

Article

Accurate Medium-Term Forecasting of Farmland Evapotranspiration Using Corrected Next-Generation Numerical Weather Prediction

Shuting Zhao, Lifeng Wu *  and Xianghui Lu * 

School of Soil and Water Conservation, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, China; zhaoshuting@whu.edu.cn

* Correspondence: lifengwu@nit.edu.cn (L.W.); xianghuilu@nit.edu.cn (X.L.)

Abstract

Accurate medium-term evapotranspiration (ET) forecasting is critical for irrigation scheduling and hydrological assessments. To address biases in numerical weather prediction (NWP) systems, we developed a hybrid GWO_XGB model integrating Extreme Gradient Boosting (XGBoost) with Gray Wolf Optimizer (GWO) for bias correction. Using the corrected data, we evaluate four hybrid models—Support Vector Machine (SVM) and XGBoost, each optimized with either GWO or Grasshopper Optimization Algorithm (GOA)—for 1- to 10-day ET forecasts across 11 farmland stations in Europe and North America (2003–2014). The results showed that the GWO_XGB model demonstrated the best comprehensive performance (average RMSE = 0.476 mm d^{−1}, R² = 0.829), while the GWO_SVM model performed the weakest (average RMSE = 0.572 mm d^{−1}, R² = 0.761). Forecast accuracy of Rs and VPD declined with lead time, with the 1-day forecasts being most accurate (RMSE range: 2.005–3.061 MJ mm d^{−1}). Using calibrated NWP data, the highest 1-day forecast accuracy was achieved (average RMSE = 0.715 mm d^{−1}), with GWO_XGB remaining the best (1–3 days average RMSE = 0.667 mm d^{−1}; 10-day cumulative forecast RMSE = 0.698 mm d^{−1}). Overall, the GWO_XGB model combined with NWP calibration provides reliable short- to medium-term ET forecasts for agricultural water management.

Keywords: Global Ensemble Forecast System version 12; machine learning; bio-inspired optimization algorithm

1. Introduction

Evapotranspiration (ET) is a critical yet often poorly constrained component of the global water cycle, with significant implications for energy balance and ecohydrological processes [1–3]. In agriculture, accurate estimation of ET is essential for efficient water resource management and irrigation scheduling [4,5]. With water scarcity affecting over half of the world’s continental areas and continuing to intensify, the demand for reliable medium-term (1–10 days) ET forecasts has become increasingly urgent [6]. For clarity, short-term forecasts in this study refer to a 1–3 day lead time, while medium-term forecasts encompass a 4–10 day lead time. Such forecasts are vital for understanding water/carbon energy interactions, optimizing irrigation planning, and supporting hydrological assessments [7–10].

To address this need for reliable forecasts, numerical weather prediction (NWP) models have been widely adopted to support forecasting of hydrological variables such as



Academic Editors: Gianni Bellocchi and Guangsheng Zhou

Received: 21 November 2025

Revised: 29 January 2026

Accepted: 30 January 2026

Published: 2 February 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 Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

floods, ET, reference evapotranspiration (ET_o), and potential evapotranspiration [10–18]. Among these, the Global Ensemble Forecast System (GEFS) represents a state-of-the-art, physics-based global model developed by the National Centers for Environmental Prediction [19]. Its latest version, GEFS v12, provides 31-member ensemble forecasts spanning the period 2005–2019. Depending on the initialization time, forecasts extend up to 35 days (0000 UTC) or 16 days (0600, 1200, and 1800 UTC) [20]. Notably, GEFS v12 features enhanced resolution and improved physical parameterizations, which lead to more accurate representations of key meteorological drivers for ET, such as solar radiation and near-surface humidity. Its ensemble framework also provides inherent uncertainty estimates, making it particularly suitable for medium-term agricultural water management applications. GEFS v12 outputs are increasingly integrated into forecasting systems for precipitation, radiation, and temperature [21–24].

However, NWP outputs frequently exhibit substantial deviations from observational data, which directly propagate into ET forecasting errors [25,26]. Biases in key meteorological drivers—such as solar radiation, temperature, and humidity—stem from multiple sources, including uncertainties in initial conditions and model parameterizations, structural limitations in representing subgrid-scale processes, spatial averaging over large areas, and systematic underestimation of precipitation [27,28]. These inaccuracies in NWP-derived meteorological inputs can lead to significant over- or under-estimation of ET, thereby undermining the reliability of irrigation scheduling and water resource assessments that depend on medium-term forecasts. Consequently, statistical post-processing is often necessary to enhance the reliability of NWP products [29]. A variety of correction techniques have been proposed, ranging from simple additive or multiplicative adjustments to more sophisticated regression and downscaling methods [9,10].

Among the advanced post-processing approaches, machine learning (ML) has emerged as a particularly powerful tool, due to recent advances in computational power and artificial intelligence. ML has demonstrated high effectiveness in estimating hydrological variables, including ET [30–33]. Moreover, heuristic optimization algorithms have demonstrated considerable potential in improving the performance of ML models by efficiently identifying optimal hyperparameters and model structures [34–38]. In particular, the Gray Wolf Optimizer (GWO) and the Grasshopper Optimization Algorithm (GOA) have been selected for this study due to their proven balance between exploration and exploitation, relatively few control parameters, and strong performance in continuous optimization problems, making them well-suited for tuning complex ML models such as XGBoost and SVM. The integration of metaheuristic algorithms such as the GWO and GOA with established ML techniques has shown promise in enhancing model accuracy and generalization [35,36].

Beyond these advanced ML-based approaches, ET estimation in practice has traditionally relied on more conventional methodologies. A common practice involves calculating a reference evapotranspiration (ET_o), such as using the Penman–Monteith equation [39], and then applying crop coefficients (K_c) to estimate crop evapotranspiration under standard conditions. However, this two-step approach not only requires extensive meteorological inputs, limiting its applicability in data-sparse regions [40,41], but also introduces uncertainty through the empirical K_c values. In contrast, eddy covariance systems directly measure ET, which reflects the real-time biophysical and environmental conditions. The objective of this study is to forecast this measured ET. Alternatively, satellite remote sensing offers a viable pathway for direct ET estimation at appropriate spatiotemporal resolutions [42].

Despite these developments, the use of NWP outputs for direct ET forecasting remains limited. Existing correction methods for raw NWP data often rely on conventional regression or linear interpolation, and few studies have explored the combination of ML models

with bio-inspired optimization algorithms to correct NWP outputs for ET forecasting. We hypothesize that such a hybrid approach can effectively mitigate systematic biases in NWP outputs (particularly in solar radiation and vapor pressure deficit), thereby significantly improving the reliability of medium-term ET forecasts. To test this hypothesis, the present study bridges the aforementioned gap with the following three objectives:

- (1) Evaluating the performance of four hybrid ML models (SVM and XGBoost coupled with GWO or GOA) in simulating ET using the Fluxnet 2015 dataset;
- (2) Applying the GWO-XGB model to correct Rs and VPD forecasts from GEFsV12 over a 1–10 day lead time;
- (3) Assessing the utility of the bias-corrected NWP data for medium-term ET forecasting across 11 farmland stations in Europe and North America.

2. Materials and Methods

2.1. Study Area

This study utilized data from 11 farmland stations across Europe and North America (Figure 1), which are part of the Fluxnet 2015 network. The dominant crop types at these stations include sugar beet, winter wheat, winter barley, potato, maize, spring barley, and rice (Table 1).

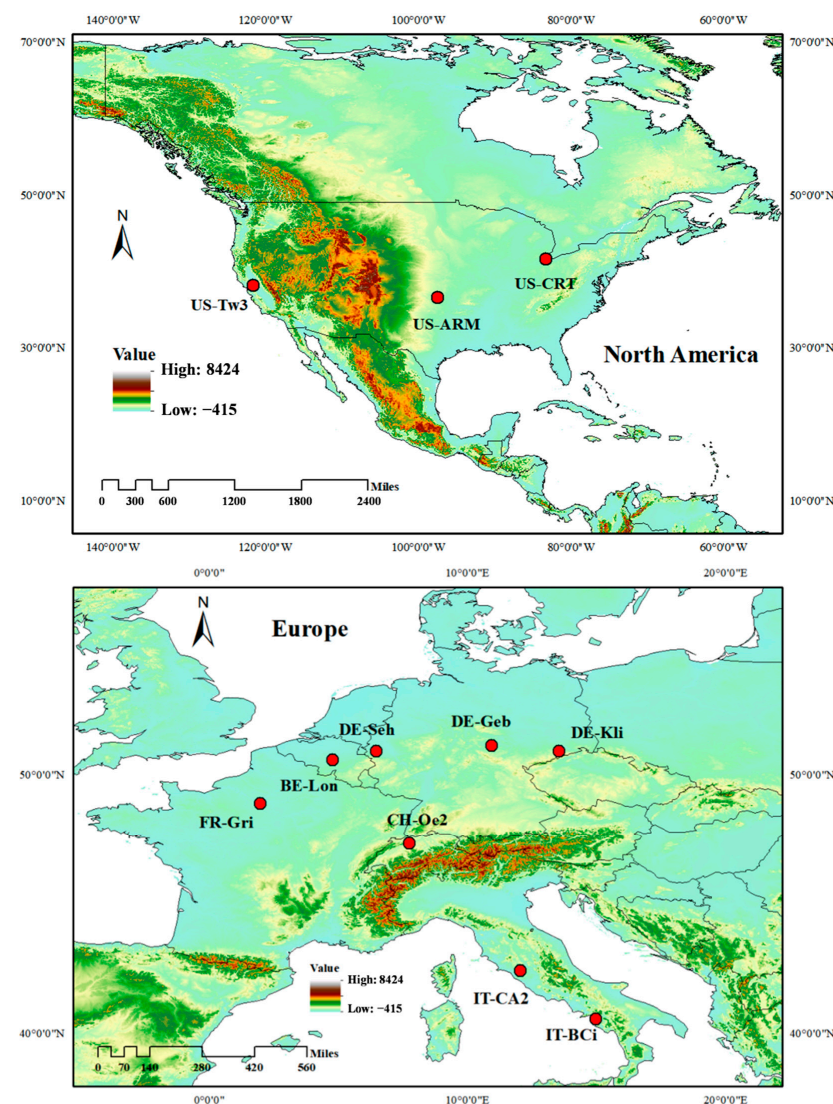


Figure 1. Location of 11 farmland flux tower stations across North America and Europe used in this study.

Table 1. The specific latitude and longitude and meteorological data information of 11 farmland stations selected in the study.

Station	Latitude (N)	Longitude (E)	Start (y)	End (y)	Elevation (m)	Rs (MJ m ⁻² d ⁻¹)	WS (m s ⁻¹)	ET (mm d ⁻¹)	Tmax (°C)	Tmin (°C)	NDVI	VPD (hPa d ⁻¹)	SWC	Crop Species
BE-Lon	50.55	4.75	2006	2014	167	120.72	3.08	27.05	15.32	7.78	0.51	3.00	33.33	Sugar beet [43]
CH-Oe2	47.29	7.73	2006	2014	452	138.04	1.66	55.14	14.22	4.87	0.51	2.80	27.26	Winter wheat, winter barley, winter rapeseed, potato, and cover crop [44]
DE-Geb	51.10	10.91	2003	2014	162	122.15	2.94	26.94	13.64	5.29	0.48	4.00	20.66	Winter wheat and potato [45]
DE-Kli	50.89	13.52	2006	2014	478	127.76	3.41	25.59	11.20	4.17	0.48	3.28	36.05	Winter barley, rapeseed, winter wheat, maize, spring barley [46]
DE-Seh	50.87	6.45	2007	2010	103	134.45	2.38	44.78	14.57	6.68	0.48	3.49	28.73	Winter wheat [47]
FR-Gri	48.84	1.95	2006	2012	125	136.88	3.12	38.44	14.92	6.79	0.48	3.89	28.08	Maize, wheat, barley, and mustard [48]
IT-BCi	40.52	14.96	2006	2014	20	182.76	2.18	58.43	22.64	13.44	0.49	5.57	27.68	Maize [49]
IT-CA2	42.38	12.03	2011	2014	200	179.94	2.72	36.03	20.76	9.63	0.56	6.57	19.84	Winter wheat [50]
US-ARM	36.61	−97.49	2006	2012	314	195.58	4.39	41.27	21.35	9.23	0.40	9.13	20.25	Winter wheat [51]
US-CRT	41.63	−83.35	2011	2013	180	156.33	3.02	52.61	15.19	6.07	0.41	4.74	50.10	Winter wheat [52]
US-Tw3	38.12	−121.65	2013	2014	−9	228.55	3.14	72.49	24.36	10.64	0.44	9.23	12.12	Rice [53]

Note: Values for meteorological variables represent site-specific long-term daily means. Abbreviations: Rs, solar radiation; WS, wind speed; Tmax/Tmin, daily maximum/minimum air temperature; NDVI, normalized difference vegetation index (derived from MODIS); VPD, vapor pressure deficit; SWC, soil water content.

North American stations are largely characterized by a temperate continental climate, with cold winters and hot summers. Over the past half-century, inappropriate irrigation and fertilization practices have led to soil salinization, affecting more than 55 million hectares of farmland. Intensive irrigation techniques often raise soil moisture levels, thereby increasing ET, particularly in irrigated croplands relative to forested areas. Three stations in North America primarily grow winter wheat and rice, where sufficient heat and precipitation support key growth stages such as heading, flowering, and grain filling.

In Europe, the temperate climate, moderated by oceanic influences, results in mild and humid conditions throughout much of the year. Generally lower temperatures and ET rates are observed due to higher latitudes. Although water and heat resources in parts of Europe are adequate for winter wheat cultivation, periodic droughts in southwestern regions have negatively impacted yields. Winter oilseed rape, typically planted after cereal harvests, serves as a break crop for managing pathogens and weeds. Widespread maize cultivation in Europe benefits from breeding efforts that have adapted varieties to local soil and climatic conditions. Given the economic and food security importance of these crops, accurate medium-term ET forecasting is essential for optimizing irrigation, reducing non-beneficial water loss, and mitigating risks related to soil degradation and drought.

2.2. Data Collection

Daily observed ET and meteorological variables, including maximum and minimum temperature ($T_{\max}$, $T_{\min}$), solar radiation (R_s), wind speed (WS), vapor pressure deficit (VPD), and soil water content (SWC), were obtained from the Fluxnet 2015 dataset (<https://fluxnet.org/data/fluxnet2015-dataset/> accessed on 22 February 2023). Additionally, the daily normalized difference vegetation index (NDVI) was derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery (MOD13Q1 product, 16-day composite at 250 m resolution, National Aeronautics and Space Administration, Greenbelt, MD, USA) to capture crop phenological dynamics. The corresponding MODIS tiles for each site were processed, and the NDVI value for the pixel containing the flux tower was extracted. A continuous daily NDVI series was then generated by applying a Savitzky–Golay filter to smooth and interpolate the composite time-series data, ensuring temporal alignment with the daily flux observations [54,55].

Forecast data for R_s , $T_{\max}$, $T_{\min}$, and specific humidity (spfh) were retrieved from the Global Ensemble Forecast System version 12 (GEFS v12, National Oceanic and Atmospheric Administration, College Park, MD, USA; <https://noaa-gefs-retrospective.s3.amazonaws.com/index.html#> accessed on 8 August 2021). Since VPD is not directly available in GEFS v12 outputs, it was derived from spfh and temperature using standard psychrometric relationships. Specifically, spfh was first converted to relative humidity (RH), which was then combined with air temperature to compute daily VPD. A summary of forecast variables is provided in Table A1 (see Appendix A).

After aligning the Fluxnet 2015 and GEFS v12 data by date, all rows with any missing values were removed, resulting in a complete dataset for the retained dates. This procedure led to the removal of a small proportion of data (<2% of the total daily records), primarily affecting a subset of stations with occasional gaps. Given the low magnitude and non-systematic spatial distribution of the missing data, we consider the impact of this deletion on overall data continuity, representativeness, and subsequent model training to be negligible.

2.3. Bias Correction and Forecasting Framework

A hybrid model was utilized to correct bias in NWP outputs, specifically the GWO_XGB model. Our modeling process entailed inputting raw NWP outputs and obtaining output values as correction results for the coupled model. NWP meteorological

variables, including Tmax, Tmin, Rs, and spfh, were used as inputs. To rigorously evaluate the model's performance and prevent temporal overfitting, we adopted a two-step data splitting strategy based on time order. First, the entire chronological dataset was divided into a training period (the first 80% of the time series) and a strictly independent testing period (the final 20%). This ensures that the model is evaluated on future data unseen during training, simulating a real forecasting scenario. Subsequently, within the training period only, we employed a rolling-origin time-series cross-validation with 5 folds to optimize the model hyperparameters and assess its stability. In this process, the training set was sequentially split into five contiguous blocks; each time, four blocks were used for training, and the subsequent block was used for validation. This method preserves the temporal order of the data and prevents information leakage from the future.

This study is divided into two parts according to the source of meteorological data. In the first part, the input of the Fluxnet 2015 dataset is used as the control group of the ET simulation, which is used to evaluate the performance of the four models selected in the study and compare it with the results of the prediction model. In the second part, we use the forecast data of NWP for ten days as model input. Since real-time, forecasted NDVI and SWC are not available from NWP systems, and given their high temporal autocorrelation over short periods, we used the most recent observed values as proxies. Specifically, NDVI_t and SWC_t represent the NDVI and SWC values of the previous day, respectively, as the input of the forecast data combination (Table 2). In Table 2, F1 represents the input combination based on the Fluxnet 2015 dataset, and GOA_SVM_F represents the model of the F1 combination input. N1 represents the input combination based on the NWP data of 1d, and GOA_SVM 1 represents the model of the N1 combination input. The meaning of other models and combinations is similar to the above.

Table 2. Input combinations and models.

Original	No.	Model				Input Combinations
		GOA_SVM	GWO_SVM	GOA_XGB	GWO_XGB	
Fluxnet 2015	F1	GOA_SVM_F	GWO_SVM_F	GOA_XGB_F	GWO_XGB_F	Rs, Ws, Tmax, Tmin, NDVI, VPD, SWC
	N1	GOA_SVM1	GWO_SVM1	GOA_XGB1	GWO_XGB1	NDVI_t, SWC_t, Tmax_1, Tmin_1, Rs_1, VPD_1
	N2	GOA_SVM2	GWO_SVM2	GOA_XGB2	GWO_XGB2	NDVI_t, SWC_t, Tmax_2, Tmin_2, Rs_2, VPD_2
	N3	GOA_SVM3	GWO_SVM3	GOA_XGB3	GWO_XGB3	NDVI_t, SWC_t, Tmax_3, Tmin_3, Rs_3, VPD_3
	N4	GOA_SVM4	GWO_SVM4	GOA_XGB4	GWO_XGB4	NDVI_t, SWC_t, Tmax_4, Tmin_4, Rs_4, VPD_4
NWP	N5	GOA_SVM5	GWO_SVM5	GOA_XGB5	GWO_XGB5	NDVI_t, SWC_t, Tmax_5, Tmin_5, Rs_5, VPD_5
	N6	GOA_SVM6	GWO_SVM6	GOA_XGB6	GWO_XGB6	NDVI_t, SWC_t, Tmax_6, Tmin_6, Rs_6, VPD_6
	N7	GOA_SVM7	GWO_SVM7	GOA_XGB7	GWO_XGB7	NDVI_t, SWC_t, Tmax_7, Tmin_7, Rs_7, VPD_7
	N8	GOA_SVM8	GWO_SVM8	GOA_XGB8	GWO_XGB8	NDVI_t, SWC_t, Tmax_8, Tmin_8, Rs_8, VPD_8
	N9	GOA_SVM9	GWO_SVM9	GOA_XGB9	GWO_XGB9	NDVI_t, SWC_t, Tmax_9, Tmin_9, Rs_9, VPD_9
	N10	GOA_SVM10	GWO_SVM10	GOA_XGB10	GWO_XGB10	NDVI_t, SWC_t, Tmax_10, Tmin_10, Rs_10, VPD_10

It is important to emphasize that our forecasting framework is not a static historical extrapolation. For each forecast lead time, the model receives future meteorological inputs from the NWP outputs, while NDVI and SWC are represented by their most recent observed values (lag 1) due to the lack of real-time forecasts. This design ensures that the forecast ET responds directly to the forecasted weather conditions on each target day, capturing day-to-day variations in radiation, temperature, and humidity. The use of lagged NDVI and SWC is justified by their high temporal autocorrelation (Section 3.1), implying that their values change slowly relative to daily weather; thus, the dominant source of ET variability in the 1–10 day range originates from the NWP-derived meteorological inputs, not from the nearly constant vegetation or soil states.

2.4. Methodology

2.4.1. Support Vector Machine (SVM)

The SVM algorithm is a supervised machine learning model for binary classification problems proposed by Vapnik et al. [56]. The basic principle of the algorithm is to solve a separation hyperplane that can correctly divide the training dataset and has the largest geometric interval, and to solve the maximum hyperplane problem, as shown in Equation (1):

$$\omega^T x + b = 0 \quad (1)$$

Solving the maximum hyperplane problem is to find the optimal ω and b . In Equation (1), ω is the normal vector that determines the direction of the hyperplane, and b is the displacement term that determines the distance between the hyperplane and the origin. x is a vector of n -dimensional sample features.

The mathematical expression of the distance from each sample to the plane is shown in Equations (2) and (3):

$$\gamma = \frac{\omega^T x + b}{\|\omega\|} \quad (2)$$

$$\|\omega\| = \sqrt{\sum_{i=1}^k \omega_i^2} \quad (3)$$

The decision function of the SVM algorithm is determined by a few support vectors, and the complexity depends on the number of support vectors rather than the spatial dimension. This not only helps us to accurately obtain key samples but also has good robustness, which is mainly reflected in the increase or decrease in non-support vector samples. There is no impact on the model. However, the SVM algorithm involves the operation of the m -order matrix when solving the quadratic programming, and m represents the number of samples. When the number of samples is too large, the operation and storage of the matrix will consume time and computer memory. In addition, the algorithm is not effective in solving multi-classification problems.

2.4.2. Extreme Gradient Boosting (XGBoost)

The XGBoost algorithm is an integrated learning algorithm based on the tree model proposed by Chen and Guestrin [57]. The algorithm iterates the t -tree model to form an additive model, which is the forecast result of the sample, as shown in Equation (4):

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (4)$$

In Equation (4), $\hat{y}_i^{(t)}$ represents the prediction result of sample i after the t -th iteration, $f_t(x_i)$ is the model of the t -th tree.

The objective function of XGBoost is the sum of the loss function and the regularization term, as shown in Equations (5) and (6):

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) \quad (5)$$

$$Obj = L + \sum_{i=1}^t \Omega(f_i) \quad (6)$$

In Equation (5), n is the sample size, y_i is the measured value, and L is the loss function. In Equation (6), $\sum_{i=1}^t \Omega(f_i)$ is the regular term, and Obj is the objective function.

The loss function is composed of the forecasted value and the measured value, which represents the deviation of the model, while the regularization term is used to prevent the model from overfitting. The Boosting algorithm follows the forward stepwise addition, and the forecasted value of the latest sample is the sum of the forecasted value of the previous sample and the forecasted value of the new model. The XGBoost model can process large-scale datasets more efficiently than the SVM model, does not occupy too much computer memory, and can automatically process missing values. However, the model cannot handle high-dimensional feature data well and is only suitable for processing structure feature data.

2.4.3. Grasshopper Optimization Algorithm (GOA)

GOA is a meta-heuristic algorithm proposed by Saremi et al. to simulate locust foraging behavior [58]. The main principle is to let the initial n locusts be randomly distributed in the multidimensional space, according to the determined mathematical formulas and parameters, and update the location of all locusts in each iteration calculation. In the coupling model of this study, the optimal position of the locust calculated by each generation will be used as the cost and gamma dynamic parameter input of the SVM and XGBoost models, and then the corresponding RMSE value of the model will be calculated and compared with the next generation. Finally, the generation with the best simulation effect is the greedy step of the algorithm. When the next position of the locust is better than the current position, it will jump to the new position. The biggest feature of the GOA algorithm is that it considers both the location of the optimal individual and the interaction of individuals within the population. However, the core parameter of the GOA algorithm adopts a linear adaptive reduction strategy. Although this parameter can prevent locusts from premature convergence during the iteration process, it cannot adapt well to complex problems, early exploration, and late convergence. In addition, in the process of location update, the fitness of locusts with poor locations will always fluctuate, which may lead to a local optimum. The mathematical expression of the locust update position in the GOA algorithm is shown in Equations (7)–(9):

$$X_i^d = c \cdot \left(\sum_{\substack{j=1 \\ j \neq i}}^N c \frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_i^d\right|\right) \frac{x_j - x_i}{|x_j - x_i|} \right) + \hat{T}_d \quad (7)$$

$$c = c_{\max} - t \frac{c_{\max} - c_{\min}}{\max_iter} \quad (8)$$

$$s(x) = f \cdot e^{\frac{-x}{l}} - e^{-x} \quad (9)$$

In Equation (7), X_i^d represents the position of the i -th locust in the d -dimensional space, ub_d and lb_d represents the upper and lower bounds of the value of the d -dimensional space, respectively, $\hat{T}_d$ representing the position of the current optimal individual in the d -dimensional space. In Equation (8), $c_{\max}$ and $c_{\min}$ is a fixed coefficient, respectively, 1 and 0.01, t represents the current number of iterations, $\max_iter$ is the maximum number of iterations set. By introducing this coefficient c , the distance between locusts can be better controlled, so that they can stay in a comfortable circle and avoid excessive aggregation in the early stage, resulting in local optimal results. In Equation (9), f is 0.5 and l is 1.5.

2.4.4. Gray Wolf Optimizer (GWO)

GWO is a meta-heuristic algorithm proposed by Mirjalili et al. [59] to simulate the hierarchy and hunting strategy of gray wolves. This algorithm divides the gray wolf into

four social classes. The head wolf (α wolf) in the wolf group is located at the highest level, which is mainly responsible for controlling the predation, habitat, and other activities of the other three wolf groups; this is the optimal solution in the GWO algorithm. The wolves in the second layer are called β wolves. They obey and assist the α wolf to dominate its lower wolf pack, and are also candidates for the α wolf, that is, the suboptimal solution in the GWO algorithm. The third layer of wolves, called δ wolves, is subordinate to the upper wolves and can manage the lower wolves. The lowest, ω wolves, are subordinate to the wolves of other social levels. They are at the lowest level, so they are not allowed to eat until the higher-ups finish. Their existence can prevent conflicts such as self-mutilation within the wolves. The GWO algorithm has strong convergence ability and high accuracy, but its convergence accuracy and speed are not ideal in the face of complex problems.

The principle of the GWO algorithm to find the optimal solution is the hunting process of the gray wolf: (1) surrounding and tracking the prey; (2) hunting prey; (3) attacking prey. The mathematical expression of the GWO algorithm is shown in Equations (10)–(14):

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p - \vec{X} \right| \quad (10)$$

$$\vec{X} = \vec{X}_p - \vec{A} \cdot \vec{D} \quad (11)$$

In Equations (10) and (11), $\vec{D}$ is the distance between the gray wolf and the prey, $\vec{X}_p$ and $\vec{X}$ are the position vectors of the prey and the gray wolf, $\vec{A}$ and $\vec{C}$ are the determined coefficients, respectively. The equations are as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (12)$$

$$\vec{C} = 2\vec{r}_2 \quad (13)$$

$$\vec{a} = 2 - l \cdot \frac{2}{\max_iter} \quad (14)$$

In Equations (12) and (13), $\vec{r}_1$ and $\vec{r}_2$ are the random number vectors generated on [0, 1]. In Equation (14), l is the current number of iterations, and $\max_iter$ is the maximum number of iterations.

According to Equation (10), α , β , and δ wolves gradually surround their prey and narrow the distance between them, while ω wolves adjust their position according to the average distance between wolves and prey at other levels. The increase in $|\vec{A}|$ in Equation (12) will make the wolf pack move away from the prey, and the corresponding $|\vec{A}|$ will be too large. When $|\vec{A}| > 1$ in Equation (12), we think that the prey is not appropriate, it will lead to the wolves for global search. When $|\vec{A}| < 1$, the wolf pack will begin to attack the prey, and the optimal solution can be obtained at this time.

2.4.5. Model Configuration and Hyperparameter Optimization

In this study, the GOA and GWO algorithms were employed to optimize the key hyperparameters of the SVM and XGBoost models to achieve optimal performance for ET forecasting. For the SVM model, the regularization parameter (C) and the kernel coefficient (γ) were optimized, with search ranges set to $[10^{-3}, 10^3]$ and $[10^{-5}, 10]$, respectively. For the XGBoost model, four hyperparameters were optimized: the number of boosting rounds ($n_estimators$, range [50, 1500]), maximum tree depth (max_depth , range [2, 20]), learning rate (eta , range $[10^{-4}, 0.3]$), and the subsample ratio of columns ($subsample$, range [0.5, 0.8]). These ranges were determined based on common practices and preliminary experiments to ensure effective exploration of the parameter space.

Both the GOA and GWO optimizers were configured with a population size of 50 and a maximum iteration count of 200. The optimization process terminated upon reaching this predefined maximum number of iterations. To mitigate overfitting during model training with the optimized hyperparameters, intrinsic regularization strategies within XGBoost—specifically, L2 regularization (with lambda fixed at 1) and a minimum child weight (fixed at 6)—were applied consistently. Early stopping was not employed during the metaheuristic optimization phase itself to ensure a comprehensive and equitable evaluation of all candidate solutions within the allocated computational budget.

All modeling was conducted in R 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) using the “e1071” (v1.7-9, Department of Statistics, TU Wien, Vienna, Austria) and “xgboost” (v1.5.0.1, DMLC group, Seattle, WA, USA) packages. GWO and GOA were implemented in R based on standard pseudo-code. Simulations ran on a Windows 11 workstation (AMD Ryzen 7 5800X, 8 cores @ 3.80 GHz, 16 GB RAM).

2.5. Evaluation Metrics

The performance of the models was evaluated using five statistical indicators. The root mean square error (RMSE) measures overall accuracy by quantifying the average magnitude of errors between measured and forecasted values; it is sensitive to outliers due to the squaring of differences, with values closer to 0 indicating higher accuracy. The coefficient of determination (R^2) evaluates the goodness-of-fit and the proportion of variance explained by the model, ranging from 0 to 1, where values closer to 1 reflect stronger explanatory power. The mean absolute error (MAE) assesses accuracy through the average absolute deviation between paired observations; it is less influenced by extreme values than RMSE and provides a direct interpretation of average error magnitude. The percent bias (PBIAS) quantifies systematic bias by showing the cumulative tendency of forecasts to over- or under-estimate observations, with values near 0 indicating minimal bias. Finally, the normalized root mean square error (NRMSE) provides a scale-independent measure of accuracy, calculated by dividing RMSE by the range of observed values, which facilitates consistent error comparisons across stations with varying ET magnitudes. Lower values of MAE, PBIAS, and NRMSE correspond to better model performance. The mathematical expressions of these five statistical indicators are shown in Equations (15)–(19):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_{obs,i} - X_{pred,i})^2}{N}} \quad (15)$$

$$R^2 = \frac{[\sum_{i=1}^N (X_{obs,i} - \bar{X}_{obs,i})(X_{pred,i} - \bar{X}_{pred,i})]^2}{\sum_{i=1}^N (X_{obs,i} - \bar{X}_{obs,i})^2 \sum_{i=1}^N (X_{pred,i} - \bar{X}_{pred,i})^2} \quad (16)$$

$$MAE = \frac{\sum_{i=1}^N |X_{obs,i} - X_{pred,i}|}{N} \quad (17)$$

$$PBIAS = \sum_{i=1}^N \frac{|X_{obs,i} - X_{pred,i}|}{X_{pred,i}} \quad (18)$$

$$NRMSE = \frac{RMSE}{\bar{X}_{pred,i}} \quad (19)$$

where N represents the total sample, $X_{obs,i}$ and $X_{pred,i}$ represent the measured value and forecasted value of the model, respectively, and $\bar{X}_{obs,i}$ and $\bar{X}_{pred,i}$ represent the average value of $X_{obs,i}$ and $X_{pred,i}$, respectively.

3. Results

3.1. ET Simulation with Fluxnet Data (Control)

The statistical results of daily ET simulations using the GOA_SVM, GOA_XGB, GWO_SVM, and GWO_XGB models across 11 farmland stations from the Fluxnet 2015 dataset are presented in Table A2 (see Appendix A). For the four models during the testing phase, the ranges of RMSE, R^2 , MAE, PBIAS, and NRMSE were 0.332–0.782 mm d^{−1}, 0.643–0.9, 0.233–0.509 mm d^{−1}, 14.5–34.9%, and 21–57.9%, respectively, indicating generally satisfactory model fitting.

Overall, the tree-based models (GOA_XGB and GWO_XGB) outperformed the kernel-based models (GOA_SVM and GWO_SVM) during the testing phase. Specifically, GWO_XGB performed the best, with average RMSE, R^2 , MAE, PBIAS, and NRMSE values of 0.476 mm d^{−1}, 0.829, 0.323 mm d^{−1}, 22.1%, and 32.9%, respectively. In contrast, the GWO_SVM model had the poorest performance, with corresponding average RMSE, R^2 , MAE, PBIAS, and NRMSE values of 0.572 mm d^{−1}, 0.761, 0.378 mm d^{−1}, 22.6%, and 39.3%, respectively. Notwithstanding, this worst model nonetheless produced appropriate results for ET simulation at several stations, including BE-Lon, CH-Oe2, DE-Geb, and DE-Seh, where RMSE ranged from 0.418 to 0.67 mm d^{−1} and R^2 from 0.792 to 0.842.

To facilitate better comparison of the performance of the four models across various stations, we present a scatter plot displaying the relationship between daily ET measured values and simulated values, as demonstrated in Figure 2 (exhibiting data from the stations BE-Lon, DE-Geb, DE-Seh, and US-Tw3). The fitting line for all models is close to the 1:1 line, with the data points from GWO_XGB clustering most tightly around it. The GOA_XGB and GWO_XGB models showed relatively balanced point distributions on both sides of the 1:1 line, whereas GOA_SVM and GWO_SVM exhibited greater dispersion. For instance, at BE-Lon, GWO_XGB yielded the smallest errors at ET < 2 mm d^{−1}, while GOA_SVM and GWO_SVM produced larger errors when ET exceeded 6 mm d^{−1}. Similarly, at DE-Seh, GWO_SVM showed considerable deviation for ET > 5 mm d^{−1}.

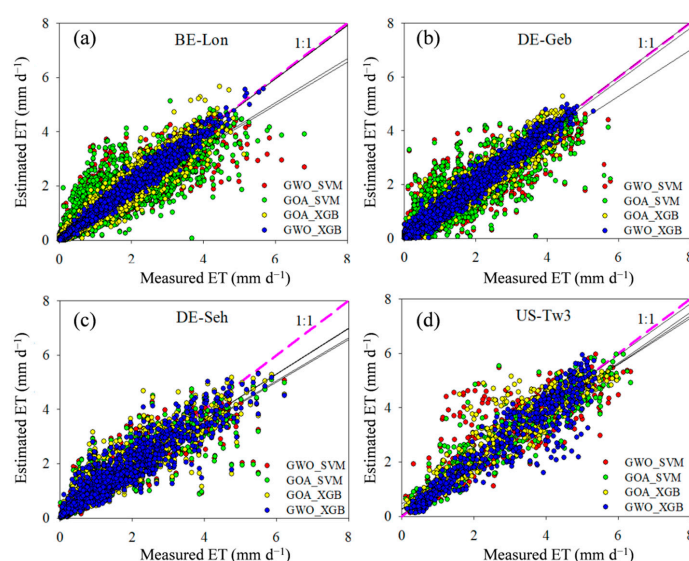


Figure 2. Comparison of simulated and observed daily ET at four representative stations (BE-Lon (a), DE-Geb (b), DE-Seh (c), and US-Tw3 (d)) using the four hybrid models (GWO_XGB, GOA_XGB, GWO_SVM, and GOA_SVM).

A boxplot analysis of model performance across all 11 farmland stations is shown in Figure 3. The results reveal notable variability among models. GWO_XGB consistently performed best, with the lowest median values for RMSE, MAE, PBIAS, and NRMSE, as well as the narrowest interquartile range and a median R^2 around 0.8, reflecting high and stable accuracy. Overall, while all four models achieved comparable overall performance, GWO_XGB ranked highest, followed by GOA_XGB and GOA_SVM, with GWO_SVM being the least accurate.

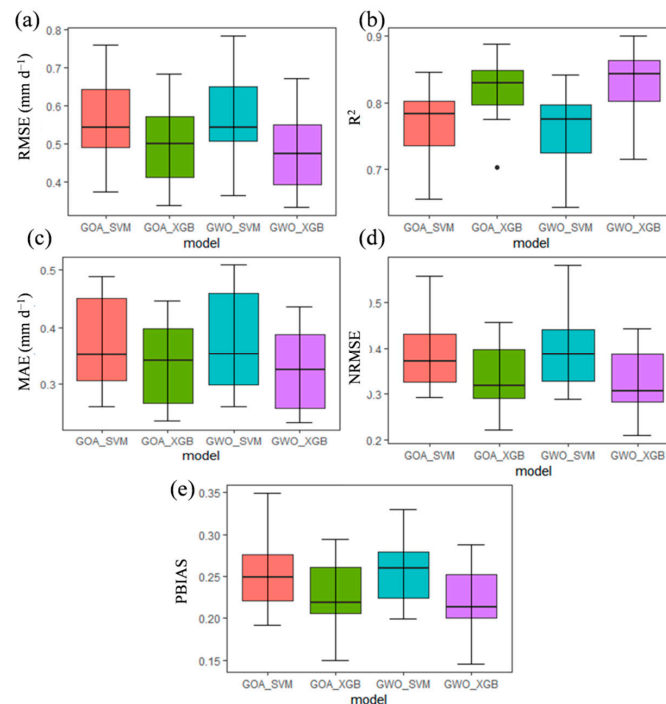


Figure 3. Model performance metrics across all 11 farmland stations for the four hybrid models (GWO_XGB, GOA_XGB, GWO_SVM, and GOA_SVM). Accuracy (RMSE (a), MAE (c)), Goodness-of-fit (R^2 (b)), Normalized accuracy (NRMSE (d)), and Bias (PBIAS (e)).

The current evaluation is based on historical meteorological data from Fluxnet 2015. For real-world irrigation applications, however, forecasting future ET is essential. NWP outputs offer a valuable basis for such forecasts, though they often contain systematic biases. Given the superior performance of GWO_XGB demonstrated in this section, we propose using this model to correct NWP forecast data before employing it for ET prediction. The following section will compare the observed-data-based simulations presented here with ET forecasts generated from bias-corrected NWP outputs, examining the feasibility of the hybrid modeling approach and the influence of forecast lead time on accuracy.

Since real-time forecasted NDVI and SWC are not available from the NWP outputs, and their short-term temporal persistence is generally high, we used the most recent observed values (i.e., lag-1) as proxies in the forecasting framework. To quantitatively justify this approach, we conducted autocorrelation function (ACF) analyses for the NDVI and SWC daily time series. As an illustrative example, Figure 4 presents the ACF plots for the DE-Geb station over lags up to 10 days. Both NDVI and SWC exhibit strong temporal autocorrelation, with the first-lag (1-day) coefficients exceeding 0.98 for NDVI and 0.96 for SWC. Even at a 10-day lag, the autocorrelation remains above 0.85 for NDVI and 0.80 for SWC. This high persistence confirms that using the previous day's observed NDVI and SWC values (NDVI_t, SWC_t) as inputs for 1- to 10-day ahead forecasts introduces only minor informational loss, thereby supporting the practicality of the proxy approach adopted in this study.

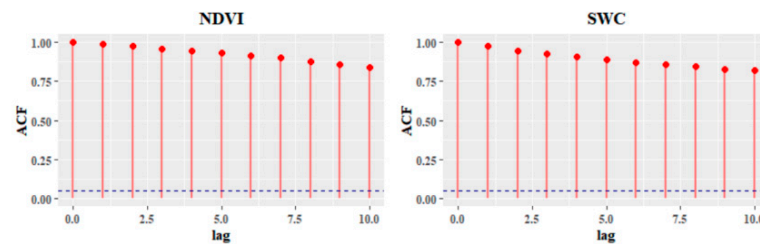


Figure 4. ACF plots of NDVI and SWC flux data at DE-Geb station.

3.2. Correction of NWP Variables

The NWP data retrieved from GEFS v12 comprises Tmax, Tmin, Rs, and VPD. Due to uncertainties in the initial conditions of NWP models and discrepancies between forecast outputs and actual measurements, the raw NWP data cannot be directly used for ET forecasting. Therefore, in this study, the four meteorological variables extracted from the NWP outputs were used as inputs, and the GWO_XGB model—which demonstrated the best simulation performance in the previous section—was employed to correct the biases in Rs and VPD, thereby improving the agreement between forecasted and observed values.

3.2.1. Solar Radiation

To enhance the utility of NWP outputs, the GWO_XGB model was applied to correct the forecast Rs data. The statistics results of the corrected Rs are presented in Table A3 (see Appendix A). Overall, the accuracy of Rs forecasts generally decreased as the forecast lead time increased, though this trend was not strictly monotonic. For instance, at stations such as BE-Lon, CH-Oe2, and DE-Geb, the accuracy on the 10d lead time was not the lowest among the 10 forecast days. Taking the BE-Lon station as an example, the RMSE, R^2 , MAE, PBIAS, and NRMSE values on 10d were $4.789 \text{ MJ m}^{-2} \text{ d}^{-1}$, 0.607, $3.622 \text{ MJ m}^{-2} \text{ d}^{-1}$, 34.8%, and 46%, respectively. The PBIAS and NRMSE values were 2.7%, 2.6%, 2.5%, and 2.7% lower, respectively, than those on 7d, which had the lowest accuracy.

The most pronounced changes in Rs accuracy occurred between 1d and 3d. The average RMSE increased by 44% from 1d to 3d, while R^2 decreased by 37.6%. Beyond 3d, the rate of deterioration slowed. For example, the average RMSE on 3d was 4.5% lower than that on 5d, with a corresponding R^2 increase of 5.2%. In contrast, the difference between 5d and 7d was much smaller, with RMSE increasing by only 0.8% and R^2 decreasing by 0.9%.

Across all stations, RMSE on 1d ranged from 2.005 to $3.061 \text{ MJ m}^{-2} \text{ d}^{-1}$, while on 10d it ranged from 4.133 to $5.55 \text{ MJ m}^{-2} \text{ d}^{-1}$. The US-Tw3 station exhibited the lowest average RMSE ($3.754 \text{ MJ m}^{-2} \text{ d}^{-1}$), whereas US-CRT showed the highest ($5.211 \text{ MJ m}^{-2} \text{ d}^{-1}$). Similarly, R^2 on 1d ranged from 0.848 to 0.952, declining to 0.486–0.795 by 10d. US-Tw3 achieved the highest average R^2 (0.825), while US-ARM had the lowest (0.543). NRMSE values on 1d ranged from 10.2% to 25.7%, with some stations (e.g., IT-BCi and US-Tw3) maintaining NRMSE below 28% across all 10 days, indicating consistently better performance. PBIAS on 1d ranged from 6.7% to 19.2%, but exceeded 20% from 2d onward at most stations (range from 20.5% to 37.8%), except for US-Tw3. These results indicate that Rs correction performance was substantially higher on 1d compared to longer lead times, with US-Tw3 exhibiting the best overall accuracy.

Figure 5 illustrates the scatter plot of simulated and measured Rs, which were corrected by the GWO_XGB model using NWP outputs from the DE-Geb, IT-BCi, IT-CA2, and US-Tw3 stations. The results show that corrected Rs on 1d was the most accurate, followed by 2d and 3d, while forecasts for 7–10d showed greater scatter. At US-Tw3, data points on 1d were more tightly clustered than at the other three stations, possibly due to the station's higher sunlight availability and greater number of days with daily Rs exceeding $30 \text{ MJ m}^{-2} \text{ d}^{-1}$.

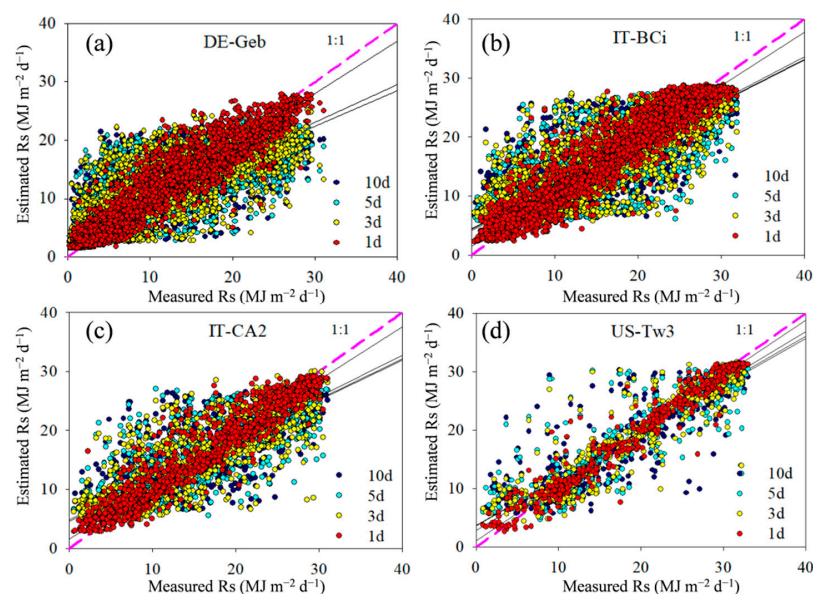


Figure 5. Evaluation of GWO_XGB-corrected R_s forecasts against observations at DE-Geb (a), IT-BCi (b), IT-CA2 (c), and US-Tw3 (d) stations based on NWP data.

Figure 6 illustrates the boxplots of the R_s statistical indicators across all 11 stations for the 10 forecast days. The R_s corrected for 1d was clearly superior to that for 2–9d, with greatly lower RMSE and higher R^2 . RMSE gradually increased from 2d to 10d, while R^2 exhibited a corresponding decline. Accuracy was similar between 5d and 10d, with 10d slightly outperforming 8d and 9d in some metrics. The mean R^2 for 1d was close to 0.9, whereas for the remaining days it remained below 0.7, indicating that the current NWP-based correction approach provides reliable R_s estimates only for 1d forecasts, with accuracy declining markedly as lead time increases.

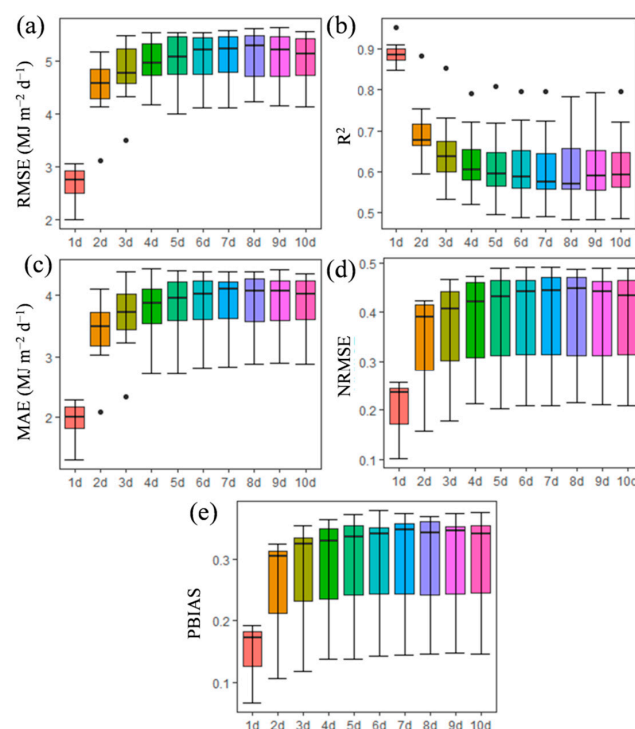


Figure 6. Accuracy performance of corrected R_s based on 10 days of NWP data at 11 farmland stations. Accuracy (RMSE (a), MAE (c)), Goodness-of-fit (R^2 (b)), Normalized accuracy (NRMSE (d)), and Bias (PBIAS (e)).

3.2.2. Vapor Pressure Deficit

The statistical results of VPD data corrected by the GWO_XGB model are summarized in Table A4 (see Appendix A). In contrast to the Rs correction, the VPD correction exhibited a clearer pattern, with accuracy generally declining as forecast lead time increased. Most stations followed this trend, with the exception of FR-Gri, US-ARM, and US-Tw3, where accuracy on 1d was slightly lower than on 2d (taking US-ARM as an example, $\text{RMSE} = 3.784 \text{ hPa d}^{-1}$ on 1d, $\text{RMSE} = 3.601 \text{ hPa d}^{-1}$ on 2d), and the accuracy of VPD data on 5d of the US-CRT station was the worst ($\text{RMSE} = 2.972 \text{ hPa d}^{-1}$, $R^2 = 0.496$).

The accuracy ranking of the other seven stations from 1 to 7d gradually decreased over days. The average RMSE and MAE values of all stations on 7d ($\text{RMSE} = 2.917 \text{ hPa d}^{-1}$, $\text{MAE} = 2.014 \text{ hPa d}^{-1}$, $R^2 = 0.471$) were 1% and 1.5% lower than those on 10d ($\text{RMSE} = 2.945 \text{ hPa d}^{-1}$, $\text{MAE} = 2.044 \text{ hPa d}^{-1}$, $R^2 = 0.462$), and the R^2 value increased by 1.9%. Additionally, the decrease rate of the accuracy of different forecast dates at each station gradually slowed down. The average RMSE values of VPD on 5d and 10d were 1.3 times and 1.4 times that of 1d, respectively. The average RMSE value of each station on 1d ranged from 1.521 to 3.784 hPa d^{-1} , and the average RMSE value on 10d ranged from 2.215 to 5.44 hPa d^{-1} . The average RMSE value of VPD at the CH-Oe2 station was the lowest (average $\text{RMSE} = 2.072 \text{ hPa d}^{-1}$), while the average RMSE value of VPD at the US-ARM station was the highest (average $\text{RMSE} = 4.794 \text{ hPa d}^{-1}$). The average R^2 value of each station on 1d ranged from 0.534 to 0.786, and the average R^2 value on 10d ranged from 0.359 to 0.62. Among them, the R^2 value at the IT-CA2 station was the highest (average $R^2 = 0.665$), and the R^2 value at the BE-Lon station was the lowest (average $R^2 = 0.426$). The average NRMSE values for each station varied between 38.9% and 74%, indicating unsatisfactory overall performance of the VPD data from the NWP. The US-Tw3 station had lower NRMSEs ranging from 23.1% to 44.5%, whereas the CH-Oe2 station had higher NRMSEs between 54.3% and 80.9%. The average PBIAS value of each station ranged from 29.3% to 52%. The PBIAS value of the US-Tw3 station was smaller (17.8–34.1%), and the PBIAS value of the CH-Oe2 station was larger (37.7–56.7%). The average MAE values of each station ranged from 1.403 to 3.416 hPa d^{-1} . The MAE value of the DE-Seh station was lower (1.111–1.502 hPa d^{-1}), and the MAE value of the US-ARM station was higher (2.586–3.856 hPa d^{-1}).

Figure 7 illustrates the scatter plot of the simulated and measured values of VPD corrected by the GWO_XGB model based on NWP data at CH-Oe2, DE-Geb, IT-CA2, and US-ARM stations. Based on the scatter distribution of model simulations for various forecast periods, it is evident that the VPD corrections produced low overall accuracy. The scatter model fit lines for the 10-day data indicate a trend of diverging from the 1:1 fit line over time, with the gap between two consecutive fit lines gradually reducing over time. Among them, the data accuracy on 1d is relatively high and the worst on 10d, which indicates that the delay of the forecast period will lead to a decrease in data accuracy, and the rate of decline will decrease accordingly. For the US-ARM station, the range of VPD values is relatively wide (0–50 hPa d^{-1}), while the data of other stations are less than 30 hPa d^{-1} , and the simulated scatter distribution of the model is more uniform than that of the CH-Oe2 station. The statistical indicators for the VPD data correction forecast over a 10-day period are presented in box plot form for 11 stations, as depicted in Figure 8. The boxplot for RMSE indicates a steady increase in the box for 1 to 10d, as expected, and the box plot for R^2 indicates a steady decrease in the box for 1 to 10d, with both trends gradually slowing down. Furthermore, the second-dimensional box in the box line diagram of R^2 exhibits a significantly lower value than the first-dimensional box (1d: 0.75, 2d: 0.65). However, it is not as prominent as the 2d box in the box line diagram of R^2 portrayed in Figure 6 (1d: 0.89, 2d: 0.69).

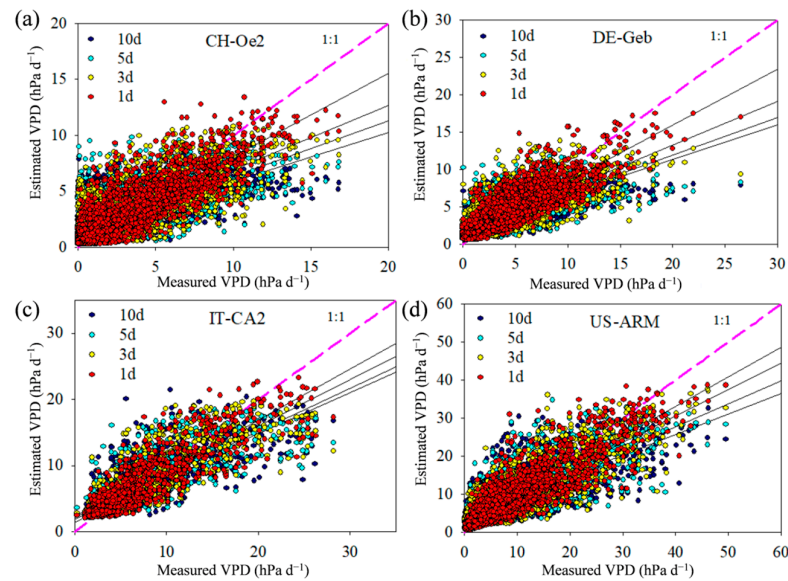


Figure 7. Scatter plot of VPD corrected by the GWO_XGB model at CH-Oe2 (a), DE-Geb (b), IT-CA2 (c), and US-ARM (d) stations based on NWP data.

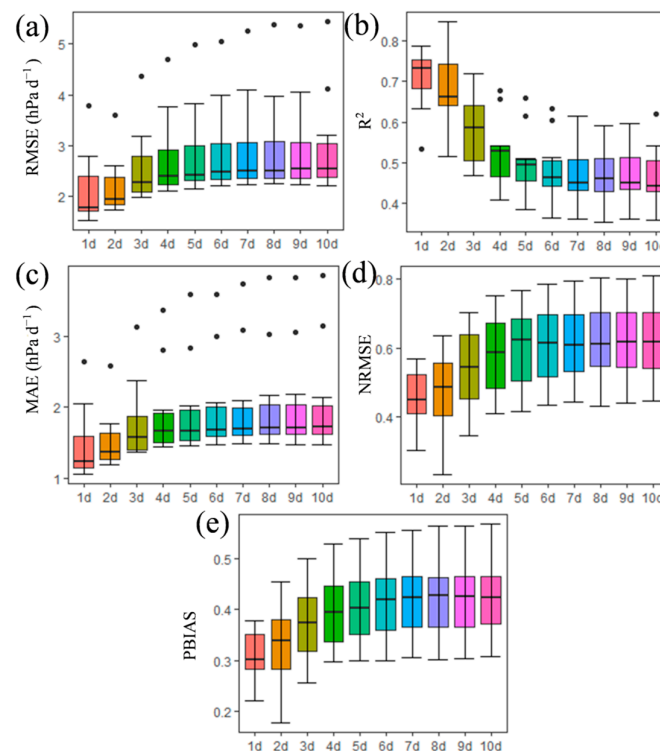


Figure 8. Accuracy performance of corrected VPD based on 10-day NWP data at 11 farmland stations. Accuracy (RMSE (a), MAE (c)), Goodness-of-fit (R^2 (b)), Normalized accuracy (NRMSE (d)), and Bias (PBIAS (e)).

Overall, both Rs and VPD corrections showed markedly better performance on 1d compared to the following 9 days. Rs correction for 5–10d exhibited instability, likely due to longer forecast horizons. Performance was generally consistent across most stations, with only a few stations, including DE-Geb, demonstrating higher simulation precision. These results indicate that the GWO_XGB model is suitable for correcting NWP data at farmland sites, and its corrected outputs can be used as inputs for subsequent ET forecasting.

3.3. Short-Term ET Forecast (1–3d)

The control group for this analysis consists of the models established in Section 3.1, which used only meteorological variables from the Fluxnet 2015 dataset. To evaluate the impact of bias-corrected forecast data on ET modeling, this study introduced a revised approach in which NWP outputs from GEFS v12 for 1–10 days ahead were first corrected (as described in Section 3.2 for R_s and VPD) and then used alongside additional variables (NDVI and SWC) from the flux dataset as inputs for ET forecasting.

The statistical results for short-term ET forecasts (1–3d) are summarized in Table A5 (see Appendix A). The average RMSE, MAE, PBIAS, and NRMSE on 1d were 14.7%, 15.4%, 16.7%, and 15.9% lower than those on 3d, and R^2 increased by 18.7%. This indicates that the forecast data on 1d significantly positively impact the accuracy of the ET model compared with 3d. However, the average accuracy of all models based on the forecast data of 2d was lower than that in 3d. The average RMSE on 2d increased by 2.4%, and R^2 decreased by 3.7%. The performance of the tree-based model was superior on 2d (average RMSE = 0.711 mm d^{-1}) compared to 3d (average RMSE = 0.715 mm d^{-1}). Using only the meteorological variables from the Fluxnet 2015 dataset as inputs (combination F1), the mean RMSE value was 0.526 mm d^{-1} for each station, which was 17.3%, 31.2%, and 29.5% lower than the RMSE values of 1–3d, respectively. Although the model's performance based on the NWP outputs was slightly degraded, the short-term ET forecast's mean accuracy (RMSE = 0.715 mm d^{-1} , $R^2 = 0.618$) still fell within an acceptable range.

Table A5 shows that the performance of GOA_SVM, GOA_XGB, and GWO_SVM models is relatively poor (the average RMSE values on 1 to 3d are 0.752 , 0.685 , and 0.757 mm d^{-1} , respectively, and the R^2 values are 0.582 , 0.648 and 0.578 , respectively), and the GWO_XGB model has the best performance (the average RMSE on 1 to 3d is 0.667 mm d^{-1} , $R^2 = 0.665$). This demonstrates the strong performance of the GWO_XGB model in forecasting ET, in agreement with the meteorological data of different forecast days. Based on the short-term ET forecast data (combination N1 to combination N3), the RMSE values of the GWO_XGB model simulated ET at each station ranged from 0.408 to 0.883 mm d^{-1} , and the R^2 values ranged from 0.549 to 0.828 . Despite the fact that the GWO_SVM model displayed the poorest performance among the four models, the average RMSE values remained at a low range of 0.673 – 0.82 mm d^{-1} . This strongly implies that the four models invented and examined in this research could generate more precise results in ET modelling. Furthermore, the precision of every model was at its highest level when tested at the IT-CA2 station. As an example, in the case of the combination N1, the RMSE was 0.423 mm d^{-1} , the R^2 was 0.697 , and the MAE was 0.308 mm d^{-1} . Additionally, all models delivered satisfactory outcomes at BE-Lon, DE-Geb, and DE-Kli stations with an average RMSE of 0.539 mm d^{-1} , R^2 of 0.721 , and MAE of 0.347 mm d^{-1} . However, the model showed lower performance at the IT-BCi station, recording an average RMSE of 0.729 mm d^{-1} , R^2 of 0.53 , and MAE of 0.549 mm d^{-1} .

Three farmland stations were chosen (BE-Lon, DE-Geb, and IT-CA2), and scatter plots of simulated and measured ET for the four models were drawn based on the 1–3d forecast data, as shown in Figure 9. The figure illustrates that the scatter distribution on 1d is much more uniform than that on the next 2 days, the fit line is closer to the 1:1 line, and the four models have a uniform performance. A clearer differentiation among the models is apparent in the scatter plot of IT-CA2_3d. The GWO_XGB model exhibits the most concentrated scatter; the remaining three models exhibit similar scatter distributions at the other stations. Figure 10 displays the statistical indicators of ET simulation from the models using short-term forecast data from each station. The boxplots demonstrate that the four models based on 1d station forecast data have the best performance. Moreover, the GOA_XGB and GWO_XGB models based on 2d data outperform their 3d models,

whereas the GOA_SVM and GWO_SVM models outperform GOA_XGB and GWO_XGB in contrast. In addition, the RMSE of the GWO_XGB model and the median line of the MAE box plot are kept at the lowest level, and the performance is the best. The RMSE is maintained at about 0.6 mm d^{-1} , showing high prediction accuracy. This further proves that the GWO_XGB model based on NWP outputs is suitable for ET forecast.

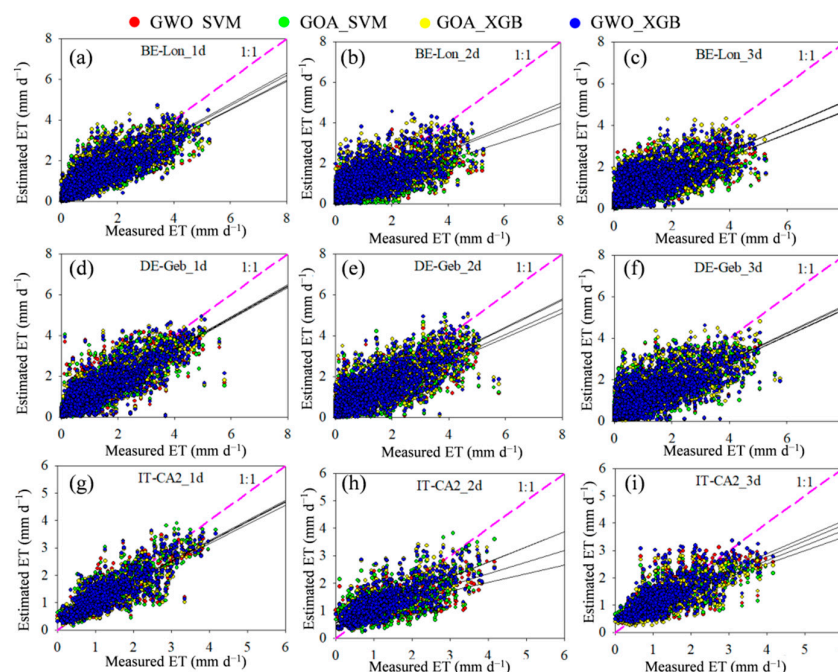


Figure 9. Short-term (1–3 day lead time) ET forecast performance of the four models at BE-Lon (a–c), DE-Geb (d–f), and IT-CA2 (g–i) stations.

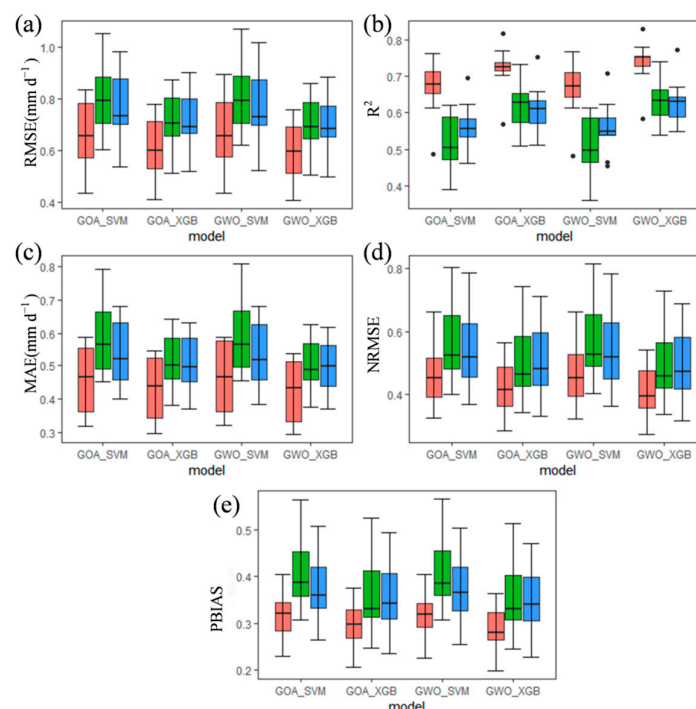


Figure 10. The accuracy performance of four models based on short-term ET forecast data at 11 farmland stations. Accuracy (RMSE (a), MAE (c)), Goodness-of-fit (R^2 (b)), Normalized accuracy (NRMSE (d)), and Bias (PBIAS (e)).

3.4. Medium-Term ET Forecast (4–10d)

The statistical results of daily ET forecasting based on NWP outputs for 4 to 10d are presented in Table A6 (see Appendix A). Across this period, the average RMSE values from 4 to 10d were 0.797, 0.757, 0.801, 0.758, 0.802, 0.802, and 0.767 mm d^{−1}, respectively. The average R² values were 0.534, 0.574, 0.528, 0.572, 0.528, 0.527, and 0.576, respectively, which indicates that the results of medium-term ET forecasting are very similar. There is a significant difference from the significant decrease in accuracy shown in Table A5. With the delay of the forecast period (4–10d), the downward trend of model accuracy becomes relatively flat. This plateau suggests that after the initial 3-day degradation, error growth becomes modest. The likely explanation is that the influence of slowly varying land surface conditions (NDVI/SWC) stabilizes predictions, while at these longer lead times, model uncertainty becomes dominated by structural error and the inherent limits of ET predictability, rather than NWP input accuracy.

It can be seen from the statistical results from 4 to 10d that there are certain fluctuations in various indicators between different forecast days, which cannot fully indicate that the forecast data accuracy of a certain day is the highest. At the IT-CA2 station, the model based on 7d forecast data exhibited the best performance (average RMSE = 0.516 mm d^{−1}, R² = 0.551, MAE = 0.38 mm d^{−1}), and the model based on 4d forecast data exhibited the worst performance (average RMSE = 0.571 mm d^{−1}, R² = 0.451, MAE = 0.43 mm d^{−1}). At the BE-Lon station, the model based on 5d forecast data showed the best performance (mean RMSE = 0.711 mm d^{−1}, R² = 0.531, MAE = 0.471 mm d^{−1}), and the model performance based on 6d forecast data was relatively poor (mean RMSE = 0.778 mm d^{−1}, R² = 0.444, MAE = 0.536 mm d^{−1}). At some stations, the model based on 10d forecast data performed better than 4d, such as the CH-Oe2 station, DE-Seh station, and FR-Gri station. The RMSE value on 4d was 2.9%, 1.8%, and 5.4% higher than that on 10d, and the R² value was 3.6%, 0.7%, and 8.2% lower. In all stations, the overall performance of the four models based on the forecast data on 5d (average RMSE = 0.757 mm d^{−1}) is better than that of the other six days (average RMSE is 0.797 mm d^{−1}, 0.801 mm d^{−1}, 0.758 mm d^{−1}, 0.802 mm d^{−1}, 0.803 mm d^{−1} and 0.767 mm d^{−1}, respectively), slightly worse than the forecast model on 3d (average RMSE = 0.746 mm d^{−1}). The analysis mentioned above shows that the model accuracy stays the same in the medium-term forecasting models, and the model accuracy is significantly lower than that of short-term forecasting.

We compared the performance of the four models at the same station (Figure 11), and the results were consistent with the performance of the models in Table A5. The model accuracy ranking was GWO_XGB (average RMSE = 0.732 mm d^{−1}, R² = 0.599) > GOA_XGB (average RMSE = 0.751 mm d^{−1}, R² = 0.581) > GOA_SVM (average RMSE = 0.824 mm d^{−1}, R² = 0.505) > GWO_SVM (average RMSE = 0.827 mm d^{−1}, R² = 0.503). Based on the forecast data from 4 to 10d, the scatter plots of ET measured values and forecasted values of these four models at the IT-CA2 station are shown in Figure 11. The figure illustrates that the scatter plots from 4d to 10d are very similar. It is evident that the fit between the forecasted and measured values is superior on certain days compared to others. Additionally, the fitting lines of all four models using 7d data are nearly identical in the graph, indicating a high level of consistency in performance. However, the scatter points of the GWO_XGB model in the figure are closer to the standard line than the other three models. The results show that the GWO_XGB model has better forecast performance than the other three models for the same input.

Figure 12 illustrates a box plot of statistical indicators for the four models based on the medium-term forecast data. There is not much variation in accuracy among models with forecast dates after 4d. The GOA_XGB and GWO_XGB models utilize 7d forecast data with comparable distribution rules, whereas the GOA_SVM and GWO_SVM models exhibit

slight fluctuations in their outcomes. In addition, the box plots of RMSE for the GWO_XGB and GOA_XGB models show a lower level of RMSE among these four models, and the box plots of R^2 also show a relatively high accuracy. Overall, it is still evident from the box plots that the GWO_XGB model has the best performance.

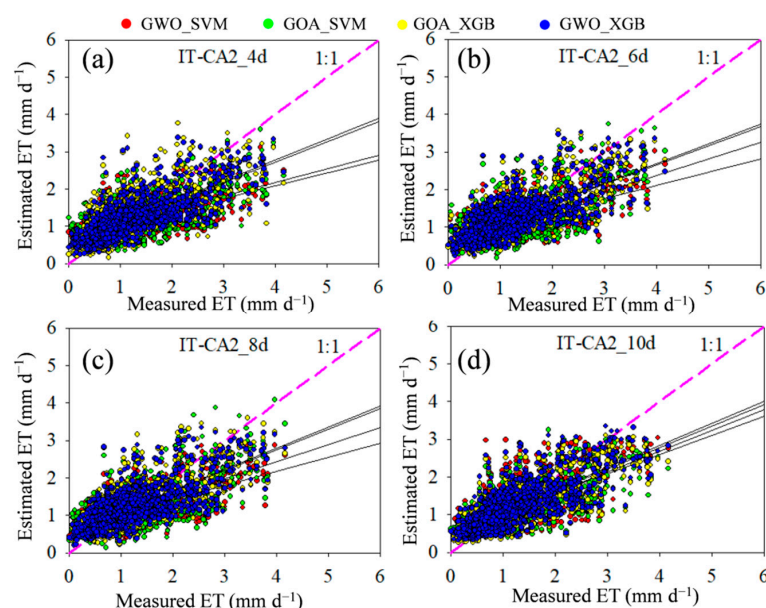


Figure 11. Scatter plot of forecasting ET at IT-CA2 stations based on medium-term ET forecast data for lead times of 4d (a), 6d (b), 8d (c), and 10d (d).

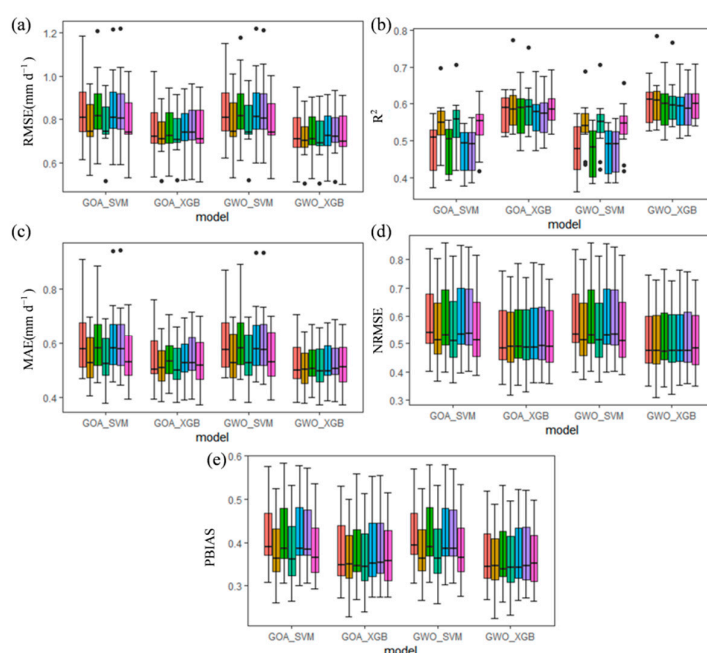


Figure 12. The accuracy performance of four models based on medium-term ET forecast data at 11 farmland stations. Accuracy (RMSE (a), MAE (c)), Goodness-of-fit (R^2 (b)), Normalized accuracy (NRMSE (d)), and Bias (PBIAS (e)).

Factor contribution plots for the model's forecast and correction process using the XGBoost model were calculated for the BE-Lon station (see Appendix A Figures A1–A10 for factor contribution plots for the other stations), as shown in Figure 13. The second layer represents the six meteorological factors input when the model predicts ET, and the third layer represents the meteorological factors input when the GWO_XGB model

corrects Rs and VPD, respectively. The figure illustrates that VPD contributes half of the total contribution to the ET forecast, while Rs contributes about a quarter of the total contribution, indicating that VPD has a critical influence on the ET forecast, followed by Rs. In contrast, the other four meteorological factors have a relatively low impact on the ET forecast, among which Tmax and Tmin have the lowest contribution. When correcting Rs and VPD, it can be seen from the value of factor contribution that Rs greatly influences both, especially the correction of Rs. The contribution of Rs data of NWP to Rs accounts for the vast majority, which is as high as 0.87. Tmax and Tmin had almost the same effect on the prediction of ET and Rs, while Tmax had a greater effect on the correction of VPD than Tmin. Figure 13 shows that the contribution value of Tmax was 0.18, which was 80% higher than that of Tmin, while the contribution of VPD was the lowest.

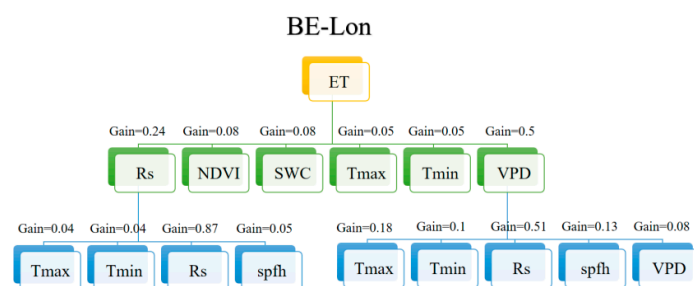


Figure 13. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at BE-Lon station.

Table 3 shows the statistical indicators of the mean values of the measured and forecasted values of the four models at 11 stations based on the NWP outputs. The RMSE, R^2 , MAE, PBIAS, and NRMSE of each station were 0.477–0.86 mm d^{−1}, 0.549–0.772, 0.361–0.622 mm d^{−1}, 22.6–47.1%, 32–69.5%, respectively. Comparing the results of Tables A3–A5, it is obvious that the ET cumulative prediction error of these 10 days is significantly lower than that of the next 9 days, but it will be slightly higher than that of 1d. Taking the DE-Geb station as an example, the RMSE and R^2 of 1d were 0.532 mm d^{−1} and 0.763, respectively. The RMSE and R^2 range from 2 to 10d were 0.661–0.733 and 0.553–0.637, respectively. The RMSE and R^2 of the ET cumulative value of the ten days were 0.615 mm d^{−1} and 0.692, respectively. The RMSE of cumulative ET increased by 15.6% compared with 1d and decreased by 7% compared with 2d, with the highest accuracy in the next 9 days. These four models' cumulative ET prediction performance is also the same as that described in the previous sections. Taking the US-ARM station as an example, the GWO_XGB model showed the highest prediction accuracy with the lowest RMSE value (RMSE = 0.638 mm d^{−1}), followed by the GOA_SVM model (RMSE = 0.649 mm d^{−1}), and the GWO_SVM model performed the worst (RMSE = 0.702 mm d^{−1}). The finding indicates that the medium-term ET forecasts are highly accurate and adaptable in the European and North American regions, especially in the farmland areas of these two continents, while the GWO_XGB model forecasts ET well for all combinations of inputs at all stations.

3.5. Implications of Irrigation Scheduling (3d, 5d, and 10d)

Based on the corrected NWP data, the present study uses a variety of optimized machine learning models to forecast medium-term ET across various farmland stations in Europe and North America. The purpose of the study is to make better use of the raw NWP data with low data accuracy and to explore a method to achieve accurate prediction of future crop ET, to achieve the purpose of planning water resources allocation and irrigation resource scheduling. The crop species in the study area of this study included maize, winter wheat, rice, and sugar beet, etc. As one of the main crops in Italy, maize needs a lot of

irrigation during the growth period. According to different growth stages, the crop needs to be irrigated within 3–5 days, and the irrigation cycle for winter wheat and other crops is also about 10 days. Therefore, combined with the local crop irrigation cycle and the forecast of ET of farmland crops, timely adjustment of crop irrigation amount at the next irrigation is conducive to improving the quality of farmland water supply, timely grasping the water demand of crops, and ensuring the safe production of farmland.

Table 3. Statistical values of the cumulative values of four models predicting ET at 11 farmland stations based on the medium-term numerical weather forecast data (1–10d) and the Fluxnet 2015 dataset.

Station	Model/Statistical Indicators	RMSE (mm d ^{−1}) Accuracy	R ² Goodness-of-Fit	MAE (mm d ^{−1}) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
BE-Lon	GOA_SVM	0.693	0.580	0.460	0.484	0.728
	GOA_XGB	0.634	0.632	0.441	0.464	0.666
	GWO_SVM	0.698	0.574	0.462	0.486	0.734
	GWO_XGB	0.620	0.650	0.428	0.450	0.651
CH-Oe2	GOA_SVM	0.893	0.651	0.654	0.337	0.460
	GOA_XGB	0.799	0.714	0.592	0.305	0.412
	GWO_SVM	0.900	0.646	0.658	0.339	0.463
	GWO_XGB	0.786	0.724	0.581	0.299	0.405
DE-Geb	GOA_SVM	0.637	0.671	0.416	0.439	0.672
	GOA_XGB	0.596	0.710	0.402	0.424	0.629
	GWO_SVM	0.640	0.669	0.419	0.442	0.675
	GWO_XGB	0.587	0.718	0.395	0.416	0.619
DE-Kli	GOA_SVM	0.682	0.502	0.441	0.490	0.757
	GOA_XGB	0.586	0.628	0.413	0.459	0.651
	GWO_SVM	0.686	0.497	0.444	0.493	0.761
	GWO_XGB	0.556	0.668	0.390	0.434	0.617
DE-Seh	GOA_SVM	0.772	0.603	0.560	0.355	0.490
	GOA_XGB	0.691	0.678	0.515	0.327	0.439
	GWO_SVM	0.770	0.604	0.559	0.355	0.489
	GWO_XGB	0.675	0.695	0.503	0.319	0.428
FR-Gri	GOA_SVM	0.714	0.575	0.503	0.371	0.528
	GOA_XGB	0.602	0.695	0.433	0.320	0.445
	GWO_SVM	0.715	0.575	0.502	0.371	0.528
	GWO_XGB	0.585	0.712	0.421	0.311	0.433
IT-BCi	GOA_SVM	0.765	0.487	0.582	0.283	0.372
	GOA_XGB	0.667	0.613	0.509	0.248	0.324
	GWO_SVM	0.775	0.472	0.591	0.287	0.377
	GWO_XGB	0.659	0.624	0.502	0.244	0.320
IT-CA2	GOA_SVM	0.495	0.621	0.376	0.297	0.390
	GOA_XGB	0.457	0.656	0.344	0.272	0.361
	GWO_SVM	0.503	0.609	0.382	0.302	0.397
	GWO_XGB	0.451	0.664	0.343	0.270	0.356
US-ARM	GOA_SVM	0.702	0.601	0.496	0.342	0.483
	GOA_XGB	0.649	0.657	0.466	0.321	0.447
	GWO_SVM	0.710	0.592	0.503	0.346	0.489
	GWO_XGB	0.638	0.667	0.462	0.318	0.440
US-CRT	GOA_SVM	0.897	0.651	0.606	0.327	0.484
	GOA_XGB	0.788	0.726	0.548	0.296	0.425
	GWO_SVM	0.975	0.579	0.649	0.350	0.526
	GWO_XGB	0.781	0.730	0.545	0.294	0.422
US-Tw3	GOA_SVM	0.892	0.734	0.687	0.270	0.350
	GOA_XGB	0.743	0.816	0.564	0.221	0.292
	GWO_SVM	0.911	0.712	0.689	0.271	0.357
	GWO_XGB	0.719	0.827	0.546	0.215	0.282

Note: Bold values indicate the best model performance for each metric: the lowest RMSE, MAE, PBIAS, and NRMSE, and the highest R².

As irrigation is limited to the crop sowing period, we chose the partially vegetated time span as the sowing period based on the NDVI values. To conduct the research, it was postulated that the 11 stations were irrigated at intervals of every 3, 5, and 10 days. In the process of data processing, taking the irrigation once every 3 days as an example, the ET of the next 3 days forecasted by the GWO_XGB model is accumulated as the amount of irrigation required for a crop, and it is compared with the measured value of ET accumulation in 3 days. In order to clearly display the data, we plotted the boxplots of the 3-day, 5-day, and 10-day cumulative ET forecasted values of these 11 farmland stations compared with the measured values (Figure 14). It is used to analyze and explain the gap between the ET forecast results obtained based on the improved method of this study and the actual results when the crops need to be irrigated within 3–10 days in the actual irrigation application.

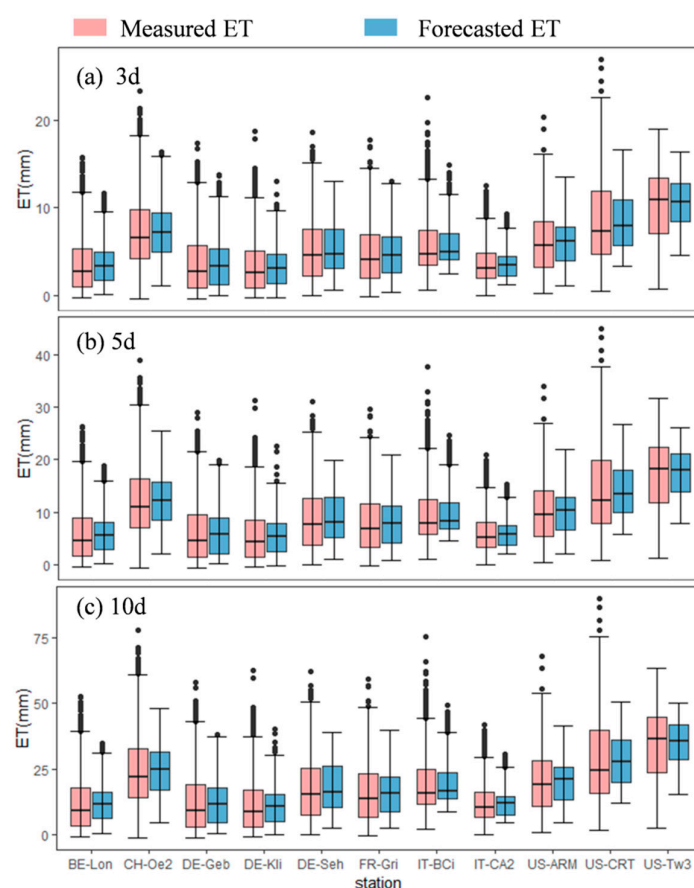


Figure 14. Comparison of forecasted and measured cumulative ET for 3 (a), 5 (b), and 10 (c) day irrigation scheduling windows across all stations.

It can be seen from Figure 14 that the difference between the forecasted value of crop ET accumulated for 3 days and the measured value of accumulated ET is small. The difference between the measured value and the forecasted value of the three irrigation cycles is ranked as 3d (average ET forecasted value = 5.03 mm, average ET measured value = 5.02 mm) < 5d (average ET forecasted value = 8.37 mm, average ET measured value = 8.38 mm) < 10d (average ET forecasted value = 16.73 mm, average ET measured value = 16.77 mm). As expected, the shorter the forecast date, the higher the accuracy of the forecasted value, so the cumulative result of the 3-day ET forecasted value is very close to the measured cumulative value. Within the 3-day irrigation period, the ET forecasted according to the improved method proposed by us is more accurate. Among them, it can be observed in Figure 14a that the forecasted value of ET at the DE-Seh station is the closest to the measured

value. The average measured value of ET is 5.24 mm, and the average forecasted value of ET is 5.24 mm. However, the ET forecasted value of the US-CRT station is quite different from the measured value. The average ET measured value is 8.57 mm, and the average ET forecasted value is 8.44 mm. Taking the DE-Seh and US-CRT stations as examples, it can be seen from Figure 14b that the average irrigation application amount forecasted at the DE-Seh station is 0.01 mm more than the actual average irrigation application amount, while the average irrigation application amount forecasted at the US-CRT station is 0.19 mm less than the actual average irrigation application amount. This is 160% and 52% higher than the cumulative irrigation difference of 3 days, respectively. In Figure 14c, the forecasted average irrigation application at the DE-Seh station is 0.01 mm more than the actual average irrigation application, and the forecasted average irrigation application at the US-CRT station is 0.42 mm less than the actual average irrigation application. This was 7.6% and 121% higher than the cumulative irrigation difference of 5 days, respectively. It can be seen from the above analysis that the difference between the 3–10d cumulative irrigation difference at the DE-Seh station is the smallest compared with other stations, especially the difference between 5d and 10d is closer, which indicates that at this station, the ET forecasted based on the improved method of this study can have a very small error even if the irrigation period is 10 days. At the US-CRT station, the forecasted ET was significantly different from the actual ET in the case of an irrigation cycle of 10 days, while in the case of an irrigation cycle of 5 days, the difference between the forecasted ET and the actual ET was significantly smaller than the difference between the measured ET and the forecasted value at the DE-Seh station compared with the cycle of 3 days. Therefore, we can conclude that in the case of long irrigation period, the ET prediction results of most stations still have high accuracy, while in some farmland stations with sufficient precipitation and high irrigation water demand, there is a relatively large error in the case of irrigation period of 10 days, but it can also have good prediction performance in the case of 3 days and 5 days. Therefore, in the practical application of irrigation scheduling, we recommend the use of appropriate prediction methods in the same area as the climate of the study station, combined with the conclusions of this study and the consideration of the crop irrigation cycle, to select the cumulative value of the forecasted days with less error.

4. Discussion

This study utilized meteorological factors from the Fluxnet 2015 dataset—including ET, Rs, WS, Tmax, Tmin, VPD, and SWC—to simulate ET using hybrid models combining the GOA and GWO with SVM and XGBoost. The results demonstrate that all four hybrid models performed effectively in the ET simulation. Notably, models combining heuristic algorithms with XGBoost showed superior performance, with the GWO_XGB model achieving the best results. Previous studies in various regions have validated the feasibility and high performance of both GWO and XGBoost for ET estimation [42,60–64]. Therefore, coupling these two robust algorithms can theoretically achieve good performance.

Building upon this methodological foundation, the study further examined the differences between actual and forecast ET across various crops and stations (Figure 15). Winter wheat emerged as the most prevalent crop (35%), benefiting from favorable climatic conditions driven by the Atlantic warm current across the European plains, while the sunnier, drier southeastern regions showed a higher concentration of maize cultivation [65]. Notably, maize at FR-Gri and rice at US-Tw3 exhibited the highest actual ET ($>2.5 \text{ mm d}^{-1}$), whereas spring barley at DE-Kli showed the lowest demand. The generally lower ET for other crops—excluding rice—can be attributed to both crop-specific traits and the high precipitation levels characteristic of many station locations. Europe's topography facilitates inland moisture penetration, leading to widespread precipitation beyond coastal areas [66],

which in turn significantly modulates regional ET patterns and supports adjustments in projected irrigation volumes for crops like maize in wetter zones [67,68].

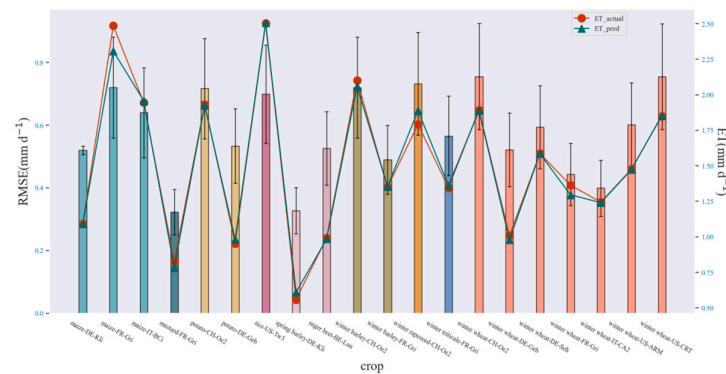


Figure 15. The forecast accuracy and ET of farmland stations planted with different crops based on the GWO_XGB model.

Regional ET patterns are closely linked to climatic conditions. For instance, seasonal ET for winter wheat and maize in the North China Plain is significantly higher than that for corresponding crops in Europe and North America, primarily due to distinct local precipitation regimes [69]. Similarly, actual ET estimates in southeastern Europe align closely with the observations in this study, although their forecasted values are generally slightly higher [70]. These comparative analyses highlight the strong geographic dependency of ET, which is consistent with the observed spatial pattern in this study, where the highest ET occurred at the southernmost station (IT-BCi)—a region characterized by ample precipitation and irrigation. Such consistencies underscore that ET variability can be largely explained by local climatic drivers.

Further analysis of forecast accuracy revealed that the performance of the GWO_XGB model varied by crop and location (Figure 15). The lowest RMSE values were observed for mustard (FR-Gri) and spring barley (DE-Kli), while winter wheat forecasts were most accurate at IT-CA2. Overall, mustard and spring wheat exhibited the smallest errors ($\sim 0.3 \text{ mm d}^{-1}$ RMSE). For other crops, station location played a clear role: maize performed best at DE-Kli ($\sim 0.6 \text{ mm d}^{-1}$), tomatoes at DE-Geb, and winter wheat at IT-CA2. This spatial pattern suggests that stations in eastern and southern Europe may offer more favorable forecasting conditions. In a broader performance context, our model achieved ET prediction RMSE below 0.8 mm d^{-1} , surpassing the $\sim 1 \text{ mm d}^{-1}$ reported in similar studies using optimized models on Fluxnet data [71,72].

Regarding input variables, forecast data for NDVI and SWC are often challenging to obtain but typically show strong temporal persistence. Autocorrelation analysis for DE-Geb station (Figure 4) confirmed high stability over lags up to ten days, justifying the use of previous-day observed values as reliable model inputs without significant accuracy loss.

In the data preprocessing phase, the Savitzky–Golay filter was applied to smooth and interpolate the MODIS composite NDVI time series into a daily sequence. This approach effectively preserves the overall trend of the phenological curve while capturing vegetation dynamics to a reasonable extent. Although NDVI can undergo rapid changes during key phenological transitions such as fast growth or harvest periods, the autocorrelation analysis conducted in this study demonstrates strong short-term persistence of NDVI over 1–10 days (lag-1 autocorrelation coefficient > 0.98), indicating limited day-to-day variability. Consequently, this does not introduce significant bias into the model inputs. Moreover, our modeling results reveal that NDVI contributes relatively little to ET forecasts, and its short-term fluctuations have minimal impact on overall prediction accuracy. These findings

further justify the use of the filtered and interpolated NDVI as a robust input within the present forecasting framework.

Among meteorological drivers, R_s contributed ~25% to ET forecasts, reinforcing its established role as a key source of prediction forecast error [73]. Nevertheless, a fascinating issue emerged in this study when adjusting the corrected meteorological combinations. While full meteorological correction might be theoretically ideal, the highest forecast accuracy was achieved by correcting only radiation and relative humidity—not temperature. Temperature is typically among the most reliably forecasted NWP variables due to well-constrained physical parameterizations in atmospheric models and its direct linkage to radiative and advective processes [74]. Consequently, systematic biases in temperature forecasts are generally small, and any residual errors often resemble random noise rather than structured bias [75]. Statistically correcting such a low-bias variable risks amplifying stochastic fluctuations—essentially fitting to noise—rather than removing meaningful systematic error. From a physical perspective, temperature influences ET indirectly through its role in VPD and aerodynamic resistance. Since our correction framework already adjusts VPD directly, which integrates temperature and humidity effects, additional temperature correction would likely introduce redundancy and potentially distort the physically consistent relationship between temperature, humidity, and VPD. Thus, we opted to correct only R_s and VPD, which exhibit clearer systematic biases and more directly drive ET variability in the short to medium term.

The dominant role of VPD in driving ET forecast errors underscores its physiological importance, yet its correction proved particularly challenging across stations. This discrepancy may be attributed to the fact that VPD is a derived variable sensitive to both temperature and humidity, whose errors can amplify nonlinearly in NWP outputs. VPD exhibits a greater impact on ET in rice than in wheat, primarily through indirect pathways mediated by net radiation, which may complicate its direct correction in statistical models [76]. Furthermore, as highlighted in studies on warming patterns [77], asymmetric warming—where nighttime temperatures rise more than daytime temperatures—can alter canopy development and phenology, thereby indirectly affecting VPD dynamics and ET responses in complex ways. These interacting factors may limit the effectiveness of purely statistical bias correction for VPD, suggesting that future approaches could benefit from incorporating physical constraints or dynamic phenological adjustments.

It is important to acknowledge a key methodological consideration. The GWO_XGB model operates as a static post-processor, trained on a fixed historical period and applied without modification to future forecasts. This approach inherently assumes stationarity in the bias characteristics of the NWP system between the training and forecasting periods. While this is a common and effective practice for retrospective analysis and short-term operational forecasting, its performance may degrade if systematic biases in the NWP outputs shift over time due to factors such as climate change, NWP model updates, or changes in local land surface conditions. The operational longevity of the model could be enhanced by implementing adaptive strategies. Future work should thus explore dynamic updating mechanisms, such as online learning, periodic retraining using a sliding time window, or the assimilation of recent observations into an ensemble correction framework, to maintain forecast reliability under non-stationary conditions.

Collectively, these results affirm that NWP data, when strategically corrected, can substantially improve ET forecasts—particularly at medium-range timescales [9,12,15]. The coupled modeling approach presented here effectively supports trend analysis and forward prediction, providing actionable insights for irrigation planning and water resource management. Nevertheless, limitations remain, including data gaps or inaccuracies for certain variables (e.g., WS). Future research could integrate higher-resolution datasets and

incorporate additional parameters such as sunshine duration or cloud cover to further refine forecast corrections and enhance model applicability across diverse agricultural and hydrological settings.

5. Conclusions

This study presents a practical framework for generating reliable medium-term ET forecasts by integrating bias-corrected NWP data with optimized machine learning models. The findings offer direct insights for refining irrigation strategies and confirm the value of hybrid modeling in agricultural water management.

The results showed that (1) The GWO_XGB model outperforms the other three evaluated models using corrected NWP outputs, whereas the GWO_SVM model demonstrates the lowest accuracy; (2) VPD and Rs are the primary parameters influencing ET forecast error, with VPD exhibiting the highest factor contribution, and Tmax and Tmin contribute the least; (3) The cumulative error of medium-term ET forecast is slightly lower than that of 1d and higher than that of 2 to 10d.

While the GWO_XGB model driven by corrected NWP outputs provides a solid foundation for medium-term ET forecasting, opportunities for improvement remain. Future work should focus on testing more advanced algorithms and extending the framework to diverse climate zones and crop types. We will further explore ensemble forecast methods and assess the practical impact of forecast errors on irrigation decisions to enhance the operational relevance of the approach.

Author Contributions: Conceptualization, L.W.; methodology, L.W. and S.Z.; software, L.W. and S.Z.; validation, L.W. and S.Z.; formal analysis, S.Z.; investigation, S.Z.; resources, S.Z.; data curation, S.Z.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z. and X.L.; visualization, S.Z.; supervision, L.W. and X.L.; project administration, L.W.; funding acquisition, L.W. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Foundation of the Education Department of Jiangxi Province (GJJ211904).

Data Availability Statement: The Fluxnet 2015 dataset is available at: <https://fluxnet.org/data/fluxnet2015-dataset/> (accessed on 22 February 2023). The GEFS v12 forecast data is available at: <https://noaa-gefs-retrospective.s3.amazonaws.com/index.html#> (accessed on 8 August 2021). The model implementation code used in this study is available on GitHub: <https://github.com/whispers28/ET-> (accessed on 28 January 2026).

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Abbreviations

The following abbreviations are used in this manuscript:

Nomenclature

Variables

ET	evapotranspiration (mm d^{-1})
Rs	solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)
WS	2-m wind speed (m s^{-1})
Tmax	maximum temperature ($^{\circ}\text{C}$)
Tmin	minimum temperature ($^{\circ}\text{C}$)
NDVI	normalized difference vegetation index
SWC	soil water content

VPD	vapor pressure deficit (hPa d^{-1})
spfh	2-m specific humidity (kg kg^{-1} dry air)
Tmax_+n, (n = 1, 2, ..., 10)	maximum temperature from GEFS v12 (K)
Tmin_+n, (n = 1, 2, ..., 10)	minimum temperature from GEFS v12 (K)
Rs_+n, (n = 1, 2, ..., 10)	solar radiation from GEFS v12 ($\text{MJ m}^{-2} \text{d}^{-1}$)
VPD_+n, (n = 1, 2, ..., 10)	vapor pressure deficit from GEFS v12 (hPa d^{-1})
SWC_t	soil water content a day ahead
NDVI_t	normalized difference vegetation index a day ahead
Abbreviations	
GOA	Grasshopper optimization algorithm
GWO	Gray wolf optimization
SVM	Support vector machine
XGB	Extreme gradient boosting
RMSE	root mean square error (mm d^{-1})
R ²	coefficient of determination
MAE	mean absolute error (mm d^{-1})
PBIAS	percent bias (%)
NRMSE	normalized root mean square error (%)
NWP	numerical weather prediction
PWF	public weather forecast

Appendix A

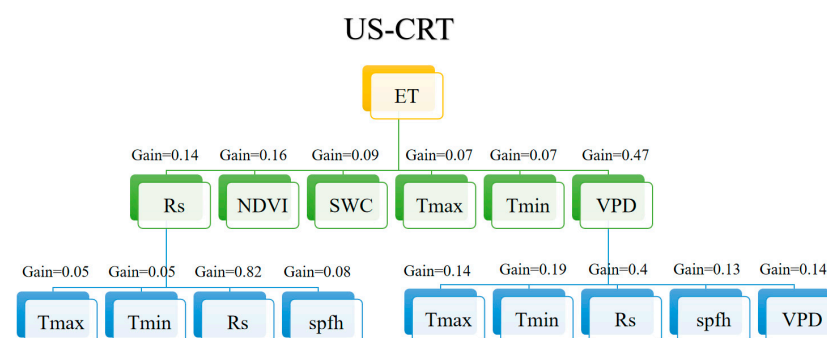


Figure A1. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at US-CRT station.

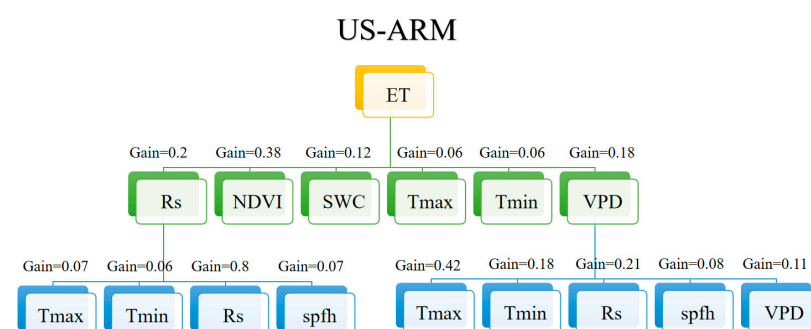


Figure A2. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at US-ARM station.

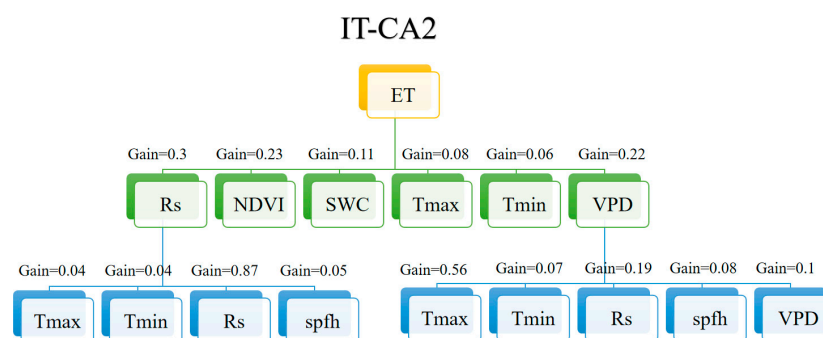


Figure A3. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at IT-CA2 station.

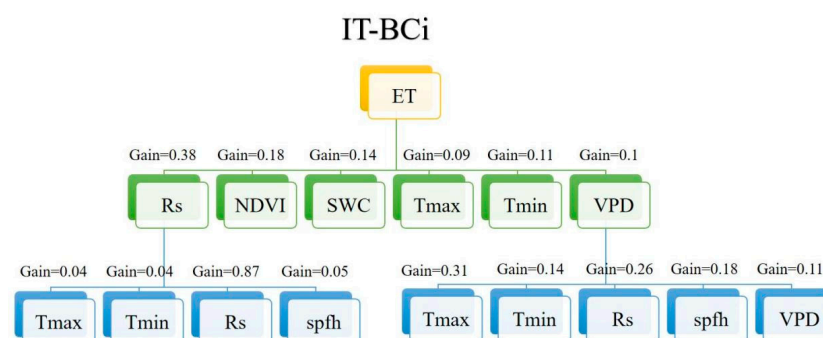


Figure A4. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at IT-BCi station.

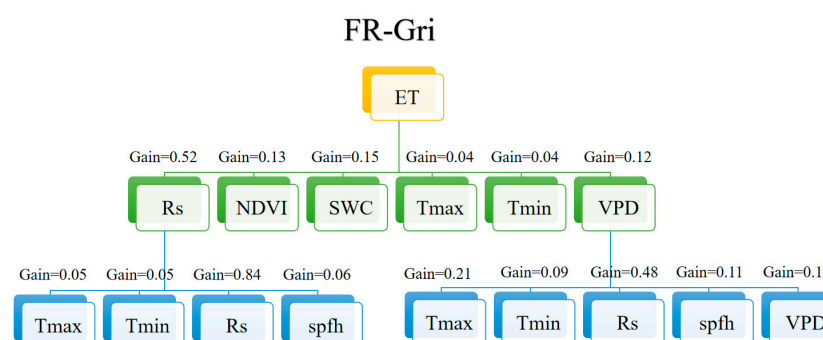


Figure A5. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at FR-Gri station.

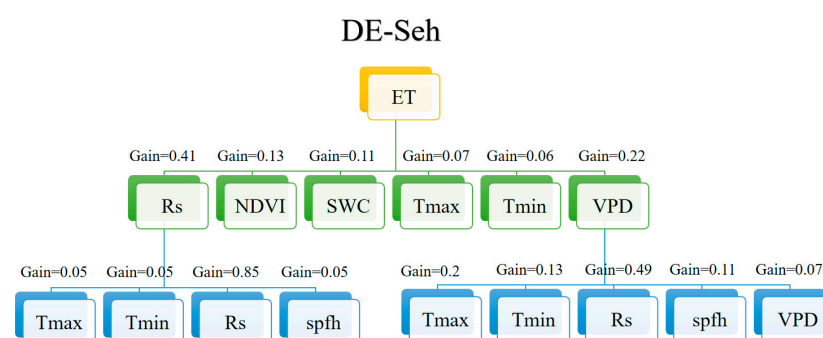


Figure A6. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at DE-Seh station.

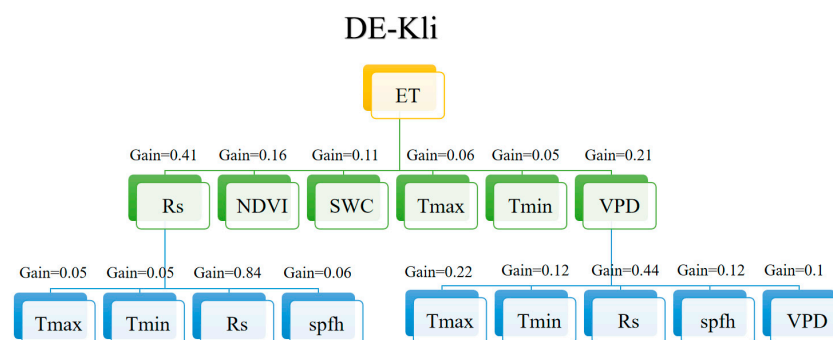


Figure A7. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at DE-Kli station.

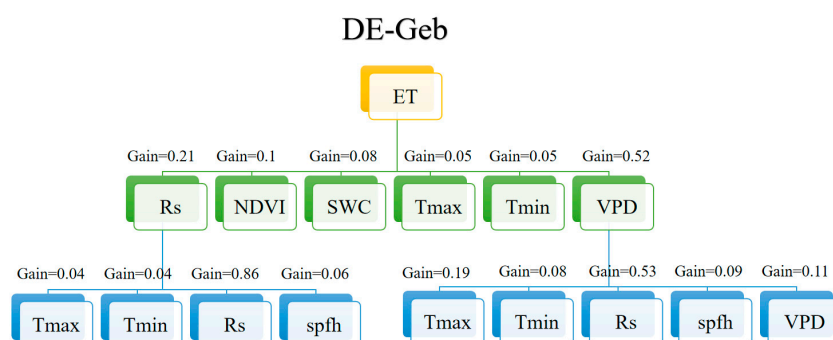


Figure A8. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at DE-Geb station.

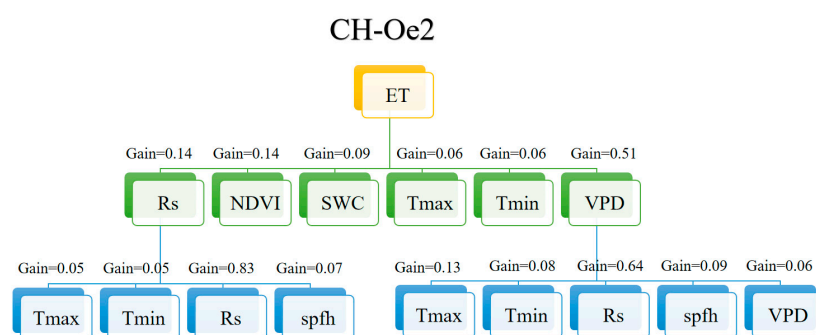


Figure A9. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at CH-Oe2 station.

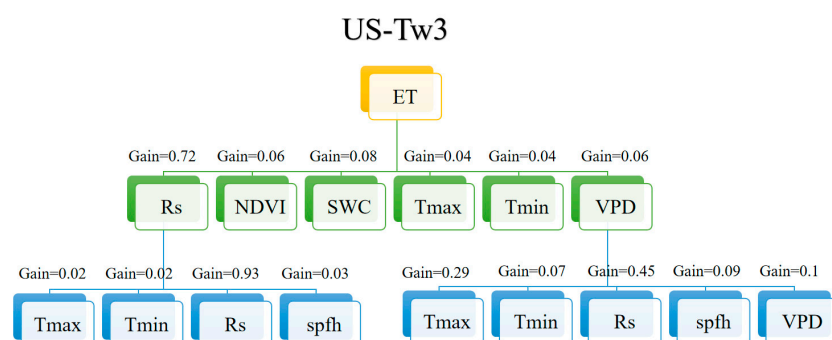


Figure A10. Meteorological factor contribution map for forecasting ET, correcting Rs and VPD using XGBoost at US-Tw3 station.

Table A1. NWP data of 11 flux stations.

Station	Rs_n	spfh_n						VPD_n	Tmax_1	Tmin_1	Rs_1	VPD_1	Tmax_2	Tmin_2	Rs_2	VPD_2	Tmax_3	Tmin_3	Rs_3	VPD_3	Tmax_4	Tmin_4	Rs_4	VPD_4
BE-Lon	567.10	0.03						1.15	280.90	279.65	120.04	3.01	281.26	279.61	120.72	3.00	281.18	279.55	120.45	2.95	281.20	279.57	120.72	3.00
CH-Oe2	603.09	0.03						1.05	279.36	277.96	137.56	2.77	279.56	277.74	138.04	2.80	279.46	277.65	137.58	2.77	279.42	277.62	138.04	2.80
DE-Geb	568.22	0.03						1.13	280.29	278.76	121.93	3.98	280.59	278.78	122.15	4.00	280.55	278.76	121.40	3.93	280.50	278.70	122.15	4.00
DE-Kli	573.25	0.02						1.16	280.46	279.05	127.75	3.26	280.79	279.01	127.76	3.28	280.76	279.01	126.32	3.24	280.71	278.96	127.76	3.28
DE-Seh	598.80	0.03						1.28	282.29	281.00	134.38	3.46	282.56	280.86	134.45	3.49	282.45	280.77	132.32	3.40	282.47	280.80	134.45	3.49
FR-Gri	612.06	0.03						1.34	282.92	282.06	136.76	3.86	283.64	282.03	136.88	3.89	283.59	281.99	135.47	3.81	283.59	281.99	136.88	3.89
IT-BCi	807.80	0.03						1.67	286.78	284.80	182.66	5.49	286.60	284.33	182.76	5.57	286.53	284.24	182.76	5.55	286.48	284.20	182.76	5.57
IT-CA2	755.84	0.03						1.47	284.66	282.83	179.96	6.49	284.81	282.69	179.94	6.57	284.81	282.70	179.82	6.50	284.73	282.60	179.94	6.57
US-ARM	811.48	0.03						2.02	287.86	286.31	194.67	9.11	291.07	286.47	195.58	9.13	291.17	286.53	194.99	9.02	291.23	286.62	195.58	9.13
US-CRT	676.61	0.03						1.51	283.21	282.19	155.98	4.64	284.34	282.24	156.33	4.74	284.43	282.27	155.98	4.64	284.40	282.28	156.33	4.74
US-Tw3	915.04	0.03						1.81	288.83	286.55	227.66	9.12	294.38	286.57	228.55	9.23	294.19	286.49	226.91	8.97	294.22	286.48	228.55	9.23
Station	Tmax_5	Tmin_5	Rs_5	VPD_5	Tmax_6	Tmin_6	Rs_6	VPD_6	Tmax_7	Tmin_7	Rs_7	VPD_7	Tmax_8	Tmin_8	Rs_8	VPD_8	Tmax_9	Tmin_9	Rs_9	VPD_9	Tmax_10	Tmin_10	Rs_10	VPD_10
BE-Lon	281.18	279.53	120.23	2.96	281.16	279.49	120.72	3.00	281.16	279.51	119.42	2.98	281.10	279.44	120.72	3.00	281.01	279.37	120.72	3.00	281.05	279.34	119.51	2.97
CH-Oe2	279.38	277.59	137.81	2.77	279.33	277.52	138.04	2.80	279.32	277.48	137.19	2.78	279.31	277.50	138.04	2.80	279.23	277.38	138.04	2.80	279.16	277.32	136.78	2.76
DE-Geb	280.49	278.71	121.34	3.92	280.41	278.61	122.15	4.00	280.45	278.63	120.81	3.92	280.45	278.63	122.15	4.00	280.32	278.49	122.15	4.00	280.30	278.47	121.52	3.92
DE-Kli	280.67	278.94	126.76	3.26	280.58	278.82	127.76	3.28	280.60	278.84	126.58	3.25	280.60	278.84	127.76	3.28	280.48	278.72	127.76	3.28	280.46	278.70	127.38	3.21
DE-Seh	282.43	280.73	133.01	3.44	282.47	280.77	134.45	3.49	282.43	280.76	132.33	3.45	282.34	280.62	134.45	3.49	282.29	280.58	134.45	3.49	282.39	280.66	133.86	3.47
FR-Gri	283.61	281.97	135.66	3.79	283.58	281.97	136.88	3.89	283.51	281.92	136.10	3.85	283.49	281.88	136.88	3.89	283.50	281.88	136.88	3.89	283.49	281.86	135.51	3.82
IT-BCi	286.42	284.13	181.95	5.51	286.43	284.15	182.76	5.57	286.45	284.15	182.77	5.52	286.45	284.16	182.76	5.57	286.47	284.16	182.76	5.57	286.40	284.11	182.31	5.54
IT-CA2	284.69	282.57	179.06	6.45	284.61	282.54	179.94	6.57	284.62	282.51	178.22	6.36	284.61	282.49	179.94	6.57	284.54	282.41	179.94	6.57	284.57	282.44	178.35	6.46
US-ARM	291.24	286.61	193.96	8.99	291.16	286.48	195.58	9.13	291.13	286.52	194.37	9.06	291.17	286.51	195.58	9.13	291.00	286.39	195.58	9.13	290.87	286.23	194.57	9.01
US-CRT	284.33	282.20	153.47	4.54	284.31	282.22	156.33	4.74	284.27	282.11	153.36	4.60	284.29	282.12	156.33	4.74	284.10	281.90	156.33	4.74	284.02	281.77	155.50	4.57
US-Tw3	294.23	286.41	227.30	9.01	294.18	286.37	228.55	9.23	294.12	286.40	226.90	9.05	294.17	286.39	228.55	9.23	294.23	286.48	228.55	9.23	294.52	286.55	225.35	9.05

Table A2. Statistical values of four models simulating ET at 11 cropland stations in the Fluxnet 2015 dataset.

Station	Model/Statistical Indicators	RMSE (mm d ^{−1}) Accuracy	R ² Goodness-of-Fit	MAE (mm d ^{−1}) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
BE-Lon	GOA_SVM	0.488	0.804	0.301	0.294	0.476
	GOA_XGB	0.439	0.842	0.272	0.266	0.428
	GWO_SVM	0.497	0.796	0.303	0.296	0.485
	GWO_XGB	0.419	0.856	0.259	0.253	0.409
CH-Oe2	GOA_SVM	0.663	0.796	0.456	0.240	0.349
	GOA_XGB	0.599	0.834	0.415	0.218	0.315
	GWO_SVM	0.670	0.792	0.464	0.244	0.352
	GWO_XGB	0.582	0.843	0.405	0.213	0.306
DE-Geb	GOA_SVM	0.433	0.846	0.261	0.273	0.452
	GOA_XGB	0.418	0.857	0.261	0.272	0.436
	GWO_SVM	0.439	0.842	0.266	0.277	0.459
	GWO_XGB	0.394	0.873	0.244	0.255	0.411
DE-Kli	GOA_SVM	0.493	0.733	0.309	0.349	0.556
	GOA_XGB	0.405	0.820	0.261	0.295	0.457
	GWO_SVM	0.513	0.712	0.293	0.330	0.579
	GWO_XGB	0.392	0.831	0.255	0.287	0.442
DE-Seh	GOA_SVM	0.542	0.800	0.351	0.223	0.344
	GOA_XGB	0.499	0.830	0.341	0.216	0.317
	GWO_SVM	0.543	0.798	0.353	0.224	0.345
	GWO_XGB	0.473	0.847	0.324	0.206	0.300
FR-Gri	GOA_SVM	0.499	0.784	0.334	0.248	0.371
	GOA_XGB	0.400	0.861	0.278	0.207	0.298
	GWO_SVM	0.520	0.765	0.349	0.259	0.387
	GWO_XGB	0.386	0.871	0.269	0.200	0.287
IT-BCi	GOA_SVM	0.621	0.655	0.445	0.219	0.306
	GOA_XGB	0.575	0.702	0.417	0.205	0.283
	GWO_SVM	0.630	0.643	0.456	0.224	0.310
	GWO_XGB	0.562	0.716	0.409	0.201	0.277
IT-CA2	GOA_SVM	0.372	0.764	0.262	0.207	0.294
	GOA_XGB	0.336	0.807	0.236	0.186	0.265
	GWO_SVM	0.364	0.776	0.261	0.206	0.287
	GWO_XGB	0.332	0.812	0.232	0.183	0.262
US-ARM	GOA_SVM	0.559	0.728	0.397	0.280	0.394
	GOA_XGB	0.507	0.774	0.362	0.256	0.358
	GWO_SVM	0.559	0.728	0.396	0.280	0.394
	GWO_XGB	0.497	0.783	0.354	0.250	0.351
US-CRT	GOA_SVM	0.759	0.737	0.488	0.264	0.410
	GOA_XGB	0.682	0.787	0.446	0.241	0.368
	GWO_SVM	0.782	0.722	0.508	0.274	0.422
	GWO_XGB	0.671	0.794	0.437	0.236	0.363
US-Tw3	GOA_SVM	0.747	0.810	0.489	0.192	0.293
	GOA_XGB	0.565	0.888	0.383	0.150	0.222
	GWO_SVM	0.776	0.796	0.509	0.200	0.304
	GWO_XGB	0.535	0.900	0.370	0.145	0.210

Note: Bold values indicate the best model performance for each metric: the lowest RMSE, MAE, PBIAS, and NRMSE, and the highest R².

Table A3. The statistical values of the GWO_XGB model used to correct Rs based on the NWP data.

Station	Date/ Statistical Indicators	RMSE (MJ m ⁻² d ⁻¹) Accuracy	R ² Goodness-of- Fit	MAE (MJ m ⁻² d ⁻¹) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
BE-Lon	1	2.460	0.896	1.807	0.174	0.236
	2	4.340	0.677	3.268	0.314	0.417
	3	4.606	0.637	3.460	0.333	0.443
	4	4.797	0.606	3.591	0.345	0.461
	5	4.859	0.596	3.637	0.350	0.467
	6	4.867	0.595	3.647	0.350	0.468
	7	4.917	0.586	3.718	0.357	0.473
	8	4.883	0.592	3.700	0.356	0.469
	9	4.813	0.604	3.646	0.350	0.463
	10	4.789	0.607	3.622	0.348	0.460
CH-Oe2	1	3.061	0.868	2.282	0.192	0.257
	2	4.830	0.671	3.685	0.310	0.406
	3	5.208	0.617	4.008	0.337	0.438
	4	5.504	0.573	4.276	0.359	0.463
	5	5.483	0.576	4.245	0.357	0.461
	6	5.541	0.567	4.315	0.363	0.466
	7	5.573	0.562	4.333	0.364	0.468
	8	5.617	0.555	4.372	0.367	0.472
	9	5.624	0.554	4.383	0.368	0.473
	10	5.550	0.566	4.318	0.363	0.466
DE-Geb	1	2.506	0.898	1.818	0.173	0.238
	2	4.334	0.695	3.251	0.309	0.412
	3	4.642	0.650	3.479	0.330	0.441
	4	4.789	0.628	3.623	0.344	0.455
	5	4.873	0.615	3.675	0.349	0.463
	6	4.808	0.625	3.629	0.345	0.457
	7	4.876	0.614	3.669	0.348	0.463
	8	4.794	0.627	3.610	0.343	0.455
	9	4.845	0.619	3.670	0.348	0.460
	10	4.871	0.615	3.684	0.350	0.463
DE-Kli	1	2.747	0.885	2.003	0.182	0.249
	2	4.581	0.680	3.486	0.317	0.416
	3	5.130	0.599	3.891	0.353	0.466
	4	5.196	0.589	4.001	0.363	0.472
	5	5.374	0.560	4.095	0.372	0.488
	6	5.401	0.556	4.166	0.378	0.490
	7	5.402	0.555	4.107	0.373	0.490
	8	5.367	0.561	4.061	0.369	0.487
	9	5.390	0.557	4.115	0.374	0.489
	10	5.376	0.559	4.126	0.375	0.488
DE-Seh	1	2.898	0.880	2.188	0.189	0.250
	2	4.911	0.656	3.755	0.324	0.424
	3	5.282	0.603	4.013	0.346	0.456
	4	5.341	0.593	4.084	0.352	0.461
	5	5.498	0.569	4.177	0.360	0.474
	6	5.374	0.588	4.075	0.352	0.464
	7	5.465	0.575	4.147	0.358	0.471
	8	5.490	0.570	4.226	0.365	0.474
	9	5.369	0.589	4.114	0.355	0.463
	10	5.417	0.581	4.144	0.358	0.467

Table A3. Cont.

Station	Date/ Statistical Indicators	RMSE (MJ m ⁻² d ⁻¹) Accuracy	R ² Goodness-of- Fit	MAE (MJ m ⁻² d ⁻¹) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
FR-Gri	1	2.835	0.875	2.149	0.182	0.240
	2	4.592	0.671	3.543	0.300	0.389
	3	4.772	0.645	3.725	0.316	0.404
	4	4.957	0.617	3.874	0.328	0.420
	5	5.078	0.598	3.957	0.335	0.430
	6	5.202	0.578	4.024	0.341	0.441
	7	5.226	0.575	4.098	0.347	0.443
	8	5.280	0.566	4.125	0.350	0.448
	9	5.201	0.579	4.071	0.345	0.441
	10	5.123	0.591	4.024	0.341	0.434
IT-BCi	1	2.507	0.910	1.825	0.116	0.159
	2	4.141	0.754	3.036	0.193	0.263
	3	4.321	0.732	3.234	0.205	0.274
	4	4.397	0.722	3.308	0.210	0.279
	5	4.414	0.720	3.311	0.210	0.280
	6	4.365	0.726	3.292	0.209	0.277
	7	4.387	0.724	3.328	0.211	0.278
	8	4.296	0.735	3.248	0.206	0.273
	9	4.376	0.725	3.313	0.210	0.278
	10	4.400	0.722	3.360	0.213	0.279
IT-CA2	1	2.567	0.903	1.905	0.123	0.165
	2	4.237	0.736	3.092	0.199	0.273
	3	4.535	0.697	3.410	0.220	0.292
	4	4.664	0.680	3.504	0.226	0.301
	5	4.650	0.682	3.534	0.228	0.300
	6	4.678	0.678	3.583	0.231	0.302
	7	4.695	0.676	3.592	0.232	0.303
	8	4.607	0.688	3.536	0.228	0.297
	9	4.615	0.686	3.537	0.228	0.298
	10	4.673	0.679	3.607	0.233	0.301
US-ARM	1	2.990	0.848	2.169	0.129	0.177
	2	4.855	0.599	3.779	0.224	0.288
	3	5.234	0.533	4.094	0.243	0.310
	4	5.299	0.522	4.125	0.245	0.314
	5	5.445	0.495	4.284	0.254	0.323
	6	5.476	0.489	4.288	0.254	0.325
	7	5.463	0.492	4.301	0.255	0.324
	8	5.502	0.484	4.312	0.256	0.326
	9	5.508	0.483	4.343	0.258	0.327
	10	5.492	0.486	4.351	0.258	0.326
US-CRT	1	2.943	0.869	2.198	0.163	0.218
	2	5.170	0.595	4.094	0.304	0.384
	3	5.485	0.544	4.379	0.325	0.407
	4	5.534	0.536	4.426	0.328	0.411
	5	5.526	0.538	4.396	0.326	0.410
	6	5.540	0.535	4.385	0.325	0.411
	7	5.482	0.546	4.384	0.325	0.407
	8	5.468	0.548	4.389	0.326	0.406
	9	5.529	0.538	4.417	0.328	0.410
	10	5.431	0.553	4.332	0.321	0.403

Table A3. Cont.

Station	Date/ Statistical Indicators	RMSE (MJ m ⁻² d ⁻¹) Accuracy	R ² Goodness-of- Fit	MAE (MJ m ⁻² d ⁻¹) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
US-Tw3	1	2.005	0.952	1.315	0.067	0.102
	2	3.112	0.884	2.094	0.106	0.158
	3	3.497	0.853	2.333	0.118	0.177
	4	4.177	0.790	2.711	0.138	0.212
	5	3.991	0.808	2.709	0.137	0.203
	6	4.117	0.796	2.814	0.143	0.209
	7	4.122	0.796	2.828	0.144	0.209
	8	4.239	0.784	2.883	0.146	0.215
	9	4.147	0.794	2.894	0.147	0.210
	10	4.133	0.795	2.886	0.146	0.210

Note: Bold values indicate the best model performance for each metric: the lowest RMSE, MAE, PBIAS, and NRMSE, and the highest R².

Table A4. The statistical values of the GWO_XGB model used to correct VPD based on the NWP data.

Station	Date/Statistical Indicators	RMSE (hPa d ⁻¹) Accuracy	R ² Goodness-of- Fit	MAE (hPa d ⁻¹) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
BE-Lon	1	1.710	0.633	1.129	0.376	0.569
	2	1.875	0.559	1.263	0.420	0.624
	3	2.058	0.469	1.364	0.454	0.685
	4	2.169	0.410	1.438	0.479	0.722
	5	2.215	0.385	1.461	0.486	0.737
	6	2.251	0.365	1.480	0.493	0.749
	7	2.254	0.363	1.489	0.496	0.750
	8	2.274	0.354	1.488	0.496	0.757
	9	2.257	0.361	1.491	0.496	0.752
	10	2.260	0.359	1.495	0.498	0.752
CH-Oe2	1	1.521	0.753	1.057	0.377	0.543
	2	1.779	0.662	1.274	0.455	0.635
	3	1.969	0.587	1.398	0.499	0.703
	4	2.104	0.528	1.481	0.529	0.751
	5	2.153	0.506	1.510	0.539	0.768
	6	2.201	0.484	1.542	0.550	0.786
	7	2.229	0.470	1.555	0.555	0.796
	8	2.250	0.460	1.578	0.563	0.803
	9	2.247	0.461	1.578	0.563	0.802
	10	2.267	0.452	1.587	0.567	0.809
DE-Geb	1	1.696	0.754	1.185	0.296	0.424
	2	1.943	0.677	1.357	0.339	0.486
	3	2.171	0.597	1.493	0.373	0.543
	4	2.312	0.542	1.593	0.398	0.578
	5	2.394	0.510	1.640	0.410	0.599
	6	2.419	0.499	1.676	0.419	0.605
	7	2.436	0.493	1.688	0.422	0.609
	8	2.446	0.488	1.714	0.429	0.612
	9	2.465	0.480	1.709	0.427	0.616
	10	2.473	0.477	1.729	0.432	0.618

Table A4. Cont.

Station	Date/Statistical Indicators	RMSE (hPa d ^{−1}) Accuracy	R ² Goodness-of-Fit	MAE (hPa d ^{−1}) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
DE-Kli	1	1.777	0.690	1.236	0.377	0.542
	2	1.943	0.629	1.362	0.416	0.593
	3	2.266	0.495	1.579	0.482	0.692
	4	2.385	0.441	1.665	0.508	0.728
	5	2.465	0.402	1.728	0.528	0.753
	6	2.495	0.388	1.744	0.532	0.762
	7	2.509	0.381	1.764	0.539	0.766
	8	2.506	0.382	1.764	0.539	0.765
	9	2.535	0.368	1.779	0.543	0.774
	10	2.545	0.364	1.787	0.546	0.777
DE-Seh	1	1.646	0.693	1.111	0.318	0.472
	2	1.749	0.653	1.189	0.341	0.501
	3	2.065	0.517	1.371	0.393	0.592
	4	2.163	0.470	1.440	0.413	0.620
	5	2.201	0.450	1.473	0.422	0.631
	6	2.243	0.430	1.493	0.428	0.643
	7	2.247	0.427	1.502	0.431	0.644
	8	2.256	0.423	1.499	0.430	0.647
	9	2.228	0.438	1.480	0.424	0.639
	10	2.215	0.443	1.475	0.423	0.635
FR-Gri	1	1.742	0.734	1.176	0.302	0.448
	2	1.720	0.740	1.212	0.312	0.442
	3	2.083	0.619	1.385	0.356	0.536
	4	2.281	0.542	1.513	0.389	0.587
	5	2.422	0.485	1.568	0.403	0.623
	6	2.475	0.462	1.640	0.422	0.636
	7	2.502	0.450	1.650	0.424	0.643
	8	2.529	0.438	1.664	0.428	0.650
	9	2.538	0.434	1.674	0.431	0.653
	10	2.534	0.436	1.656	0.426	0.652
IT-BCi	1	2.253	0.534	1.536	0.276	0.404
	2	2.294	0.516	1.583	0.284	0.412
	3	2.373	0.482	1.636	0.294	0.426
	4	2.420	0.462	1.659	0.298	0.435
	5	2.422	0.460	1.667	0.299	0.435
	6	2.431	0.457	1.671	0.300	0.437
	7	2.464	0.442	1.702	0.306	0.442
	8	2.461	0.443	1.676	0.301	0.442
	9	2.466	0.440	1.696	0.305	0.443
	10	2.500	0.425	1.713	0.308	0.449
IT-CA2	1	2.399	0.786	1.626	0.247	0.365
	2	2.602	0.748	1.765	0.269	0.396
	3	2.749	0.719	1.867	0.284	0.418
	4	2.944	0.678	1.963	0.299	0.448
	5	3.030	0.659	2.014	0.306	0.461
	6	3.146	0.632	2.062	0.314	0.479
	7	3.218	0.616	2.091	0.318	0.490
	8	3.315	0.591	2.167	0.330	0.504
	9	3.296	0.596	2.180	0.332	0.502
	10	3.196	0.620	2.137	0.325	0.486

Table A4. Cont.

Station	Date/Statistical Indicators	RMSE (hPa d ⁻¹) Accuracy	R ² Goodness-of-Fit	MAE (hPa d ⁻¹) Accuracy	PBIAS Bias	NRMSE Normalized Accuracy
US-ARM	1	3.784	0.779	2.643	0.290	0.415
	2	3.601	0.800	2.586	0.283	0.394
	3	4.359	0.707	3.126	0.342	0.477
	4	4.703	0.658	3.370	0.369	0.515
	5	4.997	0.614	3.586	0.393	0.547
	6	5.057	0.605	3.593	0.394	0.554
	7	5.253	0.574	3.733	0.409	0.575
	8	5.387	0.551	3.832	0.420	0.590
	9	5.356	0.557	3.832	0.420	0.587
	10	5.440	0.543	3.856	0.422	0.596
US-CRT	1	2.376	0.677	1.556	0.328	0.501
	2	2.466	0.652	1.628	0.343	0.520
	3	2.817	0.546	1.870	0.394	0.594
	4	2.876	0.529	1.874	0.395	0.606
	5	2.972	0.496	1.912	0.403	0.627
	6	2.913	0.514	1.946	0.410	0.614
	7	2.882	0.525	1.890	0.398	0.608
	8	2.849	0.535	1.904	0.401	0.600
	9	2.807	0.548	1.891	0.399	0.592
	10	2.862	0.533	1.904	0.401	0.603
US-Tw3	1	2.779	0.740	2.047	0.222	0.301
	2	2.136	0.847	1.644	0.178	0.231
	3	3.177	0.662	2.372	0.257	0.344
	4	3.765	0.525	2.808	0.304	0.408
	5	3.831	0.507	2.844	0.308	0.415
	6	3.992	0.464	2.993	0.324	0.433
	7	4.091	0.437	3.086	0.334	0.443
	8	3.964	0.472	3.025	0.328	0.430
	9	4.049	0.450	3.066	0.332	0.439
	10	4.106	0.433	3.143	0.341	0.445

Note: Bold values indicate the best model performance for each metric: the lowest RMSE, MAE, PBIAS, and NRMSE, and the highest R².

Table A5. Statistical values of four models simulating ET at 11 cropland stations based on the short-term NWP data (1–3d) and the Fluxnet 2015 dataset.

Station	Statistical Indicators	RMSE (mm d ⁻¹)			R ²			MAE (mm d ⁻¹)			PBIAS Bias			NRMSE Normalized Accuracy		
		Accuracy			Goodness-of-Fit			Accuracy								
	Model/Date	1d	2d	3d	1d	2d	3d	1d	2d	3d	1d	2d	3d	1d	2d	3d
BE-Lon	GOA_SVM	0.549	0.764	0.695	0.718	0.473	0.555	0.358	0.537	0.459	0.377	0.564	0.482	0.577	0.803	0.730
	GOA_XGB	0.537	0.707	0.675	0.730	0.532	0.574	0.351	0.500	0.458	0.369	0.525	0.481	0.564	0.743	0.709
	GWO_SVM	0.553	0.774	0.694	0.715	0.462	0.555	0.361	0.538	0.460	0.380	0.565	0.483	0.581	0.813	0.729
	GWO_XGB	0.515	0.693	0.652	0.751	0.551	0.601	0.338	0.488	0.446	0.355	0.513	0.468	0.541	0.728	0.685
CH-Oe2	GOA_SVM	0.805	0.943	0.942	0.708	0.601	0.603	0.589	0.698	0.679	0.303	0.360	0.350	0.415	0.486	0.486
	GOA_XGB	0.746	0.873	0.868	0.746	0.653	0.657	0.546	0.643	0.632	0.281	0.331	0.326	0.385	0.450	0.447
	GWO_SVM	0.807	0.946	0.944	0.706	0.599	0.602	0.588	0.699	0.680	0.303	0.360	0.350	0.416	0.487	0.486
	GWO_XGB	0.734	0.855	0.850	0.754	0.667	0.671	0.539	0.622	0.617	0.278	0.320	0.318	0.378	0.441	0.438
DE-Geb	GOA_SVM	0.547	0.677	0.673	0.750	0.620	0.622	0.348	0.461	0.438	0.367	0.486	0.462	0.577	0.714	0.710
	GOA_XGB	0.524	0.645	0.664	0.770	0.652	0.631	0.336	0.440	0.438	0.354	0.464	0.462	0.552	0.680	0.700
	GWO_SVM	0.543	0.683	0.673	0.754	0.614	0.622	0.344	0.466	0.438	0.363	0.492	0.462	0.572	0.721	0.710
	GWO_XGB	0.513	0.637	0.652	0.779	0.661	0.644	0.329	0.439	0.432	0.347	0.463	0.455	0.541	0.672	0.688
DE-Kli	GOA_SVM	0.596	0.686	0.706	0.612	0.491	0.462	0.365	0.453	0.458	0.405	0.503	0.508	0.662	0.762	0.784
	GOA_XGB	0.504	0.604	0.639	0.718	0.596	0.546	0.338	0.430	0.445	0.376	0.478	0.494	0.560	0.670	0.710
	GWO_SVM	0.596	0.689	0.703	0.613	0.488	0.464	0.365	0.454	0.454	0.405	0.504	0.504	0.662	0.765	0.781
	GWO_XGB	0.485	0.581	0.618	0.740	0.626	0.577	0.327	0.410	0.424	0.363	0.455	0.471	0.538	0.645	0.686
DE-Seh	GOA_SVM	0.710	0.828	0.819	0.656	0.538	0.546	0.509	0.609	0.593	0.323	0.386	0.376	0.451	0.526	0.520
	GOA_XGB	0.635	0.740	0.754	0.724	0.628	0.612	0.469	0.548	0.554	0.298	0.348	0.352	0.403	0.469	0.479
	GWO_SVM	0.709	0.832	0.817	0.658	0.535	0.547	0.508	0.608	0.595	0.323	0.386	0.378	0.450	0.528	0.519
	GWO_XGB	0.598	0.733	0.736	0.755	0.633	0.630	0.441	0.545	0.540	0.280	0.346	0.343	0.380	0.465	0.467
FR-Gri	GOA_SVM	0.615	0.793	0.729	0.675	0.468	0.549	0.437	0.568	0.507	0.323	0.420	0.375	0.454	0.586	0.539
	GOA_XGB	0.562	0.676	0.670	0.726	0.603	0.610	0.402	0.488	0.473	0.297	0.361	0.350	0.416	0.499	0.495
	GWO_SVM	0.617	0.794	0.729	0.673	0.466	0.549	0.432	0.567	0.507	0.319	0.419	0.375	0.456	0.587	0.539
	GWO_XGB	0.532	0.656	0.652	0.754	0.626	0.631	0.383	0.474	0.460	0.283	0.350	0.340	0.394	0.485	0.482
IT-BCi	GOA_SVM	0.762	0.825	0.780	0.488	0.405	0.464	0.572	0.631	0.596	0.278	0.307	0.290	0.371	0.401	0.379
	GOA_XGB	0.698	0.746	0.747	0.569	0.509	0.511	0.527	0.569	0.573	0.256	0.277	0.278	0.339	0.363	0.363
	GWO_SVM	0.767	0.827	0.788	0.482	0.402	0.454	0.578	0.633	0.602	0.281	0.308	0.293	0.373	0.402	0.383
	GWO_XGB	0.687	0.721	0.714	0.582	0.539	0.549	0.519	0.548	0.546	0.252	0.267	0.266	0.334	0.351	0.347
IT-CA2	GOA_SVM	0.438	0.604	0.538	0.678	0.389	0.521	0.320	0.454	0.400	0.252	0.358	0.316	0.345	0.476	0.425
	GOA_XGB	0.413	0.513	0.520	0.710	0.552	0.565	0.297	0.381	0.372	0.234	0.301	0.293	0.326	0.405	0.410
	GWO_SVM	0.434	0.623	0.524	0.681	0.360	0.539	0.322	0.474	0.385	0.254	0.374	0.303	0.342	0.491	0.414
	GWO_XGB	0.408	0.507	0.501	0.717	0.563	0.574	0.295	0.377	0.370	0.233	0.298	0.292	0.322	0.400	0.395
US-ARM	GOA_SVM	0.657	0.723	0.732	0.648	0.576	0.566	0.466	0.519	0.523	0.321	0.357	0.360	0.453	0.498	0.504
	GOA_XGB	0.600	0.669	0.693	0.703	0.632	0.605	0.439	0.482	0.497	0.302	0.332	0.342	0.413	0.461	0.477
	GWO_SVM	0.655	0.726	0.729	0.650	0.572	0.569	0.467	0.520	0.520	0.322	0.358	0.358	0.451	0.500	0.502
	GWO_XGB	0.595	0.663	0.685	0.708	0.638	0.614	0.433	0.482	0.498	0.298	0.332	0.343	0.410	0.456	0.471
US-CRT	GOA_SVM	0.831	1.047	0.984	0.686	0.505	0.566	0.539	0.726	0.666	0.291	0.392	0.360	0.449	0.565	0.531
	GOA_XGB	0.778	0.859	0.901	0.724	0.662	0.636	0.527	0.604	0.614	0.285	0.326	0.332	0.420	0.463	0.487
	GWO_SVM	0.895	1.059	1.016	0.638	0.498	0.538	0.586	0.716	0.677	0.316	0.387	0.366	0.483	0.572	0.549
	GWO_XGB	0.758	0.839	0.883	0.737	0.677	0.644	0.517	0.591	0.604	0.279	0.319	0.326	0.409	0.453	0.477
US-Tw3	GOA_SVM	0.834	1.052	0.938	0.763	0.616	0.695	0.584	0.793	0.673	0.229	0.311	0.264	0.327	0.413	0.368
	GOA_XGB	0.726	0.872	0.845	0.816	0.733	0.752	0.528	0.630	0.601	0.207	0.248	0.236	0.285	0.342	0.332
	GWO_SVM	0.822	1.069	0.927	0.768	0.601	0.706	0.577	0.807	0.647	0.227	0.317	0.254	0.323	0.420	0.364
	GWO_XGB	0.699	0.860	0.808	0.828	0.740	0.771	0.506	0.625	0.581	0.199	0.245	0.228	0.275	0.338	0.317

Note: Bold values indicate the best model performance for each metric: the lowest RMSE, MAE, PBIAS, and NRMSE, and the highest R².

Table A6. Statistical values of four models simulating ET at 11 cropland stations based on the medium and long-term NWP data (4–10d) and the Fluxnet 2015 dataset.

Station	Statistical Indicators	RMSE (mm d ⁻¹) Accuracy										R ² Goodness-of-Fit										MAE (mm d ⁻¹) Accuracy										PBIAS Bias						NRMSE Normalized Accuracy					
		Accuracy										Goodness-of-Fit										Accuracy										Bias						Normalized Accuracy					
		Model/Date	4d	5d	6d	7d	8d	9d	10d	4d	5d	6d	7d	8d	9d	10d	4d	5d	6d	7d	8d	9d	10d	4d	5d	6d	7d	8d	9d	10d	4d	5d	6d	7d	8d	9d	10d						
BE-Lon	GOA_SVM	0.797	0.723	0.817	0.742	0.810	0.805	0.738	0.421	0.520	0.400	0.493	0.405	0.415	0.504	0.548	0.474	0.555	0.487	0.550	0.545	0.479	0.576	0.498	0.583	0.511	0.578	0.573	0.503	0.837	0.760	0.858	0.779	0.851	0.845	0.775							
	GOA_XGB	0.722	0.705	0.750	0.701	0.750	0.744	0.697	0.512	0.535	0.474	0.540	0.474	0.481	0.546	0.504	0.472	0.532	0.474	0.527	0.528	0.467	0.530	0.496	0.559	0.498	0.553	0.555	0.491	0.759	0.740	0.787	0.737	0.788	0.782	0.732							
	GWO_SVM	0.796	0.724	0.817	0.735	0.814	0.805	0.736	0.429	0.519	0.402	0.506	0.401	0.415	0.505	0.543	0.473	0.553	0.482	0.551	0.543	0.479	0.571	0.497	0.581	0.506	0.579	0.570	0.503	0.837	0.760	0.858	0.772	0.856	0.846	0.773							
	GWO_XGB	0.710	0.693	0.730	0.691	0.726	0.721	0.693	0.528	0.551	0.502	0.552	0.506	0.514	0.550	0.493	0.466	0.506	0.467	0.498	0.496	0.464	0.518	0.489	0.531	0.490	0.523	0.521	0.488	0.746	0.728	0.766	0.726	0.763	0.757	0.728							
CH-Oe2	GOA_SVM	0.999	0.966	1.001	0.957	1.011	0.988	0.949	0.555	0.580	0.555	0.587	0.547	0.563	0.597	0.733	0.700	0.732	0.693	0.741	0.732	0.690	0.378	0.361	0.377	0.357	0.382	0.377	0.356	0.515	0.498	0.516	0.493	0.521	0.509	0.489							
	GOA_XGB	0.899	0.895	0.922	0.917	0.933	0.914	0.902	0.632	0.636	0.613	0.617	0.604	0.620	0.631	0.651	0.655	0.673	0.662	0.683	0.673	0.661	0.335	0.338	0.347	0.341	0.352	0.347	0.340	0.463	0.461	0.475	0.473	0.481	0.471	0.465							
	GWO_SVM	1.003	0.962	1.000	0.958	1.004	0.992	0.947	0.552	0.587	0.555	0.588	0.550	0.560	0.600	0.737	0.697	0.731	0.698	0.737	0.736	0.691	0.380	0.359	0.376	0.360	0.380	0.379	0.356	0.517	0.496	0.515	0.494	0.517	0.511	0.488							
	GWO_XGB	0.885	0.884	0.906	0.909	0.907	0.891	0.882	0.643	0.644	0.626	0.624	0.625	0.638	0.645	0.652	0.643	0.658	0.659	0.663	0.654	0.651	0.336	0.331	0.339	0.340	0.342	0.337	0.335	0.456	0.455	0.467	0.468	0.467	0.459	0.454							
DE-Geb	GOA_SVM	0.718	0.704	0.750	0.712	0.758	0.753	0.722	0.574	0.590	0.535	0.578	0.524	0.531	0.565	0.481	0.454	0.502	0.460	0.506	0.495	0.465	0.507	0.479	0.529	0.485	0.533	0.522	0.490	0.757	0.742	0.791	0.751	0.799	0.794	0.762							
	GOA_XGB	0.691	0.681	0.697	0.687	0.714	0.713	0.692	0.600	0.612	0.594	0.606	0.573	0.574	0.600	0.479	0.445	0.467	0.448	0.487	0.486	0.455	0.505	0.469	0.493	0.472	0.514	0.513	0.479	0.729	0.718	0.736	0.724	0.753	0.752	0.730							
	GWO_SVM	0.719	0.702	0.752	0.709	0.755	0.752	0.720	0.573	0.589	0.533	0.582	0.529	0.532	0.568	0.483	0.452	0.504	0.458	0.502	0.497	0.463	0.509	0.477	0.531	0.483	0.530	0.524	0.488	0.758	0.740	0.793	0.748	0.796	0.793	0.759							
	GWO_XGB	0.679	0.672	0.691	0.677	0.704	0.701	0.683	0.613	0.622	0.600	0.616	0.585	0.588	0.610	0.464	0.440	0.466	0.445	0.477	0.474	0.448	0.489	0.464	0.491	0.469	0.503	0.499	0.472	0.717	0.708	0.729	0.714	0.742	0.739	0.720							
DE-Kli	GOA_SVM	0.735	0.723	0.752	0.731	0.761	0.755	0.734	0.419	0.434	0.393	0.421	0.378	0.386	0.417	0.487	0.473	0.502	0.479	0.502	0.498	0.482	0.541	0.525	0.557	0.532	0.557	0.553	0.535	0.816	0.803	0.835	0.812	0.845	0.839	0.815							
	GOA_XGB	0.660	0.653	0.672	0.660	0.670	0.672	0.658	0.517	0.526	0.499	0.517	0.502	0.499	0.520	0.471	0.449	0.480	0.462	0.473	0.482	0.464	0.523	0.499	0.533	0.513	0.525	0.536	0.515	0.733	0.726	0.746	0.732	0.744	0.746	0.730							
	GWO_SVM	0.739	0.721	0.753	0.731	0.756	0.756	0.733	0.417	0.436	0.392	0.423	0.386	0.387	0.418	0.488	0.472	0.501	0.480	0.498	0.498	0.482	0.542	0.524	0.556	0.533	0.553	0.535	0.821	0.801	0.837	0.812	0.840	0.839	0.814								
	GWO_XGB	0.639	0.637	0.650	0.639	0.646	0.647	0.638	0.548	0.549	0.532	0.547	0.537	0.535	0.548	0.453	0.439	0.463	0.447	0.454	0.461	0.449	0.504	0.488	0.514	0.497	0.504	0.512	0.498	0.709	0.708	0.721	0.709	0.717	0.719	0.708							
DE-Seh	GOA_SVM	0.852	0.806	0.838	0.803	0.843	0.849	0.805	0.514	0.562	0.528	0.563	0.519	0.513	0.561	0.635	0.582	0.615	0.578	0.625	0.635	0.590	0.403	0.369	0.390	0.367	0.397	0.403	0.375	0.541	0.512	0.532	0.510	0.535	0.539	0.511							
	GOA_XGB	0.774	0.774	0.745	0.773	0.772	0.773	0.790	0.590	0.594	0.622	0.592	0.593	0.591	0.573	0.585	0.558	0.557	0.567	0.573	0.588	0.591	0.371	0.354	0.354	0.360	0.364	0.373	0.375	0.492	0.492	0.473	0.491	0.490	0.491	0.502							
	GWO_SVM	0.842	0.805	0.840	0.805	0.840	0.846	0.802	0.523	0.563	0.525	0.561	0.524	0.517	0.564	0.622	0.580	0.619	0.581	0.622	0.631	0.583	0.395	0.368	0.393	0.369	0.395	0.401	0.370	0.535	0.511	0.533	0.511	0.533	0.537	0.509							
	GWO_XGB	0.747	0.748	0.739	0.768	0.770	0.749	0.765	0.619	0.618	0.627	0.597	0.595	0.616	0.600	0.551	0.548	0.548	0.563	0.573	0.559	0.568	0.350	0.348	0.348	0.357	0.364	0.355	0.360	0.474	0.475	0.469	0.487	0.489	0.476	0.485							
FR-Gri	GOA_SVM	0.809	0.744	0.804	0.746	0.807	0.807	0.729	0.448	0.529	0.452	0.524	0.449	0.452	0.547	0.577	0.519	0.580	0.525	0.582	0.578	0.510	0.426	0.384	0.429	0.388	0.430	0.427	0.377	0.598	0.550	0.594	0.552	0.596	0.596	0.539							
	GOA_XGB	0.689	0.691	0.687	0.707	0.697	0.703	0.692	0.587	0.586	0.589	0.567	0.577	0.571	0.585	0.497	0.490	0.497	0.499	0.506	0.511	0.490	0.367	0.362	0.368	0.369	0.374	0.378	0.362	0.509	0.511	0.508	0.523	0.515	0.520	0.512							
	GWO_SVM	0.809	0.745	0.805	0.738	0.806	0.804	0.729	0.449	0.529	0.452	0.537	0.451	0.453	0.547	0.576	0.517	0.582	0.514	0.579	0.576	0.510	0.426	0.382	0.430	0.380	0.428	0.426	0.377	0.598	0.551	0.595	0.545	0.596	0.594	0.539							
	GWO_XGB	0.664	0.671	0.679	0.681	0.681	0.672	0.694	0.671	0.616	0.609	0.599	0.597	0.608	0.582	0.609	0.474	0.476	0.489	0.486	0.490	0.505	0.475	0.351	0.352	0.361	0.359	0.362	0.373	0.351	0.491	0.496	0.502	0.503	0.496	0.513	0.496						
IT-BCi	GOA_SVM	0.824	0.792	0.823	0.788	0.813	0.825	0.798	0.404	0.449	0.409	0.455	0.423	0.404	0.441	0.634	0.600	0.628	0.593	0.616	0.628	0.605	0.308	0.292	0.306	0.288	0.300	0.305	0.294	0.401	0.385	0.400	0.383	0.395	0.401	0.388							
	GOA_XGB	0.732	0.740	0.725	0.745	0.741	0.742	0.738	0.527	0.517	0.535	0.512	0.514	0.513	0.519	0.559	0.560	0.551	0.568	0.569	0.565	0.565	0.272	0.272	0.268	0.276	0.277	0.275	0.275	0.356	0.360	0.353	0.362	0.360	0.361	0.359							
	GWO_SVM	0.823	0.797	0.828	0.796	0.818	0.828	0.804	0.406	0.442	0.400	0.443	0.415	0.400	0.432	0.631	0.604	0.634	0.603	0.621	0.630	0.611	0.307	0.294	0.308	0.293	0.302	0.306	0.297	0.400	0.387	0.403	0.387	0.398	0.403	0.391							
	GWO_XGB	0.719	0.729	0.711	0.723	0.726	0.736	0.721	0.542	0.529	0.553	0.537	0.533	0.521	0.540	0.552	0.552	0.539	0.551	0.555	0.557	0.552	0.268	0.268	0.262	0.268	0.270	0.271	0.268	0.350	0.355	0.346	0.352	0.333	0.358	0.351							
IT-CA2	GOA_SVM	0.614	0.543	0.597	0.518	0.594	0.592	0.530	0.373	0.513	0.410	0.550	0.416	0.423	0.528	0.471	0.407	0.455	0.380	0.457	0.446	0.393	0.372	0.321	0.359	0.300	0.361	0.352	0.310	0.484	0.429	0.471	0.409	0.468	0.467	0.418							
	GOA_XGB	0.534	0.515	0.539	0.520	0.520	0.525	0.513	0.517	0.549	0.505	0.544	0.540	0.532	0.567	0.394	0.384	0.414	0.383	0.391	0.395	0.373	0.311	0.303	0.327	0.302	0.308	0.311	0.295	0.421	0.407	0.425</											

References

- Pereira, L.S.; Allen, R.G.; Smith, M.; Raes, D. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* **2015**, *147*, 4–20. [\[CrossRef\]](#)
- Kool, D.; Agam, N.; Lazarovitch, N.; Heitman, J.L.; Sauer, T.J.; Ben-Gal, A. A review of approaches for evapotranspiration partitioning. *Agric. For. Meteorol.* **2014**, *184*, 56–70. [\[CrossRef\]](#)
- Wang, K.; Dickinson, R.E. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.* **2012**, *50*, 2011RG000373. [\[CrossRef\]](#)
- Littell, J.S.; Peterson, D.L.; Riley, K.L.; Liu, Y.; Luce, C.H. A review of the relationships between drought and forest fire in the United States. *Glob. Change Biol.* **2016**, *22*, 2353–2369. [\[CrossRef\]](#)
- Zhang, Y.; Peña-Arancibia, J.L.; McVicar, T.R.; Chiew, F.H.S.; Vaze, J.; Liu, C.; Lu, X.; Zheng, H.; Wang, Y.; Liu, Y.Y.; et al. Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci. Rep.* **2016**, *6*, 19124. [\[CrossRef\]](#)
- Newman, B.D.; Wilcox, B.P.; Archer, S.R.; Breshears, D.D.; Dahm, C.N.; Duffy, C.J.; McDowell, N.G.; Phillips, F.M.; Scanlon, B.R.; Vivoni, E.R. Ecohydrology of water-limited environments: A scientific vision. *Water Resour. Res.* **2006**, *42*, 2005WR004141. [\[CrossRef\]](#)
- Rhee, J.; Im, J. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agric. For. Meteorol.* **2017**, *237–238*, 105–122. [\[CrossRef\]](#)
- Talsma, C.J.; Good, S.P.; Jimenez, C.; Martens, B.; Fisher, J.B.; Miralles, D.G.; McCabe, M.F.; Purdy, A.J. Partitioning of evapotranspiration in remote sensing-based models. *Agric. For. Meteorol.* **2018**, *260–261*, 131–143. [\[CrossRef\]](#)
- Medina, H.; Tian, D. Comparison of probabilistic post-processing approaches for improving numerical weather prediction-based daily and weekly reference evapotranspiration forecasts. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 1011–1030. [\[CrossRef\]](#)
- Medina, H.; Tian, D.; Srivastava, P.; Pelosi, A.; Chirico, G.B. Medium-range reference evapotranspiration forecasts for the contiguous United States based on multi-model numerical weather predictions. *J. Hydrol.* **2018**, *562*, 502–517. [\[CrossRef\]](#)
- Vitart, F.; Molteni, F. Dynamical Extended-Range Prediction of Early Monsoon Rainfall over India. *Mon. Weather Rev.* **2010**, *137*, 1480–1492. [\[CrossRef\]](#)
- Perera, K.C.; Western, A.W.; Nawarathna, B.; George, B. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. *Agric. For. Meteorol.* **2014**, *194*, 50–63. [\[CrossRef\]](#)
- Tian, D.; Wood, E.F.; Yuan, X. CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1477–1490. [\[CrossRef\]](#)
- Jain, S.K.; Mani, P.; Jain, S.K.; Prakash, P.; Singh, V.P.; Tullios, D.; Kumar, S.; Agarwal, S.P.; Dimri, A.P. A Brief review of flood forecasting techniques and their applications. *Int. J. River Basin Manag.* **2018**, *3*, 329–344. [\[CrossRef\]](#)
- Ha, W.S.; Diak, G.R.; Krajewski, W.F. Estimating Near Real-Time Hourly Evapotranspiration Using Numerical Weather Prediction Model Output and GOES Remote Sensing Data in Iowa. *Remote Sens.* **2020**, *12*, 2337. [\[CrossRef\]](#)
- Jang, J.C.; Sohn, E.H.; Park, K.H.; Lee, S. Estimation of Daily Potential Evapotranspiration in Real-Time from GK2A/AMI Data Using Artificial Neural Network for the Korean Peninsula. *Hydrology* **2021**, *8*, 129. [\[CrossRef\]](#)
- Zhao, S.; Xiang, Y.; Wu, L.; Zhang, F.; Li, Z. Improvement of Daily Reference Evapotranspiration Forecast Accuracy Based on Machine Learning in the Poyang Lake Basin. *Taiwan Water Conserv.* **2023**, *71*, 33–52.
- Zhao, S.; Qiu, Y.; Yan, Z.; Wu, L.; Qiu, R.; Luo, Y. The potential of AI global weather models for reference evapotranspiration forecasting: A comparison with numerical weather prediction models. *J. Hydrol.* **2026**, *664*, 134363. [\[CrossRef\]](#)
- Shastri, H.; Ghosh, S.; Karmakar, S. Improving Global Forecast System of extreme precipitation events with regional statistical model: Application of quantile-based probabilistic forecasts. *J. Geophys. Res. Atmos.* **2017**, *122*, 1617–1634. [\[CrossRef\]](#)
- Hamill, T.M.; Whitaker, J.S.; Shlyaeva, A.; Bates, G.; Fredrick, S.; Pegion, P.; Sinsky, E.; Zhu, Y.; Tallapragada, V.; Guan, H.; et al. The Reanalysis for the Global Ensemble Forecast System, Version 12. *Mon. Weather Rev.* **2022**, *150*, 59–79. [\[CrossRef\]](#)
- Harrison, L.; Landsfeld, M.; Husak, G.; Davenport, F.; Shukla, S.; Turner, W.; Peterson, P.; Funk, C. Advancing early warning capabilities with CHIRPS-compatible NCEP GEFS precipitation forecasts. *Sci. Data* **2022**, *9*, 375. [\[CrossRef\]](#) [\[PubMed\]](#)
- Saminathan, S.; Medina, H.; Mitra, S.; Tian, D. Improving short to medium range GEFS precipitation forecast in India. *J. Hydrol.* **2021**, *598*, 126431. [\[CrossRef\]](#)
- Duan, G.; Wu, L.; Liu, F.; Wang, Y.; Wu, S. Improvement in solar-radiation forecasting based on evolutionary KNEA method and numerical weather prediction. *Sustainability* **2022**, *14*, 6824. [\[CrossRef\]](#)
- Eidipour, A.; Akhondali, A.M.; Maddah, M.A. Ensemble forecasting and flood warning in Dez Dam Reservoir using GEFSv12 Re-forecast temperature and precipitation data. *Irrig. Water Eng.* **2023**, *53*, 308–310.
- Abhilash, S.; Sahai, A.K.; Borah, N.; Chattopadhyay, R.; Joseph, S.; Sharmila, S.; De, S.; Goswami, B.N. Does bias correction in the forecasted SST improve the extended range prediction skill of active-break spells of Indian summer monsoon rainfall? *Atmos. Sci. Lett.* **2014**, *15*, 114–119. [\[CrossRef\]](#)

26. Roulin, E.; Vannitsem, S. Postprocessing of Ensemble Precipitation Predictions with Extended Logistic Regression Based on Hindcasts. *Mon. Weather Rev.* **2012**, *140*, 874–888. [\[CrossRef\]](#)
27. Maraun, D.; Wetterhall, F.; Ireson, A.M.; Chandler, R.E.; Kendon, E.J.; Widmann, M.; Brienen, S.; Rust, H.W.; Sauter, T.; Themeßl, M.; et al. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.* **2010**, *48*, 2009RG000314. [\[CrossRef\]](#)
28. Fan, J.; Wu, L.; Zheng, J.; Zhang, F. Medium-range forecasting of daily reference evapotranspiration across China using numerical weather prediction outputs downscaled by extreme gradient boosting. *J. Hydrol.* **2021**, *601*, 126664. [\[CrossRef\]](#)
29. Bakker, K.; Whan, K.; Knap, W.; Schmeits, M. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. *Sol. Energy* **2019**, *191*, 138–150. [\[CrossRef\]](#)
30. Granata, F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric. Water Manag.* **2019**, *217*, 303–315. [\[CrossRef\]](#)
31. Dong, J.; Zhu, Y.; Jia, X.; Shao, M.; Han, X.; Qiao, J.; Bai, C.; Tang, X. Nation-scale reference evapotranspiration estimation by using deep learning and classical machine learning models in China. *J. Hydrol.* **2022**, *604*, 127207. [\[CrossRef\]](#)
32. Min, Y.C.; Huang, Y.F.; Chai, H.K. Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agric. Water Manag.* **2022**, *261*, 107343.
33. Bellido-Jiménez, J.A.; Estévez, J.; García-Marín, A.P. New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. *Agric. Water Manag.* **2021**, *245*, 106558. [\[CrossRef\]](#)
34. Dkr, A.; Rb, B.; Jq, C.; Ja, B. Using ensembles of adaptive neuro-fuzzy inference system and optimization algorithms to predict reference evapotranspiration in subtropical climatic zones. *J. Hydrol.* **2020**, *591*, 125509. [\[CrossRef\]](#)
35. Adnan, R.M.; Mostafa, R.; Islam, A.; Kisi, O.; Kuriqi, A.; Heddam, S. Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. *Comput. Electron. Agric.* **2021**, *191*, 106541. [\[CrossRef\]](#)
36. Tao, H.L.A.K. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. *Agric. Water Manag.* **2018**, *208*, 140–151. [\[CrossRef\]](#)
37. Dong, J.; Zeng, W.; Wu, L.; Huang, J.; Gaiser, T.; Srivastava, A.K. Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions of China. *Eng. Appl. Artif. Intel.* **2023**, *117*, 105579. [\[CrossRef\]](#)
38. Wu, L.; Zhou, H.; Ma, X.; Fan, J.; Zhang, F. Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: Application in contrasting climates of China. *J. Hydrol.* **2019**, *577*, 123960. [\[CrossRef\]](#)
39. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*, FAO Irrigation and Drainage Paper 56; FAO: Rome, Italy, 1998; Volume 300, 50p.
40. Babaeian, E.; Paheding, S.; Siddique, N.; Devabhaktuni, V.K.; Tuller, M. Short-and mid-term forecasts of actual evapotranspiration with deep learning. *J. Hydrol.* **2022**, *612*, 128078. [\[CrossRef\]](#)
41. Talebmorad, H.; Ahmadnejad, A.; Eslamian, S.; Ostad-Ali-Askari, K.; Singh, V.P. Evaluation of uncertainty in evapotranspiration values by FAO56-Penman-Monteith and Hargreaves-Samani methods. *Int. J. Hydrol. Sci. Technol.* **2020**, *10*, 135–147. [\[CrossRef\]](#)
42. Hadadi, F.; Moazen-zadeh, R.; Mohammadi, B. Estimation of actual evapotranspiration: A novel hybrid method based on remote sensing and artificial intelligence. *J. Hydrol.* **2022**, *609*, 127774. [\[CrossRef\]](#)
43. Moureaux, C.; Debaq, A.; Bodson, B.; Heinesch, B.; Aubinet, M. Annual net ecosystem carbon exchange by a sugar beet crop. *Agric. For. Meteorol.* **2006**, *139*, 25–39. [\[CrossRef\]](#)
44. Dietiker, D.; Buchmann, N.; Eugster, W. Testing the ability of the DNDC model to predict CO₂ and water vapour fluxes of a Swiss cropland site. *Agric. Ecosyst. Environ.* **2010**, *139*, 396–401. [\[CrossRef\]](#)
45. Anthoni, P.M.; Knohl, A.; Rebmann, C.; Freibauer, A.; Mund, M.; Ziegler, W.; Kolle, O.; Schulze, E.-D. Forest and agricultural land-use-dependent CO₂ exchange in Thuringia, Germany. *Glob. Change Biol.* **2010**, *10*, 2005–2019. [\[CrossRef\]](#)
46. Prescher, A.K.; Grünwald, T.; Bernhofer, C. Land use regulates carbon budgets in eastern Germany: From NEE to NBP. *Agric. For. Meteorol.* **2010**, *150*, 1016–1025. [\[CrossRef\]](#)
47. Schmidt, M.; Reichenau, T.G.; Fiener, P.; Schneider, K. The carbon budget of a winter wheat field: An eddy covariance analysis of seasonal and inter-annual variability. *Agric. For. Meteorol.* **2012**, *165*, 114–126. [\[CrossRef\]](#)
48. Loubet, B.; Laville, P.; Lehuger, S.; Larmanou, E.; Fléhard, C.; Mascher, N.; Genermont, S.; Roche, R.; Ferrara, R.M.; Stella, P.; et al. Carbon, nitrogen and Greenhouse gases budgets over a four years crop rotation in northern France. *Plant Soil* **2011**, *343*, 109–137. [\[CrossRef\]](#)
49. Vitale, L.; Di Tommasi, P.; D Urso, G.; Magliulo, V. The response of ecosystem carbon fluxes to LAI and environmental drivers in a maize crop grown in two contrasting seasons. *Int. J. Biometeorol.* **2016**, *60*, 411–420. [\[CrossRef\]](#)

50. Sabbatini, S.; Arriga, N.; Bertolini, T.; Castaldi, S.; Chiti, T.; Consalvo, C.; Njakou Djomo, S.; Gioli, B.; Matteucci, G.; Papale, D. Greenhouse gas balance of cropland conversion to bioenergy poplar short-rotation coppice. *Biogeosciences* **2016**, *13*, 95–113. [\[CrossRef\]](#)
51. Raz-Yaseef, N.; Billesbach, D.P.; Fischer, M.L.; Biraud, S.C.; Gunter, S.A.; Bradford, J.A.; Torn, M.S. Vulnerability of crops and native grasses to summer drying in the U.S. Southern Great Plains. *Agric. Ecosyst. Environ.* **2015**, *213*, 209–218. [\[CrossRef\]](#)
52. Lu, Y.; Williams, I.N.; Bagley, J.E.; Torn, M.S.; Kueppers, L.M. Representing winter wheat in the Community Land Model (version 4.5). *Geosci. Model Dev.* **2017**, *10*, 1873–1888. [\[CrossRef\]](#)
53. Chamberlain, S.D.; Verfaillie, J.; Eichelmann, E.; Hemes, K.S.; Baldocchi, D.D. Evaluation of Density Corrections to Methane Fluxes Measured by Open-Path Eddy Covariance over Contrasting Landscapes. *Bound.-Layer Meteorol.* **2017**, *165*, 197–210. [\[CrossRef\]](#)
54. Schafer, R.W. What is a savitzky-golay filter? [lecture notes]. *IEEE Signal Proc. Mag.* **2011**, *28*, 111–117. [\[CrossRef\]](#)
55. Chen, J.; Jönsson, P.; Tamura, M.; Gu, Z.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [\[CrossRef\]](#)
56. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
57. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
58. Saremi, S.; Mirjalili, S.; Lewis, A. Grasshopper Optimisation Algorithm: Theory and application. *Adv. Eng. Softw.* **2017**, *105*, 30–47. [\[CrossRef\]](#)
59. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [\[CrossRef\]](#)
60. Han, Y.; Wu, J.; Zhai, B.; Pan, Y.; Huang, G.; Wu, L.; Zeng, W. Coupling a Bat Algorithm with XGBoost to Estimate Reference Evapotranspiration in the Arid and Semiarid Regions of China. *Adv. Meteorol.* **2019**, *2019*, 9575782. [\[CrossRef\]](#)
61. Yu, J.; Zheng, W.; Xu, L.; Zhangzhong, L.; Zhang, G.; Shan, F. A PSO-XGBoost Model for Estimating Daily Reference Evapotranspiration in the Solar Greenhouse. *Intell. Autom. Soft Comput.* **2020**, *26*, 990–1003. [\[CrossRef\]](#)
62. Zhang, J.; Yang, S.; Wang, J.; Zeng, R.; Zhang, S.; Bai, Y.; Zhang, J. Evapotranspiration Partitioning for Croplands Based on Eddy Covariance Measurements and Machine Learning Models. *Agronomy* **2025**, *15*, 512. [\[CrossRef\]](#)
63. Dong, S.; Ma, Q.; Yu, C.; Li, L.; Liu, H.; Cui, G.; Qiu, H.; Yang, S.; Wang, G. Comparative Analysis of Crop Coefficient Approaches and Machine Learning Models for Predicting Water Requirements in Three Major Crops in Coastal Saline-Alkali Land. *Agronomy* **2025**, *15*, 492. [\[CrossRef\]](#)
64. Yan, S.; Wu, L.; Fan, J.; Zhang, F.; Zou, Y.; Wu, Y. A novel hybrid WOA-XGB model for estimating daily reference evapotranspiration using local and external meteorological data: Applications in arid and humid regions of China. *Agric. Water Manag.* **2021**, *244*, 106594. [\[CrossRef\]](#)
65. Olesen, J.E.; Trnka, M.; Kersebaum, K.C.; Skjelvåg, A.O.; Seguin, B.; Peltonen-Sainio, P.; Rossi, F.; Kozyra, J.; Micale, F. Impacts and adaptation of European crop production systems to climate change. *Eur. J. Agron.* **2011**, *34*, 96–112. [\[CrossRef\]](#)
66. Millán, M.M. Extreme hydrometeorological events and climate change predictions in Europe. *J. Hydrol.* **2014**, *518*, 206–224. [\[CrossRef\]](#)
67. He, Z.; Zhang, T.; Liu, X.; Shang, X. Water-Yield Relationship Responses of Maize to Ridge-Furrow Planting Systems Coupled with Multiple Irrigation Levels in China's Horqin Sandy Land. *Agronomy* **2018**, *8*, 221. [\[CrossRef\]](#)
68. Koukouli, P.; Georgiou, P.; Karpouzos, D. Assessment of the Impacts of Climate Change Scenarios on Maize Yield and Irrigation Water Using the CropSyst Model: An Application in Northern Greece. *Agronomy* **2025**, *15*, 638. [\[CrossRef\]](#)
69. Zhang, X.; Chen, S.; Sun, H.; Shao, L.; Wang, Y. Changes in evapotranspiration over irrigated winter wheat and maize in North China Plain over three decades. *Agric. Water Manag.* **2011**, *98*, 1097–1104. [\[CrossRef\]](#)
70. Nistor, M.M.; Ronchetti, F.; Corsini, A.; Cheval, S.; Dezsi, T. Crop evapotranspiration variation under climate change in south east europe during 1991–2050. *Carpathian J. Earth Environ. Sci.* **2017**, *12*, 571–582.
71. Wang, K.; Dickinson, R.E.; Wild, M.; Liang, S. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. *J. Geophys. Res. Atmos.* **2010**, *115*, 2009JD013671. [\[CrossRef\]](#)
72. Ma, S.; Churkina, G.; Wieland, R.; Gessler, A. Optimization and evaluation of the ANTHRO-BGC model for winter crops in Europe. *Ecol. Model.* **2011**, *222*, 3662–3679. [\[CrossRef\]](#)
73. Flores-Velazquez, J.; Akrami, M.; Villagrán, E. The Role of Radiation in the Modelling of Crop Evapotranspiration from Open Field to Indoor Crops. *Agronomy* **2022**, *12*, 2593. [\[CrossRef\]](#)
74. Zhao, S.; Wu, L.; Xiang, Y.; Zhang, F. Forecasting short-term methane based on corrected numerical weather prediction outputs. *J. Clean. Prod.* **2024**, *460*, 142500. [\[CrossRef\]](#)
75. Liu, Y.; Starzyk, J.A.; Zhu, Z. Optimized approximation algorithm in neural networks without overfitting. *IEEE Trans. Neural Netw.* **2008**, *19*, 983–995.

76. Qiu, R.; Liu, C.; Cui, N.; Wu, Y.; Wang, Z.; Li, G. Evapotranspiration estimation using a modified Priestley-Taylor model in a rice-wheat rotation system. *Agric. Water Manag.* **2019**, *224*, 105755. [[CrossRef](#)]
77. Qiu, R.; Katul, G.G.; Wang, J.; Xu, J.; Kang, S.; Liu, C.; Zhang, B.; Li, L.; Cajucom, E.P. Differential response of rice evapotranspiration to varying patterns of warming. *Agric. For. Meteorol.* **2021**, *298–299*, 108293. [[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.