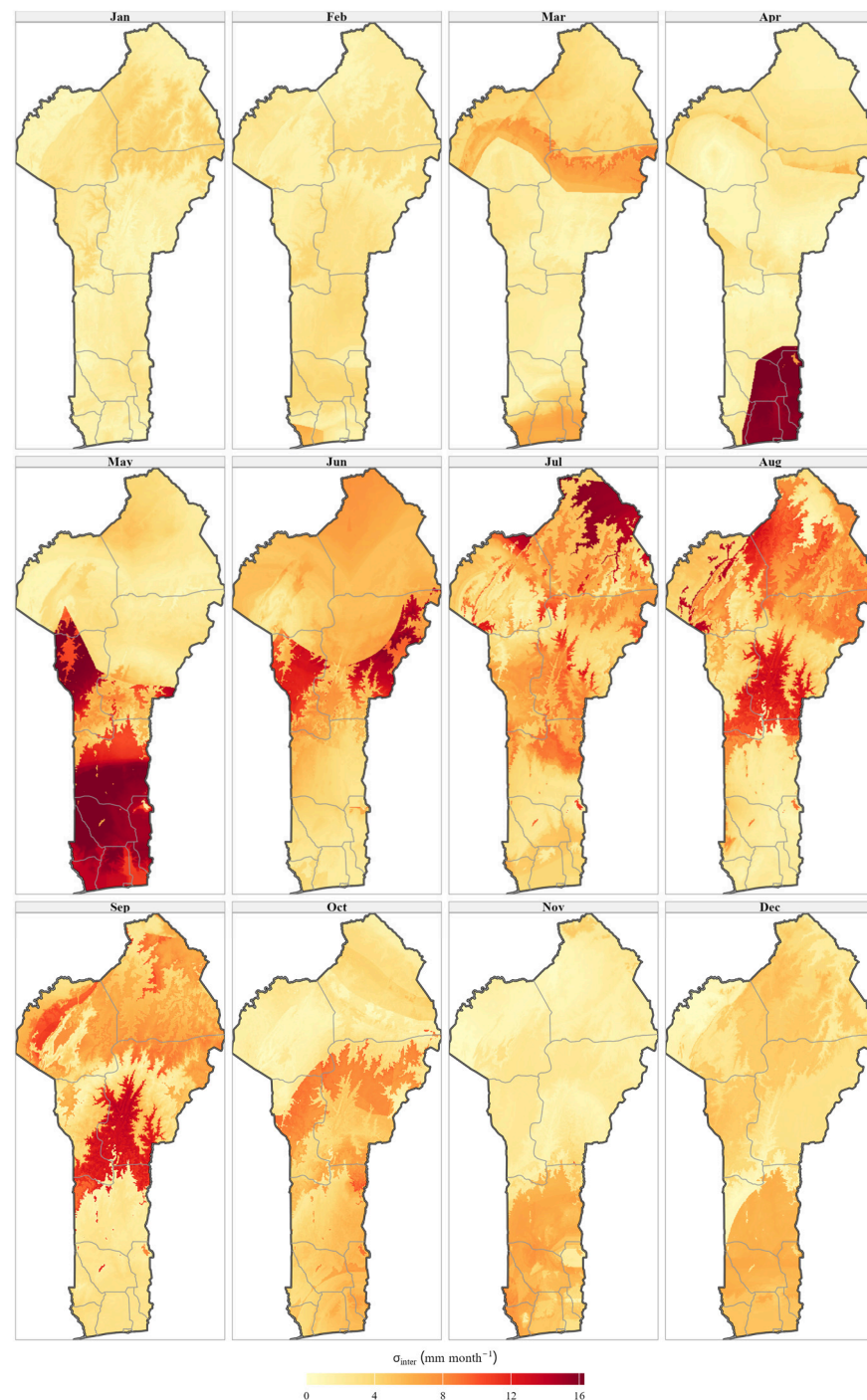


Figure 8. Spatial distribution of the BMA inter-model standard deviation (σ_{inter} , mm month^{-1}) as a measure of prediction uncertainty for Benin, 2025. σ_{inter} is the weighted standard deviation across the 7 ML models. Higher values indicate greater model disagreement. The 12 panels correspond to January–December.

4. Discussion

Evapotranspiration is highly dependent on the interaction of climatic, geographic, biological, and soil factors [6,36]. Thus, the choice of variables to be used for prediction is critical to modeling performance and consequently accuracy. In this study, we investigated the capability of seven models in predicting ET_0 in Benin using 10 remote-sensing predictors: MODIS LST, six Sentinel-2 optical vegetation indices (NDVI, EVI, NDMI, NDWI, MSI,

NDRE), elevation, and cyclic month encoding, all strictly independent of the TerraClimate variables used to construct the FAO-56 PM target.

Among the 10 predictor variables, MODIS LST was the dominant predictor across all seven models, with a mean relative importance of 100% (Figure 4). Because LST is a spatially continuous surface-state variable, the 1 km ET_0 maps carry genuine spatial structure, resolving the apparent tension between temporal importance scores and spatial map validity. LST integrates net radiation, soil moisture, and aerodynamic resistance—the primary controls on Penman–Monteith evaporative demand. Cyclic month encoding (month_sin) ranked second (mean: 54%), encoding intra-annual solar-geometry seasonality. Among Sentinel-2 optical indices, MSI was the most informative (mean: 33%), followed by NDVI (24%), NDMI (21%), NDWI (20%), NDRE (20%), and EVI (18%). Elevation retained moderate importance (mean: 14%), reflecting topographic controls across Benin’s relief gradient [6,50].

The cyclic month encoding captured the intra-annual structure of ET_0 variability driven by solar geometry and radiation regime [51]. Elevation retained moderate importance, especially in tree-based models, reflecting its role as a proxy for topographic controls on temperature, cloudiness, and wind patterns across Benin’s three climatic zones [6,50].

Of the seven models evaluated, the BMA ensemble achieved the best overall performance (RMSE = 7.0%, R^2 = 0.802), followed by Cubist (RMSE = 7.3%, R^2 = 0.787). This ranking is consistent with findings by Dias et al. [52], who reported Cubist as top-performing for ET_0 prediction in Brazil. The advantage of Cubist lies in its rule-based partitioning combined with local linear regression, particularly effective across Benin’s three climatic zones [52–54]. XGBoost also performed well; its boosting framework effectively corrects residual errors from successive weak learners.

The LOSO spatial cross-validation (Table 3) reveals the mechanism underlying this weighting: while Cubist recorded the best temporal test RMSE (7.3%), its LOSO RMSE was the highest among complex models (11.08 mm, R^2 = 0.440), when trained on the five remaining stations. This spatial instability reflects the local rule-based architecture of Cubist, which overfits to the within-station covariate distribution and generalizes poorly to new geographic locations. In contrast, the BMA ensemble’s conservative weighting of Cubist (8.7%) in favour of DT (41.8%) and KNN/SVM ($\approx 20\%$) produced the most spatially robust predictions (LOSO RMSE = 8.21 mm, R^2 = 0.722). These results highlight a key property of the EM-BMA framework: by learning within-model residual variance during training, it implicitly detects and down-weights models with high spatial variability in their errors, without requiring explicit spatial cross-validation.

The predictive performance achieved by the seven ML models (RMSE = 7.3–9.2%, R^2 = 0.662–0.787) relies exclusively on satellite-observable predictors, with no meteorological input variables. This constraint is methodologically deliberate: incorporating variables such as solar radiation, wind speed, or vapour pressure deficit, which enter directly into the FAO-56 PM equation used to construct the ET_0 target, as ML predictors would introduce a circular dependency that artificially inflates apparent model skill without reflecting genuine predictive capacity. The satellite-only framework therefore provides a more rigorous and physically interpretable benchmark. The RMSE range of 7.0–9.2% is consistent with results reported by Yonaba et al. [34] and Landaras et al. [35] for ML-based ET_0 estimation in West Africa using limited climate datasets, confirming that the approach achieves competitive accuracy under data-sparse conditions. It should further be acknowledged that the ET_0 target variable itself carries inherent uncertainty, as solar radiation, wind speed, and vapor pressure deficit were sourced from TerraClimate, a global monthly reanalysis product that, despite its relatively high spatial resolution, retains systematic biases relative to direct station measurements [40]. This propagated uncertainty from the target construction con-

stitutes a practical lower bound on the predictive accuracy attainable by any ML model trained on these data, and partly explains the residual RMSE floor observed across all algorithms. Similar remote-sensing frameworks using MODIS LST as the primary thermal predictor have been successfully applied across diverse climatic regions [20,22,25], further supporting the spatial generalizability of this strategy.

The predictive performance achieved in this study must also be interpreted in light of the structural data scarcity that characterizes Benin's meteorological monitoring landscape, a constraint that simultaneously motivates and justifies the remote-sensing-based ML modeling framework adopted here. With only six operational synoptic stations covering a national territory of 114,763 km², the average density corresponds to approximately one station per 19,000 km², far below the World Meteorological Organization (WMO) recommended density for semi-arid and tropical regions. Beyond their limited number, many weather stations in Benin and across West Africa are affected by chronic underfunding, instrument failures, and unequal temporal coverage, resulting in fragmented and incomplete observation records [10–12]. This structural limitation is directly evident in the present dataset: NCEI station archives yielded only temperature observations (Tmax, Tmin, Tmean), while solar radiation, wind speed, and vapor pressure deficit, all mandatory inputs to the FAO-56 PM equation, were unavailable at the station level and had to be sourced from TerraClimate gridded reanalysis data. The small pool of labelled observations ($n = 216$ station-month records in training; $n = 144$ in the independent test set) further limits the statistical power of model comparisons and restricts the ability to capture the full spatial heterogeneity of ET₀ across Benin's three climatic zones via conventional spatial cross-validation. It is precisely this combination of sparse station coverage, inadequate in both number and functional reliability, and incomplete variable reporting that creates the observational vacuum that satellite-based modeling is uniquely positioned to fill. Freely available, globally consistent satellite observations (MODIS, Sentinel-2) provide spatially continuous, temporally regular measurements that can substitute for the dense ground-based networks absent in Benin and much of sub-Saharan West Africa, enabling national-scale ET₀ mapping without additional ground infrastructure. From this perspective, the BMA inter-model uncertainty maps produced here are not merely diagnostic outputs but actionable tools: by spatially identifying where satellite-derived predictions are least constrained by observational evidence, notably in the northern Sudanian zone during the dry season, they provide direct guidance on where the rehabilitation or installation of weather stations would most effectively improve the accuracy of future ET₀ products across the country.

From a statistical perspective, the training set comprises 216 monthly observations (six stations $\times$ 36 months, 2017–2019) and the test set 144 observations (six stations $\times$ 24 months, 2020–2021). While these sample sizes are small relative to typical ML benchmarks, they are inherent to the monitoring infrastructure of a country where the average station density is one station per 19,127 km² (territory: 114,763 km²; six synoptic stations), among the lowest in West Africa. This scarcity is itself the primary justification for a remote-sensing-based modeling approach: where ground observations are sparse, satellite-derived predictors provide the only spatially continuous information available. Cross-validation experiments (LOSO; Table 3) confirm that despite the small sample, the ensemble framework achieves meaningful spatial generalization ($R^2 \geq 0.72$). Nevertheless, the risk of overfitting is real for complex non-linear models such as XGBoost and Cubist; this is directly reflected in their poor LOSO performance relative to the temporal test set.

The systematic bias across the models is positive (Bias% = +1.6–3.1%; Table 2). Incorporating optical vegetation indices helps reduce systematic errors, although a positive bias persists on the 2020–2021 test set. However, a residual nonlinear bias structure persists:

inspection of residuals versus observed ET_0 confirms that all models slightly underestimate the highest ET_0 values ($>135 \text{ mm month}^{-1}$) at Kandi and Natitingou during the dry season peak, while marginally overestimating at mid-range values. This heteroscedastic pattern suggests that wind speed and vapour pressure deficit, not captured by any of the satellite-derived predictors, remain the primary unaccounted controls during extreme dry-season episodes [55–59].

The BMA ensemble approach provided several advantages over individual model predictions. Following the Expectation-Maximization BMA framework of Picard et al. [36], combining all seven models produced a BMA ensemble ($RMSE = 7.0\%$, $R^2 = 0.802$) that outperforms all individual models, confirming the value of the EM-based ensemble approach. A closer examination of the BMA posterior weights reveals an important subtlety regarding the relationship between individual model performance and BMA weight assignment. Notably, Cubist, the best-performing individual model on the test set ($RMSE = 7.3\%$, $R^2 = 0.787$), received only 8.7% of the BMA weight, while DT received the highest weight (41.8%) despite ranking second on test RMSE. This apparent paradox is resolved by the EM mechanism: weights are assigned in proportion to the Gaussian likelihood of each observation under each model's component, which depends jointly on prediction accuracy and within-model residual variance (σ_k). Cubist exhibits a relatively high training-phase residual standard deviation ($\sigma = 10.5 \text{ mm month}^{-1}$), producing a wide Gaussian component that generates lower per-observation likelihoods even when predictions are accurate. DT, by contrast, combines moderate accuracy with a tighter residual distribution ($\sigma = 7.9 \text{ mm month}^{-1}$), yielding higher average likelihoods and thus greater EM responsibility. KNN ($\sigma = 4.2$) and SVM ($\sigma = 3.8$) receive substantial weights (25.4% and 21.7%) by virtue of their low within-model variance, reflecting consistently precise predictions across the training period. XGBoost presents an additional nuance: despite having the lowest within-model variance ($\sigma = 2.9$), it receives only 2.4% weight. This is likely because its training predictions closely mirrored those of KNN and SVM, models already assigned high EM responsibility, leaving little residual distributional space for XGBoost to explain. The near-zero weights assigned to RF ($\sigma = 11.6$) and LR ($\sigma = 11.0$) effectively reduce the operational BMA to a five-model ensemble: these two models, characterized by the highest within-model variances, generate the broadest and least informative Gaussian components and are consequently outcompeted in the EM likelihood competition. This weight structure underscores a key distinction between BMA and simple test-set ranking: BMA rewards models that are reliably precise across the training distribution, not merely accurate on a holdout set. Second, the BMA inter-model standard deviation (σ_{inter}) provided the first spatially explicit, quantitative characterization of ET_0 prediction uncertainty across Benin. The finding that σ_{inter} is systematically higher in the northern Sudanian zone and during the dry season is physically interpretable: this region combines higher ET_0 values with greater spatial heterogeneity in LST and elevation, generating more divergent model responses. The spatial pattern of CV_{inter} further confirms that relative prediction uncertainty increases toward climatically extreme conditions, consistent with broader findings in ensemble ML literature [37,38,60,61]. Beyond its diagnostic role, σ_{inter} constitutes a spatially explicit validation layer that should accompany the ET_0 maps in practical applications. Grid cells where σ_{inter} remains low reflect strong consensus among the five models that received non-negligible BMA weight, and can be used with high confidence for irrigation scheduling and agricultural water balance assessments. Conversely, cells where σ_{inter} is elevated, particularly in the northern Sudanian zone during the dry season, indicate that satellite-derived predictors generate divergent model responses; these areas should be interpreted with caution and cross-referenced with ground-truth data before operational deployment. In the absence of spatial cross-validation, σ_{inter} thus serves as a first-order proxy for spatial

prediction reliability, guiding end-users toward areas where model estimates can be acted upon versus areas where additional validation is required.

The spatial distribution of monthly variations in ET_0 showed that the central and northern parts of Benin experience the highest ET_0 rates, corresponding to the driest tropical regions of the country, which are characterized by longer dry episodes and exposure to the Harmattan, a hot, dry wind from the Sahara that blows during the dry season [62,63]. Conversely, lower ET_0 rates were predicted in the southern regions, especially near the coast. This spatial pattern is consistent with the study of Obada et al. [14], who documented a general decreasing trend in ET_0 at Cotonou. The ET_0 information generated in this study constitutes the climatic basis for estimating crop water requirements under standard conditions ($ET_c = K_c \times ET_0$) [6], and spatially explicit ET_0 maps can therefore support irrigation planning, crop-water assessment and more efficient water use, as demonstrated in northern Benin by Bouraima et al. [64] and at the broader West African scale by Gbode et al. [65].

The 2025 prediction maps were generated by applying models trained on 2017–2021 data to the same predictor set acquired for 2025. This approach implicitly assumes that the statistical relationships between satellite predictors and ET_0 , calibrated over the 2017–2021 training period, remain stable through 2025 (temporal stationarity assumption). While intra-annual seasonality captured by LST and cyclic month encodings is expected to be stable, gradual land-cover changes (urbanization, agricultural expansion) or multi-year rainfall anomalies in West Africa could shift the predictor space beyond the training domain for specific locations. The BMA inter-model standard deviation map (Figure 8) implicitly identifies regions where such extrapolation uncertainty is largest: areas with high σ_{inter} exhibit both high ET_0 variability and high inter-model disagreement, and should be interpreted with additional caution when used for long-term planning beyond 2025.

Several limitations should be considered when interpreting the results of this study. First, the modeling framework relies on only six synoptic stations covering a national territory of 114,763 km² (one station per 19,127 km²), which is insufficient to capture the full spatial variability of ET_0 across Benin's three climatic zones. Although LOSO cross-validation (Table 3) confirms spatial generalizability for the ensemble, performance degraded notably for the northern Sudanian station at Kandi (LOSO RMSE ≥ 10 mm for all models), likely reflecting the extreme seasonal amplitude of the Harmattan dry season that is underrepresented in the training data. Second, the satellite feature space lacks continuous dynamic indicators for wind speed and vapor pressure deficit, resulting in a structural underestimation of the highest peak ET_0 values (>135 mm month^{−1}) observed during the northern dry season (November–March). Third, the ET_0 target variable is a hybrid product: station temperature observations are combined with TerraClimate reanalysis for solar radiation, wind speed, and vapor pressure deficit. This propagates reanalysis uncertainties into the training target, bounding the achievable validation accuracy independently of model skill. Fourth, MICE-based imputation of fragmented temperature records introduces minor statistical smoothing artifacts into the early training distributions, potentially underestimating inter-monthly variability at stations with the highest missing-data rates (Cotonou: 6.8%). Fifth, the temporal stationarity assumption underlying the 2025 projection maps has not been formally tested; users are advised to treat the 2025 maps as a best-available estimate under observed 2017–2021 climatic conditions, cross-checking against local hydro-agronomic records where available.

5. Conclusions

Reference evapotranspiration is a complex process governed by interacting thermal, radiative, aerodynamic, and seasonal controls, which makes its prediction and spatial-

ization particularly challenging in data-sparse environments such as West Africa. In this study, seven machine-learning models were evaluated for the prediction and mapping of monthly ET_0 in Benin using 10 remote-sensing predictors, MODIS LST, six Sentinel-2 optical vegetation indices (NDVI, EVI, NDMI, NDWI, MSI, NDRE), elevation, and cyclic month encoding, a predictor set entirely independent of the TerraClimate variables used to construct the FAO-56 PM target. To ensure temporal consistency across the full 2017–2021 modeling period, only satellite predictor sources with uninterrupted global coverage were retained.

Among the seven models evaluated, the BMA ensemble achieved the highest accuracy on the test set (RMSE = 7.0% of mean ET_0 , $R^2 = 0.802$), followed by Cubist (RMSE = 7.3%, $R^2 = 0.787$) and DT (RMSE = 7.5%, $R^2 = 0.776$). All models showed positive predictive performance ($R^2 > 0.66$). Variable importance analysis revealed that MODIS LST was the dominant predictor across all models (mean relative importance: 100%), with cyclic month encoding (month_sin) ranking second (mean: 54%), followed by Sentinel-2 optical indices (MSI: 33%, NDVI: 24%). The dominance of LST, a spatially variable predictor, supports the spatial structure of the ET_0 maps.

Bayesian Model Averaging (BMA) of all seven models, with posterior weights estimated via the EM algorithm (Picard et al. [36]), produced spatially coherent monthly ET_0 maps at 1 km resolution for Benin for 2025, reproducing the expected north–south gradient and seasonal cycle. The BMA inter-model standard deviation provided the first spatially explicit quantification of ET_0 prediction uncertainty in Benin, identifying the northern Sudanian zone during the dry season as the region of highest uncertainty. This finding provides actionable guidance for future monitoring efforts: additional weather stations and ground-based validation data in the northern Sudanian zone would most effectively reduce prediction uncertainty in this spatially important region.

For practitioners and water management authorities in Benin, the 2025 BMA monthly ET_0 maps and their associated inter-model uncertainty layers are directly actionable. Monthly BMA ET_0 estimates for the dry season (November–March) in the Sudanian zone (northern Benin) should be used to calibrate irrigation scheduling for rice and vegetable crops in the Niger and Mekrou river valleys, where the highest ET_0 rates were observed (≥ 135 mm month^{−1}). Grid cells where the BMA inter-model standard deviation exceeds 10 mm month^{−1} should be flagged as high-uncertainty zones and cross-validated against any available local agrometeorological station data before operational use. The availability of annual ET_0 maps at 1 km resolution enables a transition from point-based to spatially continuous irrigation water demand estimation, which is urgently needed for the 2030 water security objectives of Benin’s Plan de Développement du Secteur Agricole (PDSA).

Author Contributions: Conceptualization, B.C.F.M. and M.M.; methodology, B.C.F.M. and M.M.; software, B.C.F.M. and M.M.; validation, B.C.F.M., M.M. and V.R.H.; formal analysis, B.C.F.M. and M.M.; investigation, B.C.F.M.; resources, V.R.H. and S.A.R.M.A.; data curation, B.C.F.M.; writing—original draft preparation, B.C.F.M., M.M. and C.A.O.; writing—review and editing, S.A.R.M.A., V.R.H. and C.A.O.; visualization, B.C.F.M.; supervision, V.R.H. and S.A.R.M.A.; project administration, V.R.H.; funding acquisition, NA. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the findings of this study are publicly available. The satellite data used in this study are accessible through Google Earth Engine (<https://earthengine.google.com/>, accessed on 23 June 2026), while the meteorological data are publicly available from the National Centers for Environmental Information (NCEI) of the National Oceanic and Atmospheric Administration (NOAA) at <https://www.ncei.noaa.gov/>, accessed on 23 June 2026. No new datasets were generated during this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

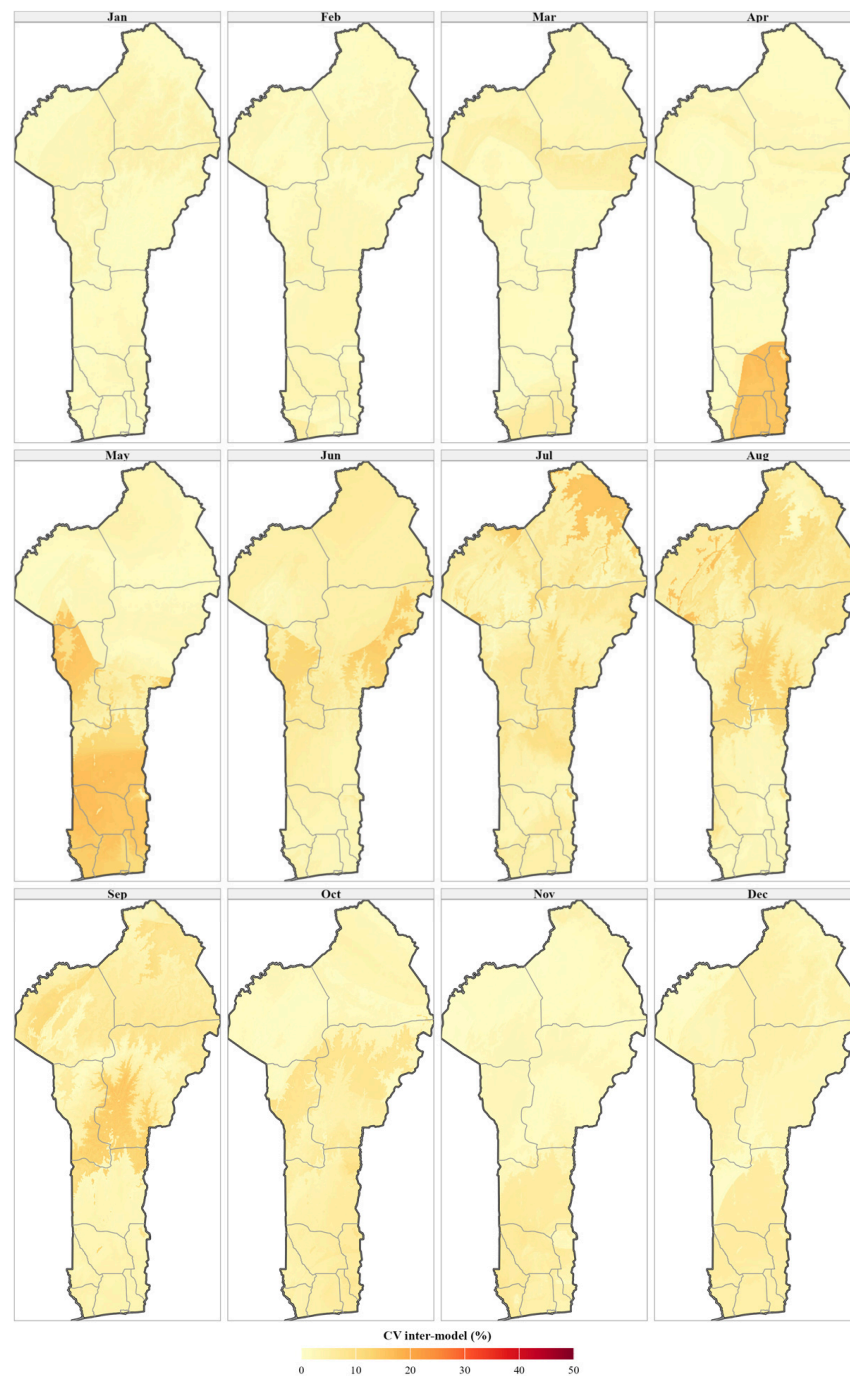


Figure A1. Spatial distribution of the BMA inter-model coefficient of variation (CV_{inter} , %) for Benin, 2025 (1 km resolution). $CV_{inter} = (\sigma_{inter} / ET_0^{BMA}) \times 100$ is a scale-invariant measure of relative inter-model spread, expressing prediction uncertainty as a percentage of the local BMA ensemble mean ET_0 . Higher values indicate greater relative model disagreement. Unlike the absolute standard deviation (σ_{inter} , Figure 8), CV_{inter} is highest during the wet season (June–September) in the southern Guinean zone, where ET_0 values are lowest, and lowest in the northern Sudanian zone during the dry season peak, where high ET_0 magnitudes reduce the relative spread. The 12 panels correspond to January–December.

Table A1. Predictor variables used as inputs to the machine learning models (10 variables: MODIS LST, Sentinel-2 optical indices, SRTM elevation, and temporal encoding).

Variable	Unit	Abbreviation	Source
MODIS daytime LST	°C	LST	MODIS MOD11A2 (Terra)
NDVI	-	NDVI	Sentinel-2 surface reflectance (GEE)
EVI	-	EVI	Sentinel-2 surface reflectance (GEE)
NDMI	-	NDMI	Sentinel-2 surface reflectance (GEE)
NDWI	-	NDWI	Sentinel-2 surface reflectance (GEE)
MSI	-	MSI	Sentinel-2 surface reflectance (GEE)
NDRE	-	NDRE	Sentinel-2 surface reflectance (GEE)
Elevation	m	elev	SRTM (static)
Month (sine component)	-	month_sin	Derived (cyclic encoding)
Month (cosine component)	-	month_cos	Derived (cyclic encoding)

Table A2. Characteristics of the six synoptic stations with mean annual temperature for 2017–2021.

Station	Longitude (°E)	Latitude (°N)	Elevation (m)	T _{mean} (°C)	T _{max} (°C)	T _{min} (°C)
Bohicon	2.07	7.17	167	27.31	33.29	23.58
Cotonou	2.38	6.35	4	27.73	31.79	24.48
Kandi	2.93	11.13	292	28.12	34.93	22.51
Natitingou	1.48	10.32	460	27.06	34.11	21.19
Parakou	2.62	9.35	393	27.27	33.47	22.25
Savè	2.47	8.03	198	27.61	34.10	23.03

Table A3. Hyperparameters of the seven models.

Model	Hyperparameter	Value
DT	cp	0.0078
KNN	k	5
	Distance	2
	Kernel	optimal
LR	(Intercept)	Fitted from training data
RF	mtry	5
SVM	sigma	0.127
	C	2
XGBoost	nrounds	144
	max_depth	6
	eta	0.05
	colsample_bytree	0.8
	subsample	0.8
Cubist	committees	20
Cubist	neighbors	5

Table A4. Pearson correlation matrix of the predictor variables used in the models ($n = 1080$ station-month observations, 2007–2021). Bold values indicate $|r| > 0.70$.

	LST_Day	NDVI	EVI	LAI	FPAR	Elevation	sin (m)	cos (m)
LST_day	—	−0.48	−0.44	−0.25	−0.14	0.25	0.33	0.20
NDVI	−0.48	—	0.97	0.77	0.72	0.23	−0.54	−0.33
EVI	−0.44	0.97	—	0.76	0.72	0.26	−0.53	−0.41
LAI	−0.25	0.77	0.76	—	0.96	0.49	−0.26	−0.18
FPAR	−0.14	0.72	0.72	0.96	—	0.46	−0.24	−0.12
Elevation	0.25	0.23	0.26	0.49	0.46	—	−0.00	−0.00
sin (m)	0.33	−0.54	−0.53	−0.26	−0.24	−0.00	—	−0.00
cos (m)	0.20	−0.33	−0.41	−0.18	−0.12	−0.00	−0.00	—

References

- He, C.; Liu, Z.; Wu, J.; Pan, X.; Fang, Z.; Li, J.; Bryan, B.A. Future global urban water scarcity and potential solutions. *Nat. Commun.* **2021**, *12*, 4667. [\[CrossRef\]](#)
- Jägermeyr, J.; Gerten, D.; Schaphoff, S.; Heinke, J.; Lucht, W.; Rockström, J. Integrated crop water management might sustainably halve the global food gap. *Environ. Res. Lett.* **2016**, *11*, 025002. [\[CrossRef\]](#)
- Kummu, M.; Guillaume, J.; de Moel, H.; Eisner, S.; Flörke, M.; Porkka, M.; Siebert, S.; Veldkamp, T.I.E.; Ward, P.J. The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. *Sci. Rep.* **2016**, *6*, 38495. [\[CrossRef\]](#) [\[PubMed\]](#)
- Douville, H.; Ribes, A.; Decharme, B.; Alkama, R.; Sheffield, J. Anthropogenic influence on multidecadal changes in reconstructed global evapotranspiration. *Nat. Clim. Change* **2013**, *3*, 59–62. [\[CrossRef\]](#)
- Mao, J.; Fu, W.; Shi, X.; Ricciuto, D.M.; Fisher, J.B.; Dickinson, R.E.; Wei, Y.; Shem, W.; Piao, S.; Wang, K.; et al. Disentangling climatic and anthropogenic controls on global terrestrial evapotranspiration trends. *Environ. Res. Lett.* **2015**, *10*, 094008. [\[CrossRef\]](#)
- Allen, R.G.; Pereira, L.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO Irrigation and Drainage Paper 56; Food and Agriculture Organization: Rome, Italy, 1998.
- Guo, D.; Westra, S.; Maier, H.R. Sensitivity of potential evapotranspiration to changes in climate variables for different Australian climatic zones. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 2107–2126. [\[CrossRef\]](#)
- Jian, S.; Wang, A.; Su, C.; Wang, K. Prediction of future spatial and temporal evolution trends of reference evapotranspiration in the Yellow River Basin, China. *Remote Sens.* **2022**, *14*, 5674. [\[CrossRef\]](#)
- World Development Indicators—Agriculture, Forestry, and Fishing, Value Added (% of GDP), Benin. Available online: <https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=BJ> (accessed on 31 March 2023).
- Vodounou, J.B.K.; Doubogan, Y.O. Agriculture paysanne et stratégies d'adaptation au changement climatique au Nord-Bénin. *Eur. J. Geogr.* **2016**. [\[CrossRef\]](#)
- Ogoujalé, E. *Changement Climatique dans le Bénin Méridionale et Central: Indicateurs, Scenarios et Perspectives de la Sécurité Alimentaire*; Université d'Abomey-Calavi: Abomey-Calavi, Benin, 2006.
- Agossou, D.S.M.; Tossou, C.R.; Vissoh, V.P.; Agbossou, K.E. Perception des perturbations climatiques, savoirs locaux et stratégies d'adaptation des producteurs agricoles béninois. *Afr. Crop Sci. J.* **2012**, *20*, 565–588.
- UNDP (United Nations Development Programme). *La Maîtrise de L'eau pour Renforcer les Moyens de Subsistance au Bénin*; UNDP: Cotonou, Bénin, 2019; Available online: <https://www.undp.org/fr/benin/news> (accessed on 21 March 2023).
- Obada, E.; Alamou, E.; Chabi, A.; Zandagba, J.; Afouda, A. Trends and changes in recent and future Penman-Monteith potential evapotranspiration in Benin (West Africa). *Hydrology* **2017**, *4*, 38. [\[CrossRef\]](#)
- Hargreaves, G.L.; Hargreaves, G.H.; Riley, J.P. Agricultural benefits for Senegal River basin. *J. Irrig. Drain. Eng.* **1985**, *111*, 113–124. [\[CrossRef\]](#)
- Jensen, M.E.; Haise, H.R. Estimating evapotranspiration from solar radiation. *J. Irrig. Drain. Div.* **1963**, *89*, 15–41. [\[CrossRef\]](#)
- Thorntwaite, C.W. An approach toward a rational classification of climate. *Geogr. Rev.* **1948**, *38*, 55–94. [\[CrossRef\]](#) [\[PubMed\]](#)
- Courault, D.; Clastre, P.; Guinot, J.-P.; Seguin, B. Analyse des sécheresses de 1988 à 1990 en France à partir de l'analyse combinée de données satellitaires NOAA-AVHRR et d'un modèle agrométéorologique. *Agronomie* **1994**, *14*, 41–56.
- Oliosio, A.; Jacob, F. Estimation de l'évapotranspiration à partir de mesures de télédétection. *La Houille Blanche* **2002**, *1*, 62–67.
- Biggs, T.W.; Marshall, M.; Messina, A. Mapping daily and seasonal evapotranspiration from irrigated crops using global climate grids and satellite imagery: Automation and methods comparison. *Water Resour. Res.* **2016**, *52*, 7311–7326. [\[CrossRef\]](#)

21. Farmer, W.; Strzepek, K.; Schlosser, C.A.; Droogers, P.; Gao, X. *A Method for Calculating Reference Evapotranspiration on Daily Time Scales*; Joint Program Report Series Report 195; MIT Global Change Program: Cambridge, MA, USA, 2011.
22. Wang, K.; Li, Z.; Cribb, M. Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: A new method to determine the Priestley–Taylor parameter. *Remote Sens. Environ.* **2006**, *102*, 293–305. [\[CrossRef\]](#)
23. Agrawal, Y.; Kumar, M.; Ananthakrishnan, S.; Kumarapuram, G. Evapotranspiration modeling using different tree based ensembled machine learning algorithm. *Water Resour. Manag.* **2022**, *36*, 1025–1042. [\[CrossRef\]](#)
24. Ruiz-Álvarez, M.; Gomariz-Castillo, F.; Alonso-Sarría, F. Evapotranspiration response to climate change in semi-arid areas: Using random forest as multi-model ensemble method. *Water* **2021**, *13*, 222. [\[CrossRef\]](#)
25. Hao, P.; Di, L.; Guo, L. Estimation of crop evapotranspiration from MODIS data by combining random forest and trapezoidal models. *Agric. Water Manag.* **2022**, *259*, 107249. [\[CrossRef\]](#)
26. Ferreira, L.B.; da Cunha, F.F. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agric. Water Manag.* **2020**, *234*, 106113. [\[CrossRef\]](#)
27. Kumar, M.; Raghuwanshi, N.S.; Singh, R.; Wallender, W.W.; Pruitt, W.O. Estimating evapotranspiration using artificial neural network. *J. Irrig. Drain. Eng.* **2002**, *128*, 224–233. [\[CrossRef\]](#)
28. El-Shafie, A.; Najah, A.; Alsulami, H.M.; Jahanbani, H. Optimized neural network prediction model for potential evapotranspiration utilizing ensemble procedure. *Water Resour. Manag.* **2014**, *28*, 947–967. [\[CrossRef\]](#)
29. Kisi, O. Least squares support vector machine for modeling daily reference evapotranspiration. *Irrig. Sci.* **2013**, *31*, 611–619. [\[CrossRef\]](#)
30. Wen, X.; Si, J.; He, Z.; Wu, J.; Shao, H.; Yu, H. Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resour. Manag.* **2015**, *29*, 3195–3209. [\[CrossRef\]](#)
31. Shiri, J.; Dierickx, W.; Baba, A.P.-A.; Neamati, S.; Ghorbani, M.A. Estimating daily pan evaporation from climatic data of the state of Illinois, USA using adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN). *Hydrol. Res.* **2011**, *42*, 491–502. [\[CrossRef\]](#)
32. Wu, L.; Fan, J. Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. *PLoS ONE* **2019**, *14*, e0217520. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Reis, M.M.; da Silva, A.J.; Zullo Junior, J.; Santos, L.D.T.; Azevedo, A.M.; Lopes, É.M.G. Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data. *Comput. Electron. Agric.* **2019**, *165*, 104937. [\[CrossRef\]](#)
34. Yonaba, R.; Kiema, A.; Tazen, F.; Belemtougri, A.; Cissé, M.; Mounirou, L.A.; Bodian, A.; Koïta, M.; Karambiri, H. Accuracy and interpretability of machine learning-based approaches for daily ETo estimation under semi-arid climate in the West African Sahel. *Earth Sci. Inform.* **2025**, *18*, 87. [\[CrossRef\]](#)
35. Landeras, G.; Bekoe, E.; Ampofo, J.; Logah, F.; Diop, M.; Cisse, M.; Shiri, J. New alternatives for reference evapotranspiration estimation in West Africa using limited weather data and ancillary data supply strategies. *Theor. Appl. Climatol.* **2018**, *132*, 701–716. [\[CrossRef\]](#)
36. Picard, N.; Besic, N.; Meliho, M.; Sainte-Marie, J.; Mortier, F.; Legay, M. Bayesian model averaging of climate-dependent forest models using Expectation–Maximization. *Ecol. Model.* **2025**, *510*, 111355. [\[CrossRef\]](#)
37. Chen, Y.; Yuan, W.; Xia, J.; Fisher, J.B.; Dong, W.; Zhang, X.; Liang, S.; Ye, A.; Cai, W.; Feng, J. Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. *J. Hydrol.* **2015**, *528*, 537–549. [\[CrossRef\]](#)
38. Yang, Y.; Sun, H.; Xue, J.; Liu, Y.; Liu, L.; Yan, D.; Gui, D. Estimating evapotranspiration by coupling Bayesian model averaging methods with machine learning algorithms. *Environ. Monit. Assess.* **2021**, *193*, 156. [\[CrossRef\]](#) [\[PubMed\]](#)
39. World Development Indicators—Population Total, Benin. Available online: <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=BJ> (accessed on 21 March 2023).
40. Abatzoglou, J.T.; Dobrowski, S.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [\[CrossRef\]](#)
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
43. Fix, E.; Hodges, J.L. *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*; Technical Report 4; USAF School of Aviation Medicine: Randolph Field, TX, USA, 1951.
44. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
45. Quinlan, J.R. Learning with continuous classes. In *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, Hobart, Tasmania, 16–18 November 1992; pp. 343–348.
46. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference*, New York, NY, USA; ACM: New York, NY, USA, 2016; pp. 785–794.

47. Van Buuren, S.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [\[CrossRef\]](#)
48. Fan, J.; Yue, W.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* **2018**, *263*, 225–241. [\[CrossRef\]](#)
49. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **2008**, *28*, 1–26. [\[CrossRef\]](#)
50. Allen, R.G.; Pereira, L.S.; Howell, T.A.; Jensen, M.E. Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agric. Water Manag.* **2011**, *98*, 899–920. [\[CrossRef\]](#)
51. Vicente-Serrano, S.M.; Azorin-Molina, C.; Sanchez-Lorenzo, A.; Revuelto, J.; Morán-Tejeda, E.; López-Moreno, J.I.; Espejo, F. Sensitivity of reference evapotranspiration to changes in meteorological parameters in Spain (1961–2011). *Water Resour. Res.* **2014**, *50*, 8458–8480. [\[CrossRef\]](#)
52. Dias, S.H.B.; Filgueiras, R.; Fernandes Filho, E.I.; Arcanjo, G.S.; da Silva, G.H.; Mantovani, E.C.; da Cunha, F.F. Reference evapotranspiration of Brazil modeled with machine learning techniques and remote sensing. *PLoS ONE* **2021**, *16*, e0245834. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Frondana, G. Empirical Comparison of 16 Regression Algorithms on 59 Datasets. Master’s Thesis, Universidade Estadual de Campinas, Campinas, Brazil, 2017.
54. Althoff, D.; Bazame, H.C.; Filgueiras, R.; Dias, S.H.B. Heuristic methods applied in reference evapotranspiration modeling. *Ciênc. Agrotecnol.* **2018**, *42*, 314–324. [\[CrossRef\]](#)
55. McVicar, T.R.; Roderick, M.L.; Donohue, R.J.; Li, L.T.; Van Niel, T.G.; Thomas, A.; Grieser, J.; Jhajharia, D.; Himri, Y.; Mahowald, N.M.; et al. Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *J. Hydrol.* **2012**, *416–417*, 182–205. [\[CrossRef\]](#)
56. Ahmadi, A.; Daccache, A.; Snyder, R.L.; Suvočarev, K. Meteorological driving forces of reference evapotranspiration and their trends in California. *Sci. Total Environ.* **2022**, *849*, 157823. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Yonaba, R.; Tazen, F.; Cissé, M.; Mounirou, L.A.; Belemtougri, A.; Ouedraogo, V.A.; Koïta, M.; Niang, D.; Karambiri, H.; Yacouba, H. Trends, sensitivity and estimation of daily reference evapotranspiration ET₀ using limited climate data: Regional focus on Burkina Faso in the West African Sahel. *Theor. Appl. Climatol.* **2023**, *153*, 947–974. [\[CrossRef\]](#)
58. Nkiaka, E.; Bryant, R.G.; Dembélé, M.; Yonaba, R.; Aigbedion, I.P.; Karambiri, H. Quantifying the effects of climate and environmental changes on evapotranspiration variability in the Sahel. *J. Hydrol.* **2024**, *642*, 131874. [\[CrossRef\]](#)
59. Novick, K.A.; Ficklin, D.L.; Stoy, P.C.; Williams, C.A.; Bohrer, G.; Oishi, A.C.; Papuga, S.A.; Blanken, P.D.; Noormets, A.; Sulman, B.N.; et al. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nat. Clim. Change* **2016**, *6*, 1023–1027. [\[CrossRef\]](#)
60. Hoppe, H.; Dietrich, P.; Marzahn, P.; Weiß, T.; Nitzsche, C.; von Lukas, U.F.; Wengerek, T.; Borg, E. Transferability of machine learning models for crop classification in remote sensing imagery using a new test methodology: A study on phenological, temporal, and spatial influences. *Remote Sens.* **2024**, *16*, 1493. [\[CrossRef\]](#)
61. Stock, A. Choosing blocks for spatial cross-validation: Lessons from a marine remote sensing case study. *Front. Remote Sens.* **2025**, *6*, 1531097. [\[CrossRef\]](#)
62. Jenik, J.; Hall, J.B. The Ecological effects of the Harmattan wind in the Djebobo Massif (Togo Mountains, Ghana). *J. Ecol.* **1966**, *54*, 767–779. [\[CrossRef\]](#)
63. Lyngsie, G.; Awadzi, T.; Breuning-Madsen, H. Origin of Harmattan dust settled in Northern Ghana—Long transported or local dust? *Geoderma* **2011**, *167–168*, 351–359. [\[CrossRef\]](#)
64. Bouraima, A.K.; Weihong, L.; Chaofu, W.; Varis, O. Irrigation water requirements of rice using CropWat model in Northern Benin. *Int. J. Agric. Biol. Eng.* **2015**, *8*, 58–64.
65. Gbode, I.E.; Diro, G.T.; Intsiful, J.D.; Dudhia, J. Current conditions and projected changes in crop water demand, irrigation requirement, and water availability over west africa. *Atmosphere* **2022**, *13*, 1155. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.