International Journal of Digital Earth

ISSN: 1753-8947 (Print) 1753-8955 (Online) Journal homepage: www.tandfonline.com/journals/tjde20

Estimating actual evapotranspiration of soil-vegetation system by combining a physics-based model and machine learning

Yiqing Zhang , Li Jia , Chaolei Zheng , Massimo Menenti , Guangcheng Hu , Jing Lu & Qiting Chen

To cite this article: Yiqing Zhang , Li Jia , Chaolei Zheng , Massimo Menenti , Guangcheng Hu , Jing Lu & Qiting Chen (2026) Estimating actual evapotranspiration of soil-vegetation system by combining a physics-based model and machine learning, International Journal of Digital Earth, 19:1, 2624207, DOI: 10.1080/17538947.2026.2624207

To link to this article: https://doi.org/10.1080/17538947.2026.2624207

INTERNATIONAL JOURNAL OF DIGITAL EARTH
2026, VOL. 19, NO. 1, 2624207
https://doi.org/10.1080/17538947.2026.2624207

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Estimating actual evapotranspiration of soil-vegetation system by combining a physics-based model and machine learning

Yiqing Zhang[a,b], Li Jia[a], Chaolei Zheng[a], Massimo Menenti[a,c,d], Guangcheng Hu[a], Jing Lu[a,e] and Qiting Chen[a]

[a]State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, People's Republic of China; [b]University of Chinese Academy of Sciences, Beijing, People's Republic of China; [c]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands; [d]Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, People's Republic of China; [e]Hainan Aerospace Information Research Institute, Wenchang, People's Republic of China

ABSTRACT

A synergistic integration of physics-based and data-driven approaches has emerged as promising research field for terrestrial evapotranspiration (ET) estimation, enabling robust modeling of land-atmosphere interactions. This study proposes a hybrid model by integrating machine learning (ML)-based canopy surface resistance ($r_{s,c}$) estimation into the Shuttleworth-Wallace (S-W) dual-source scheme under the ETMonitor framework, replacing traditional physics-based $r_{s,c}$ parameterization. Three ML algorithms, Random Forest (RF), Gradient Boosting Regression Tree (GBRT) and Deep Neural Network (DNN) were tested in the hybrid model. A reference dataset of $r_{s,c}$ was derived by inverting S-W dual-source model with in-situ flux measurements. The model was trained on 179 global flux tower sites and independently validated on 45 sites. Three full ML-based models based on DNN, GBRT and RF, were also developed to estimate ET directly for comparison. The DNN-integrated hybrid model outperformed the original physics-based model, with Kling-Gupta Efficiency (KGE) increasing from 0.7 to 0.84 and coefficient of determination (R^2) increasing from 0.66 to 0.72. The three full ML models showed comparable performance to the hybrid models. Notably, the physics-ML hybrid framework balances physical interpretability with data-driven efficiency, minimizing reliance on prior knowledge and avoiding over-parameterization.

ARTICLE HISTORY
Received 25 June 2025
Accepted 25 January 2026

KEYWORDS
Evapotranspiration; hybrid model; deep learning; Shuttleworth-Wallace dual-source model; ETMonitor; canopy surface resistance

1. Introduction

Terrestrial evapotranspiration (ET) is a critical component of the hydrological and energy cycles, and pivotal for addressing hydrological challenges and water resources management (Fang et al. 2010; Good, Noone, and Bowen 2015; Morillas et al. 2013; Zhang et al. 2019). Accurate ET estimation is of great importance for advancing understanding of the hydrological cycle and characterising extreme drought events (Allen et al. 1998; Anderson et al. 2011; Otkin et al. 2016). However, this remains challenging due to complex feedback mechanisms, limited insights from local-scale studies, and inherent heterogeneity of terrain and climate (Allen et al. 2011; Fisher et al. 2017; Liang et al. 2019). Since ET cannot be directly retrieved from satellite-based observations, physics-based and data-driven models have been developed, driven by biophysical variables retrievable from satellite observations (Jia et al. 2009; Jung et al. 2011; Jung et al. 2019; Mu et al. 2007, Su 2002; Zhao et al., 2024).

Physics-based models have a robust physical foundation for ET estimation and can deliver high accuracy at the local and regional scales. Representative methods include surface energy balance (SEB)-based (single and dual-source) models (Bastiaanssen et al. 1998; Menenti and Choudhury 1993; Norman, Kustas, and Humes 1995; Su 2002), combined approaches based on the Penman-Monteith (*P*-M) equation (Allen et al. 1998, 2006; Mu, Zhao, and Running 2011), models based on Priestley-Taylor (*P*-T) equation

CONTACT Li Jia jiali@aircas.ac.cn; Chaolei Zheng zhengcl@aircas.ac.cn State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
Supplemental data for this article can be accessed online at https://doi.org/10.1080/17538947.2026.2624207.

(Fisher, Tu, and Baldocchi 2008; Miralles et al. 2011; Priestley and Taylor 1972), and triangle (or trapezoid) feature space methods (Jiang and Islam 1999; Moran et al. 1994; Tang, Li, and Tang 2010). SEB-based algorithms, integrating vegetation, land surface temperature (LST) and radiation data, are mainly applied in agricultural regions with high-resolution remote sensing (Feng et al. 2023; Jia et al. 2009). However, their applications are constrained by several key limitations: they are highly sensitive to the parameterisation of surface roughness length and vegetation cover, with small deviations in these parameters leading to significant ET estimation errors (Gao et al. 2017; Long and Singh 2012). Feature space-based methods (triangle/trapezoid) estimate ET by leveraging physically meaningful LST-vegetation index relationships (Jiang and Islam 2001; Moran et al. 1994). Their core advantage lies in simplicity and data accessibility. They require only remotely retrievable LST and vegetation-related variables, avoiding complex parameterisation of resistance or energy partitioning processes (Jiang and Islam 1999; Tang, Li, and Tang 2010). This makes them computationally efficient and well-suited for large-scale ET mapping, especially in regions with limited in-situ meteorological data (Yang and Shang 2013). However, they depend on cloud-free LST, suffer from unstable LST-vegetation relationships across regions/seasons, overlook non-temperature/vegetation drivers, and face subjective region-specific boundary calibration, limiting transferability (Li, Jia, and Lu 2015a; Wang, Li, and Cribb 2006). *P*-T methods offer physically grounded ET estimation with parameter parsimony (only α coefficient), low data demand, and computational efficiency, which is ideal for regions with limited meteorological data (Yao et al. 2017). However, they overlook vapour diffusion constraints, rely on region/season-dependent α calibration, and suffer from biases under water-limited conditions (e.g. droughts), limiting adaptability across diverse ecosystems (Wang et al. 2024; Zhang et al. 2021). *P*-M equation-based methods enable physically rigorous ET estimation, explicitly incorporating energy balance and vapour diffusion processes with resistance parameterisations, ensuring high accuracy at local and regional scales (Cleugh et al. 2007; Mu, Zhao, and Running 2011). However, they have been established on the basis of local experiments and application to global scale is challenging (Allen et al. 1998; Yang et al. 2015). Previous studies have shown that the parameterisation of resistances, especially the canopy surface resistance ($r_{s,c}$), has significant impact on ET estimation (Hu et al. 2013; Wei et al. 2020; Zhao et al. 2020). Current $r_{s,c}$ parameterisations are empirical/semi-empirical formulations relying on researchers' prior knowledge, making it difficult to capture all the influences and interactions among $r_{s,c}$ drivers (Ball, Woodrow, and Berry 1987; Jarvis 1976; Leuning et al. 2008). In numerous *P*-M equation-based models for ET estimation (either one-source or two-source models), the Jarvis-type model is widely used to parameterise $r_{s,c}$, with uncertainties stemming from two primary sources. First, the Jarvis-type equation adopts a multiplicative combination of parameters, implicitly assigning equal weight to multiple environmental stressors while overlooking interactions between variables and the differential importance of physiological responses. Second, the pre-specified values of empirical parameters can impede large-scale extrapolation. These processes are frequently characterised based on limited experimental studies, rendering their applicability to diverse vegetation and climate regimes uncertain.

The data-driven methods, including simple empirical regression models and machine learning (ML) algorithms, predict ET by establishing statistical relationships between the target variable and observable land surface variables/parameters (Jung et al. 2011, Jung et al. 2019). ML methods extract inherent patterns from data, building models to learn relationships between target variable and explanatory variables with minimal prior knowledge. Due to their superior predictive power, ML methods are increasingly used to derive patterns from the geoscience data for land surface process modelling and parameter retrieval (Jung et al. 2019; Reichstein et al. 2019; Yuan et al. 2020). However, full ML models lack physical constraints, and their performance is highly dependent on the quality, representativeness and coverage of training datasets (Wang, Gao, and Wang 2022; Yin et al. 2021). Moreover, the choice of ML algorithms and explanatory variables also has an impact on ET estimates (Carter and Liang 2019; Liu et al. 2021).

Hybrid models, combining physics-based models and ML algorithms, leverage the strengths of both approaches, offering robust frameworks for modelling complex biophysical processes (Karpatne et al. 2017). Previous studies have demonstrated that the merits of hybrid modelling in ET estimation (Hu et al. 2021; Koppa et al. 2022; Zhao et al. 2019), but most focused on the single-layer models that do not account for vegetation-soil interactions. Recent work has integrated ML-based estimators with dual-source models, for example, Shang et al. (2023) developed a hybrid model combining two-layer *P*-M equations and the Light Gradient Boosting Machine algorithm to estimate regional ET over the Tibetan Plateau. However,

these hybrid models are trained using limited ET observation sites covering narrow land cover types, with unclear performance in broader applications. Moreover, most studies adopted a single ML algorithm and fixed explanatory variables, lacking systematic evaluation of how different hybrid variants influence ET simulations. It remains unclear whether, and to what extent, hybrid models retain the advantages of both physics-based model and ML algorithms for global ET estimation across diverse scenarios.

The well-established ETMonitor model (Hu and Jia 2015; Zheng et al. 2019, Zheng, Jia, and Hu 2022) uses the Shuttleworth-Wallace (S-W) dual-source scheme (Shuttleworth and Wallace 1985) to estimate ET of soil-vegetation system by partitioning it into plant transpiration and soil evaporation. To facilitate the global application, Zheng, Jia, and Hu (2022) calibrated $r_{s,c}$ by optimising the most sensitive parameters using global flux tower measurements, extrapolating these parameters via plant function types (PFTs) and climate regimes. This calibration improves $r_{s,c}$ parameterisation for more accurate ET estimation but computationally intensive and time-consuming. ML offers a promising solution by distiling relationships from extensive global in-situ observations into data-driven $r_{s,c}$ parameterisation. ML-based $r_{s,c}$ predictors eliminate the need for pre-defined equation structure, instead using data-driven learning to infer functional relationship between the covariates (i.e. environmental factors) and the target variable (i.e. $r_{s,c}$). They inherently assign adaptive weights to $r_{s,c}$ drivers, capturing complex non-linear and interactive physiological responses—including unobserved processes.

Thus, this study uses an unprecedented number of global in-situ observations to develop a novel, assumption-free ML-based $r_{s,c}$ algorithm. By integrating ML into the physics-based ET model ETMonitor, we aim to develop a hybrid model that integrates physical interpretability with data-driven representation of poorly understood inherent processes. The main objectives are: (1) to develop a ML-based estimation approach and provide a drop-in replacement for physics-based parameterisation of canopy surface resistance; (2) to integrate this ML-based $r_{s,c}$ approach into ETMonitor, establishing a hybrid framework fusing physical interpretability with the data-adaptive power of ML; (3) to compare the hybrid model with purely data-driven (i.e. full ML) model and the purely physics-based model, thereby evaluating the pros and cons of each strategies; (4) to interpret hybrid model prediction and quantify how different ML algorithms and input explanatory variables influence ET estimation accuracy.

2. Methodology

In this study, a hybrid model (ETM-SW-ML) is proposed, which integrates ML-based $r_{s,c}$ estimation with the S-W scheme embedded in ETMonitor model (ETM-SW) to estimate ET of soil-vegetation system. ML methods were used in two different configurations: one is directly trained to estimate ET (referred to as the full ML model), and the other is trained to first estimate $r_{s,c}$ before ET is calculated using the ETMonitor framework (referred to as the hybrid model). To further evaluate model performance, the hybrid model is compared with the physics-based model (ETM-SW, see Section 2.1) and full ML models (see Section 2.2). For the physics-based model, two variants of ETMonitor are evaluated: the default version (ETM-SW_default) in which parameters involved in the $r_{s,c}$ parameterisation were set according to Hu and Jia (2015), and the optimised version (ETM-SW_Zheng) proposed by Zheng, Jia, and Hu (2022). The set of explanatory variables for full ML models to estimate actual ET was selected based on a priori knowledge that accounts for the key controls of evapotranspiration process, including net radiation (Rn), leaf area index (LAI), plant function type (PFT), surface soil moisture (SM), vapour pressure deficit (VPD), air temperature (Ta), wind speed (WS), and atmospheric pressure (Pa). The selection of explanatory variables for $r_{s,c}$ estimation in the hybrid model was determined through experiments with different variable combinations (note: solar radiation was used to replace net radiation). Since direct in-situ measurement of $r_{s,c}$ are unavailable, $r_{s,c}$ values and a reference dataset were developed by inverting latent heat flux (LE) from EC measurements based on the ETM-SW framework (see Section 2.3).

A summary of the models is listed in Table 1, with detailed descriptions provided in the subsequent sections. The workflow of this study is illustrated in Figure 1.

2.1. Physics-based model

2.1.1. Shuttleworth-Wallace scheme in ETMonitor model

In ETMonitor, the Shuttleworth-Wallace dual-source model (Shuttleworth and Wallace 1985) has been combined with the Jarvis-type surface resistances parameterisations to estimate the total ET of soil-

Table 1. Overview of the model scenarios for ET estimation developed and compared in this study.

Model type	Model name	Physical framework	Method for $r_{s,c}$ estimation
Hybrid model	ETM-SW-RF	ETM-SW	RF
	ETM-SW-GBRT	ETM-SW	GBRT
	ETM-SW-DNN	ETM-SW	DNN
Full ML model	RF	/	/
	GBRT	/	/
	DNN	/	/
Physics-based model	ETM-SW_default	ETM-SW	Jarvis-type model with default parameter setting (Hu and Jia 2015)
	ETM-SW_Zheng	ETM-SW	Jarvis-type model with optimised parameter setting (Zheng, Jia, and Hu 2022)

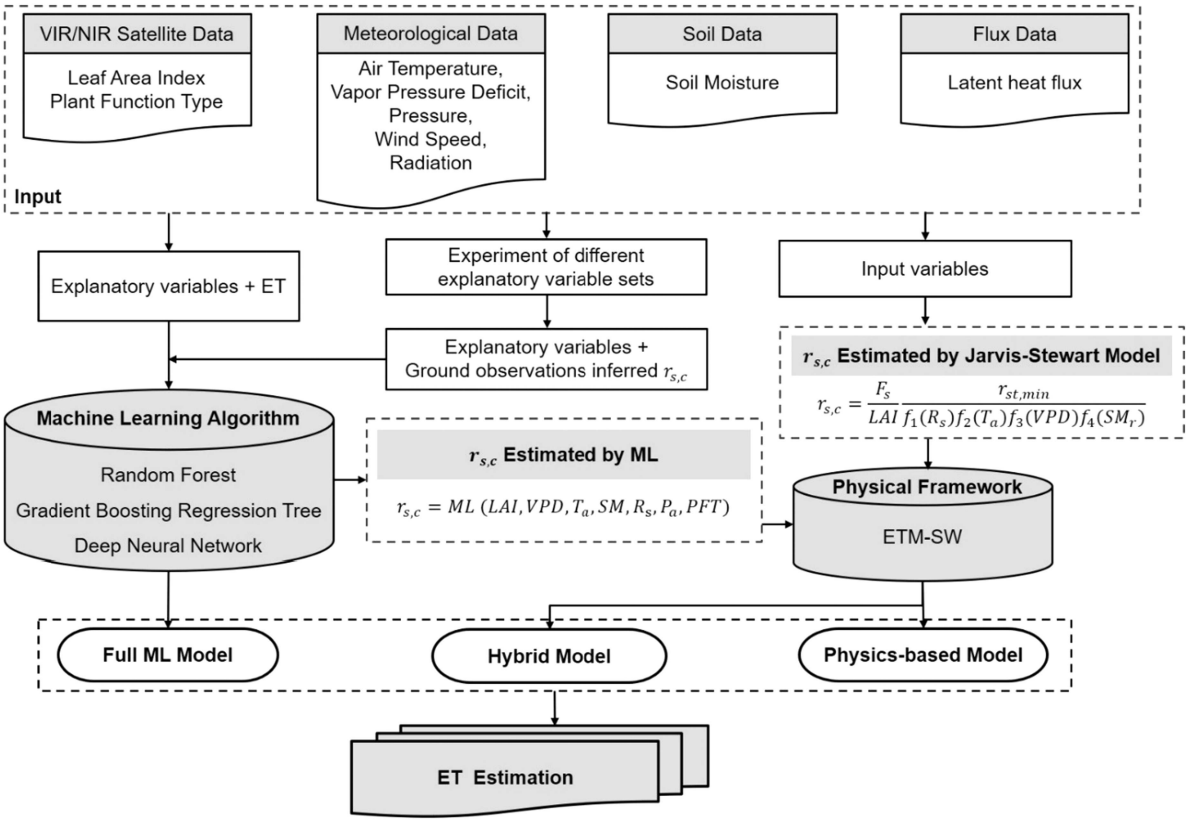

Figure 1. Architecture of hybrid model for ET estimation and workflow of the model comparison.

vegetation system that is the sum of soil evaporation and vegetation transpiration. The S-W model accounts for the energy balance of the vegetation and soil components separately, which assumes that the water vapour flux arriving at the reference height in the atmosphere is the sum of water vapour evaporated from the vegetation and soil layers. The basic equations of the S-W model can be written as:

$$ET = C_c \cdot PM_C + C_s \cdot PM_s \tag{1}$$

$$PM_C = \frac{\Delta(Rn - G) + [\rho C_p VPD - \Delta r_{a,c}(Rn_s - G)]/(r_{a,a} + r_{a,c})}{\Delta + \gamma[1 + r_{s,c}/(r_{a,a} + r_{a,c})]} \tag{2}$$

$$PM_s = \frac{\Delta(Rn - G) + (\rho C_p VPD - \Delta r_{a,s} Rn_c)/(r_{a,a} + r_{a,s})}{\Delta + \gamma[1 + r_{s,s}/(r_{a,a} + r_{a,s})]} \tag{3}$$

$$C_c = \{1 + R_c \cdot R_a/[R_s \cdot (R_c + R_a)]\}^{-1} \tag{4}$$

$$C_s = \{1 + R_s \cdot R_a/[R_c \cdot (+R_a)]\}^{-1} \tag{5}$$

$$R_a = (\Delta + \gamma)\cdot r_{a,a} \tag{6}$$

$$R_s = (\Delta + \gamma)\cdot r_{a,s} + \gamma\cdot r_{s,s} \tag{7}$$

$$R_c = (\Delta + \gamma)\cdot r_{a,c} + \gamma\cdot r_{s,c} \tag{8}$$

where ET (W m^{-2}) is the total ET of the soil-vegetation system; PM_C and PM_s (W m^{-2}) are variables related to the vegetation transpiration and the soil evaporation, respectively; C_c and C_s are surface and aerodynamic resistance coefficients (unitless) for vegetation transpiration and soil evaporation; G (W m^{-2}) is soil heat flux at soil surface; R_n (W m^{-2}) is net radiation flux; Rn_c and Rn_s (W m^{-2}) are net radiation fluxes intercepted by the vegetation canopy and received at the soil surface, respectively; Δ is slope of the saturated vapour pressure curve (Pa K^{-1}); γ (Pa K^{-1}) is the psychrometric constant; ρ (kg m^{-3}) is the air density; C_p (J kg^{-1} K^{-1}) is the constant pressure specific heat; $r_{a,a}$ (s m^{-1}) is the aerodynamic resistance; $r_{a,c}$ and $r_{a,s}$ (s m^{-1}) are the canopy and soil boundary layer resistance, respectively; $r_{s,c}$ and $r_{s,s}$ are the surface resistance of canopy and soil (s m^{-1}), respectively. λ, Δ, γ, ρ, C_p, and VPD are determined following Allen et al. (1998), $r_{a,a}$, $r_{a,c}$, $r_{a,s}$, $r_{s,s}$ R_n, Rn_c and Rn_s are estimated following the ETMonitor model (Hu and Jia 2015; Zheng, Jia, and Hu 2022). The canopy resistance $r_{s,c}$ is estimated by Jarvis-type parameterisation considering several environmental factors and described in Section 2.1.2.

2.1.2. Parameterisation of canopy surface resistance

In ETMonitor, canopy surface resistance is calculated by scaling leaf stomatal resistance with LAI using the Jarvis-type model (Hu and Jia 2015; Jarvis 1976). Leaf stomatal resistance is estimated using a multiplicative model by setting the minimum stomatal resistance ($r_{st,\,\min}$) according to PFT and by considering the constraints of several environmental stress factors on $r_{st,\,\min}$. In ETMonitor, the Jarvis-type equation is written as (Hu and Jia 2015):

$$r_{s,c} = \frac{1}{LAI_{eff}}\cdot\frac{r_{st,\,\min}}{f_1(R_s)f_2(T_a)f_3(VPD)f_4(SM_r)} \tag{9}$$

where $r_{st,\,\min}$ is the minimum leaf stomatal resistance (s m^{-1}) without any environmental constrains; LAI_{eff} is the effective canopy leaf area index and estimated by an empirical equation following Allen et al. 2006; $f_1(R_s)$, $f_2(T_a)$, $f_3(VPD)$ and $f_4(SM_r)$ are functions that quantifies stomatal response to environmental factors R_s (W m^{-2}), T_a (°C), VPD (hPa) and root zone soil water content (SM_r) (cm^3 cm^{-3}), respectively.

The stress functions f_i (i = 1, 2, 3, 4), varying between 0 and 1 (0 for infinite stress; 1 for no stress with fully open stomata) are expressed as (Mo et al. 2004; Stewart 1988):

$$f_1(R_s) = 1 - \exp\left(-\frac{R_s}{k_{R_s}}\right) \tag{10}$$

$$f_2(T_a) = \left(\frac{T_a - T_{\max}}{T_{opt} - T_{\min}}\right)\left(\frac{T_{\max} - T_a}{T_{\max} - T_{opt}}\right)^{(T_{\max}-T_{opt})/(T_{opt}-T_{\min})} \tag{11}$$

$$f_3(VPD) = 1 - k_{VPD}\,VPD \tag{12}$$

$$f_4(SM_r) = \begin{cases} 1 & SM_r > SM_{fc} \\ \frac{SM_r - SM_w}{SM_{fc} - SM_w} & SM_w \le SM_r \le SM_{fc} \\ 0 & SM_r < SM_w \end{cases} \tag{13}$$

where k_{R_s} is taken as 500 W m^{-2}; $T_{\max}$, $T_{\min}$ and T_{opt} are the maximum, minimum, and optimum air temperatures influencing the stomatal activity, respectively; k_{VPD} is the fitting parameter depending on PFT that describes the sensitivity of stomatal resistance to VPD; SM_{fc} and SM_w represent the thresholds of SM at

the field capacity and the wilting point, respectively, which are obtained from the soil properties map (Zhang, Schaap, and Zha 2018).

In the earlier version of ETMonitor (ETM-SW_default; Hu and Jia 2015), fixed parameter values (as in Equations (5-8)) are assigned to each PFT. Considering the comprehensive influence of climate and plant conditions, Zheng, Jia, and Hu (2022) optimised the most sensitive parameters ($r_{st,\,\min}$ and k_{VPD}) using in-situ observations from flux tower sites distributed globally to calibrate the $r_{s,c}$. They ran the ETMonitor model with thousands of parameter-settings at each flux tower site, determining parameter values for different climate regimes and PFTs by minimising the difference between modelled and observed ET. Details of this optimised ETMonitor version (ETM-SW_Zheng) and its parameter calibration protocol can be found in Zheng, Jia, and Hu (2022). ET estimates from both ETM-SW_default and ETM-SW_Zheng were subsequently compared with the hybrid models.

2.2. Machine learning models for actual ET estimation

ML algorithms were used in two different ways, one way is trained to estimate actual ET directly (full ML model), as well as trained to estimate $r_{s,c}$ first and then calculate ET based on the ETMonitor framework (hybrid model). In this section, we introduce how the full ML model for actual ET estimation is developed.

2.2.1. Machine learning algorithms

Three ML algorithms, including Random Forest (RF; Breiman 2001), Gradient Boosting Regression Tree (GBRT; Friedman 2002) and Deep Neural Network (DNN; Chollet 2017), were selected for full ML model development, legveraging their well-documented strengths and proven applicability in geoscience and hydrology.

Proposed by Breiman (2001), RF uses the Bootstrap Aggregating (Bagging) framework to ensemble multiple decision trees, fallen under the category of Ensemble Learning (EL). It has been extensively applied to regression and forecasting tasks due to its stability and ability to handle high-dimensional data (Feng et al. 2017; Lu et al. 2018; Yin et al. 2021).

Introduced by Friedman (2002), GBRT is a widely used ensemble method with robust predictive performance, having demonstrated cutting-edge capability in geoscience-focused ET estimation (Fan et al. 2018; Lu et al. 2018; Wang et al. 2021). Like RF model, GBRT integrate weak learners (decision stumps or regression trees); however, its key distinction lies in sequentially fitting new trees to the residuals of previous trees, thereby minimising prediction biases, whereas RF primarily reduces variances.

DNN is a prominent deep learning algorithm built upon shallow neural networks, characterised by a multilayer perceptron structure. Unlike shallow networks (typically 1-2 layers), DNN achieves input-to-target transformation through a deep sequence of layers (tens to hundreds of layers) (Chollet 2017). Substantial previous studies have demonstrated its excellent ability to predict a variety of hydrological variables (Ienco et al. 2019; Saggi and Jain 2019; Yuan et al. 2020). A key advantage over traditional ML methods is its insensitivity to outliers, which enhances its reliability in ET estimation.

2.2.2. Full ML model development for ET estimation

Three full ML algorithms (RF, GBRT, DNN) were trained with ET as the target variable. The explanatory variables included Rn, LAI, SM, VPD, Ta, WS, Pa, and PFT. PFT is regarded as categorical variable, while the other variables were numerical.

The architectures of the RF and GBRT models were configured by identifying the optimal values of a set of hyperparameters. For RF, the hyperparameter configuration was determined by evaluating the individual effects of each hyperparameter and then integrating their optimal values. GBRT was tuned using a Bayesian optimiser to identify the optimal hyperparameter values that maximised model performance. For DNN, the key hyperparameters included the number of hidden layers, the number of neurons per layer, activation function, learning rate, and optimiser. We systematically trained a series of DNNs where the number of hidden layers ranged from 1 to 15, and the number of neurons per layer varied as powers of two (ranged from 8 to 128). As a result, the DNN configuration with 9 hidden layers (128 neurons in each hidden layer) yielded the optimal performance on the given dataset. The 'Rectified Linear Units' (ReLU)

activation function was adopted for the hidden layers, learning rate was set to 0.001, and the 'Adam' optimiser was used to compile the DNN model. The hyperparameter values for the three full ML models for ET estimation are presented in Table 2.

Flux tower measurements of ET (i.e. latent heat flux) and meteorological variables served as the core data source for training and validating the full ML models. These data (from 224 sites spanning divers land cover and climate regimes, see Section 3.1 for details) were split into two subsets: 80% of the total sample (from 179 sites) was used to train the ML models and parameters calibration, while the remaining 20% (from the rest 45 sites) severed as independent validation using metrics described in Section 2.4. The training dataset was randomly divided into ten smaller subsets for ten-fold cross validation. Each ML algorithm was tuned by varying hyperparameters to identify the optimal parameter set that minimised root mean square error (RMSE). These optimal parameter sets were then applied to their respective ML algorithm. While computationally intensive, this approach maximised sample utilisation— a key advantage for mitigating over-fitting. ML models were further validated using the independent validation set to assess the generalisation capability. All ML models were implemented in Python using the sklearn and tensorflow library packages.

2.3. Physics-ML hybrid model

Physics-ML hybrid models were developed by replacing the $r_{s,c}$ parameterisation (Equation 9) in ETMonitor with estimation of $r_{s,c}$ by ML algorithms. Therefore, three ML models for $r_{s,c}$ estimation were first developed with the three ML algorithms given in Section 2.2.1, then they are implemented in the ETMonitor model to obtain the respective physics-ML hybrid models, i.e. ETM-SW-RF, ETM-SW-GBRT and ETM-SW-DNN.

Since there are no available $r_{s,c}$ observations to train the ML algorithms, we generated the $r_{s,c}$ reference dataset by inverting the ETM-SW formulation using the measurements of LE (response) and the environmental variables (relative humidity (RH), T_a, P_a, WS, SM, R_n, G, i.e. forcing). Firstly, we calculated daily values of the forcing variables and parameters (except for $r_{s,c}$) in Equations (1-3) using in-situ measurements. Then the $r_{s,c}$ values were determined by minimising difference between the ET measurements and the ET estimates using Equations (1-8) with $r_{s,c}$ varying within 0-50000 s m^{-1}. The reference $r_{s,c}$ was taken as the target variable to build ML models for $r_{s,c}$ estimation using RF, GBRT and DNN algorithms. At last, by substituting the predicted $r_{s,c}$ value from ML models back into the ETM-SW formulation (Equations (1-3)) the improved daily ET estimation can be obtained.

Solar radiation (energy), LAI (vegetation) and SM (available soil water) are key controls on plant transpiration via their regulation of $r_{s,c}$, while VPD and Ta determine the efficiency of vapour evaporation from the canopy surface. These environmental factors should therefore be accounted in the ML-based $r_{s,c}$ estimation. Previous studies have reported that the accuracy of hybrid models improved when more explanatory variables are used for $r_{s,c}$ estimation (Chen et al. 2022). To evaluate the importance of

Table 2. Hyperparameters and their values used in the three full ML models for evapotranspiration estimation.

Hyperparameter	RF	GBRT	DNN
n_estimators	128	120	–
max_depth	22	24	–
max_features	Sqrt	Sqrt	–
min_samples_split	2	–	–
min_samples_leaf	1	–	–
criterion	squared_error	friedman_mse	–
min_impurity_decrease	–	7	–
subsample	–	0.9	–
loss	–	squared_error	mse
learning_rate	–	0.002	0.001
num_hidden_layers	–	–	9
num_neurons	–	–	128
epochs	–	–	50
batch_size	–	–	4096
activation	–	–	relu
optimiser	–	–	adam

explanatory variables in ML-based $r_{s,c}$ estimation, nine experiments with varying input variable combinations were conducted (Table 3). In the Jarvis-type $r_{s,c}$ parameterisation, LAI, VPD, T_a, R_s, SM (Test6) serve as the core drivers of $r_{s,c}$. From Test1 to Test5, one of these five variables was sequentially excluded to assess its impact on hybrid model performance. Test7 and Test8 introduced two additional explanatory variables (PFT and P_a) not explicitly considered in the Jarvis-type model, designed to deepen insight into how $r_{s,c}$ responds to the environmental factors. Test9, which includes the most explanatory variables, comprises LAI, VPD, Ta, Rs, SM, Pa, and PFT. While Pa is not typically included in widely-used $r_{s,c}$ parameterisations, it reflects environmental conditions (e.g. elevation) that may influence stomatal closure, thus justifying its inclusion in our test.

In the physics-ML hybrid models, the hyperparameter optimisation process for each ML algorithm is the same as in the full ML models (see Section 2.2). The hyperparameter values of each ML algorithm for $r_{s,c}$ estimation are shown in Table 4. We measured the wall-clock time and RAM cost of ML training and prediction, with detailed information of the computation cost provided in **Supplementary Text S1** and **Supplementary Table S1**.

2.4. Model validation and interpretability analysis

We used the in-situ measurement subset excluded from model training (from 45 sites, accounting for 20% of the total sites) to validate both the full ML model and the physics-ML hybrid model for ET estimation. To assess model performance, we adopted the following commonly used metrics: Kling-Gupta efficiency (KGE) (Gupta et al. 2009), coefficient of determination (R^2), the bias, and the RMSE. To interpret the model estimates, we conducted SHapley Additive exPlanation (SHAP) analysis to calculate Shapley values. Shapley value, proposed to calculate model predictions in cooperative game theory, is among the most widely used metrics for feature attribution (Wang et al. 2021). SHAP theory conceptualises model

Table 3. Experimental designs of varying explanatory variable combinations for ML-based $r_{s,c}$ estimation.

Test	Explanatory variables
Test1	VPD, Ta, Rs, SM
Test2	LAI, Ta, Rs, SM
Test3	LAI, VPD, Rs, SM
Test4	LAI, VPD, Ta, SM
Test5	LAI, VPD, Ta, Rs
Test6	LAI, VPD, Ta, Rs, SM
Test7	LAI, VPD, Ta, Rs, SM, PFT
Test8	LAI, VPD, Ta, Rs, SM, Pa
Test9	LAI, VPD, Ta, Rs, SM, Pa, PFT

Table 4. Hyperparameters and their values used for the three ML-based models for $r_{s,c}$ estimation.

Hyperparameter	RF	GBRT	DNN
n_estimators	196	270	–
max_depth	22	40	–
max_features	Sqrt	Sqrt	–
min_samples_split	2	–	–
min_samples_leaf	1	–	–
criterion	squared_error	friedman_mse	–
min_impurity_decrease	–	7	–
subsample	–	0.9	–
loss	–	squared_error	mse
learning_rate	–	0.004	0.001
num_hidden_layers	–	–	8
num_neurons	–	–	128
epochs	–	–	150
batch_size	–	–	4096
activation	–	–	relu
optimiser	–	–	adam

estimates as the sum of marginal contributions from individual input features (Lundberg, Erion, and Lee 2018). This involved generating diverse input samples and tracking how model estimates change with features variation. We used SHAP library in Python to compute the Shapley values for each explanatory variable in the hybrid and full ML models, quantifying their contributions and enhancing the interpretability of ET estimates.

3. Data

3.1. Flux data

Flux tower data were used to train and validate the hybrid and full ML models, as well as for ET validation of the full physics-based models. Data were collected from a total of 224 flux tower sites globally, including 148 from the Fluxnet2015 dataset (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/), 34 from the Ameriflux network (http://ameriflux.lbl.gov/), 10 from the Euroflux network (http://www.europe-fluxdata.eu/), 9 from the Ozflux network (https://ozflux.org.au/), 8 from the Asiaflux network (http://www.asiaflux.net/), 6 from the Chinaflux network (http://chinaflux.org/) and 9 from the HiWATER (Heihe Watershed Allied Telemetry Experimental Research) network (https://data.tpdc.ac.cn/). These sites span a wide range of climate regimes, from tropical to polar regions, and encompass thirteen plant functional types (PFTs) based on the International Geosphere-Biosphere Programme vegetation classification scheme (Loveland et al. 1999): evergreen needle leaf forests (ENF), evergreen broadleaf forest (EBF), deciduous needle leaf forest (DNF), deciduous broadleaf forest (DBF), mixed forest (MF), closed shrub lands (CSH), open shrub lands (OSH), woody savannas (WSA), savannas (SAV), grasslands (GRA), permanent wetlands (WET), croplands (CRO) and cropland/natural vegetation mosaics (CVM). The spatial distribution of these 224 sites is presented in Figure 2, with detailed site information provided in **Supplementary Table S2**.

In-situ observations include measurements of LE (from eddy covariance system), radiation fluxes, Ta, VPD, WS, Pa and SM (from automatic weather station). For networks providing daily datasets (Fluxnet2015, Ameriflux, Euroflux, Chinaflux), we excluded samples with sensor failure, incomplete records, negative values of LE and VPD. Data from rainy days and subsequent day were also excluded to avoid effects of rain interception (Medlyn et al. 2011). For other networks with only half-hourly measurements (Ozflux, Asiaflux, HiWATER), we first removed data affected by sensor malfunction,

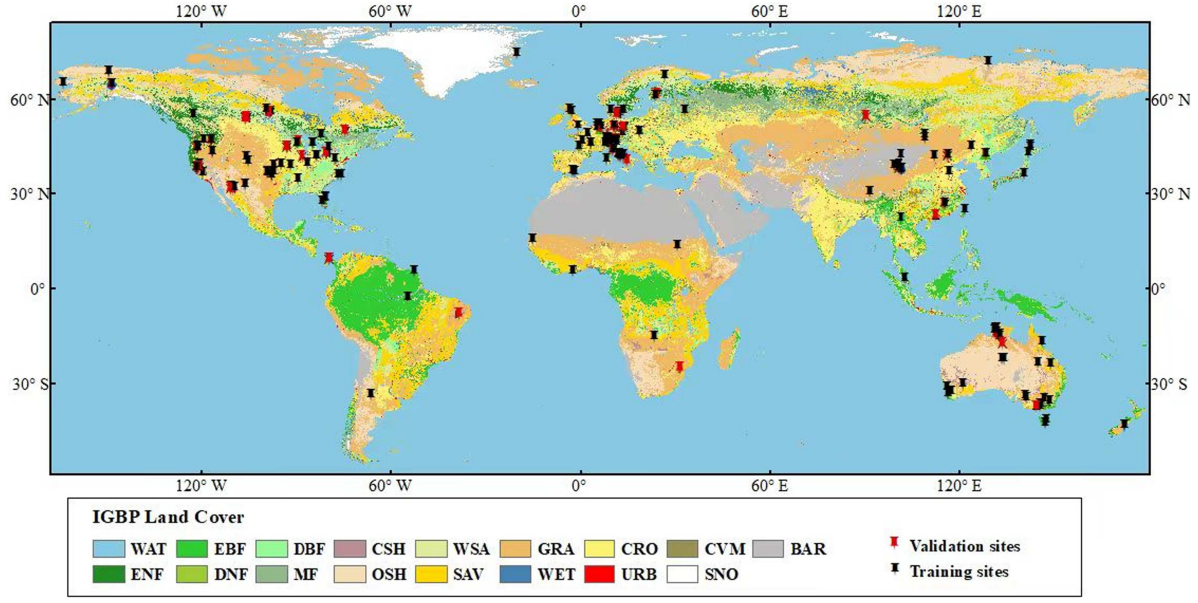

Figure 2. Locations of selected flux tower sites over IGBP-classified land cover map. Different coloured pushpins denote training sites and validation sites, respectively.

incomplete records, or negative LE values (Liu et al. 2018; Xu et al. 2020). Days with fewer than 40 valid 30-min observations were then excluded, and observations from remaining days were aggregated into daily values. Daily records from rainy, snowy days, and the subsequent day were also removed.

The original dataset was split into two subsets: 80% of the total samples (179 sites) were used for ML model training and parameters calibration, while the remaining 20% (45 sites) severed as an independent validation set. The selection of training sites and independent validation sites adheres to criteria for uniform spatial distribution and sufficient representativeness of plant functional types. The independent validation set, excluded from the ML training process, allows evaluation of the generalisation and spatial simulation capabilities of both physics-ML hybrid and ML models. Details of site selection for model training and validation are shown in Table 5.

3.2. Remote sensing data

Remote sensing data used in this study include LAI and the global maps of soil properties (Table 6). The 8-day LAI data with a 500-m spatial resolution were taken from the Global Land Surface Satellite (GLASS) products (Xiao et al. 2014), and used for $r_{s,c}$ and ET estimation. Based on the location of each site, we extracted the LAI value for corresponding pixels and reconstructed them to daily resolution by linear interpolation to ensure consistency with the ET estimates. The global map of soil hydraulic properties used in the ETMonitor model was derived from a hierarchical parameterisation of a water retention model (Zhang, Schaap, and Zha 2018).

4. Results

4.1. Interpretability of the ML models

We analysed the correlations between variables to explore their dependencies (Figure 3). Factors influencing ET process were categorised into four groups: energy (radiation and Ta), water supply (SM), atmospheric water demand (VPD), and aerodynamic factor (WS). Generally, $r_{s,c}$ showed relatively variable yet weak correlations with these factors (Figure 3a), suggesting complex nonlinear responses of stomatal biological processes to synergistic variations among these factors. ML approaches are well-suited to address this, given their advantages in representing biophysical processes and identifying model parameters (de Bézenac, Pajot, and Gallinari 2019; Jiang et al. 2024; Kraft et al. 2020; Rasp, Pritchard, and Gentine 2018). In addition to the above factors, vegetation distribution and condition, represented by LAI, also play important roles in determining transpiration, as plants extract SM from the root-zone via their root systems. $r_{s,c}$ showed a stronger correlation with VPD and LAI than with other factors, highlighting their

Table 5. List of the selected sites for ML models training and independent validation.

PFT	Total number of sites	Number of sites for training	Number of sites for validation
DBF	23	18	5
DNF	3	2	1
EBF	19	17	2
ENF	43	35	8
MF	9	6	3
CRO	30	24	6
GRA	56	45	11
OSH	13	10	3
WET	9	7	2
SAV	7	5	2
WSA	7	5	2
CSH	4	4	0
CVM	1	1	0
ALL	224	179	45

Table 6. Remote sensing data used in this study.

Source	Variable	Spatio-temporal resolution	Periods	Reference
GLASS	LAI	500 m, 8-day	1996-2021 (different across sites)	Xiao et al. (2014)
Global maps of soil properties	SM_{fc}, SM_w	1 km, –	–	Zhang, Schaap, and Zha (2018)

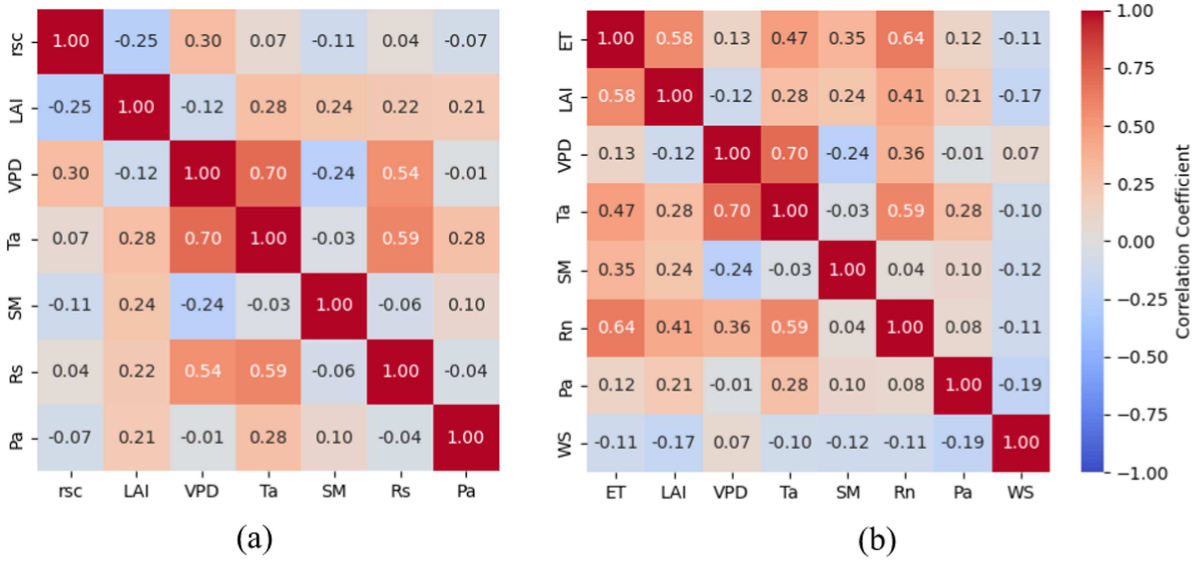

Figure 3. The heat map of the correlation coefficients between the target variables and the influencing factors in the hybrid model and full ML models: (a) for $r_{s,c}$; (b) for total ET. The values in the grid represent the correlation coefficient, and the colour of the grid visualises the correlation coefficient value, with red representing positive correlation and blue representing negative correlation.

critical role in regulating $r_{s,c}$. ET showed strong correlations with Rn, LAI and Ta (Figure 3b). In addition, high correlations exist among some influencing factors (e.g. VPD and Ta), suggesting potential information redundancy. Notably, ET showed a negative correlation with WS, which contradicts evapotranspiration mechanisms. This is probably because the relationship between ET and WS at different sites is significantly affected by local environmental conditions, and uncertainties in WS measurements may further introduce ambiguity in interpreting this relationship.

To evaluate the importance of explanatory variables in ML-based $r_{s,c}$ estimation, nine experiments with varying input variable combinations for $r_{s,c}$ (see Table 3) were implemented to develop the three hybrid models (i.e. ETM-SW-DNN, ETM-SW-GBRT, ETM-SW-RF), and Figure 4 compares the performance of these models for ET estimation against in-situ measurements. Overall, model performance improved with more input explanatory variables, characterised by higher KGE and R^2 (Figure 4a,b), and lower RMSE and absolute bias (Figure 4c,d). Observable performance difference exists, with the hybrid model ETM-SW-DNN incorporating all seven explanatory variables achieving the highest KGE (0.84). Results from Test1 and Test5 (where LAI and SM were excluded, respectively) showed significant performance declines across all three hybrid models. In Test8 and Test9, the addition of Pa enhanced ET estimation performance for ETM-SW-DNN, while results varied for ETM-SW-GBRT and ETM-SW-RF, indicating that contribution of Pa to $r_{s,c}$ estimation requires further investigation. All seven variables (i.e. the combination in Test 9) were included in the final version of the hybrid models.

To interpret the model predictions, we conducted the SHAP analysis for both hybrid and full ML models. Figure 5 shows the proportion of the Shapley values for each explanatory variable, quantifying their contributions to target variable estimation. Contributions of each explanatory variable to $r_{s,c}$ and ET estimation vary across models with different ML configurations. In full ML models, LAI and Rn are the top contributors (Figure 5a); while in hybrid models, LAI and VPD dominate $r_{s,c}$ estimation thereby ET (Figure 5b). These findings align with the correlation analysis shown in Figure 3. In all these models, PFT had minimal influence on the variation of $r_{s,c}$ and ET, indicating that PFT-derived information may be redundant with other explanatory variables.

4.2. Validation with in-situ measurements

We first evaluated the performance of three ML algorithms (DNN, GBRT, RF) in predicting $r_{s,c}$ (Figure 6). Overall, the DNN performed better than GBRT and RF, with higher KGE and R^2, and lower RMSE. As a

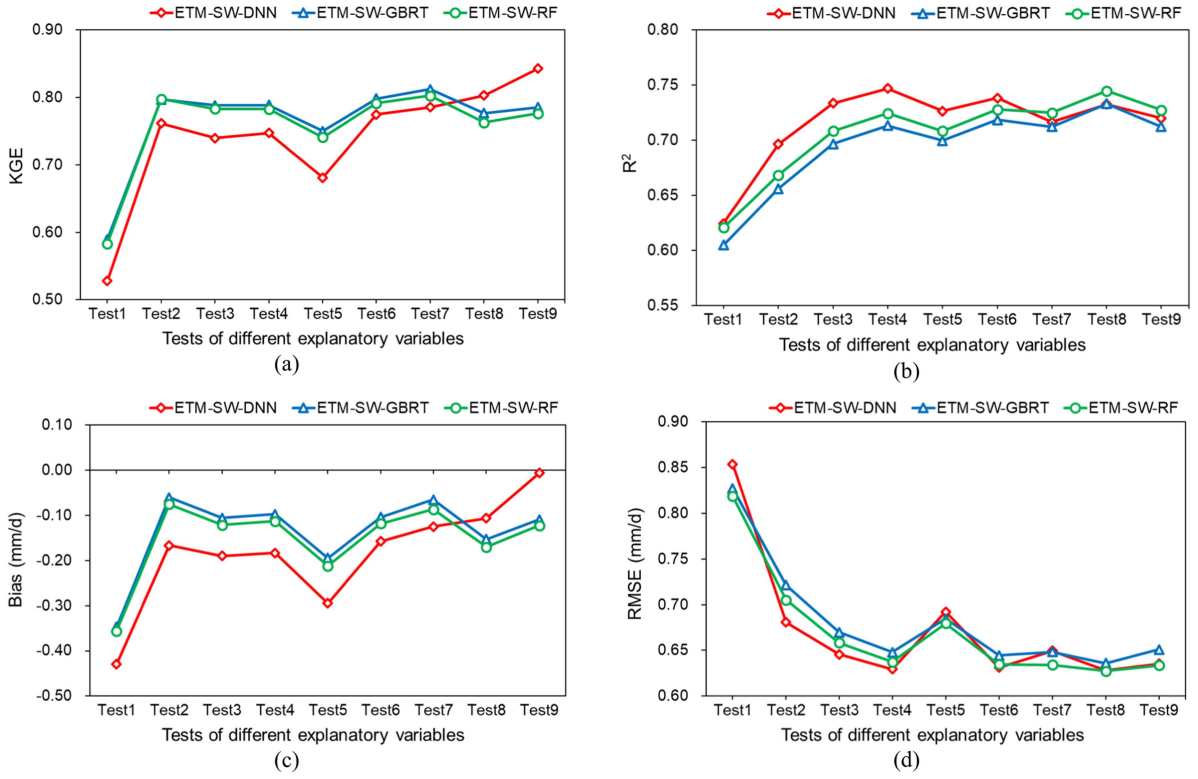

Figure 4. Performance comparison of ET estimates from the three hybrid models (ETM-SW-DNN, ETM-SW-GBRT, ETM-SW-RF) with different explanatory variables combinations: (a) KGE; (b) R^2; (c) bias; (d) RMSE. The unit of bias and RMSE is mm/d, R^2 and KGE are unitless.

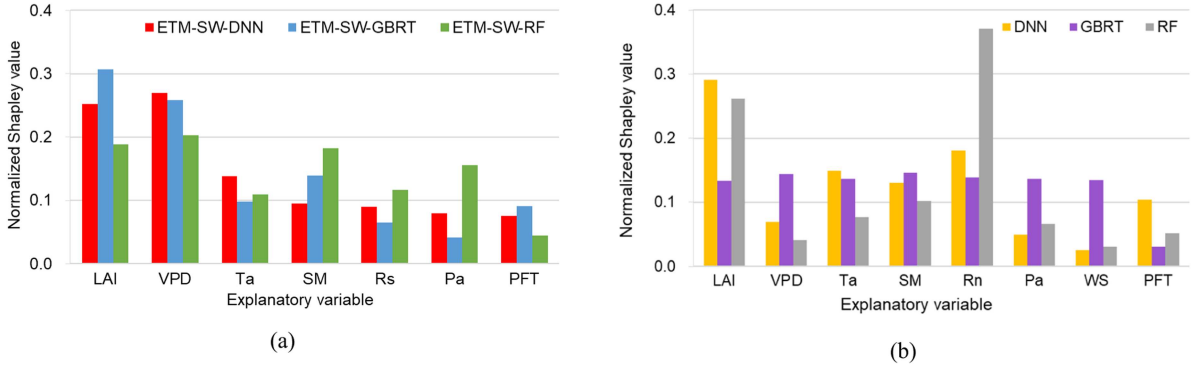

Figure 5. Contributions of explanatory variables based on Shapley values to: (a) $r_{s,c}$ estimation in the three hybrid models; (b) ET estimation in the three full ML models.

result, the ETM-SW-DNN also achieved the best performance in ET estimation (Figure 7), highlighting the importance of improving $r_{s,c}$ predictions accuracy, as higher $r_{s,c}$ accuracy contributes to better ET estimation. Bias analysis indicates that all three ML methods underestimate $r_{s,c}$, especially for the extremely large values. This prediction shrinkage issue is common in many regression models, which often sacrifice accuracy in predicting extreme values to improve the overall fitting performance, as their objective is to minimise overall error (i.e. RMSE). Using complex non-linear models (e.g. DNN) to capture non-linear relationships in data can alleviate the inaccuracies in predicting extreme values (Chollet 2017).

We evaluated the performance of three hybrid models (ETM-SW-DNN, ETM-SW-GBRT, ETM-SW-RF), three full ML models (DNN, GBRT, RF) and two physics-based models (ETM-SW_default and ETM-SW_Zheng) by comparing daily ET estimates with in-situ measurements from the 45 independent sites

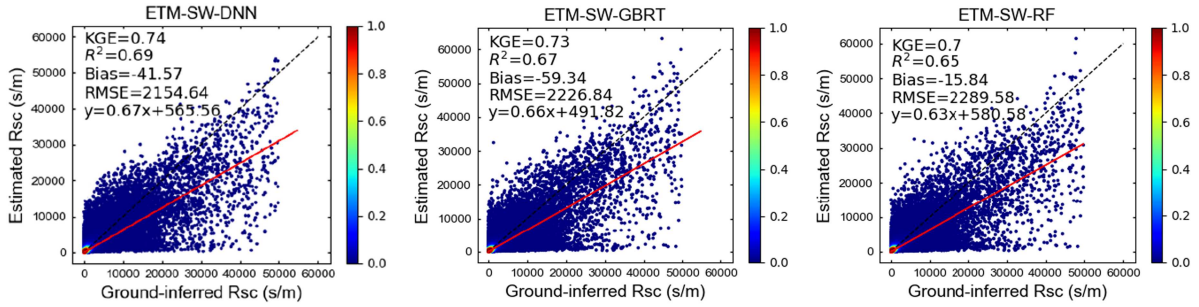

Figure 6. Validation results of three hybrid models in predicting $r_{s,c}$. The units of bias and RMSE are s/m, R^2 and KGE are unitless. The black dotted line is the 1:1 line, and the red solid line is the regression line.

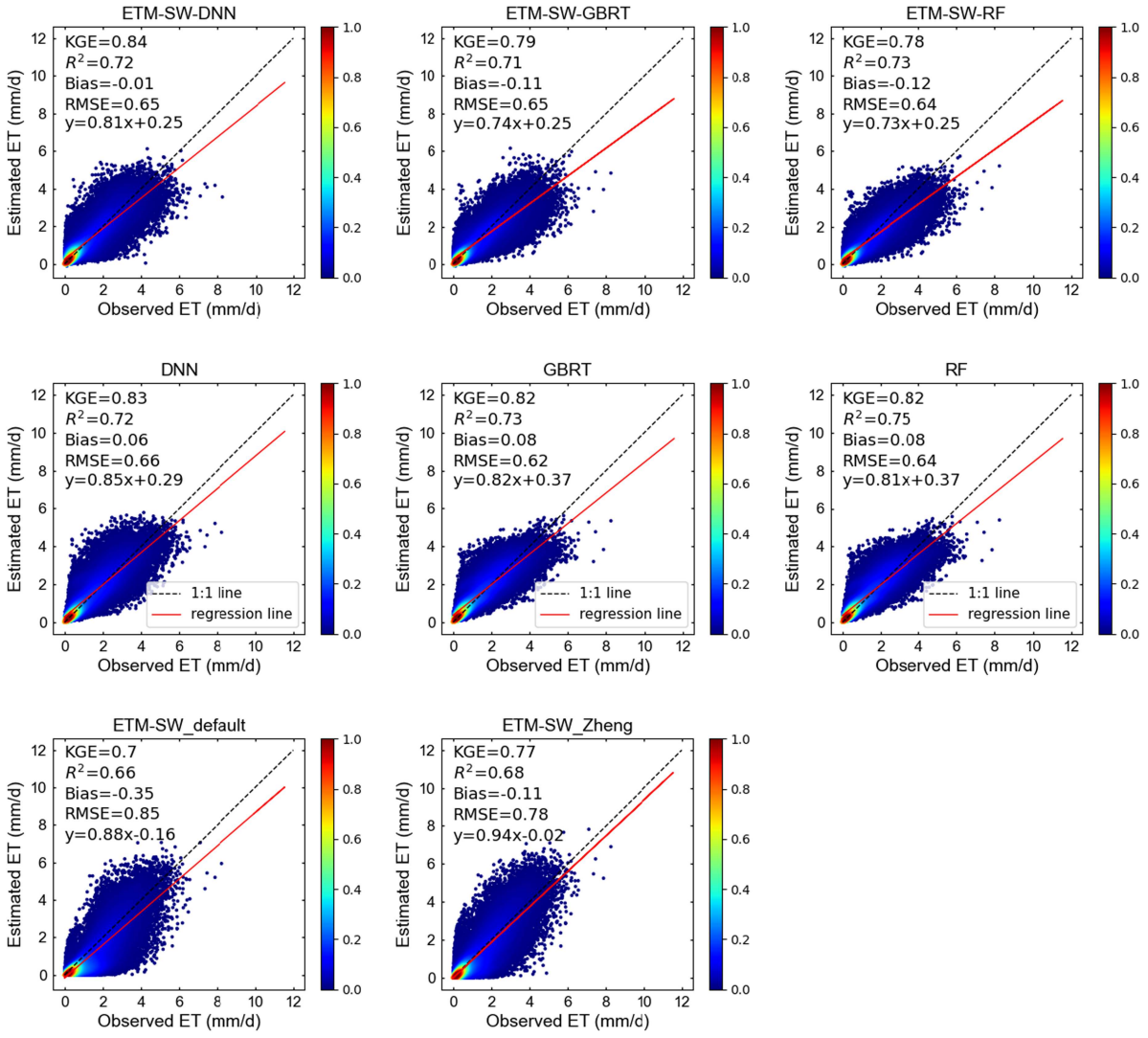

Figure 7. Validation of model predicted ET against in-situ measurements: three hybrid models (first row), three full ML models (second row), and two process-based models (third row). Units: bias and RMSE in mm/d; R^2 and KGE are unitless. Black dotted line - 1:1 line; red solid line = regression line.

(Figure 7). The results indicated that the hybrid models improve the performance of the physics-based model in ET estimation, and show comparable accuracy with the full ML models. ETM-SW-DNN achieved the highest KGE (0.84), followed by the three full ML models. In addition, the full ML models tend to overestimate at low values and underestimate at high values, which is particularly evident in the

scatter plot of RF and GBRT. From the bias result we can find that ETM-SW_default obtained strong negative bias due to the high $r_{st,min}$ value assigned to some PFTs (e.g. CRO), and the optimised version ETM-SW_Zheng has alleviated this problem thanks to the parameter calibration. After integrating the ML-based $r_{s,c}$ formulation into ETMonitor, the underestimation of ET was significantly reduced, suggesting the importance of improving the $r_{s,c}$ computation and the superiority of ML in $r_{s,c}$ simulation.

Boxplots of the four evaluation metrics for ET estimation accuracy of each model at independent validation sites were shown in Figure 8. Overall, the six new models (three hybrid models and three full ML models) performed significantly better than ETM-SW_default and slightly better than ETM-SW_Zheng, characterised by larger KGE and R^2, and smaller bias and RMSE. For KGE, hybrid models produced fewer outliers and consistent ET estimation accuracy across sites, suggesting greater stability when applied under varying conditions. Results for R^2 and RMSE limited differences between hybrid and full ML models, i.e. median R^2 values ranged from 0.77 to 0.81, and median RMSE values from 0.53 to 0.58. Bias values, however, reveal that physics-based and hybrid models tend to underestimate daily ET, whereas full ML models tend to overestimate it. Notably, replacing the Jarvis-type parameterisation scheme of $r_{s,c}$ with the DNN algorithm effectively mitigated the underestimation of daily ET by the ETM-SW_default model (the median bias value decreased from -0.42 to -0.04).

4.3. Model performance in different plant function types

To investigate the ability of the hybrid models to estimate ET across different plant functional types, we carried out the separate validation for each PFT using the four metrics (KGE, R^2, bias and RMSE) (Figure 9). The number of sites per PFT is detailed in Table 5, with specific independent validation statistics for eleven PFTs are shown in Figure 9. In general, RF, GBRT and DNN showed similar performance for most PFTs when directly estimating ET. In contrast, when integrated into the physics-based model, ETM-SW-DNN outperformed ETM-SW-RF and ETM-SW-GBRT. Compared to the physics-based models, hybrid models performed better for most PFTs, with the exception of DNF, CRO and GRA. For PTFs with limited training sites (DNF, SAV, WSA and MF), hybrid models achieved a higher accuracy of ET estimates than the full ML models, with significant differences in metrics across models. For PTFs with abundant training sites (GRA, CRO, ENF and EBF), the full ML models gave better results, and differences in the ET estimation accuracy among models were relatively small.

To further explore how the quantity of training data for each PFT influence model performance, we divided the validation dataset into three types based on the number of sites per PFT: Type 1 includes PFTs with fewer than 10 training sites (DNF, MF, SAV, WSA and WET); Type 2 includes PFTs with 10-30 training sites (DBF, EBF, OSH and CRO); and Type 3 includes PFTs with more than 30 training sites (ENF and GRA). We compared the performance of ETM-SW-DNN with that of full ML models (DNN, GBRT and RF) and the physics-based model (ETM-SW_Zheng) for each PTF type (Figure 10). Additionally, time

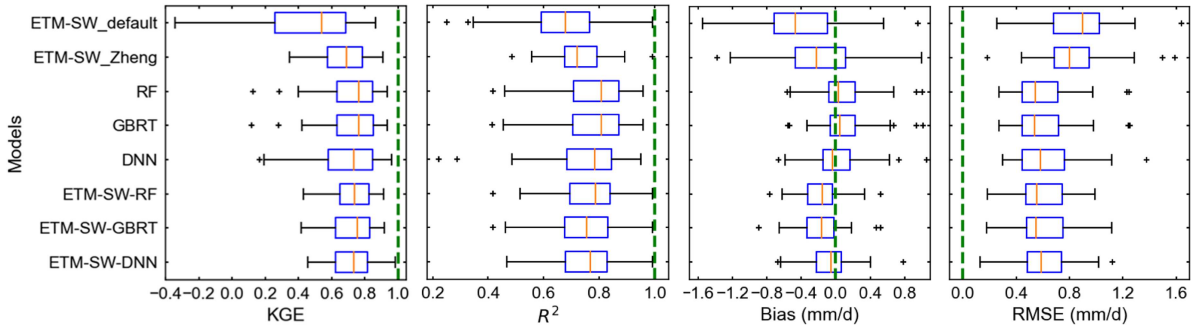

Figure 8. Comparison of ET estimates from three hybrid models (ETM-SW-DNN, ETM-SW-GBRT, ETM-SW-RF), three full ML models (DNN, GBRT, RF) and two full physics-based models (ETM-SW_default and ETM-SW_Zheng) against in-situ measurements. Metrics (left to right): KGE, R^2, bias and RMSE; Units: bias and RMSE in mm/d; R^2 and KGE are unitless. Green dotted lines = best performance of each metric.

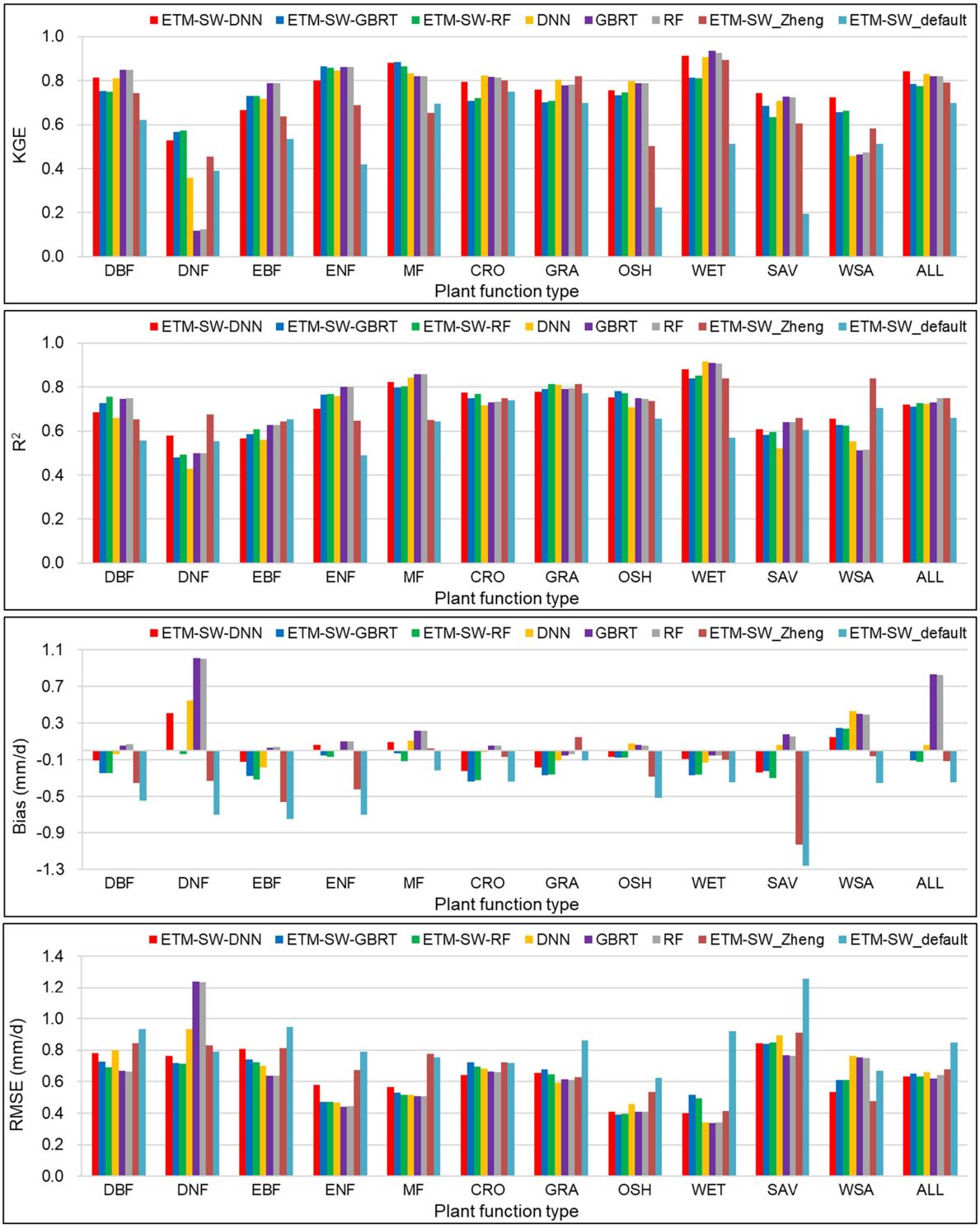

Figure 9. Independent validation of ET estimates from six new models and the two earlier versions of ETMonitor across different plant functional types. All: all plant functional types. Unit: bias and RMSE in mm/d; R^2 and KGE are unitless.

series (during the growing season of one year) from three selected sites per type are shown in Figure 10 to investigate the model's ability to capture seasonal ET variation. For Type 1, ETM-SW-DNN (KGE = 0.79) performed significantly better than other models (KGEs < 0.75) based on KGE results (Figure 10A1). ET estimation at the US-Ton (WSA) site differed significantly among models (Figure 10B1), with ETM-SW-DNN showing better agreement with in-situ observations than the full ML models. For Types 2 and 3, full

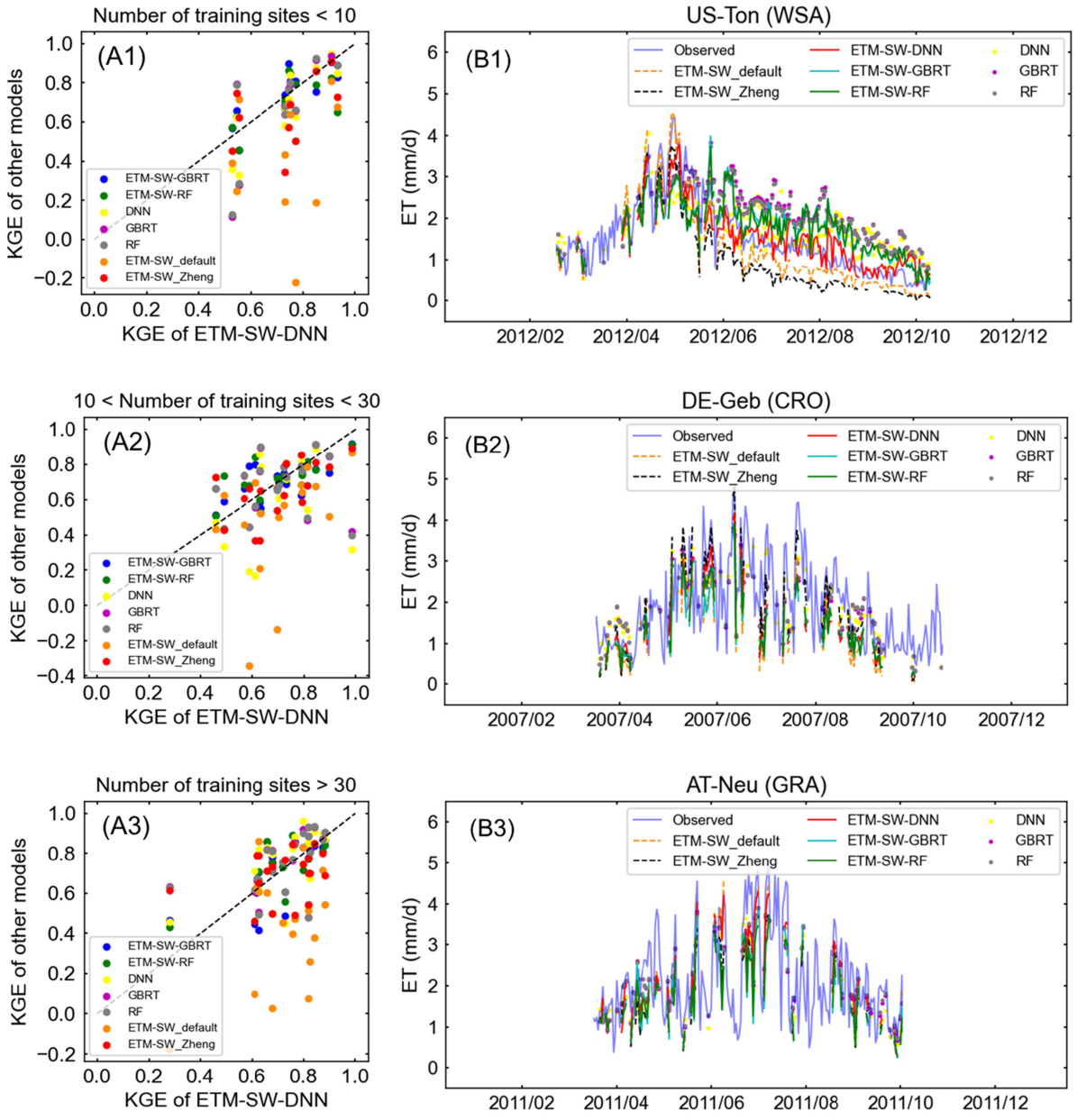

Figure 10. (A) Comparison of model validation results (KGE) between ETM-SW-DNN and other models (ETM-SW-GBRT, ETM-SW-RF, DNN, GBRT, RF, ETM-SW_default and ETM-SW_Zheng) for three site types grouped by abundance of training sites (A1: < 10 training sites; A2: 10–30 training sites; A3: > 30 training sites). (B) Time series of estimated ET from the five models and measured ET at three independent validation sites selected per type, representing different PFTs.

ML models also achieved accurate ET estimation (Figure 10A2, A3), with little difference in the time series of ET estimates among models (Figure 10B2, B3).

5. Discussion

5.1. Advantages of ML-based $r_{s,c}$ estimation and hybrid modelling

This study developed multiple hybrid models by integrating different ML algorithms into the physics-based model ETMonitor, with the core goal of improving the $r_{s,c}$ estimation, thereby improve the accuracy and efficiency of ET estimation. We applied ML algorithms to replace the widely-used Jarvis-type parameterisation for, as $r_{s,c}$ is the most influential parameter for ET estimation using ETMonitor-like models, and existing

parameterisation schemes are plagued by large uncertainties (Hu et al. 2013; Wei et al. 2020; Zhao et al. 2020). The Jarvis-type formula involves numerous parameters, most of which are derived from limited literature or in-situ observations, making it challenging to capture complex climate and vegetation conditions (Li et al. 2015b; Wei et al. 2020). While Zheng, Jia, and Hu (2022) improved ET estimation by calibrating $r_{s,c}$ parameters using in-situ flux data and extrapolating to the global scale by incorporating climate regimes and PFTs, this approach requires extensive expert experience. Additionally, the parameter optimisation is complicated and time consuming, involving thousands of model runs to identify optimal parameter combinations.

ML methods provide a robust solution for $r_{s,c}$ simulation by avoiding uncertainties associated with pre-assigned empirical parameters and parameterisations of environmental stress factors (Li et al. 2015b; Reichstein et al. 2019). ML-based $r_{s,c}$ predictors do not rely on pre-defined functional forms; instead, they use data-driven learners to uncover intrinsic relationships between covariates and target variable ($r_{s,c}$). This capability may deepen insights into unobserved biophysical processes and responses. Moreover, as Earth-system and climate models become increasingly complex, accurately estimating ET using traditional methods requires excessive computational resources. ML methods have acceptable training time requirements, and once trained, ML-based $r_{s,c}$ predictors enable fast and accurate $r_{s,c}$ estimation—an advantage for large-scale or long-term simulations.

Integrating the ML-based $r_{s,c}$ predictor into the physics-based model has the potential to curb over-parameterisation, reduce computation times, and improve accuracy in process representation in hybrid schemes. Previous studies in the field of climate sciences have demonstrated the advantages of hybrid modelling, using ML sub-models to represent different processes (de Bézenac, Pajot, and Gallinari 2019; Rasp, Pritchard, and Gentine 2018) or to improve model parameter discovery (Chen et al. 2022; Koppa et al. 2022; Kraft et al. 2020; Shang et al. 2023; Zhao et al. 2019). Hybrid models can balance the strengths of bio-physical process models and ML methods while overcoming the limitations of both effectively. Combining the physics-based model with a ML method can achieve accurate ET estimation, as well as reducing the need for priori knowledge and computational cost, since this approach extracts information from available observations while respecting bio-physical laws. For certain PFTs (e.g. GRA, CRO, ENF and EBF) with a large amount of training data, the performance of full ML models produced the best results (see Figure 9). However, for PFTs with limited training data, the performance of full ML models decreased significantly due to their excessive dependence on the quantity and quality of training data. This indicates that their performance is unstable for out-of-sample locations. Hybrid models can still achieve relatively stable performance in locations with limited in-situ measurements due to the consideration of physical constraints.

5.2. Selection of ML algorithms and impact of training data

We used three ML algorithms (RF, GBRT and DNN) to estimate $r_{s,c}$ and compared multiple hybrid model versions. The results showed that the hybrid model using DNN performed better than those using RF or GBRT to estimate $r_{s,c}$. Liu et al. (2021) have suggested that artificial neural network (ANN) is more suitable for modelling stomatal conductance in cropland ET estimation than other ML algorithms, such as K-nearest neighbour algorithm, random forest, support vector machine, extreme gradient boosting algorithm, and long short-term memory. Our study produced similar results, as DNNs are ANNs with two or more hidden layers, highlighting the advantages of (deep) neural network algorithms for model parameter retrieval. The main advantage of DNNs is their insensitivity to outliers (Chollet 2017), an important feature that improves the ability and performance of the hybrid model when applied in different situations. While deep learning can achieve a remarkable performance, it should be noted that careful tuning is required to avoid local optimisation, which can lead to overfitting (Ma et al. 2019). In this study, the hybrid models were carefully tuned by selecting appropriate hyperparameters to maximise accuracy and stability, thereby addressing the challenges of overfitting and vanishing/exploding gradients.

In addition, the representativeness of the training data and the selection of explanatory variables have great impact on the ML model performance. Overall, full ML models can achieve comparable accuracy to ETM-SW-DNN in estimating ET. For certain PFTs (e.g. GRA, CRO, ENF and EBF) with a large amount of training data, full ML models gave the best results (Figure 9). However, in PFTs with limited sites, the performance of full ML models decreased significantly due to their excessive dependence on the quantity and quality of training data. This indicates that their performance is unstable for out-of-sample locations. Hybrid models can still achieve relatively

stable performance in locations with limited in-situ measurements due to the consideration of physical constraints. Previous studies suggest that $r_{s,c}$ is mainly affected by vegetation properties (e.g. LAI or NDVI), moisture (e.g. RH and SM) and VPD (Li et al. 2025). A model performance would substantially decrease in the absence of LAI (e.g. Chen et al. 2022; Gan et al. 2018; Leuning et al. 2008). Our study also found that the accuracy of ET estimates decreased significantly when LAI and SM were excluded from the $r_{s,c}$ simulation (Figure 4). Based on the experimental results of different explanatory variable combinations, we adopted seven explanatory variables (including LAI, VPD, Ta, Rs, SM, Pa, and PFT) to estimate $r_{s,c}$. The interpretability analysis results suggest that LAI and VPD play a leading role in $r_{s,c}$ estimation, followed by Ta and SM (Figure 5). This is consistent with the findings of previous studies (Akbar et al. 2019; Douville et al. 2016; Gentine et al. 2012). Furthermore, the contribution of PFT is relatively small in the hybrid and full ML models (Figure 5), and its absence slightly affects model performance (Figure 4). One possible explanation is that the classification information provided by PFT is redundant, as other variables (e.g. LAI and VPD) may partly reflect the differences across PFTs.

5.3. Uncertainties and perspective of this study

There are still some limitations and uncertainties that need further attention despite the hybrid model has shown better performance than other models in ET estimation. The $r_{s,c}$ cannot be observed directly and the reference dataset required for ML training was derived by the inversion of ETM-SW formulation (Equations 1-8) using in-situ measurements of ET and meteorological variables from flux tower sites. Measurements errors will propagate to the $r_{s,c}$ estimates in model inversion and ultimately affect the performance of the hybrid model.

Firstly, uncertainties come from the biophysical framework. In the S-W scheme, the complexity of radiative and convective energy exchanges in the soil and vegetation layer (Gao et al. 2017), and errors in the net radiation flux decomposition (Larsen and Kershaw 1996; Sinclair 2006) can lead to biases. Although the improved ETMonitor model has been reported to provide a more accurate ET estimation, further modifications are required as other parameters (e.g. $r_{s,s}$) significantly affect the model performance in estimating soil evaporation (Chen, Huang, and McBean 2019; Wei et al. 2020). More specifically, measurements of surface soil moisture (SM_{surf}) varied slightly in depths across sites: at some sites the SM_{surf} was measured at 4 cm, while at other sites a slightly deeper depths (5 cm) was used. In the inversion process, SM_{surf} was used to calculate the soil surface resistance, a factor that also influences model performance (Chen, Huang, and McBean 2019; Zheng, Jia, and Hu 2022).

Secondly, uncertainties arise from the input data of the ML models. The target and explanatory variables used for ML model training and calibration were collected from EC flux observations. Due to large-scale eddies and the advection effect, the uncertainty in ET measurements by EC is approximately 5-30% (Foken 2008; Mahrt 2010). The ML model performance largely depends on the quality and representativeness of the training data, which can lead to unpredictable errors and uncertainties in the absence of sufficient EC observations (Denisko and Hoffman 2018). To achieve reliable model estimates, we collected data from as many sites as possible, carefully controlling the data quality and ensuring sufficient representativeness of typical PFTs in the selection of training and validation sites.

Nevertheless, this study clearly demonstrates that the developed hybrid models can accurately estimate ET at site scale. Full ML models also showed a comparable performance in ET simulation, particularly for land cover types with abundant observations. Notably, while regional and global ET benchmarks are well-established (Jung et al. 2019; Koppa et al. 2022; Zhang et al. 2025), our subsequent research will focus on enhancing the hybrid model's capability to produce spatio-temporal continuous ET estimates using satellite remote sensing data. Future work will further involve mapping the spatial variability of terrestrial ET at regional scale and investigating the reliability of these spatial patterns.

6. Conclusion

This study proposes a hybrid evapotranspiration (ET) prediction framework that integrate machine learning (ML)-derived canopy surface resistance ($r_{s,c}$) into the physics-based ETMonitor model. This

replaces the Jarvis-type $r_{s,c}$ parameterisation, which is prone to uncertainty, with DNN, GBRT and RF algorithms. This results in three hybrid models (ETM-SW-RF, ETM-SW-GBRT, ETM-SW-DNN). Three full ML-based models based on DNN, GBRT and RF, were also developed for direct ET estimation, and they were compared with the hybrid models. To validate the framework, the comprehensive evaluations were conducted using in-situ observations from 224 global flux tower sites, which cover diverse ecosystems, climate zones and PFTs. The full ML-based models and hybrid models were trained using data from 179 sites and were independently validated at 45 sites, with additional assessments across PFTs and training data quantity gradients. The key findings are as follows:

(1) The hybrid models outperformed the original physics-based models (ETM-SW_default and ETM-SW_Zheng) with ETM-SW-DNN achieving the best performance with a KGE of 0.84. The full ML models (DNN, GBRT, RF) for ET estimation show comparable performance to the hybrid models.
(2) Hybrid models exhibit superior stability: they mitigate the under- or over-estimation biases commonly found in full ML models, produce fewer outliers and maintain reliable performance for PFTs with limited training data. In contrast, full ML models only excel for PFTs with abundant observations, highlighting their dependence on data quantity and quality.
(3) The hybrid models show better performance when more explanatory variables are used, indicating the importance of using comprehensive input variables in ML training. Model interpretability analysis using SHAP identifies leaf area index (LAI), vapour pressure deficit (VPD), and soil moisture (SM) as the dominant drivers of $r_{s,c}$. This highlights the biological and environmental relevance of the ML-derived $r_{s,c}$ formulation rather than relying on empirical parameterisations and parameters.
(4) Incorporating ML into a physics-based framework strikes a balance between process constraints and data-driven efficiency. Hybrid models reduce computational overhead compared to complex physical parameterisation, minimise reliance on prior knowledge and avoid issues of over-parameterisation.

Future research will focus on integrating satellite remote sensing data to generate continuous spatio-temporal ET products, mapping spatial variability of ET, and validating it against regional/global ET benchmarks. The hybrid framework's ability to combine physical rigour with ML flexibility offers a promising tool for improving ET simulation in Earth system models and supporting water cycle research.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) (Grant No. 42090014, 42171039, 42271394) and Hainan Provincial Natural Science Foundation of China (No. 325MS147). We acknowledge the data support from NASA Earthdata Search (https://search.earthdata.nasa.gov/search). We also acknowledge flux observation networks (FLUXNET, AmeriFlux, EuroFlux, OzFlux, AsiaFlux, ChinaFlux, HiWATER) that provide flux tower datasets to validate our ET estimation.

Author contributions

CRediT: **Yiqing Zhang:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Li Jia:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing; **Chaolei Zheng:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Visualization, Writing – review & editing; **Massimo Menenti:** Methodology, Validation, Visualization, Writing – review & editing; **Guangcheng Hu:** Conceptualization, Methodology, Resources, Validation; **Jing Lu:** Funding acquisition, Methodology, Validation, Visualization, Writing – review & editing; **Qiting Chen:** Methodology, Validation, Visualization, Writing – review & editing.

Disclosure statement

All authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Akbar, R., D. J. Short Gianotti, G. D. Salvucci , and D. Entekhabi. 2019. "Mapped Hydroclimatology of Evapotranspiration and Drainage Runoff Using SMAP Brightness Temperature Observations and Precipitation Information." *Water Resources Research* 55 (April): 3391–3413. https://doi.org/10.1029/2018WR024459.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements. 56Rome: FAO Irrigation and Drainage Paper.

Allen, R. G., L. S. Pereira, T. A. Howell , and M. E. Jensen. 2011. "Evapotranspiration Information Reporting: I. Factors Governing Measurement Accuracy." *Agricultural Water Management* 98 (6): 899–920. https://doi.org/10.1016/j.agwat.2010.12.015.

Allen, R. G., W. O. Pruitt, J. L. Wright, T. A. Howell, F. Ventura, R. Snyder, and M. Smith. 2006. "A Recommendation on Standardized Surface Resistance for Hourly Calculation of Reference ETo By the FAO56 Penman-Monteith Method." *Agricultural Water Management* 8(1): 1–22. https://doi.org/10.1016/j.agwat.2005.03.007.

Anderson, M. C., W. P. Kustas, J. M. Norman, C. R. Hain, J. R. Mecikalski, L. Schultz, M. P. González-Dugo, et al. 2011. "Mapping Daily Evapotranspiration at Field to Continental Scales Using Geostationary and Polar Orbiting Satellite Imagery." *Hydrology and Earth System Sciences* 15 (1): 223–239. https://doi.org/10.5194/hess-15-223-2011.

Ball, J. T., I. E. Woodrow, and J. A. Berry. 1987. A Model Predicting Stomatal Conductance and Its Contribution to the Control of Photosynthesis under Different Environmental Conditions. In *Progress in Photosynthesis Research*4(pp):221–222.https://doi.org/10.1007/978-94-017-0519-6_48

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag. 1998. "A Remote Sensing Surface Energy Balance Algorithm for Land (SEBAL). 1. Formulation." *Journal of Hydrology* 212–213(December): 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. https://doi.org/10.1023/A:1010933404324.

Carter, C. , and S. Liang. 2019. "Evaluation of Ten Machine Learning Methods for Estimating Terrestrial Evapotranspiration from Remote Sensing." *International Journal of Applied Earth Observation and Geoinformation* 78 (6): 86–92. https://doi.org/10.1016/j.jag.2019.01.020.

Chen, H., A. Jiang, J. Huang, H. Li, E. Mcbean, V. P. Singh, J. Zhang, Z. Lan, J. Gao , and Z. Zhou. 2022. "An Enhanced Shuttleworth-Wallace Model for Simulation of Evapotranspiration and Its Components." *Agricultural and Forest Meteorology* 313 (15): 108769. https://doi.org/10.1016/j.agrformet.2021.108769.

Chen, H., J. Huang , and E. McBean. 2019. "Partitioning of Daily Evapotranspiration Using a Modified Shuttleworth-Wallace Model, Random Forest and Support Vector Regression, for a Cabbage Farmland." *Agricultural Water Management* 228 (20): 105923. https://doi.org/10.1016/j.agwat.2019.105923.

Chollet, F.o. 2017. *Deep Learning with Python*. Manning Publications.

Cleugh, H. A., R. Leuning, Q. Mu , and S. W. Running. 2007. "Regional Evaporation Estimates from Flux Tower and MODIS Satellite Data." *Remote Sensing of Environment* 106 (3): 285–304. https://doi.org/10.1016/j.rse.2006.07.007.

de Bézenac, E., A. Pajot , and P. Gallinari. 2019. "Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge." *Journal of Statistical Mechanics: Theory and Experiment* 2019 (December): 124009. https://doi.org/10.1088/1742-5468/ab3195.

Denisko, D. , and M. M. Hoffman. 2018. "Classification and Interaction in Random Forests." *Proceedings of the National Academy of Sciences* 115 (8): 1690–1692. https://doi.org/10.1073/pnas.1800256115.

Douville, H., J. Colin, E. Krug, J. Cattiaux , and S. Thao. 2016. "Midlatitude Daily Summer Temperatures Reshaped By Soil Moisture under Climate Change." *Geophysical Research Letters* 43 (2): 812–818. https://doi.org/10.1002/2015GL066222.

Fan, J., W. Yue, L. Wu, F. Zhang, H. Cai, X. Wang, X. Lu , and Y. Xiang. 2018. "Evaluation of SVM, ELM and Four Tree-Based Ensemble Models for Predicting Daily Reference Evapotranspiration Using Limited Meteorological Data in Different Climates of China." *Agricultural and Forest Meteorology* 263 (15): 225–241. https://doi.org/10.1016/j.agrformet.2018.08.019.

Fang, Q. X., L. Ma, T. Green, Q. Yu, T. Wang , and L. R. Ahuja. 2010. "Water Resources and Water Use Efficiency in the North China Plain: Current Status and Agronomic Management Options." *Agricultural Water Management* 97 (8): 1102–1116. https://doi.org/10.1016/j.agwat.2010.01.008.

Feng, J., W. Wang, T. Che , and F. Xu. 2023. "Performance of the Improved Two-Source Energy Balance Model for Estimating Evapotranspiration over the Heterogeneous Surface." *Agricultural Water Management* 278 (2): 108159. https://doi.org/10.1016/j.agwat.2023.108159.

Feng, Y., N. Cui, D. Gong, Q. Zhang , and Q. Zhao. 2017. "Evaluation of Random Forests and Generalized Regression Neural Networks for Daily Reference Evapotranspiration Modelling." *Agricultural Water Management* 193 (C): 163–173. https://doi.org/10.1016/j.agwat.2017.08.003.

Fisher, J. B., F. Melton, E. Middleton, C. Hain, M. Anderson, R. Allen, M. F. McCabe, et al. 2017. "The Future of Evapotranspiration: Global Requirements for Ecosystem Functioning, Carbon and Climate Feedbacks, Agricultural

Management, and Water Resources." *Water Resources Research* 53 (4): 2618–2626. https://doi.org/10.1002/2016WR020175.
Fisher, J. B., K. Tu , and D. D. Baldocchi. 2008. "Global Estimates of the Land-Atmosphere Water Flux Based on Monthly AVHRR and ISLSCP-II Data, Validated at 16 FLUXNET Sites." *Remote Sensing of Environment* 112 (3): 901–919. https://doi.org/10.1016/j.rse.2007.06.025.
Foken, T. 2008. "The Energy Balance Closure Problem: An Overview." *Ecological applications* 18 (6): 1351–1367. https://doi.org/10.1890/06-0922.1.
Friedman, J. H. 2002. "Stochastic Gradient Boosting." *Computational Statistics & Data Analysis* 38 (4): 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.
Gan, R., Y. Zhang, H. Shi, Y. Yang, D. Eamus, L. Cheng, F. H. S. Chiew , and Q. Yu. 2018. "Use of Satellite Leaf Area Index Estimating Evapotranspira Tion and Gross Assimilation for Australian Ecosystems." *Ecohydrology* 11 (5): e1974. https://doi.org/10.1002/eco.1974.
Gao, Z., H. Liu, G. G. Katul , and T. Foken. 2017. "Non-Closure of the Surface Energy Balance Explained By Phase Difference between Vertical Velocity and Scalars of Large Atmospheric Eddies." *Environmental Research Letters* 12 (3): 034025. https://doi.org/10.1088/1748-9326/aa625b.
Gentine, P., P. D'Odorico, B. R. Lintner, G. Sivandran, and G. Salvucci. 2012. "Interdependence of Climate, Soil, and Vegetation as Constrained By the Budyko Curve." *Geophysical Research Letters* 39(19): L19404. https://doi.org/10.1029/2012GL053492.
Good, S. P., D. Noone , and G. Bowen. 2015. "Hydrologic Connectivity Constrains Partitioning of Global Terrestrial Water Fluxes." Science (New York, N.Y.) 349 (6244): 175–177. https://doi.org/10.1126/science.aaa5931.
Gupta, H. V., H. Kling, K. K. Yilmaz , and G. F. Martinez. 2009. "Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling." *Journal of Hydrology* 377 (1): 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.
Hu, Z., S. Li, G. Yu, X. Sun , and L. Zhang. 2013. "Modeling Evapotranspiration By Combing a Two-Source Model, a Leaf Stomatal Model, and a Light-Use Efficiency Model." *Journal of Hydrology* 501 (25): 186–192. https://doi.org/10.1016/j.jhydrol.2013.08.006.
Hu, X., L. Shi, G. Lin , and L. Lin. 2021. "Comparison of Physical-Based, Data-Driven and Hybrid Modeling Approaches for Evapotranspiration Estimation." *Journal of Hydrology* 601 (October): 126592. https://doi.org/10.1016/j.jhydrol.2021.126592.
Hu, G. , and L. Jia. 2015. "Monitoring of Evapotranspiration in a Semi-Arid Inland River Basin By Combining Microwave and Optical Remote Sensing Observations." *Remote Sensing* 7 (3): 3056–3087. https://doi.org/10.3390/rs70303056.
Ienco, D., R. Interdonato, R. Gaetano , and D. Ho Tong Minh. 2019. "Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for Land Cover Mapping Via a Multi-Source Deep Learning Architecture." *ISPRS Journal of Photogrammetry and Remote Sensing* 158 (1): 11–22. https://doi.org/10.1016/j.isprsjprs.2019.09.016.
Jarvis, P. G.. 1976. "The Interpretation of Variations in Leaf Water Potential and Stomatal Conductance Found in Canopies in the Field." *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 273 (927): 563–610. https://doi.org/10.1098/rstb.1976.0035.
Jia, L., G. Xi, S. Liu, C. Huang, Y. Yan , and G. Liu. 2009. "Regional Estimation of Daily to Annual Regional Evapotranspiration with MODIS Data in the Yellow River Delta Wetland." *Hydrology and Earth System Sciences* 13 (10): 1775–1787. https://doi.org/10.5194/hess-13-1775-2009.
Jiang, L. , and S. Islam. 1999. "A Methodology for Estimation of Surface Evapotranspiration over Large Areas Using Remote Sensing Observations." *Geophysical Research Letters* 26 (17): 2773–2776. https://doi.org/10.1029/1999GL006049.
Jiang, L. , and S. Islam. 2001. "Estimation of Surface Evaporation Map over Southern Great Plains Using Remote Sensing Data." *Water Resources Research* 37 (2): 329–340. https://doi.org/10.1029/2000WR900255.
Jiang, S., L.-b. Sweet, G. Blougouras, A. Brenning, W. Li, M. Reichstein, J. Denzler, et al. 2024. "How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences." *Earth's Future* 12(7): e2024EF004540. https://doi.org/10.1029/2024EF004540.
Jung, M., M. Reichstein, H. A. Margolis, A. Cescatti, A. Richardson, A. Arain, A. Arneth, C. Bernhofer, D. Bonal , and J. Chen. 2011. "Global Patterns of Land-Atmosphere Fluxes of Carbon Dioxide, Latent Heat, and Sensible Heat Derived from Eddy Covariance, Satellite, and Meteorological Observations." *Journal of Geophysical Research: Biogeosciences* 116 (September): G00J07. https://doi.org/10.1029/2010JG001566.
Jung, M., S. Koirala, U. Weber, K. Ichii , and M. Reichstein. 2019. "The FLUXCOM Ensemble of Global Land-Atmosphere Energy Fluxes." *Scientific Data* 6 (1): 74. https://doi.org/10.1038/s41597-019-0076-8.
Karpatne, A., G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova , and V. Kumar. 2017. "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data." *IEEE Transactions on Knowledge and Data Engineering* 29 (10): 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168.
Koppa, A., D. Rains, P. Hulsman, R. Poyatos , and D. G. Miralles. 2022. "A Deep Learning-Based Hybrid Model of Global Terrestrial Evaporation." *Nature Communication* 13 (1): 1912. https://doi.org/10.1038/s41467-022-29543-7.
Kraft, B., M. Jung, M. Körner , and M. Reichstein. 2020. "Hybrid Modeling: Fusion of a Deep Learning Approach and a Physics-Based Model for Global Hydrological Modeling." *The International Archives of Photogrammetry,*

Remote Sensing and Spatial Information Sciences XLIII-B2-2020 (August): 1537–1544. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1537-2020.

Larsen, D. R. , and J. A. Kershaw. 1996. "Influence of Canopy Structure Assumptions on Predictions from Beer's Law. a Comparison of Deterministic and Stochastic Simulations." *Agricultural and Forest Meteorology* 81 (1): 61–77. https://doi.org/10.1016/0168-1923(95)02307-0.

Leuning, R., Y. Q. Zhang, A. Rajaud, H. Cleugh, and K. Tu. 2008. "A Simple Surface Conductance Model to Estimate Regional Evaporation Using Modis Leaf Area Index and the Penman-Monteith Equation." *Water Resources Research* 4(1): W10419. https://doi.org/10.1029/2007wr006562.

Li, Z., L. Jia , and J. Lu. 2015a. "On Uncertainties of the Priestley-Taylor/LST-Fc Feature Space Method to Estimate Evapotranspiration: Case Study in an Arid/Semiarid Region in Northwest China." *Remote Sensing* 7 (1): 447–466. https://doi.org/10.3390/rs70100447.

Li, S., L. Zhang, S. Kang, L. Tong, T. Du, X. Hao , and P. Zhao. 2015b. "Comparison of Several Surface Resistance Models for Estimating Crop Evapotranspiration over the Entire Growing Season in Arid Regions." *Agricultural and Forest Meteorology* 208 (5): 1–15. https://doi.org/10.1016/j.agrformet.2015.04.002.

Li, Z., Q. Yuan, L. Yue, H. Shen , and L. Zhang. 2025. "Large-Scale Soil Moisture Retrieval under Limited in Situ Data Coupling Physical and Machine Learning Model." *Journal of Remote Sensing* 5: 0367. https://doi.org/10.34133/remotesensing.0367.

Liang, S., D. Wang, T. He , and Y. Yu. 2019. "Remote Sensing of Earth'S Energy Budget: Synthesis and Review." *International Journal of Digital Earth* 12 (7): 737–780. https://doi.org/10.1080/17538947.2019.1597189.

Liu, Y., S. Zhang, J. Zhang, L. Tang , and Y. Bai. 2021. "Assessment and Comparison of Six Machine Learning Models in Estimating Evapotranspiration over Croplands Using Remote Sensing and Meteorological Factors." *Remote Sensing* 13 (19): 3838. https://doi.org/10.3390/rs13193838.

Liu, S., L. Xin, Z. Xu, and CTao Ren. 2018. "The Heihe Integrated Observatory Network: a Basin-Scale Land Surface Processes Observatory in China." *Vadose Zone Journal* 17(1): 1–21. https://doi.org/10.2136/vzj2018.04.0072.

Long, D., and V. P. Singh. 2012. "A Modified Surface Energy Balance Algorithm for Land (M-SEBAL) Based on a Trapezoidal Framework." *Water Resources Research* 48(2): W02528. https://doi.org/10.1029/2011WR010607.

Loveland, T. R., Z. L. Zhu, D. O. Ohlen, J. F. Brown , and L. M. Yang. 1999. "An Analysis of IGBP Global Land-Cover Characterization Process." *Photogrammetric Engineering and Remote Sensing* 65 (9): 1021–1032.

Lu, X., Y. Ju, L. Wu, J. Fan, F. Zhang , and Z. Li. 2018. "Daily Pan Evaporation Modeling from Local and Cross-Station Data Using Three Tree-Based Machine Learning Models." *Journal of Hydrology* 566 (2): 668–684. https://doi.org/10.1016/j.jhydrol.2018.09.055.

Lundberg, S. M., G. G. Erion, and S. - I. J. A. Lee. 2018. "Consistent Individualized Feature Attribution for Tree Ensembles." *arXiv: Learning*, pp (February):154–196. https://doi.org/10.48550/arXiv.1802.03888.

Ma, L., Y. Liu, X. Zhang, Y. Ye, G. Yin , and B. A. Johnson. 2019. "Deep Learning in Remote Sensing Applications: a Meta-Analysis and Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 152 (June): 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Mahrt, L.. 2010. "Computing Turbulent Fluxes Near the Surface: Needed Improvements." *Agricultural and Forest Meteorology* 150 (4): 501–509. https://doi.org/10.1016/j.agrformet.2010.01.015.

Medlyn, B. E., R. A. Duursma, D. Eamus, D. S. Ellsworth, I. C. Prentice, C. V. M. Barton, K. Y. Crous, P. de Angelis, M. Freeman , and L. Wingate. 2011. "Reconciling the Optimal and Empirical Approaches to Modelling Stomatal Conductance." *Global Change Biology* 17 (6): 2134–2144. https://doi.org/10.1111/j.1365-2486.2010.02375.x.

Menenti, M. , and B. J. Choudhury. 1993. "Parameterization of Land Surface Evaporation By Means of Location Dependent Potential Evaporation and Surface Temperature Range. Exchange Processess at the Land Surface for a Range of Space and Time Scales." *IAHS Publication* 212 (January): 561–568. https://doi.org/10.1007/978-1-4612-1494-6.

Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters , and A. J. Dolman. 2011. "Global Land-Surface Evaporation Estimated from Satellite-Based Observations." *Hydrology and Earth System Sciences* 15 (2): 453–469. https://doi.org/10.5194/hess-15-453-2011.

Mo, X., S. Liu, Z. Lin , and W. Zhao. 2004. "Simulating Temporal and Spatial Variation of Evapotranspiration over the Lushi Basin." *Journal of Hydrology* 285 (1): 125–142. https://doi.org/10.1016/j.jhydrol.2003.08.013.

Moran, M. S., T. R. Clarke, Y. Inoue , and A. Vidal. 1994. "Estimating Crop Water Deficit Using the Relationship between surface–air Temperature and Spectral Vegetation Index." *Remote Sensing of Environment* 49 (3): 246–363. https://doi.org/10.1016/0034-4257(94)90020-5.

Morillas, L., R. Leuning, L. Villagarcía, M. García, P. Serrano-Ortiz , and F. Domingo. 2013. "Improving Evapotranspiration Estimates in Mediterranean Drylands: the Role of Soil Evaporation." *Water Resources Research* 49 (10): 6572–6586. https://doi.org/10.1002/wrcr.20468.

Mu, Q., F. A. Heinsch, M. Zhao , and S. W. Running. 2007. "Development of a Global Evapotranspiration Algorithm Based on MODIS and Global Meteorology Data." *Remote Sensing of Environment* 111 (4): 519–536. https://doi.org/10.1016/j.rse.2007.04.015.

Mu, Q., M. Zhao , and S. W. Running. 2011. "Improvements to a MODIS Global Terrestrial Evapotranspiration Algorithm." *Remote Sensing of Environment* 115 (8): 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019.

Norman, J. M., W. P. Kustas , and K. S. Humes. 1995. "Source Approach for Estimating Soil and Vegetation Energy Fluxes in Observations of Directional Radiometric Surface Temperature." *Agricultural and Forest Meteorology* 77 (3–4): 263–293. https://doi.org/10.1016/0168-1923(95)02265-Y.

Otkin, J. A., M. C. Anderson, C. Hain, M. Svoboda, D. Johnson, R. Mueller, T. Tadesse, B. Wardlow , and J. Brown. 2016. "Assessing the Evolution of Soil Moisture and Vegetation Conditions during the 2012 United States Flash Drought." *Agricultural and Forest Meteorology* 218-219 (March): 230–242. https://doi.org/10.1016/j.agrformet.2015.12.065.

Priestley, C., and R. Taylor. 1972. "On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters." *Monthly Weather Review* 100(2): 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.

Rasp, S., M. S. Pritchard , and P. Gentine. 2018. "Deep Learning to Represent Subgrid Processes in Climate Models." *Proceedings of the National Academy of Sciences of the United States of America* 115 (39): 9684–9689. https://doi.org/10.1073/pnas.1810286115.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais , and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." Nature (London) 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Saggi, M. K. , and S. Jain. 2019. "Reference Evapotranspiration Estimation and Modeling of the Punjab Northern India Using Deep Learning." *Computers and Electronics in Agriculture* 156 (January): 387–398. https://doi.org/10.1016/j.compag.2018.11.031.

Shang, K., Y. Yao, Z. Di, K. Jia, X. Zhang, J. B. Fisher, J. Chen, et al. 2023. "Coupling Physical Constraints with Machine Learning for Satellite-Derived Evapotranspiration of the Tibetan Plateau." *Remote Sensing of Environment* 289 (May): 113519. https://doi.org/10.1016/j.rse.2023.113519.

Shuttleworth, W. J. , and J. S. Wallace. 1985. "Evaporation from Sparse Crops-An Energy Combination Theory." *Quarterly Journal of the Royal Meteorological Society* 111 (469): 839–855. https://doi.org/10.1002/qj.49711146910.

Sinclair, T. R.. 2006. "A Reminder of the Limitations in Using Beer's Law to Estimate Daily Radiation Interception By Vegetation." *Crop Science* 46 (6): 2343–2347. https://doi.org/10.2135/cropsci2006.01.0044.

Stewart, J. B.. 1988. "Modelling Surface Conductance of Pine Forest." *Agricultural and Forest Meteorology* 43 (1): 19–35. https://doi.org/10.1016/0168-1923(88)90003-2.

Su, Z.. 2002. "The Surface Energy Balance System (SEBS) for Estimation of Turbulent Heat Flux." *Hydrology and Earth System Sciences* 6 (1): 85–99. https://doi.org/10.5194/hess-6-85-2002.

Tang, R., Z. -L. Li , and B. Tang. 2010. "An Application of the Ts–VI Triangle Method with Enhanced Edges Determination for Evapotranspiration Estimation from MODIS Data in Arid and Semi-Arid Regions: Implementation and Validation." *Remote Sensing of Environment* 114 (3): 540–551. https://doi.org/10.1016/j.rse.2009.10.012.

Wang, K., Z. Li , and M. Cribb. 2006. "Estimation of Evaporative Fraction from a Combination of Day and Night Land Surface Temperatures and NDVI: a New Method to Determine the Priestley–Taylor Parameter." *Remote Sensing of Environment* 102 (3–4): 293–305. https://doi.org/10.1016/j.rse.2006.02.007.

Wang, J., B. Xue, Y. Wang, Y. A. G. Wang, D. Long, and J. Huang. 2024. "Estimates of the Priestley-Taylor Coefficient Based on FLUXNET Data at Multiple Spatiotemporal Scales." *Journal of Hydrology* 629(4): 130636. https://doi.org/10.1016/j.jhydrol.2024.130636.

Wang, X., B. Gao , and X. -S. Wang. 2022. "Investigating the Ability of Deep Learning on Actual Evapotranspiration Estimation in the Scarcely Observed Region." *Journal of Hydrology* 607 (8): 127506. https://doi.org/10.1016/j.jhydrol.2022.127506.

Wang, L., Y. Zhang, Y. Yao, Z. Xiao, K. Shang, X. Guo, J. Yang, S. Xue , and J. Wang. 2021. "GBRT-Based Estimation of Terrestrial Latent Heat Flux in the Haihe River Basin from Satellite and Reanalysis Datasets." *Remote Sensing* 13 (6): 1054. https://doi.org/10.3390/rs13061054.

Wei, G., L. Zhou, Hw Liu, Q. Tian, L. Ding , and X. Ran. 2020. "Improving Evapotranspiration Model Performance By Treating Energy Imbalance and Interaction." *Water Resources Research* 56 (9): e2020WR027367. https://doi.org/10.1029/2020WR027367.

Xiao, Z., S. Liang, J. Wang, P. Chen, X. Yin, L. Zhang , and J. Song. 2014. "Use of General Regression Neural Networks for Generating the GLASS Leaf Area Index Product from Time-Series MODIS Surface Reflectance." *IEEE Transactions on Geoscience and Remote Sensing* 52 (1): 209–223. https://doi.org/10.1109/TGRS.2013.2237780.

Xu, Z., S. Liu, Z. Zhu, J. Zhou , and X. He. 2020. "Exploring Evapotranspiration Changes in a Typical Endorheic Basin through the Integrated Observatory Network." *Agricultural and Forest Meteorology* 290 (1): 108010. https://doi.org/10.1016/j.agrformet.2020.108010.

Yang, Y. , and S. Shang. 2013. "A Hybrid Dual-Source Scheme and Trapezoid Framwork Based Evapotranspiration Model (HTEM) Using Satellite Images: Algorithm and Model Test." *Journal of Geophysical Research Atmospheres* 118 (5): 2284–2300. https://doi.org/10.1002/jgrd.50259.

Yang, Y., D. Long, H. Guan, W. Liang, C. Simmons , and O. Batelaan. 2015. "Comparison of Three Dual-Source Remote Sensing Evapotranspiration Models during the MUSOEXE-12 Campaign: Revisit of Model Physics." *Water Resources Research* 51 (5): 3145–3165. https://doi.org/10.1002/2014WR015619.

Yao, Y., S. Liang, J. Yu, S. Zhao, Y. Lin, K. Jia, X. Zhang, et al. 2017. "Differences in Estimating Terrestrial Water Flux from Three Satellite-Based Priestley-Taylor Algorithms." *International Journal of Applied Earth Observation and Geoinformation* 56 (April): 1–12. https://doi.org/10.1016/j.jag.2016.10.009.

Yin, L., F. Tao, Y. Chen, F. Liu , and J. Hu. 2021. "Improving Terrestrial Evapotranspiration Estimation Across China during 2000-2018 with Machine Learning Methods." *Journal of Hydrology* 600 (4): 126538. https://doi.org/10.1016/j.jhydrol.2021.126538.

Yuan, Q., H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, et al. 2020. "Deep Learning in Environmental Remote Sensing: Achievements and Challenges." *Remote Sensing of Environment* 241 (11): 111716. https://doi.org/10.1016/j.rse.2020.111716.

Zhang, K., G. Zhu, J. Ma, Y. Yang, S. Shang , and C. Gu. 2019. "Parameter Analysis and Estimates for the MODIS Evapotranspiration Algorithm and Multiscale Verification." *Water Resources Research* 55 (3): 2211–2231. https://doi.org/10.1029/2018WR023485.

Zhang, L., M. Marshall, A. Nelson, and A. Wrieling. 2021. "A Global Assessment of PT-JPL Soil Evaporation in Agroecosystems with Optical, Thermal, and Microwave Satellite Data." *Agricultural and Forest Meteorology* 306(6): 108455. https://doi.org/10.1016/j.agrformet.2021.108455.

Zhang, Y., M. G. Schaap , and Y. Zha. 2018. "A High-Resolution Global Map of Soil Hydraulic Properties Produced By a Hierarchical Parameterization of a Physically Based Water Retention Model." *Water Resources Research* 54 (5): 9774–9790. https://doi.org/10.1029/2018WR023539.

Zhang, L., A. Vrieling, M. Marshall , and A. Nelson. 2025. "A Macroscale Evapotranspiration Benchmark Based on Spatial and Temporal Enhancements to the Budyko Framework." *Journal of Hydrology* 661 (A): 133625. https://doi.org/10.1016/j.jhydrol.2025.133625.

Zhao, T., S. Wang, C. Ouyang, M. Chen, C. Liu, J. Zhang, and L. Yu, et al. 2024. "Artificial Intelligence for Geoscience: Progress, Challenges, and Perspectives." *The Innovation* 5(5): 100691. https://doi.org/10.1016/j.xinn.2024.100691.

Zhao, W., G. Qiu, Y. Xiong, T. P. U. Kyaw, and B. Chen. 2020. "Uncertainties Caused By Resistances in Evapotranspiration Estimation Using High-Density Eddy Covariance Measurements." *Journal of Hydrometeorology* 21(6): 1349–1365. https://doi.org/10.1175/jhm-d-19-0191.1.

Zhao, W., P. Gentine, M. Reichstein, Y. Zhang, S. Zhou, Y. Wen, C. Lin, X. Li , and G. Qiu. 2019. "Physics-Constrained Machine Learning of Evapotranspiration." *Geophysical Research Letters* 46 (24): 14496–14507. https://doi.org/10.1029/2019GL085291.

Zheng, C., L. Jia, G. Hu , and J. Lu. 2019. "Earth Observations-Based Evapotranspiration in Northeastern Thailand." *Remote Sensing* 11 (2): 138. https://doi.org/10.3390/rs11020138.

Zheng, C., L. Jia , and G. Hu. 2022. "Global Land Surface Evapotranspiration Monitoring By ETMonitor Model Driven By Multi-Source Satellite Earth Observations." *Journal of Hydrology* 613 (B): 128444. https://doi.org/10.1016/j.jhydrol.2022.128444.