

Article

Integrating Meteorological and Remote Sensing Data to Simulate Cropland Nocturnal Evapotranspiration Using Machine Learning

Jiaojiao Huang ¹ , Sha Zhang ¹, Jiahua Zhang ^{1,2} , Xin Zheng ¹ , Xianye Meng ¹, Shanshan Yang ¹ and Yun Bai ^{3,*} 

¹ Space Information and Big Earth Data Research Center, College of Computer Science and Technology, Qingdao University, Qingdao 266071, China; 2021020712@qdu.edu.cn (J.H.); zhangsha@qdu.edu.cn (S.Z.); zhangjh@radi.ac.cn (J.Z.); 2021020707@qdu.edu.cn (X.Z.); 2021023970@qdu.edu.cn (X.M.); yangss@qdu.edu.cn (S.Y.)

² Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

³ Hebei Technology Innovation Center for Remote Sensing Identification of Environmental Change, School of Geographic Sciences, Hebei Normal University, Shijiazhuang 050024, China

* Correspondence: baiyun@radi.ac.cn

Abstract: Evapotranspiration (ET) represents a significant component of the global water flux cycle, yet nocturnal evapotranspiration (ET_n) is often neglected, leading to underestimation of global evapotranspiration. As for cropland, accurate modeling of ET_n is essential for rational water management and is important for sustainable agriculture development. We used random forest (RF) to simulate ET_n at 16 globally distributed cropland eddy covariance flux sites along with remote sensing and meteorological factors. The recursive feature elimination method was used to remove unimportant variables. We also simulated the ET_n of C₃ and C₄ crops separately. The trained RF resulted in a determination coefficient (R²) (root mean square error (RMSE)) of 0.82 (7.30 W m⁻²) on the testing dataset. C₃ and C₄ crops on the testing dataset resulted in an R² (RMSE) of 0.86 (5.59 W m⁻²) and 0.55 (4.86 W m⁻²) for the two types of crops. We also showed that net radiation is the dominant factor in regulating ET_n, followed by 2 m horizontal wind speed and vapor pressure deficit (VPD), and these three meteorological factors showed a significant positive correlation with ET_n. This research demonstrates that RF can simulate ET_n from crops economically and accurately, providing a methodological basis for improving global ET_n simulations.

Keywords: nocturnal evapotranspiration; ecological remote sensing; machine learning; random forest; characteristics analysis



Citation: Huang, J.; Zhang, S.; Zhang, J.; Zheng, X.; Meng, X.; Yang, S.; Bai, Y. Integrating Meteorological and Remote Sensing Data to Simulate Cropland Nocturnal Evapotranspiration Using Machine Learning. *Sustainability* **2024**, *16*, 1987. <https://doi.org/10.3390/su16051987>

Academic Editors: Xinchang Zhang, Ying Sun and Yongjian Ruan

Received: 6 February 2024

Revised: 23 February 2024

Accepted: 25 February 2024

Published: 28 February 2024



Copyright: © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Evapotranspiration (ET) is an indispensable part of the global hydrological cycle, which has an impact on regional soil and climate [1,2]. Agroecosystems play an essential part in terrestrial ecosystems, and it is estimated that about 90% of the water resources consumed by agroecosystems are in the form of ET. High-accuracy modeling of ET is essential for determining irrigation demand, formulating irrigation strategies, and developing agricultural water management.

However, in current estimates and simulations of terrestrial water loss, daytime evapotranspiration is considered to be dominant, while nocturnal evapotranspiration (ET_n) is generally not addressed [3]. Particularly at the leaf level, the traditional consideration that leaf stomata close at night, combined with the low evaporative demand of plants at night, leads to the widespread belief that plant water vapor fluxes at night are negligible [4]. Nevertheless, the incomplete closing of the stomata combined with the subsequent prevalence of nocturnal transpiration at the leaf and crop scales is also increasingly evidenced [5,6]. Existing studies indicate that 6.3% to 9.1% of ET typically occurs at night

in terrestrial ecosystems [2,7], and the proportion can be as high as 25% to 30% in dry ecosystems [8,9]. In cropland ecosystems, nighttime evapotranspiration of different crops also accounts for a portion of the evapotranspiration, e.g., tomatoes, beans, and cotton can evaporate between 3% and 23% at night [3,5,10]. In summary, ET_n contributes to the quantification of evapotranspiration, and ignoring ET_n will lead to an underestimation of total ecosystem evapotranspiration. For cropland, accurate simulation of ET_n can help with developing more effective irrigation schemes for irrigated fields and provide a scientific basis for irrigation.

Currently, studies on the factors influencing ET_n and its mechanisms are not sufficiently advanced. The key drivers of nocturnal evapotranspiration may be somewhat different from daytime evapotranspiration [11], and there are differences in the ranking of the importance of environmental factors of evapotranspiration revealed by studies conducted in different study areas, making it difficult for existing methods to accurately model ET_n. Zeppel et al. [12] overviewed the roles of various factors on nighttime plant water loss, which provided a theoretical reference for the later studies on the influencing factors of ET_n. Meanwhile, the study proposed that the nocturnal water loss varies greatly between different plant types and functional groups. Tolk et al. [3] suggested that environmental variables such as wind speed (WS), temperature, humidity, and crop ET_n have some correlation. Padrón et al. [2] performed a characterization of global nocturnal water loss and found that higher air temperature, VPD, WS, and soil moisture were more likely to lead to higher nocturnal water loss. Groh et al. [13] obtained a similar conclusion in the estimation of ET_n from two different grassland ecosystems in Germany. Qiwen Liao et al. [14] found, in a study of ET_n on the Tibetan Plateau, that ET_n is mainly driven by temperature differences and WS at low-altitude regions. These studies have shown the complex influence of environmental factors on ET_n from agricultural fields as well as other ecosystems, which are difficult to describe using a single physical definition.

For the last few years, the rapid development of machine learning algorithms and neural network technologies and the trend of cross-fertilization of disciplines have also facilitated the introduction of these methodological models into other fields. Nowadays, machine learning models have been applied to the simulation of evapotranspiration in crops, plains, and watersheds [15–19], and the established models for estimating ET have achieved a more satisfactory accuracy. In 2019 Zhao et al. presented a machine learning model based on physical constraints to simulate ET on a global scale, and the model can also be applied to extreme weather conditions [20]. In 2021, Yan Liu et al. [21,22] improved the accuracy of the Penman–Monteith equation using artificial neural networks (ANNs) as well as remote sensing vegetation indexes, and also achieved high accuracy in simulating ET from cropland using six machine learning algorithms. Meanwhile, Jang et al. [19] pointed out that ANNs can perform local optimization of potential evapotranspiration in the Korean Peninsula more accurately than the MODIS data model. Yue Jia et al. [23] concluded that the optimized extreme learning machine (ELM) model has higher simulation accuracy than traditional empirical models (e.g., the Priestley–Taylor model) in the estimation of ET of spring maize in China. In 2021, Zhang et al. [24] discovered that using RF algorithms in combination with MODIS and flux station data can generate relatively reliable ground-based ET datasets. Meanwhile, additional studies [25,26] also demonstrated that RF can simulate ET more efficiently than other machine learning models. In conclusion, artificial intelligence models have the advantages of high efficiency, accuracy, and generalization. Therefore, in the context that the driving mechanism of ET_n is still unclear, machine learning methods are available for simulating the complicated non-linear relationship between influencing factors and ET_n, and modeling ET_n at the cropland scale.

This study is dedicated to the development of a methodology that could accurately simulate ET_n in cropland, contributing to agricultural water management and sustainable development, and could provide a theoretical basis for future ET_n simulation studies on a global scale using machine models. Consequently, this study used machine learning algorithms to simulate ET_n at 16 globally distributed cropland eddy covariance flux sites

and analyzed the factors affecting ETn. The objectives of this study consist of the following two main points: (a) to establish a random forest (RF)-based model to simulate the ETn of cropland at the global scale hoping to provide a feasible solution for ETn simulation of cropland; (b) to analyze the degree of effect of different environmental drivers on ETn by using several feature assessment methods.

2. Materials and Methods

2.1. Materials

2.1.1. Eddy Covariance Flux Site Data

The meteorological data used in this study were mainly derived from 16 cropland vortex-related sites (CRO, Croplands) in FLUXNET2015, with an interannual span of 2001–2014. In this study, latent heat flux data (LE_F_MDS) at hourly scales (both hourly and half-hourly scales) and atmospheric temperature (TA), wind speed (WS), precipitation (P), vapor pressure deficit (VPD), carbon dioxide (CO₂), net radiation (Ra), relative humidity (RH), soil heat flux (G) data at the same time scales were used.

For data preprocessing, data with incident shortwave radiation (SW_IN) values below 5.0 W·m^{−2} were selected because the study was on nighttime evaporation. In the second step, FLUXNET2015 latent heat flux missing values were null-filled by the edge distribution sampling method (MDS), so the observed or high-quality null-filled data with an LE_F_MDS_QC field of 0 or 1 could be used as valid verification data for this experiment [27], and further outlier identification and rejection was conducted on the basis of the original data. In the third step, the response of plant evapotranspiration to precipitation had a lag, so the sum of precipitation for the first 8 days was taken for precipitation (P) [28]. In the fourth step, the energy closure factor (Ra) was calculated as (LE + H)/(Rn − G), where LE is the latent heat flux, H is the sensible heat flux, and G is the soil heat flux. Since several studies have shown the prevalence of energy non-closure at flux sites using the eddy correlation technique [29–31], for samples with Ra values outside the range of 0.85 to 1.15, we used the sum of LE, H, and G instead of Rn to retain more valid data. In the fifth step, in order to eliminate potential differences in tower height on wind speed measurements, we calculated the 2 m horizontal wind speed (WS_{2m}) with the follow equation [32]:

$$WS_{2m} = \frac{WS * 4.87}{\ln(67.8 * zw - 5.42)} \quad (1)$$

where WS is wind speed, zw is the tower height. In the sixth step, it has been mentioned that the opening and closing of plant stomata varies with the sun [33], so in this study, we calculate the hour angle at sunset (ω_{set}) of the day from the local latitude and date, which is calculated as follows [34]:

$$\omega_{set} = \arccos(-\tan \varphi \tan \delta) \quad (2)$$

$$\delta = 23.5 \sin \left(\frac{DoY + 284}{365} 2\pi \right) \times \frac{\pi}{180} \quad (3)$$

where φ is local latitude, DoY is the Julian day. In the seventh step, to eliminate the effect of extreme values on the experiment, we used data with LE fall within the 5th and 95th percentile of all LE samples. Ultimately, the data from the 16 flux sites employed in this study, which are distributed globally, are shown in Figure 1. For detailed information on these 16 cropland sites, refer to Table 1.

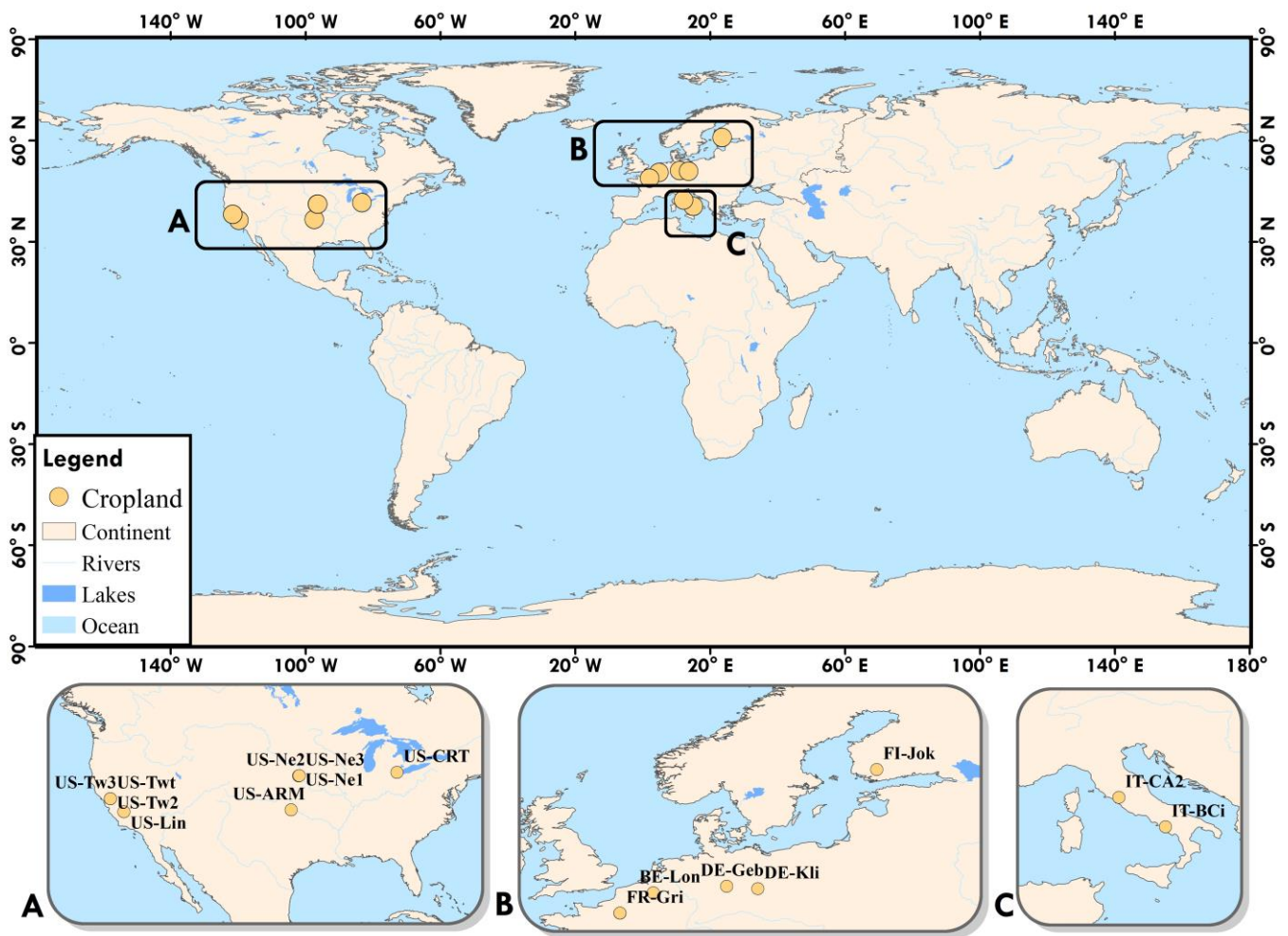


Figure 1. Global distribution of the 16 cropland sites.

Table 1. Detailed information of the 16 cropland sites.

Code	Name	Latitude	Longitude	Tower Height	Crop Rotation Period		Citation
					C ₃	C ₄	
BE-Lon	Lonzee	50.5516	4.7461	2.7	2004–2014	2012	[35]
DE-Geb	Gebesee	51.1001	10.9143	6	2001–2014	—	[36]
DE-Kli	Klingenberg	50.8929	13.5225	7.5	2004–2014	2006–2007, 2012	[37]
FI-Jok	Jokioinen	60.8986	23.5135	3	2001, 2002	—	[38]
FR-Gri	Grignon	48.8442	1.9519	2.8	2004–2014	2005–2008, 2011	[39]
IT-BCi	Borgo Cioffi	40.5238	14.9574	3.8	2004–2014	2004–2011	[40]
IT-CA2	Castel d’Asso 2	42.3772	12.026	5	2011–2014	—	
US-ARM	ARM Southern Great Plains site–Lamont	36.6058	−97.4888	60	2003–2012	2005, 2008	[41]
US-CRT	Curtice Walter–Berger cropland	41.6285	−83.3471	2	2011–2013	—	[42]
US-Lin	Lindcove Orange Orchard	36.3566	−119.8423	9.18	2009–2010	—	
US-Ne1	Mead–irrigated continuous maize site	41.1651	−96.4766	6.2	—	2001–2013	[43]
US-Ne2	Mead–irrigated maize–soybean rotation site	41.1649	−96.4701	6.2	2002, 2004, 2006, 2008	2001, 2003, 2005, 2007, 2009–2013	[44]

Table 1. Cont.

Code	Name	Latitude	Longitude	Tower Height	Crop Rotation Period C ₃	C ₄	Citation
US-Ne3	Mead–rainfed maize–soybean rotation site	41.1797	−96.4397	6.2	2002, 2004, 2006, 2008, 2010, 2012	2001, 2003, 2005, 2007, 2009, 2011, 2013	[44]
US-Tw2	Twitchell Corn	38.1047	−121.6433	5.15	— —	2012–2013	[45]
US-Tw3	Twitchell Alfalfa	38.1159	−121.6467	2.8	2012–2014	— —	[46]
US-Twt	Twitchell Island	38.1087	−121.653	3.15	2009–2014	— —	[47]

2.1.2. Remote Sensing Data

Four MODIS data products were used in this study, including MOD13Q1, MYD13Q1, MOD21A1N, and MOD21A1D. MOD13Q1 and MYD13Q1 are the 250 m vegetation index products synthesized by Terra and Aqua satellites 16 d, respectively. We extract the normalized vegetation index (NDVI) and the enhanced vegetation index (EVI) from these two products to reflect surface vegetation information. MOD21A1N and MOD21A1D are 1 km and daily synthetic products, providing the nighttime and daytime surface temperature (LST), respectively. All MODIS data were retrieved from the Application for Extracting and Exploring Analysis Ready Samples tool on the NASA website (AppEEARS ([nasa.gov](https://earthdata.nasa.gov/appears), accessed on 16 May 2022)).

We took the 16-day scale data, spliced them into an 8-day scale, and then used linear interpolation to populate the daily scale data into hourly scales, i.e., if the data at the two endpoints (t_1 , a), (t_2 , b), are known, the value (c) located at t in the interval $[t_1, t_2]$ can be computed as:

$$c = a + (t - t_1) \frac{b - a}{t_2 - t_1} \quad (4)$$

Finally, the experimental dataset was obtained by extracting the available data from the processed MODIS data through the time series of the flux dataset. Meanwhile, considering the physical properties of evapotranspiration, we also added the temperature difference between the atmosphere and the surface (ΔT_{AS}), which was obtained by subtracting LST from TA. The variables derived from these products are shown in Table 2.

Table 2. The table of variables used in this study.

Variable	Abbreviation	Data Products	Temporal Resolution
Surface Latent Heat	LE	FLUXNET2015	Hourly
Air Temperature	TA	FLUXNET2015	Hourly
Vapor Pressure Deficit	VPD	FLUXNET2015	Hourly
Precipitation	P	FLUXNET2015	Hourly
Wind Speed	WS	FLUXNET2015	Hourly
2 m Horizontal Wind Speed	WS _{2m}	FLUXNET2015	Hourly
Carbon Dioxide	CO ₂	FLUXNET2015	Hourly
Relative Humidity	RH	FLUXNET2015	Hourly
Net Radiation	Rn	FLUXNET2015	Hourly
Soil Heat Flux	G	FLUXNET2015	Hourly
Hour Angle at Sunset	ω_{set}	FLUXNET2015	Hourly
Normalized Difference Vegetation Index	NDVI	MOD13Q1, MYD13Q1	16-day, interpolated to hourly
Enhanced Vegetation Index	EVI	MOD13Q1, MYD13Q1	16-day, interpolated to hourly
Land Surface Temperature	LST	MOD21A1N	Daily, interpolated to hourly
Diurnal Temperature Difference at the Land Surface	LST_Differ	MOD21A1D, MOD21A1N	Daily, interpolated to hourly
Temperature Difference between Atmosphere and Surface	ΔT_{AS}	FLUXNET2015, MOD21A1N	Daily, interpolated to hourly

2.2. Methods

Figure 2 shows the technical flowchart of this study. First, after obtaining the original data of 16 cropland sites around the world, we performed data preprocessing operations to obtain the data required for the experiment. In the second step, we used RF combined with the recursive feature elimination method (RFE) to construct the best machine learning model to simulate ETn, and then simulated ETn of C₃, C₄ crops. In the third step, we used the random forest method, and the Shapley additive explanation method to analyze the importance of factors affecting ETn.

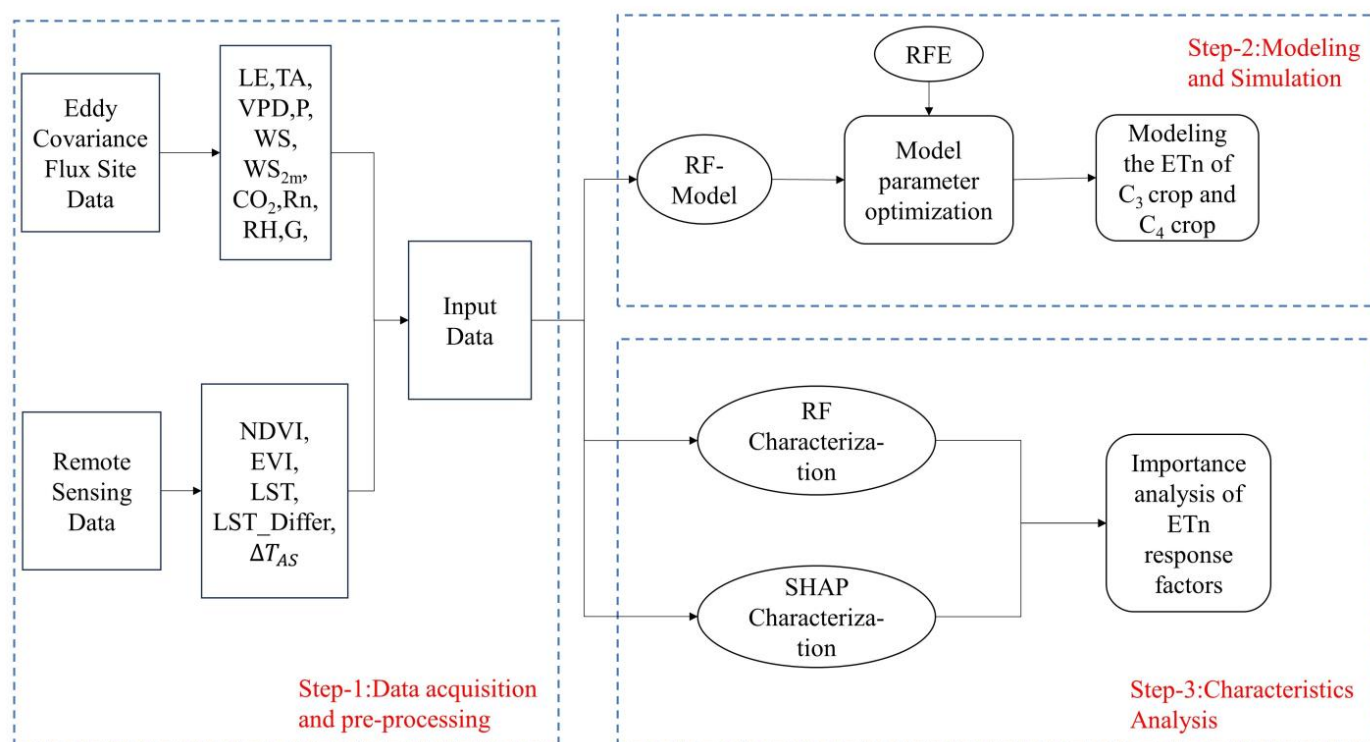


Figure 2. Flow chart of this study. TA is air temperature, VPD is vapor pressure deficit, P is precipitation, WS is wind speed, WS_{2m} is 2 m horizontal wind speed, CO₂ is carbon dioxide, RH is relative humidity, Rn is net radiation, G is soil heat flux, NDVI is normalized difference vegetation index, EVI is enhanced vegetation index, LST is land surface temperature, LST_Differ is diurnal temperature difference at the land surface, and ΔT_{AS} is the temperature difference between the atmosphere and the surface. RF is random forest, RFE is recursive feature elimination, SHAP is the Shapley additive explanation method and ETn is the nocturnal evapotranspiration. Rectangles are data, rounded rectangles are processes, and ellipses are models or methods.

2.2.1. Random Forest

Random forest (RF), proposed by Breiman [48] in 2001, is an integrated decision tree based algorithm. When RF is used as a regression, the basic theory is to generate multiple decision trees by randomly selecting samples with put-back, and then average the decision results of multiple trees as the final output to make the decision tree generalization error converge to obtain better prediction results. One of the important features of RF is the random extraction of feature variables to avoid a strong correlation between feature variables and results. Therefore, RF has the advantage of being able to evaluate the importance of features and handle high-dimensional feature data.

The dataset used in this study was randomly divided into 60% training sets, 20% validation sets, and 20% test sets based on years. For example, if a site had 10 years of data, then six years of data were randomly selected as the training set, two years as the validation set, and the remaining two years as the test set. And if a site had only one year of valid

data, the data from that site were used as the training set. This division method avoids both the poor model generalization performance due to completely random division and the time dependence arising from sequential division. Finally, 131,670 datasets and 62,743 datasets were used for the training set and validation set, respectively and 46,923 datasets were used for testing. The division of the training and test sets is shown in Table 3.

Table 3. Segmentation results for the training dataset and the test dataset *.

Year	SITE															
	BE-Lon	DE-Geb	DE-Kli	FI-Jok	FR-Gri	IT-BCi	IT-CA2	US-ARM	US-CRT	US-Lin	US-Ne1	US-Ne2	US-Ne3	US-Tw2	US-Tw3	US-Twt
2001	---	•	---	▲	---	---	---	---	---	---	•	▲	■	---	---	---
2002	---	•	---	■	---	---	---	---	---	---	▲	■	■	---	---	---
2003	---	▲	---	•	---	---	---	•	---	---	■	■	■	---	---	---
2004	•	■	■	---	•	---	---	•	---	---	•	▲	•	---	---	---
2005	▲	■	---	---	•	---	---	■	---	---	▲	▲	•	---	---	---
2006	•	•	---	---	■	---	---	•	---	---	•	•	▲	---	---	---
2007	▲	•	▲	---	■	---	---	•	---	---	•	•	•	---	---	---
2008	■	•	•	---	•	---	---	•	---	---	▲	•	•	---	---	---
2009	•	▲	•	---	•	---	---	▲	---	■	•	•	•	---	---	■
2010	•	■	•	---	•	•	---	•	---	•	■	•	▲	---	---	•
2011	•	---	•	---	---	---	■	■	•	---	■	•	▲	---	---	•
2012	■	---	•	---	▲	---	•	▲	▲	---	•	■	•	•	---	▲
2013	---	---	▲	---	■	---	---	---	■	---	•	•	•	■	•	■
2014	•	---	■	---	•	---	---	---	---	---	---	---	---	---	■	•

* • indicates that the data in the year are used as the training set; ■ indicates that the data in the year are used as the validation set; ▲ indicates that the data in the year are used as the test set; — indicates that there are no valid data for that year.

RF has two important parameters: the maximum depth of the trees (max_depth) and the number of decision trees (n_estimators). We set the max_depth parameter in the range of 1 to 25 and n_estimators in the interval of [100, 800] and tuned the parameters using the random search combined with the grid search method. After tuning, we found that n_estimators = 356 and max_depth = 10 was the RF model structure. To mitigate the risk of potential overfitting, we visualized the variation in model accuracy with parameters, as depicted in Figure 3. Figure 3a delineates the variation in R^2 concerning the max_depth parameter for both the training and validation sets, with n_estimators fixed at 356. As max_depth increases, the R^2 of both sets experiences a moderate increment; however, the R^2 of the validation set ceases to ascend when max_depth reaches 11. Notably, the discrepancy between the training and validation set R^2 is minimized when max_depth equals 9, thus identifying 9 as the optimal max_depth parameter. Figure 3b illustrates the R^2 graph of the training and validation sets concerning n_estimators with max_depth set to 9. While the overall trend appears relatively stable, the R^2 of the validation set peaks when n_estimators = 350, with the smallest difference observed between the R^2 values of the training and validation sets. Consequently, following parameter tuning, we designate the configuration with n_estimators = 356 and max_depth = 9 as the optimal structure for the random forest model.

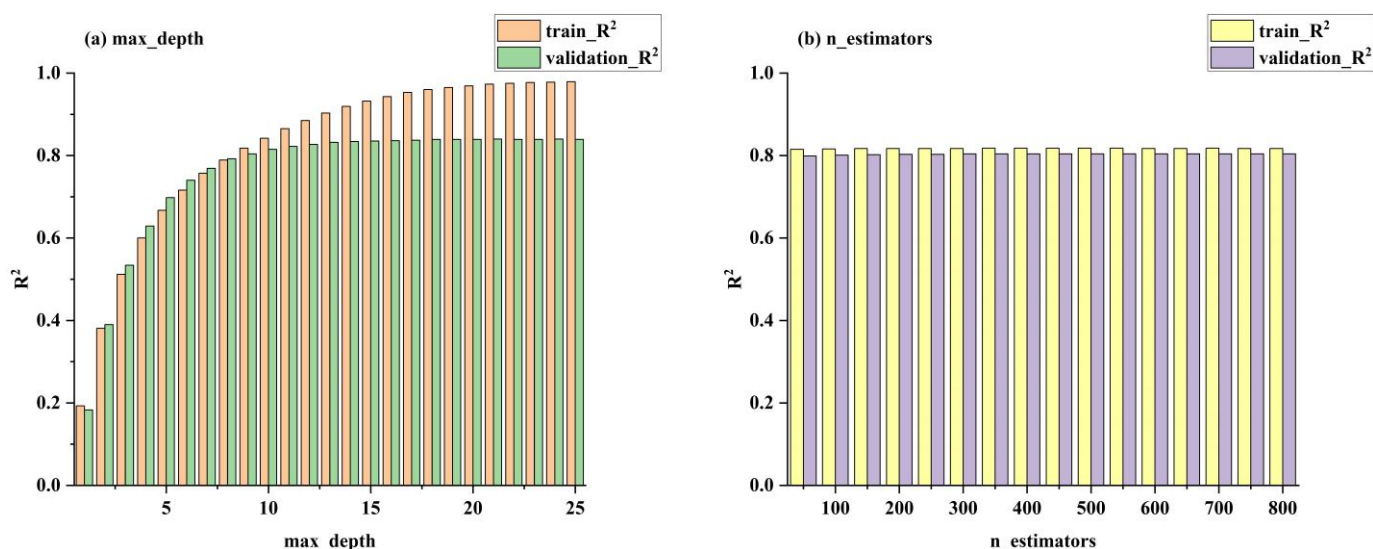


Figure 3. Variation in training and validation set R^2 values with parameters. (a) max_depth ranges from 0 to 25; (b) n_estimator ranges from 50 to 800.

2.2.2. Recursive Feature Elimination

In machine learning, when dealing with high-dimensional datasets, the problem of dimensionality catastrophe may be encountered, which reduces the generalization ability of the model and leads to overfitting problems [49]. Recursive feature elimination (RFE) is a commonly used method and tool in feature engineering, and has important applications in remote sensing, bioinformatics, power analysis, etc. [50–53]. The principle of the RFE algorithm is to repeatedly construct the model and train it, select the feature with the smallest or largest weight after each round of training, eliminate this feature, and then repeat the process until all the features are traversed, and ultimately arrive at the best feature subset [49]. Using the RFE method may result in reduced model accuracy, in order to avoid overfitting, yet the cost is relatively small and the performance degradation is insignificant.

We used three different combinations of variables to explore the impact of different input variables on model accuracy: the first with hour angle at sunset and meteorological data; the second with hour angle at sunset, meteorological data, and vegetation index data; the third with a combination of input variables obtained by recursive feature elimination. Table 4 shows the different combinations of inputs.

Table 4. The different input combinations of the RF algorithm *.

Number	Input Variables
a	ω_{set} , TA, VPD, P, WS, WS_{2m} , CO_2 , RH, LST, LST_Differ, ΔT_{AS} , Rn, G
b	ω_{set} , TA, VPD, P, WS, WS_{2m} , CO_2 , RH, LST, LST_Differ, ΔT_{AS} , Rn, G, EVI, NDVI
c	(obtained by RFE)

* ω_{set} is hour angle at sunset, TA is air temperature, VPD is vapor pressure deficit, P is precipitation, WS is wind speed, CO_2 is carbon dioxide, RH is relative humidity, LST is land surface temperature, LST_Differ is diurnal temperature difference at the land surface, ΔT_{AS} is the temperature difference between the atmosphere and the surface, Rn is net radiation, G is soil heat flux, EVI is the enhanced vegetation index, and NDVI is the normalized difference vegetation index.

2.2.3. Correlation Coefficient Method

The Pearson correlation coefficient method uses the Pearson correlation coefficient (R) as the evaluation criterion, which can be used to show the degree of correlation between the dependent variable and the independent variable. The value is between $[-1, 1]$, with the absolute value closer to 1 indicating a higher correlation, and the absolute value closer

to 0 indicating a lower correlation [54]. In this study, we analyzed multiple environmental drivers of ETn using the Pearson correlation coefficient method.

2.2.4. Shapley Additive Explanation Method

The Shapley additive explanations (SHAP) method, which measures the contribution of features to the final output by the Shapley value of each feature, is an additive attribution algorithm based on game theory that can be used to interpret machine learning models [55]. Its positive and negative direction, as well as the magnitude of its absolute value, represent the positive and negative effect of the feature on the simulated values as well as the magnitude of its contribution. In other words, when the SHAP plot shows a higher |SHAP| mean, it indicates that the feature has the greatest impact on the target simulated values. The contribution of each feature can be quantified and visualized intuitively using the SHAP method.

2.3. Model Evaluation

Three model performance assessment metrics were employed in this study: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The three evaluation indicators are calculated as:

$$R^2 = \frac{\sum_{i=1}^n [(y_i - \bar{y}_i)(x_i - \bar{x}_i)]^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2 \sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (7)$$

where n is the sample size, x_i is the observed value, $\bar{x}_i$ is the mean of the observations, y_i is the simulated value, and $\bar{y}_i$ is the mean of the simulated values.

R^2 assesses the degree of conformity between simulated and actual values in the regression model. An R^2 value that is closer to 1 represents higher simulation accuracy. RMSE and MAE are used to capture the error between the model prediction data and the raw data. The smaller these two metrics are, the more accurate the model simulation is [56].

3. Results

3.1. Comparison of RF Model with Different Input Variables

ETn was modeled and simulated for 16 stations of cropland type using the random forest algorithm, with input variables consisting of three combinations of meteorological data and vegetation index data. As shown in Figure 4, using various combinations of input variables resulted in different accuracies of the model. Using only meteorological factors as model input variables yielded relatively low simulation accuracies. The R^2 value was 0.8 for both the training and validation sets, while the RMSE (MAE) for the two datasets were 6.73 W m^{-2} (4.32 W m^{-2}) and 8.98 W m^{-2} (5.29 W m^{-2}), respectively. For the test set, the R^2 was 0.78, the RMSE was 7.79 W m^{-2} , and the MAE was 4.66 W m^{-2} . In contrast, by adding the vegetation index as a combination of model input variables, the model simulation was significantly improved, with the R^2 of the training set raised to 0.82, and the R^2 of the validation set also raised to 0.83. The R^2 of the test set was raised from 0.78 to 0.82, the RMSE was 7.36 W m^{-2} , and the MAE was 4.39 W m^{-2} . All three evaluation metrics exhibited a certain enhancement.

It is evident that RF can capture the complicated non-linear relationship from environmental factors to ETn very well and can achieve high simulation accuracy. In conclusion, RF can be selected as an effective model for further analysis of ETn in cropland.

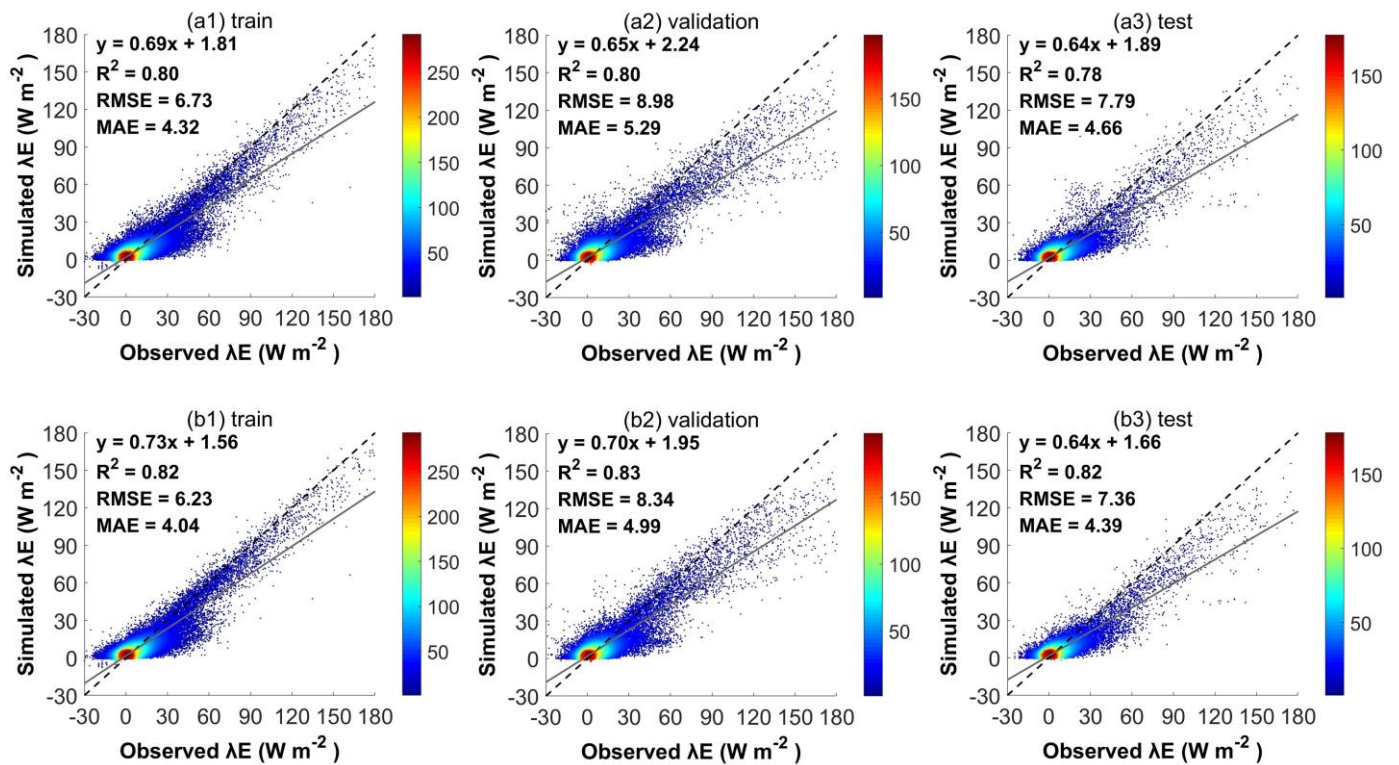


Figure 4. Performance of the RF model when simulating ETn using input variables combination (a). (a1–a3) and using input variables combination (b). (b1–b3). The solid gray line is the fitted line, and the black dashed line is the 1:1 line.

We further used the RFE method to filter the variables to obtain the input combination c to model the ETn, the variables selected were ω_{set} , TA, VPD, P, WS, WS_{2m} , RH, LST, R_n , G, EVI, and NDVI. Figure 5 shows the simulation effect of this combination of variable inputs. The overall impact of the model exhibits minor variations when compared to combination b. Although R^2 remains unchanged, there is a noticeable reduction in errors: the RMSE for the validation set decreases from $8.34 W m^{-2}$ to $8.29 W m^{-2}$, and the MAE decreases from $4.99 W m^{-2}$ to $4.96 W m^{-2}$. Furthermore, there is a reduction of $0.06 W m^{-2}$ in RMSE and $0.03 W m^{-2}$ in MAE for the test set. It is evident that this feature selection method proves effective, and the chosen feature combination contributes to enhancing the accuracy of the model simulation. Detailed information on the simulation results for these three combinations of input variables can be found in Appendix A (Table A1).

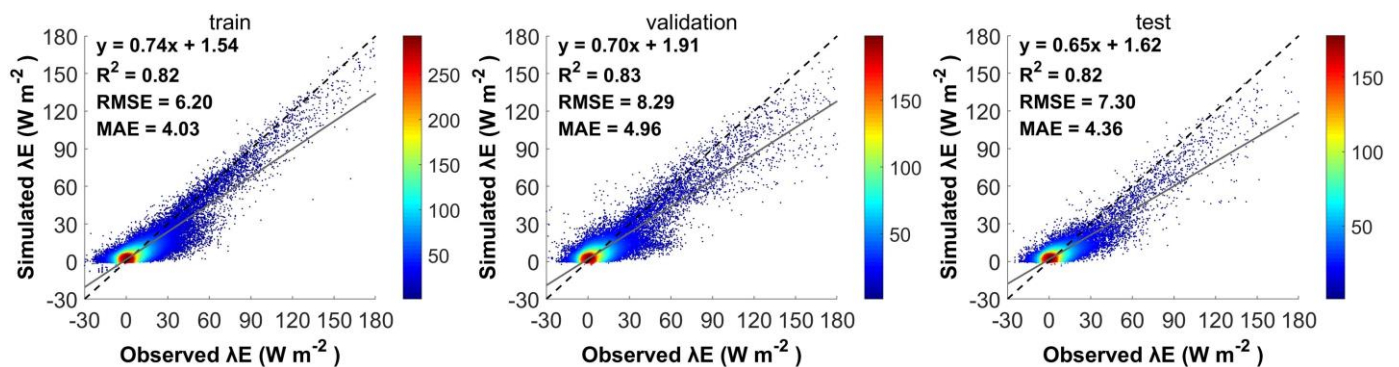


Figure 5. Performance of the RF model when simulating ETn using variables obtained by the RFE method.

3.2. Simulation of C_3 and C_4 Crops by the RF Model

C_3 and C_4 crops in the water–carbon coupling process have different physiological characteristics [57]. To further deepen the study of ETn from cropland, we modeled C_3 and C_4 crops individually and also validated them on a daily scale. The model input variables were obtained by recursive feature elimination and included ω_{set} , TA, VPD, P, WS, WS_{2m} , RH, LST, Rn, G, EVI, and NDVI. Parameter optimization was carried out on the RF model to obtain the optimum model parameters for the C_3 crop ($n_{estimators} = 369$ and $max_depth = 7$) and for the C_4 crop ($n_{estimators} = 285$ and $max_depth = 7$). The model simulation results are shown in Figure 6. The maximum simulated ETn value for C_3 crops has a maximum value of 120 W m^{-2} , while the maximum value for C_4 crops is 40 W m^{-2} . It was clear that the ETn values of C_3 plants were generally greater than those of C_4 crops, which was related to the species type of C_3 crops. In addition to typical crops such as soybean and wheat, C_3 crops also include crops such as rice. Some C_3 sites contained rice paddies within them, and the ETn of the paddies was greater than that of the drylands.

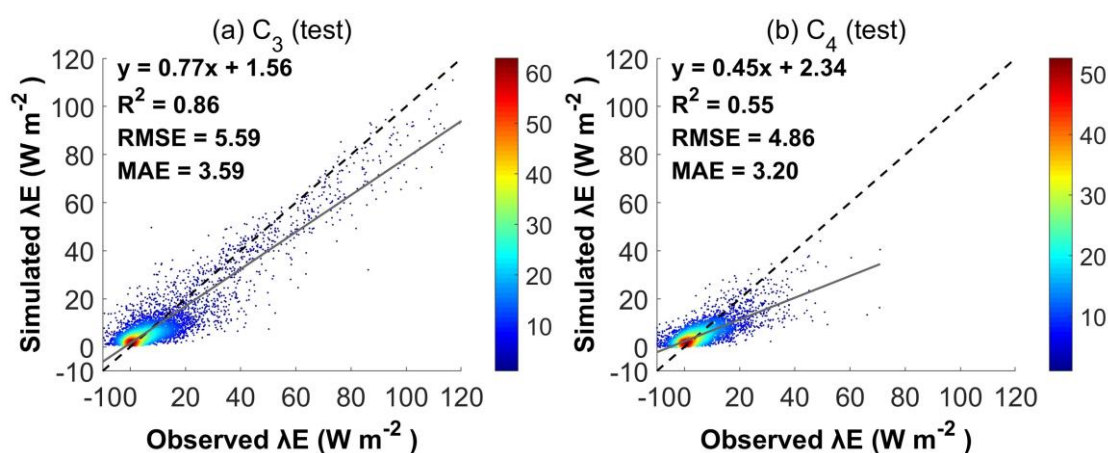


Figure 6. Performance of the RF model when simulating ETn on C_3 and C_4 crops.

The ETn simulation for the C_3 crop had a value of 0.86 for R^2 on the test set, and values of 5.59 and 3.59 for RMSE and MAE, respectively. The simulation results for C_4 crop showed R^2 , RMSE, and MAE metrics of 0.55, 4.86 W m^{-2} , and 3.2 W m^{-2} , respectively. It can be seen that although RF can effectively simulate the ETn of C_3 and C_4 , there were some differences in their simulation effects. We discuss possible reasons for this in Section 4.1.

3.3. Characteristics Analysis

3.3.1. Random Forest Characterization

To further analyze the factors affecting ETn, we analyzed the degree of influence of features using two methods. Features included ω_{set} , TA, VPD, P, WS, WS_{2m} , CO_2 , RH, LST, LST_Differ, ΔT_{AS} , Rn, G, EVI, and NDVI.

The importance of the influence of each factor on ETn was analyzed using the random forest algorithm instead of the traditional mathematical and statistical methods. The results are shown in Figure 7 and Table 5, where the relative importance is the algorithm's self-generated importance score for the corresponding factor, while the ranking is the ranking obtained from the associated importance score. The findings reveal that, in the simulation of ETn across the 16 sites, Rn and WS_{2m} emerge with notably high importance, securing relative importance scores of 0.26 and 0.12, respectively, occupying the first and second positions. VPD and WS claim the third and fourth positions, while NDVI holds the fifth position. Notably, temperature difference between the surface and the atmosphere is ranked in the last position. The importance scores of the other variables are very close to each other and lie in the middle part.

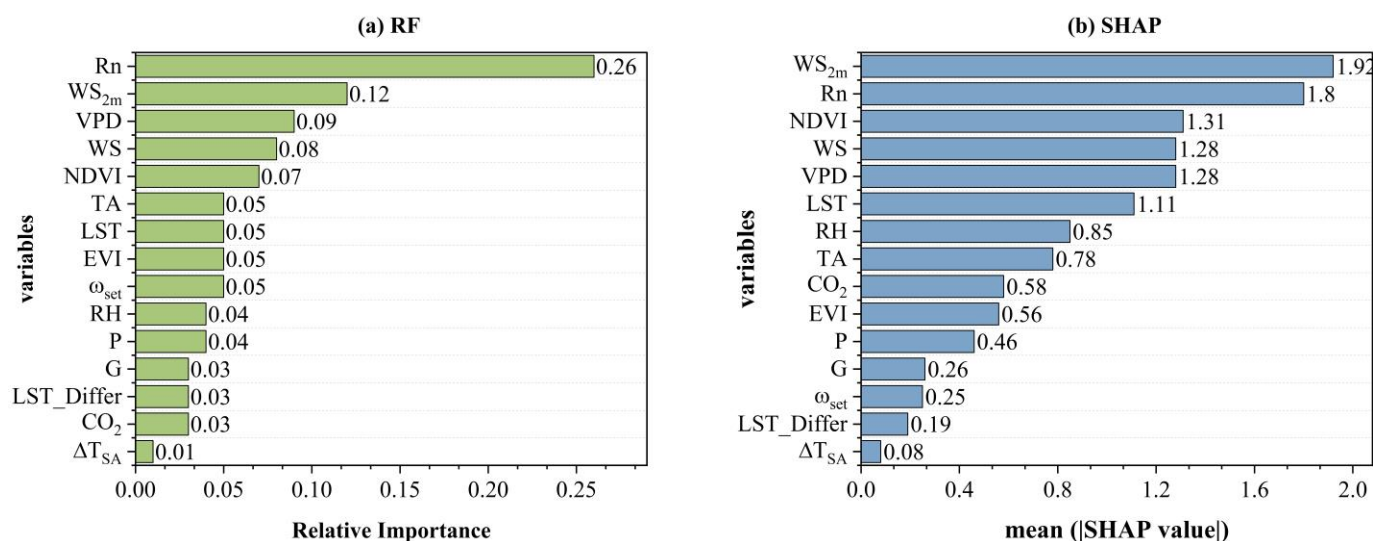


Figure 7. Characteristics analysis conducted using the random forest and SHAP methods. **(a)** Relative importance of ETn with each environmental factor obtained by the random forest algorithm; **(b)** the value of mean (|SHAP value|) of each environmental factor obtained by the SHAP algorithm.

Table 5. Feature correlations and composite rankings obtained by random forest and the SHAP method.

Variables	Rank		Ranking Average
	RF	SHAP	
Rn	1	2	1.5
WS _{2m}	2	1	1.5
VPD	3	4	3.5
WS	4	4	4
NDVI	5	3	4
TA	6	8	7
EVI	6	10	8
LST	6	6	6
RH	10	7	8.5
ω_{set}	6	13	9.5
P	10	11	10.5
CO ₂	12	9	10.5
G	12	12	12
LST_Differ	12	14	13
ΔT_{AS}	15	15	15

3.3.2. The SHAP for Characterization

We also analyzed the influences using the SHAP method, reflecting the magnitude of the contribution of each influence on the simulation results through the value of mean (|SHAP value|), and visualized the results in the form of bar charts (Figure 7). Although there are some differences with the results of the characterization of the random forest, the overall results are somewhat similar. WS_{2m} and Rn show higher values, with values of 1.92 and 1.8 for WS_{2m} and Rn, respectively. NDVI, WS, and VPD have similar performances, while at the same time ΔT_{AS} is similarly in last place. Interestingly, WS_{2m}, which is also a correlate of wind speed, manifests a greater significance than WS. The reasons for this disparity will be explored in Section 4.5.

Since different methods presented some differences in the analysis of ETn influencing factors, we combined the three methods calculated the mean values, and ranked the ranking of the obtained influencing factors (Table 5). Overall, Rn, WS_{2m}, and VPD are the most important influencing factors. However, the G, LST_Differ, and ΔT_{AS} showed a low ranking

in both experiments. Vegetation index data (EVI and NDVI) showed moderate but more important rankings.

4. Discussion

4.1. Possible Reasons for Differences in RF Modeling of C_3 , C_4 Crops

Although the RF simulations for both C_3 crops and C_4 crops presented the desired accuracy, there were some differences in the simulation results between the two. This was primarily due to differences in the volume of data and the number of sites. Of the 16 flux sites, 14 flux sites belonged to the C_3 crop with a total of 216,497 datasets, while 9 sites belonged to the C_4 crop with a total of 162,405 datasets. The difference in the amount of data makes it difficult to ensure high simulation accuracy and generalization ability.

Due to the structural differences between C_3 and C_4 crops, their responses to different environmental factors (CO_2 , radiation, temperature, precipitation, etc.) may differ to some extent [58–60]. Therefore, from the perspective of model building, in the futuristic simulation of ETn in cropland, corresponding parameterization schemes should be established for different crop types to effectively reduce the uncertainty of simulation results.

4.2. Differences in Simulation Effectiveness of Random Forest Models between Sites

We validated 11 sites on the test set on the daily scale, and the results are shown in Table 6. It can be seen that there are significant differences in the model simulation performance between different eddy covariance sites, with DE-Geb, DE-Kil, and FI-Jok having the relatively lowest R^2 values of around 0.06, followed by BE-Lon and US-CRT with an R^2 of around 0.3. US-Twt has the largest R^2 ($R^2 = 0.82$) and the worst RMSE and MAE performance (RMSE = 11.19 W m^{-2} , MAE = 8.12 W m^{-2}). The other four sites had similar accuracies, with similar R^2 (0.6 ± 0.04) and RMSE values (4.0 ± 0.3).

Table 6. Model simulation effects on the test set across different sites.

Site	R^2	RMSE	MAE
BE-Lon	0.29	1.18	0.85
DE-Geb	0.06	1.57	1.18
DE-Kli	0.06	3.15	2.23
FI-Jok	0.08	1.64	1.22
FR-Gri	0.45	4.12	2.97
US-ARM	0.57	4.27	2.80
US-CRT	0.34	4.93	3.78
US-Ne1	0.64	4.27	3.15
US-Ne2	0.64	3.60	2.48
US-Ne3	0.61	3.72	2.62
US-Twt	0.82	11.19	8.12

We conducted an in-depth analysis of the input data to elucidate the factors contributing to the observed disparities, and the data distribution is shown in Figure 8. Notably, with the exception of a few outliers, each site in the test set exhibits a substantial variation in both data distribution and volume. The LE data of the sites with the lowest R^2 are in a very narrow interval, basically distributed around 0 W m^{-2} , and thus have a low RMSE. While the sites with higher R^2 can reach a maximum value of around 40 W m^{-2} and have a larger amount of data. Noteworthy is the observation that, among these sites, US-CRT has the least amount of data, consequently presenting a lower R^2 . For the US-Twt site, the data have a wider distribution ($[-0.6, 116]$ in the 95% confidence interval) and overall large values ($[12, 56]$ in the quadratic range).

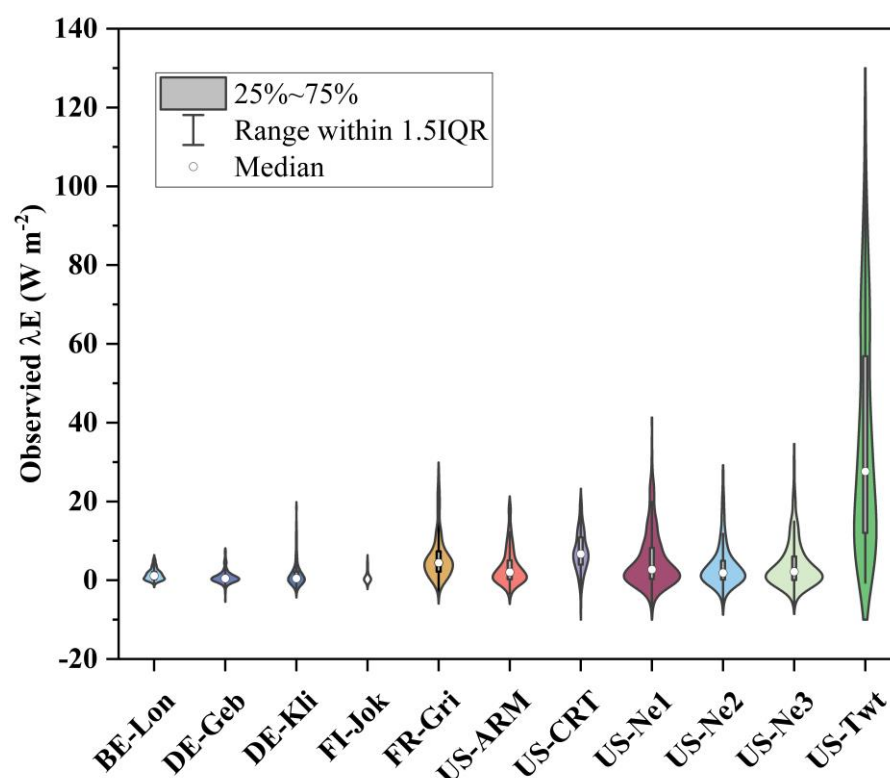


Figure 8. Violin plots of the data distribution for each site in the test set. The grey boxes represent the interquartile range, the black lines represent the range of values with 95% confidence intervals, and the white dots represent the median, with the wider portion indicating that there are relatively more data points for that value.

4.3. Impact of Different Spatial Resolution Data on the Model

To explore the impact of diverse spatial resolutions on the model's performance, we utilized NDVI and EVI data at a 500 m spatial scale sourced from MODIS's MOD13A1 and MYD13A1 products, which provide vegetation index data at the five hundred meter scale. We used the same training, validation, and test sets, with input variables being combinations of variables obtained by recursive feature elimination and substituted the original 250 m scale EVI and NDVI data with their counterparts at a 500 m scale. Modeling simulations were conducted using the RF model, and the simulation result are shown in Figure 9.

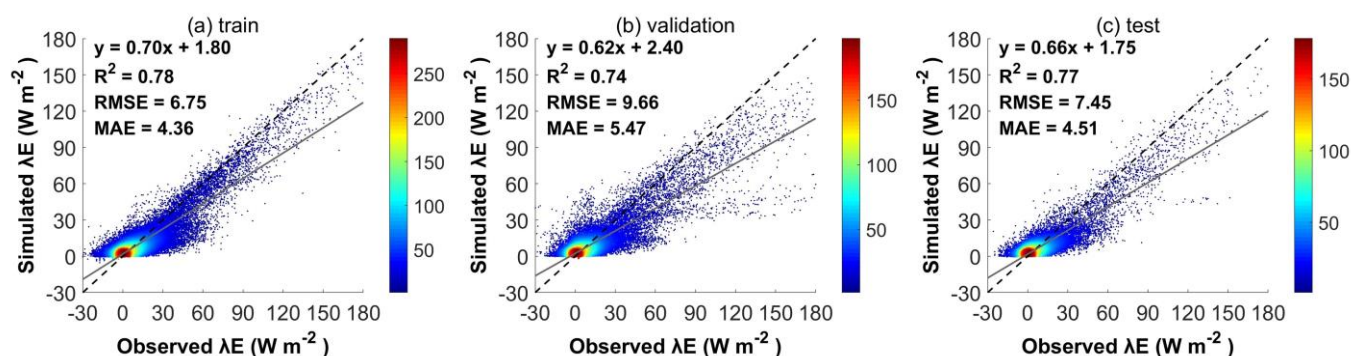


Figure 9. Performance of the RF model when simulating ETn on a 500 m spatial resolution data product.

The analysis reveals a reduction in the R^2 value of the validation set, declining from 0.83 to 0.74, and an enhancement in the RMSE (MAE) metrics, changing from 8.29 W m^{-2}

(4.96 W m^{-2}) to 9.67 W m^{-2} (5.47 W m^{-2}). The test set also exhibits some degradation in accuracy, with R^2 decreasing from 0.82 to 0.77, while RMSE and MAE also improving somewhat. These findings indicate a substantial decline in simulation accuracy when employing the 500 m data product. Higher spatial resolution MODIS data products have lower error with ground truth observations, as well as higher accuracy in reflecting localized vegetation change [61,62], and are more conducive to matching with site-scale flux data. Consequently, the model achieved a relatively optimal simulation accuracy using 250 m resolution vegetation index data.

4.4. Differences in ET_n Simulation by Different Machine Learning Algorithms

To explore whether there are significant differences in the simulation of ET_n in farmland by different machine models, we used three different types of machine learning algorithm models to conduct comparative experiments. XGBoost [63] is a type of algorithm belonging to the gradient boosting algorithm, which improves the performance by integrating multiple decision trees. KNN [64] is a basic supervised learning algorithm, which is based on the neighborhood's voting mechanism through the measure of the distance between the instances for regression prediction. ANN [65] is a multilayer neural network based on which modelling of nonlinear relationships, achieved through connections between neurons.

To ensure the consistency of the experiments, we used the same training, validation, and test sets as RF, and the input variables were derived from the combinations obtained through the RF recursive feature elimination method (combination c.). Figure 10 presents the simulation results of four distinct machine learning models. Notably, there are minor variations in the accuracy of these models, with ANN and RF achieving more similar simulation accuracies: ANN has an R^2 of 0.82, an RMSE of 6.71, and an MAE of 4.15. XGBoost attains an R^2 of 0.8, and its RMSE and MAE closely align with those of RF, with values of 7.01 and 4.35, respectively. In contrast, KNN exhibits lower model simulation accuracy, with an R^2 of 0.79. The RMSE (MAE) for KNN is 7.49 (4.39). It is apparent that there are slight simulation accuracy differences among various machine learning models for farmland ET_n. Nevertheless, these differences fall within acceptable limits and do not significantly impact the overall results. We applied the recursive feature selection method to perform feature selection on three machine learning models separately, and the results revealed no significant improvement in model accuracy (Table A2, Appendix B). Therefore, the differences in model performance may arise from both structural disparities in the machine learning models themselves, along with the fine-tuning of hyperparameters. The performance of the three models on the training and validation datasets can be found in Figure A1, Appendix B.

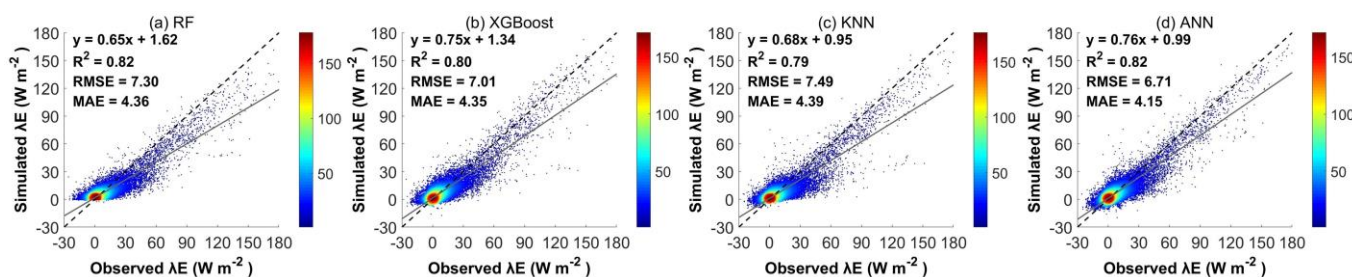


Figure 10. Performance of RF, XGBoost, KNN, and ANN in simulating ET_n on the test set.

4.5. Differences in Random Forest Feature Importance Assessment and SHAP Interpretation

Random forest feature importance assessment is achieved by calculating the impact of features when the decision tree splits nodes. This approach provides a relative ranking of features, showing which features are more important for overall model performance. However, it may ignore the interactions between features and some bias may exist in the case of highly correlated features [66]. In contrast, the SHAP method employs Shapley

values based on game theory, calculates the specific contribution of each feature for each sample, and takes into account the interactions between features [67]. However, its computational resource requirements are higher, especially when dealing with a high-dimensional feature space. Taken together, the introduction of SHAP can enhance the interpretability of the model. Combining the feature importance from random forest, we can obtain a more comprehensive decision reference for feature analysis.

4.6. Response of ETn to Rn, WS_{2m}, and VPD

Through characteristics analysis, we found that Rn, WS_{2m}, and VPD were the most important factors affecting ETn. Here, we analyzed the relationship between the three meteorological factors and ETn on a daily scale by grouping the test set data in terms of station and data. The results are shown in Figure 11.

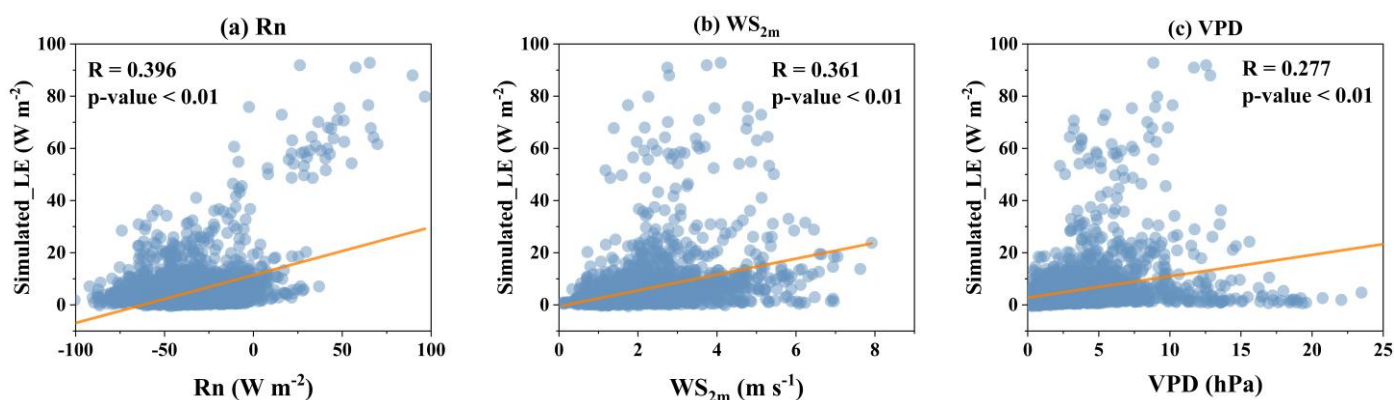


Figure 11. Scatter plot of the ETn modeled by RF with Rn, WS_{2m}, and VPD on the test dataset. The orange lines show the fitted linear regressions.

Rn is the main source of energy exchanged turbulently between crops and soil [68], reflecting the energy difference between the absorption of solar radiation by the surface and the emission of radiation to the atmosphere and space. As a central element of the surface energy balance, Rn has a direct impact on the temperature distribution and energy allocation at the surface. In both soil and plant evapotranspiration, the distribution of water and heat in the system is affected by the energy provided by radiation, which in turn affects ET changes [69]. Simultaneously, Guo et al. [70] found that on a seasonal scale, nighttime water loss in maize is mainly influenced by site bulk surface conductance and Rn. Nonetheless, Rn has often been neglected in established studies on the analysis of ETn's influencing factors. We visualized the correlation between ETn and Rn and found that there was a sufficient positive correlation between ETn and Rn, and it increased with the increase of Rn ($R = 0.396$, $p\text{-value} < 0.01$).

Both previous studies and the experimental results of this experiment showed that there was a close correlation between wind speed and evapotranspiration [12], so a discussion of the relationship between WS and ETn was necessary. Wind velocity is a means to promote soil moisture evaporation and plant evapotranspiration through air movement. Therefore, theoretically, the higher WS, the greater ET [71]. In general, surface wind speeds tend to be slower and exhibit an ascending trend with increasing height [32]. In other words, variations exist in the wind speeds recorded by flux stations at different elevations. Therefore, the consideration of 2 m standard wind speeds, in contrast to the raw wind speed (WS), accounts for the influence of tower height and underscores a significant level of importance and relevance. Our analysis of ETn and WS in cropland revealed, in general agreement with previous studies, that there was a certain degree of increase in ETn as WS increased.

Previous studies have shown that VPD is the major factor driving nighttime water loss in plants [72]. Damian Cirelli et al.'s study of nocturnal stomatal conductivity in poplar trees

showed a significant negative correlation between nocturnal stomatal conductivity and VPD [73], as did similar findings from other studies [8,74]. However, Siddiq and colleagues' study of forests showed that if climate change causes an increase in nocturnal VPD, forests will consume more water through ETn [75]. Our study found that VPD showed a weak positive correlation with ETn ($R = 0.277$) and passed the significance test (p -value < 0.01). However, at VPD to about 10 hPa, ETn did not show a significant elevation with increasing VPD, and similar VPD threshold effects were observed in other studies [76–78].

The sites in the test set are categorized into four climate types according to the Köppen–Geiger climate classification. Specifically, BE-Lon, DE-Geb, DE-Kli, FR-Gri, and US-ARM fall under the temperate oceanic climate (Cfb). US-Twt is characterized by a Mediterranean hot-summer climate (Csa), while US-CRT, US-Ne1, US-Ne2, and US-Ne3 exhibit a hot-summer humid continental climate (Dfa). FI-Jok is a warm-summer humid continental climate (Dfb). We visualized the response of ETn to the three meteorological factors under each of the four climate types (Figure 12).

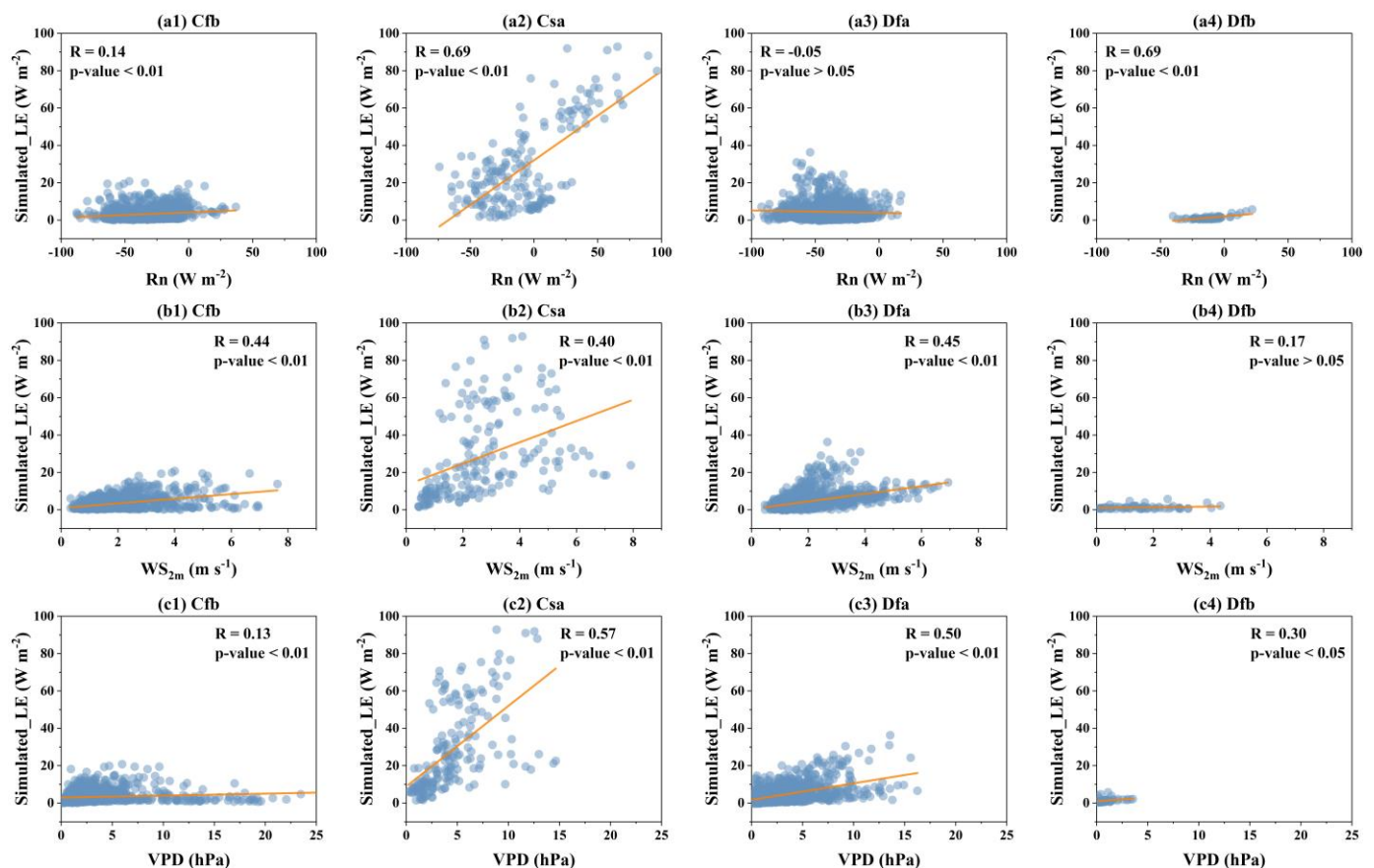


Figure 12. Scatter plot of the ETn modeled by RF with Rn, WS_{2m}, and VPD under different climatic conditions on the test dataset. (a1–a4) is the relationship of net radiation to ETn; (b1–b4) is the relationship of WS_{2m} to ETn; and (c1–c4) is the relationship of VPD to ETn. Cfb is a temperate oceanic climate; Csa is a Mediterranean hot-summer climate; Dfa is a hot-summer humid continental climate; Dfb is a warm-summer humid continental climate.

For net radiation, there is also a trend of positive correlation on individual climate types, Csa and Dfb climates in particular show more significant positive correlations. Conversely, the Dfa climate displays a weak negative correlation, albeit not a statistically significant one. This may be due to the fact that temperate continental humid climate receives less solar radiation at the surface in winter, and the Rn is usually negative, resulting in lower surface temperatures, which reduces the amount of water available for evapotranspiration by condensation of water in the soil, while Dfa tends to be drier [79],

and the rate of ET of water is slower, so that it may show a certain inhibition. Notably, Sullivan et al. [80] suggest that, compared to energy limitations (R_n), North American evapotranspiration is more sensitive to temperature and VPD. The effect of WS on ET_n shows a significant positive correlation in the Cfb, Csa, and Dfa climate types. However, in the Dfb climate type, the result shows a relatively weak correlation and the relationship is not statistically significant due to other potential factors such as small sample size or wind speed itself having a weak effect on ET_n . VPD, on the other hand, showed an overall positive correlation, and a more significant positive correlation under the Csa and Dfa climates, while the Cfb climate, which is characterized by four wet and mild seasons, makes VPD have less influence on evapotranspiration. In summary, despite the differences in the main influencing factors under different climate types, which are mainly related to the differences in each climate type, in general R_n , WS, and VPD show a close correlation with ET_n .

4.7. Impact of Data on Model Simulation Accuracy

The data employed for this research mainly include eddy covariance flux site data and MODIS data, although the original data were screened and quality controlled during the experimental process, the final experimental data still have some impact on the model simulation effect.

First, significant results are currently available for estimating global evapotranspiration using the vorticity covariance method; however, the effects of weak turbulence at night have resulted in nighttime measurements at FLUXNET sites being generally unreliable, with up to more than 50% of flux site data lacking [81]. However, nocturnal turbulent motions are limited in computational simulations by buoyancy stratification, and to accurately simulate such turbulence, models with high mesh resolution need to be used, which imposes a demand for huge computational resources, an inherent limitation of the data [82]. And accurate measurement of EC latent heat fluxes at flux sites may be more difficult due to reduced ground–atmosphere coupling [2]. Although flux sites located in cropland are less economical to maintain than flux sites in other vegetation types, the sensors themselves are subject to a certain amount of error, which can lead to common problems such as extreme noise in the data and have an impact on the accuracy of model simulations.

Secondly, the data we employed in this study were stitched from two types of data (flux site data and remote sensing data). Differences in surface resolution may generate some errors in the stitching. Moreover, the extent and direction of the data available from the flux site for the area represented varies greatly over time from site to site [83,84]. This could lead to errors in the measurement of climate conditions, vegetation indexes, etc., at the site when using FLUXNET data and MODIS data, which do not match each other better. It has been noted that finer species mapping is needed to explore data matching quantitatively [85], and this needs to be followed up with more in-depth studies.

4.8. Future Research Directions

How to develop a more generalized model to effectively simulate ET_n is the main direction of our future research. We consider here two main aspects to optimize the model.

On the one hand, we will augment the experimental sample by obtaining more valid crop field observations. Given the inherent limitations of data products in terms of temporal and spatial resolution, exploring data fusion methods and using multi-source remote sensing data could contribute to improving the spatiotemporal coverage of ET_n [86–89].

On the other hand, hybrid models based on a biophysical framework can be developed. Although pure machine learning models can fully utilize the data, they lack certain physical constraints and interpretability. It has been shown that the simulation accuracy can be improved by constructing a hybrid model by combining a biophysical framework with a machine learning model [20]. Especially in the case of limited samples, compared with the traditional pure machine learning method, simulating the intermediate parameters in the

biophysical process through the machine learning model can effectively reduce the model complexity and thus improve the generalization and stability of the model [90].

5. Conclusions

Accurate simulation of nocturnal evapotranspiration from croplands is significant for agricultural water-saving irrigation and food security. In this study, the ET_n of 16 cropland flux sites was simulated using random forest combined with a recursive feature elimination algorithm, and the dominant drivers of ET_n were analyzed using multiple feature analysis approaches. After analyzing and discussing the results, we can draw the following conclusions:

1. RF proves to be an effective tool for simulating ET_n in cropland using a combination of hour angle at sunset, meteorological, and vegetation index data as inputs ($R^2 = 0.82$, RMSE = 7.36 W m^{-2} , and MAE = 4.39 W m^{-2} for testing dataset);
2. The selected features (ω_{set} , TA, VPD, P, WS, WS_{2m}, RH, LST, Rn, G, EVI, and NDVI) through recursive feature elimination (RFE) contribute to improve model simulations ($R^2 = 0.82$, RMSE = 7.30 W m^{-2} , and MAE = 4.36 W m^{-2} for testing dataset);
3. Although the accuracy of RF simulation for C₃ and C₄ crops had some differences, the overall simulation accuracy remained within an acceptable range;
4. Among the various drivers of ET_n, Rn emerged as the primary influencing factor, followed by WS_{2m}, and VPD; the relationship between Rn, WS_{2m}, VPD and ET_n showed a positive correlation and they all passed the significance test ($p\text{-value} < 0.01$).

Overall, the methodology proposed in this study performed well in accurately modeling cropland ET_n and provided for an in-depth analysis of the relevant impact factors. In the future, we will strive to accurately simulate ET_n on a global scale by using more advanced data products and adopting more effective modeling methods.

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

Funding: This research was funded by the Excellent Young Scientist Fund of Natural Science Foundation of Hebei Province (D2023205012), the National Natural Science Foundation of China (42101382 and 42201407), the Shandong Provincial Natural Science Foundation (ZR2020QD016 and ZR2022QD120).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the study can be downloaded through the corresponding link provided in Section 2.1.

Acknowledgments: The authors would like to thank the editor and all anonymous reviewers for their valuable comments and helpful suggestions.

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

Appendix A

Table A1. Performance of the RF model when simulating ETn in different input variables. Refer to the letter (a, b, and c) used for indicating input variables in Table 4.

Input Variables	Training Dataset			Validation Dataset			Test Dataset		
	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
a	0.80	6.73	4.32	0.80	8.98	5.29	0.78	7.79	4.66
b	0.82	6.23	4.04	0.83	8.34	4.99	0.82	7.36	4.39
c	0.82	6.2	4.03	0.83	8.29	4.96	0.82	7.30	4.36

Appendix B

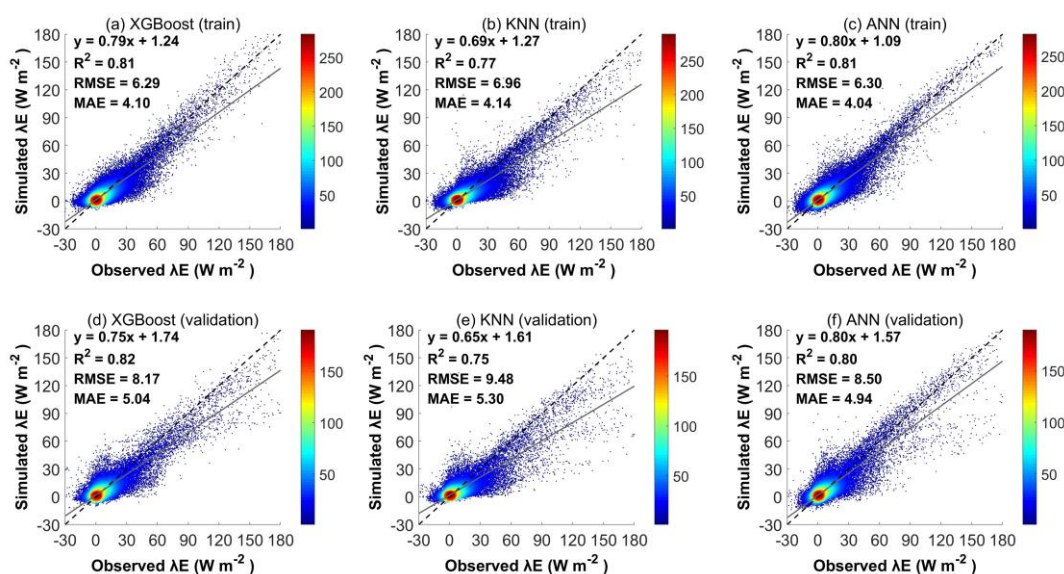


Figure A1. Performance of three different models (XGBoost, KNN, and ANN) in simulating ETn on training and validation datasets.

Table A2. Performance of models after using feature selection for XGBoost, KNN, and ANN*.

Machine Learning	Input Variables	Training Dataset			Validation Dataset			Test Dataset		
		R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
XGBoost	combination c.	0.81	6.29	4.10	0.82	8.17	5.04	0.8	7.01	4.35
	choice variables 1.	0.83	6.00	3.90	0.83	7.76	4.72	0.82	6.63	4.04
KNN	combination c.	0.77	6.96	4.14	0.75	9.48	5.30	0.79	7.49	4.39
	choice variables 2.	0.78	6.84	4.07	0.77	9.06	5.12	0.79	7.22	4.21
ANN	combination c.	0.81	6.30	4.04	0.80	8.50	4.94	0.82	6.71	4.15
	choice variables 3.	0.82	6.17	3.98	0.80	8.53	5.10	0.83	6.45	4.12

* Combination c. obtained through the random forest recursive feature elimination method included ω_{set} , TA, VPD, P, WS, WS_{2m}, RH, LST, Rn, G, EVI, and NDVI. Choice variables 1, 2, and 3 are the best variable combinations obtained from XGBoost, KNN, and ANN after recursive feature selection, respectively. The choice variables 1. include ω_{set} , TA, VPD, P, WS, WS_{2m}, CO₂, RH, Rn, NDVI, LST, and G; the choice variables 2. include ω_{set} , TA, WS, WS_{2m}, RH, Rn, NDVI, and G; the choice variables 3. include ω_{set} , TA, VPD, P, WS, WS_{2m}, CO₂, RH, LST, LST_Differ, Rn, G, EVI, and NDVI.

References

- 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.* **2010**, *115*, D20112. [\[CrossRef\]](#)
- Padrón, R.S.; Gudmundsson, L.; Michel, D.; Seneviratne, S.I. Terrestrial water loss at night: Global relevance from observations and climate models. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 793–807. [\[CrossRef\]](#)
- Tolk, J.A.; Howell, T.A.; Evett, S.R. Nighttime evapotranspiration from alfalfa and cotton in a semiarid climate. *Agron. J.* **2006**, *98*, 730–736. [\[CrossRef\]](#)

4. Rolando, J.L.; Ramirez, D.A.; Yactayo, W.; Monneveux, P.; Quiroz, R. Leaf greenness as a drought tolerance related trait in potato (*Solanum tuberosum* L.). *Environ. Exp. Bot.* **2015**, *110*, 27–35. [\[CrossRef\]](#)
5. de Dios, V.R.; Roy, J.; Ferrio, J.P.; Alday, J.G.; Landais, D.; Milcu, A.; Gessler, A. Processes driving nocturnal transpiration and implications for estimating land evapotranspiration. *Sci. Rep.* **2015**, *5*, 10975. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Schoppach, R.; Claverie, E.; Sadok, W. Genotype-dependent influence of night-time vapour pressure deficit on night-time transpiration and daytime gas exchange in wheat. *Funct. Plant Biol.* **2014**, *41*, 963–971. [\[CrossRef\]](#)
7. Novick, K.A.; Oren, R.; Stoy, P.C.; Siqueira, M.B.S.; Katul, G.G. Nocturnal evapotranspiration in eddy-covariance records from three co-located ecosystems in the Southeastern U.S.: Implications for annual fluxes. *Agric. For. Meteorol.* **2009**, *149*, 1491–1504. [\[CrossRef\]](#)
8. Bucci, S.J.; Scholz, F.G.; Goldstein, G.; Meinzer, F.C.; Hinojosa, J.A.; Hoffmann, W.A.; Franco, A.C. Processes preventing nocturnal equilibration between leaf and soil water potential in tropical savanna woody species. *Tree Physiol.* **2004**, *24*, 1119–1127. [\[CrossRef\]](#)
9. Ogle, K.; Lucas, R.W.; Bentley, L.P.; Cable, J.M.; Barron-Gafford, G.A.; Griffith, A.; Ignace, D.; Jenerette, G.D.; Tyler, A.; Huxman, T.E. Differential daytime and night-time stomatal behavior in plants from North American deserts. *New Phytol.* **2012**, *194*, 464–476. [\[CrossRef\]](#)
10. Caird, M.A.; Richards, J.H.; Donovan, L.A. Nighttime stomatal conductance and transpiration in C₃ and C₄ plants. *Plant Physiol.* **2007**, *143*, 4–10. [\[CrossRef\]](#)
11. Zeppel, M.J.B.; Lewis, J.D.; Chaszar, B.; Smith, R.A.; Medlyn, B.E.; Huxman, T.E.; Tissue, D.T. Nocturnal stomatal conductance responses to rising [CO₂], temperature and drought. *New Phytol.* **2012**, *193*, 929–938. [\[CrossRef\]](#)
12. Zeppel, M.J.; Lewis, J.D.; Phillips, N.G.; Tissue, D.T. Consequences of nocturnal water loss: A synthesis of regulating factors and implications for capacitance, embolism and use in models. *Tree Physiol.* **2014**, *34*, 1047–1055. [\[CrossRef\]](#)
13. Groh, J.; Pütz, T.; Gerke, H.H.; Vanderborght, J.; Vereecken, H. Quantification and Prediction of Nighttime Evapotranspiration for Two Distinct Grassland Ecosystems. *Water Resour. Res.* **2019**, *55*, 2961–2975. [\[CrossRef\]](#)
14. Liao, Q.; Li, X.; Shi, F.; Deng, Y.; Wang, P.; Wu, T.; Wei, J.; Zuo, F. Diurnal Evapotranspiration and Its Controlling Factors of Alpine Ecosystems during the Growing Season in Northeast Qinghai-Tibet Plateau. *Water* **2022**, *14*, 700. [\[CrossRef\]](#)
15. Yin, T.; He, W.; Yan, C.; Liu, S.; Liu, E. Effects of plastic mulching on surface of no-till straw mulching on soil water and temperature. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 78–87.
16. Patil, A.P.; Deka, P.C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Comput. Electron. Agric.* **2016**, *121*, 385–392. [\[CrossRef\]](#)
17. Mehdizadeh, S.; Behmanesh, J.; Khalili, K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput. Electron. Agric.* **2017**, *139*, 103–114. [\[CrossRef\]](#)
18. Üneş, F.; Kaya, Y.Z.; Mamak, M. Daily reference evapotranspiration prediction based on climatic conditions applying different data mining techniques and empirical equations. *Theor. Appl. Climatol.* **2020**, *141*, 763–773. [\[CrossRef\]](#)
19. 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. [\[CrossRef\]](#)
20. Zhao, W.L.; Gentine, P.; Reichstein, M.; Zhang, Y.; Zhou, S.; Wen, Y.; Lin, C.; Li, X.; Qiu, G.Y. Physics-Constrained Machine Learning of Evapotranspiration. *Geophys. Res. Lett.* **2019**, *46*, 14496–14507. [\[CrossRef\]](#)
21. Liu, Y.; Zhang, S.; Zhang, J.; Tang, L.; Bai, Y. Using Artificial Neural Network Algorithm and Remote Sensing Vegetation Index Improves the Accuracy of the Penman-Monteith Equation to Estimate Cropland Evapotranspiration. *Appl. Sci.* **2021**, *11*, 8649. [\[CrossRef\]](#)
22. Liu, Y.; Zhang, S.; Zhang, J.; Tang, L.; Bai, Y. Assessment and Comparison of Six Machine Learning Models in Estimating Evapotranspiration over Croplands Using Remote Sensing and Meteorological Factors. *Remote Sens.* **2021**, *13*, 3838. [\[CrossRef\]](#)
23. Jia, Y.; Su, Y.; Zhang, R.; Zhang, Z.; Lu, Y.; Shi, D.; Xu, C.; Huang, D. Optimization of an extreme learning machine model with the sparrow search algorithm to estimate spring maize evapotranspiration with film mulching in the semiarid regions of China. *Comput. Electron. Agric.* **2022**, *201*, 107298. [\[CrossRef\]](#)
24. Zhang, C.; Luo, G.; Hellwich, O.; Chen, C.; Zhang, W.; Xie, M.; He, H.; Shi, H.; Wang, Y. A framework for estimating actual evapotranspiration at weather stations without flux observations by combining data from MODIS and flux towers through a machine learning approach. *J. Hydrol.* **2021**, *603*, 127047. [\[CrossRef\]](#)
25. Hao, P.; Di, L.; Yu, E.; Guo, L.; Sun, Z.; Zhao, H. Using machine learning and trapezoidal model to derive All-weather ET from Remote sensing Images and Meteorological Data. In Proceedings of the 2021 9th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Shenzhen, China, 26–29 July 2021; pp. 1–4.
26. Hu, X.; Shi, L.; Lin, G.; Lin, L. Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. *J. Hydrol.* **2021**, *601*, 126592. [\[CrossRef\]](#)
27. Pastorello, G.; Trotta, C.; Canfora, E.; Chu, H.; Christianson, D.; Cheah, Y.-W.; Poindexter, C.; Chen, J.; Elbashandy, A.; Humphrey, M. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci. Data* **2020**, *7*, 225. [\[CrossRef\]](#)
28. Zhang, Y.; Leuning, R.; Hutley, L.B.; Beringer, J.; McHugh, I.; Walker, J.P. Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05° spatial resolution. *Water Resour. Res.* **2010**, *46*, 242–253. [\[CrossRef\]](#)
29. Yanzhao, Z.; Xin, L. Progress in the energy closure of eddy covariance systems. *Adv. Earth Sci.* **2018**, *33*, 898.
30. Wilson, K.; Goldstein, A.; Falge, E.; Aubinet, M.; Baldocchi, D.; Berbigier, P.; Bernhofer, C.; Ceulemans, R.; Dolman, H.; Field, C.J.A.; et al. Energy balance closure at FLUXNET sites. *Agric. For. Meteorol.* **2002**, *113*, 223–243. [\[CrossRef\]](#)

31. Anderson, R.G.; Wang, D. Energy budget closure observed in paired Eddy Covariance towers with increased and continuous daily turbulence. *Agric. For. Meteorol.* **2014**, *184*, 204–209. [\[CrossRef\]](#)
32. 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, p. D05109.
33. Resco de Dios, V.; Loik, M.E.; Smith, R.; Aspinwall, M.J.; Tissue, D.T. Genetic variation in circadian regulation of nocturnal stomatal conductance enhances carbon assimilation and growth. *Plant Cell Environ.* **2016**, *39*, 3–11. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Bai, Y.; Zhang, J.; Zhang, S.; Yao, F.; Magliulo, V. A remote sensing-based two-leaf canopy conductance model: Global optimization and applications in modeling gross primary productivity and evapotranspiration of crops. *Remote Sens. Environ.* **2018**, *215*, 411–437. [\[CrossRef\]](#)
35. Moureaux, C.; Debacq, A.; Bodson, B.; Heinesch, B.; Aubinet, M.J.A.; Meteorology, F. Annual net ecosystem carbon exchange by a sugar beet crop. *Agric. For. Meteorol.* **2006**, *139*, 25–39. [\[CrossRef\]](#)
36. 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. Chang. Biol.* **2004**, *10*, 2005–2019. [\[CrossRef\]](#)
37. Brust, K.; Hehn, M.; Bernhofer, C. Comparative analysis of matter and energy fluxes determined by Bowen Ratio and Eddy Covariance techniques at a crop site in eastern Germany. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 7–12 April 2012; p. 8006.
38. Lohila, A.; Aurela, M.; Tuovinen, J.P.; Laurila, T. Annual CO₂ exchange of a peat field growing spring barley or perennial forage grass. *J. Geophys. Res.-Atmos.* **2004**, *109*, 18116. [\[CrossRef\]](#)
39. Loubet, B.; Laville, P.; Lehuger, S.; Larmanou, E.; Fléhard, C.; Mascher, N.; Genermont, S.; Roche, R.; Ferrara, R.M.; Stella, P.J.P.; et al. Carbon, nitrogen and Greenhouse gases budgets over a four years crop rotation in northern France. *Plant Soil* **2011**, *343*, 109–137. [\[CrossRef\]](#)
40. Ranucci, S.; Bertolini, T.; Vitale, L.; Di Tommasi, P.; Ottaiano, L.; Oliva, M.; Amato, U.; Fierro, A.; Magliulo, V.J.P. The influence of management and environmental variables on soil N₂O emissions in a crop system in Southern Italy. *Plant Soil* **2011**, *343*, 83–96. [\[CrossRef\]](#)
41. 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. [\[CrossRef\]](#)
42. Chu, H.; Chen, J.; Gottgens, J.F.; Ouyang, Z.; John, R.; Czajkowski, K.; Becker, R. Net ecosystem methane and carbon dioxide exchanges in a Lake Erie coastal marsh and a nearby cropland. *J. Geophys. Res. Biogeosci.* **2014**, *119*, 722–740. [\[CrossRef\]](#)
43. Verma, S.B.; Dobermann, A.; Cassman, K.G.; Walters, D.T.; Knops, J.M.; Arkebauer, T.J.; Suyker, A.E.; Burba, G.G.; Amos, B.; Yang, H.J.A.; et al. Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems. *Agric. For. Meteorol.* **2005**, *131*, 77–96. [\[CrossRef\]](#)
44. Suyker, A.E.; Verma, S.B. Gross primary production and ecosystem respiration of irrigated and rainfed maize–soybean cropping systems over 8 years. *Agric. For. Meteorol.* **2012**, *165*, 12–24. [\[CrossRef\]](#)
45. Knox, S.H.; Sturtevant, C.; Matthes, J.H.; Koteen, L.; Verfaillie, J.; Baldocchi, D.J.G.c.b. Agricultural peatland restoration: Effects of land-use change on greenhouse gas (CO₂ and CH₄) fluxes in the Sacramento-San Joaquin Delta. *Glob. Chang. Biol.* **2015**, *21*, 750–765. [\[CrossRef\]](#)
46. Baldocchi, D.; Sturtevant, C.; Contributors, F.J.A.; Meteorology, F. Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration? *Agric. For. Meteorol.* **2015**, *207*, 117–126. [\[CrossRef\]](#)
47. Hatala, J.A.; Detto, M.; Baldocchi, D.D. Gross ecosystem photosynthesis causes a diurnal pattern in methane emission from rice. *Geophys. Res. Lett.* **2012**, *39*, 06409. [\[CrossRef\]](#)
48. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
49. Jeon, H.; Oh, S. Hybrid-Recursive Feature Elimination for Efficient Feature Selection. *Appl. Sci.* **2020**, *10*, 3211. [\[CrossRef\]](#)
50. Saeys, Y.; Inza, I.; Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [\[CrossRef\]](#)
51. Ustebay, S.; Turgut, Z.; Aydin, M.A. Intrusion Detection System with Recursive Feature Elimination by using Random Forest and Deep Learning Classifier. In Proceedings of the International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, 3–4 December 2018; pp. 71–76.
52. Yoosefzadeh-Najafabadi, M.; Earl, H.J.; Tulpan, D.; Sulik, J.; Eskandari, M. Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield from Hyperspectral Reflectance in Soybean. *Front. Plant Sci.* **2020**, *11*, 624273. [\[CrossRef\]](#) [\[PubMed\]](#)
53. dos Santos, R.A.; Mantovani, E.C.; Fernandes-Filho, E.I.; Filgueiras, R.; Lourenço, R.D.S.; Bufon, V.B.; Neale, C.M.U. Modeling Actual Evapotranspiration with MSI-Sentinel Images and Machine Learning Algorithms. *Atmosphere* **2022**, *13*, 1518. [\[CrossRef\]](#)
54. Xin, K.; Zhao, J.; Wang, T.; Gao, W. Supporting Design to Develop Rural Revitalization through Investigating Village Microclimate Environments: A Case Study of Typical Villages in Northwest China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 8310. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
56. Yu, H.; Wen, X.; Li, B.; Yang, Z.; Wu, M.; Ma, Y. Uncertainty analysis of artificial intelligence modeling daily reference evapotranspiration in the northwest end of China. *Comput. Electron. Agric.* **2020**, *176*, 105653. [\[CrossRef\]](#)

57. Wang, X.; Lei, H.; Li, J.; Huo, Z.; Zhang, Y.; Qu, Y. Estimating evapotranspiration and yield of wheat and maize croplands through a remote sensing-based model. *Agric. Water Manag.* **2023**, *282*, 108294. [\[CrossRef\]](#)
58. Mera, R.J.; Niyogi, D.; Buol, G.S.; Wilkerson, G.G.; Semazzi, F.H.M. Potential individual versus simultaneous climate change effects on soybean (C₃) and maize (C₄) crops: An agrotechnology model based study. *Glob. Planet. Chang.* **2006**, *54*, 163–182. [\[CrossRef\]](#)
59. Still, C.J.; Berry, J.A.; Collatz, G.J.; DeFries, R.S. Global distribution of C₃ and C₄ vegetation: Carbon cycle implications. *Glob. Biogeochem. Cycles* **2003**, *17*, 6–1–6–14. [\[CrossRef\]](#)
60. Sutherlin, C.E.; Brunsell, N.A.; de Oliveira, G.; Crews, T.E.; DeHaan, L.R.; Vico, G. Contrasting physiological and environmental controls of evapotranspiration over Kernza perennial crop, annual crops, and C₄ and mixed C₃/C₄ grasslands. *Sustainability* **2019**, *11*, 1640. [\[CrossRef\]](#)
61. Mei, L.; Bao, G.; Tong, S.; Yin, S.; Bao, Y.; Jiang, K.; Hong, Y.; Tuya, A.; Huang, X. Elevation-dependent response of spring phenology to climate and its legacy effect on vegetation growth in the mountains of northwest Mongolia. *Ecol. Indic.* **2021**, *126*, 107640. [\[CrossRef\]](#)
62. Ruan, Z.; Kuang, Y.; He, Y.; Zhen, W.; Ding, S. Detecting Vegetation Change in the Pearl River Delta Region Based on Time Series Segmentation and Residual Trend Analysis (TSS-RESTREND) and MODIS NDVI. *Remote Sens.* **2020**, *12*, 4049. [\[CrossRef\]](#)
63. Ester, M.; Kriegel, H.P.; Xu, X. 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; p. 785. [\[CrossRef\]](#)
64. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [\[CrossRef\]](#)
65. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386. [\[CrossRef\]](#)
66. Loecher, M. Debiasing SHAP scores in random forests. *ASTA Adv. Stat. Anal.* **2023**, 1–14. [\[CrossRef\]](#)
67. Kim, Y.; Kim, Y. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustain. Cities Soc.* **2022**, *79*, 103677. [\[CrossRef\]](#)
68. Zhang, J.; Li, S.; Wang, J.; Chen, Z. Estimation of Evapotranspiration from the People's Victory Irrigation District Based on the Data Mining Sharpener Model. *Agronomy* **2023**, *13*, 3082. [\[CrossRef\]](#)
69. Zhang, Y.; Kang, S.; Ward, E.J.; Ding, R.; Zhang, X.; Zheng, R. Evapotranspiration components determined by sap flow and microlysimetry techniques of a vineyard in northwest China: Dynamics and influential factors. *Agric. Water Manag.* **2011**, *98*, 1207–1214. [\[CrossRef\]](#)
70. Guo, X.A.; Xiao, J.F.; Zha, T.S.; Shang, G.F.; Liu, P.; Jin, C.; Zhang, Y.C. Dynamics and biophysical controls of nocturnal water loss in a winter wheat-summer maize rotation cropland: A multi-temporal scale analysis. *Agric. For. Meteorol.* **2023**, *342*, 109701. [\[CrossRef\]](#)
71. Kukal, M.S.; Irmak, S. Nocturnal transpiration in field crops: Implications for temporal aggregation and diurnal weighing of vapor pressure deficit. *Agric. Water Manag.* **2022**, *266*, 107578. [\[CrossRef\]](#)
72. Massmann, A.; Gentine, P.; Lin, C. When Does Vapor Pressure Deficit Drive or Reduce Evapotranspiration? *J. Adv. Model. Earth Syst.* **2019**, *11*, 3305–3320. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Cirelli, D.; Equiza, M.A.; Lieffers, V.J.; Tyree, M.T. Populus species from diverse habitats maintain high night-time conductance under drought. *Tree Physiol.* **2016**, *36*, 229–242. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Chowdhury, F.I.; Arteaga, C.; Alam, M.S.; Alam, I.; Resco de Dios, V. Drivers of nocturnal stomatal conductance in C₃ and C₄ plants. *Sci. Total Environ.* **2022**, *814*, 151952. [\[CrossRef\]](#)
75. Siddiq, Z.; Cao, K.-F. Nocturnal transpiration in 18 broadleaf timber species under a tropical seasonal climate. *For. Ecol. Manag.* **2018**, *418*, 47–54. [\[CrossRef\]](#)
76. Chen, D.; Wang, Y.; Liu, S.; Wei, X.; Wang, X. Response of relative sap flow to meteorological factors under different soil moisture conditions in rainfed jujube (*Ziziphus jujuba* Mill.) plantations in semiarid Northwest China. *Agric. Water Manag.* **2014**, *136*, 23–33. [\[CrossRef\]](#)
77. Wang, X.; Guan, H.; Huo, Z.; Guo, P.; Du, J.; Wang, W. Maize transpiration and water productivity of two irrigated fields with varying groundwater depths in an arid area. *Agric. For. Meteorol.* **2020**, *281*, 107849. [\[CrossRef\]](#)
78. Tie, Q.; Hu, H.; Tian, F.; Guan, H.; Lin, H. Environmental and physiological controls on sap flow in a subhumid mountainous catchment in North China. *Agric. For. Meteorol.* **2017**, *240–241*, 46–57. [\[CrossRef\]](#)
79. Dilinuer, T.; Yao, J.-Q.; Chen, J.; Mao, W.-Y.; Yang, L.-M.; Yeernaer, H.; Chen, Y.-H. Regional drying and wetting trends over Central Asia based on Köppen climate classification in 1961–2015. *Adv. Clim. Chang. Res.* **2021**, *12*, 363–372. [\[CrossRef\]](#)
80. Sullivan, R.C.; Kotamarthi, V.R.; Feng, Y. Recovering Evapotranspiration Trends from Biased CMIP5 Simulations and Sensitivity to Changing Climate over North America. *J. Hydrometeorol.* **2019**, *20*, 1619–1633. [\[CrossRef\]](#)
81. Falge, E.; Baldocchi, D.; Olson, R.; Anthoni, P.; Aubinet, M.; Bernhofer, C.; Burba, G.; Ceulemans, R.; Clement, R.; Dolman, H.; et al. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. For. Meteorol.* **2001**, *107*, 43–69. [\[CrossRef\]](#)
82. Zhou, B. Large-eddy Simulation of the Nighttime Stable Atmospheric Boundary Layer. Ph.D. Thesis, UC Berkeley, Berkeley, CA, USA, 2012.
83. Yuan, D.; Zhang, S.; Li, H.; Zhang, J.; Yang, S.; Bai, Y. Improving the Gross Primary Productivity Estimate by Simulating the Maximum Carboxylation Rate of the Crop Using Machine Learning Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4413115. [\[CrossRef\]](#)

84. Chu, H.; Luo, X.; Ouyang, Z.; Chan, W.S.; Dengel, S.; Biraud, S.C.; Torn, M.S.; Metzger, S.; Kumar, J.; Arain, M.A. Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. *Agric. For. Meteorol.* **2021**, *301*, 108350. [[CrossRef](#)]
85. Huang, L.; Liu, M.; Yao, N. Evaluation of Ecosystem Water Use Efficiency Based on Coupled and Uncoupled Remote Sensing Products for Maize and Soybean. *Remote Sens.* **2023**, *15*, 4922. [[CrossRef](#)]
86. Pan, S.; Pan, N.; Tian, H.; Friedlingstein, P.; Sitch, S.; Shi, H.; Arora, V.K.; Haverd, V.; Jain, A.K.; Kato, E.; et al. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 1485–1509. [[CrossRef](#)]
87. Ma, Y.; Liu, S.; Song, L.; Xu, Z.; Liu, Y.; Xu, T.; Zhu, Z. Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data. *Remote Sens. Environ.* **2018**, *216*, 715–734. [[CrossRef](#)]
88. Li, Y.; Huang, C.; Kustas, W.P.; Nieto, H.; Sun, L.; Hou, J. Evapotranspiration Partitioning at Field Scales Using TSEB and Multi-Satellite Data Fusion in The Middle Reaches of Heihe River Basin, Northwest China. *Remote Sens.* **2020**, *12*, 3223. [[CrossRef](#)]
89. Bhattarai, N.; Wagle, P. Recent Advances in Remote Sensing of Evapotranspiration. *Remote Sens.* **2021**, *13*, 4260. [[CrossRef](#)]
90. Kong, D.; Yuan, D.; Li, H.; Zhang, J.; Yang, S.; Li, Y.; Bai, Y.; Zhang, S. Improving the Estimation of Gross Primary Productivity across Global Biomes by Modeling Light Use Efficiency through Machine Learning. *Remote Sens.* **2023**, *15*, 2086. [[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.