



A hybrid Penman-Monteith and machine learning model for simulating evapotranspiration and its components

Han Chen ^{a,b,c,*}, Stephen Good ^{a,b,*}, Kelly Caylor ^d, Richard P. Fiorella ^e, Lixin Wang ^f

^a Biological & Ecological Engineering, Oregon State University (OSU), Corvallis, OR, USA

^b Water Resources Graduate Program, Oregon State University, Corvallis, OR 97331, USA

^c Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, 300072 Tianjin, China

^d Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, CA 93016, USA

^e Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

^f Department of Earth and Environmental Sciences, Indiana University Indianapolis, Indianapolis, IN 46202, USA

ARTICLE INFO

This manuscript was handled by Amilcare Porporato, Editor-in-Chief, with the assistance of Jun Yin, Associate Editor

Keywords:

Hybrid dual-source evapotranspiration model
Surface conductance
Penman-Monteith
Residual neural network
NEON flux sites

ABSTRACT

Integrating physical processes with machine learning has advanced evapotranspiration (*ET*) simulation, yet most hybrid models fail to partition total *ET* into its components: soil evaporation (*E*) and vegetation transpiration (*T*). This study introduces Residual Neural Network–Penman–Monteith (RNN-PM), a novel hybrid dual-source *ET* model designed to overcome this limitation. The model synergizes the physically-based Penman–Monteith framework with three specialized residual neural networks trained to estimate key conductance parameters (canopy conductance, soil surface conductance, and aerodynamic conductance). This explicit parameterization allows for the direct partitioning of total *ET*. Validation at National Ecological Observatory Network (NEON) flux sites using high-frequency partitioned *E* and *T* shows that RNN-PM reliably reproduces *ET* and the transpiration fraction (*T/ET*). For *ET*, the model achieves an average Kling–Gupta efficiency (KGE) of 0.89 and a root-mean-square error (RMSE) of 0.55 mm/day; for *T/ET*, the KGE is 0.87 with an RMSE of 0.06. Furthermore, RNN-PM demonstrates robust generalization, accurately simulating *ET* and its components well beyond the initial training dataset, even under extreme climatic conditions. This study extended the analysis by comparing the RNN-PM model with seven established dual-source *ET* models. The results indicate that RNN-PM outperforms both conventional machine learning models and purely physical process-based models in simulating *ET* components in most cases. Among the purely physical process-based dual-source models, those based on surface temperature decomposition showed improved performance as the leaf area index (*LAI*) decreased when evaluated against high-frequency *ET* component datasets. In contrast, the performance of conductance-based dual-source models declined with decreasing *LAI*. Although purely machine learning-based models can produce relatively accurate simulations of *ET* components, they often exhibit limited generalization capability, an issue that the RNN-PM model effectively overcomes. Ultimately, the RNN-PM model represents a significant advance in simulating *ET* components, offering a novel and scalable approach for improving the representation of land–atmosphere interactions in Earth system models.

1. Introduction

Evapotranspiration (*ET*), which comprises soil evaporation (*E*) and vegetation transpiration (*T*), represents a critical link between the hydrological cycle and the surface energy balance (Wang & Dickinson, 2012; Yang et al., 2023). Differentiating between *E* and *T* helps in assessing plant health, soil moisture status, and atmosphere-land interactions (Lee et al., 2012; Kool et al., 2014; Anderegg et al., 2019;

Nelson et al., 2020; Massari et al., 2021, 2022). The partitioning of *ET* into individual components is also essential for informed water management strategies, enhancing agricultural practices, and adapting to a higher vapor pressure deficit (VPD) atmosphere, making it a significant focus in environmental and agricultural sciences (Wei et al., 2017).

ET models based on physical process mechanisms are generally categorized into single-source and dual-source energy balance models (Timmermans et al., 2007; Bich Ngoc Tran et al., 2023). Dual-source

* Corresponding authors at: Biological & Ecological Engineering, Oregon State University (OSU), Corvallis, OR, USA.

E-mail addresses: han.chen@oregonstate.edu (H. Chen), stephen.good@oregonstate.edu (S. Good).

models are particularly valued in the research community for their capacity to separately estimate *ET* components (Yang et al., 2015; Kool et al., 2021; Jaafar et al., 2022). Most dual-source energy balance models utilize conductance parameters to represent the turbulent diffusion of water vapor from vegetation and soil surfaces to a reference height (e.g., Kustas et al., 2012; Aouade et al., 2019; Chen et al., 2022, 2023). These parameters typically include aerodynamic conductance, canopy conductance, and soil surface conductance (Liu et al., 2007; Kustas et al., 2012; Chen et al., 2023; Mu et al., 2007, 2011). Accurately estimating the conductance parameters is crucial for the successful simulation of *ET* and its components using dual-source energy balance models. However, reliably estimating conductance parameters presents significant challenges due to the complex turbulent diffusion processes within the atmospheric boundary layer (ABL) (Van Ulden & Holtslag, 1985; Chen et al., 2025). For instance, the estimation of aerodynamic conductance is complicated by the parameterization of thermal and momentum roughness lengths (Chappell et al., 2010). Similarly, the parameterization of stomatal conductance presents considerable challenges due to the need to incorporate various environmental stress factors, such as temperature, soil moisture, and vapor pressure deficit, which influence stomatal behavior (Farquhar & Sharkey, 1982). Classical stomatal conductance models, such as the Jarvis model, represent environmental variables as scalar functions ranging from 0 to 1 to quantify their influence on stomatal conductance (Jarvis, 1976). However, these models often neglect the interactive effects among different environmental factors (Ball et al., 1987; Medlyn et al., 2011). Consequently, the uncertainties associated with conductance parameters in the traditional two-source models can significantly affect the accuracy of the *ET* and components predictions.

In recent years, machine learning has been widely applied to the simulation of key hydrological, meteorological, and geophysical parameters. Its strong nonlinear fitting capabilities allow for accurate prediction of surface variables governed by complex physical processes (e.g., Torres et al., 2011; Pan et al., 2020; Chen et al., 2022). The distinct advantage of machine learning models lies in their independence from physical assumptions and empirical parameter settings, markedly reducing simulation uncertainties (Reichstein et al., 2019; Nourani et al., 2019, 2025). Therefore, the application of machine learning to simulate conductance parameters in traditional *ET* models, coupled with its integration into dual-source energy balance frameworks, represents a novel and promising approach for improving the accuracy of *ET* simulations. To date, several studies have developed *ET* models that couple physical process representations with machine learning techniques (e.g., Zhao et al., 2019; Shang et al., 2023). However, although these hybrid models are effective in simulating total *ET*, they are unable to partition *ET* into its individual components. For example, Zhao et al. (2019) developed a hybrid *ET* model that integrates the Penman equation with neural network techniques. Validation conducted at 82 FLUXNET sites worldwide demonstrated that the hybrid model outperforms purely machine learning-based approaches in simulating *ET*. Similarly, Koppa et al. (2022) proposed a global land surface *ET* model that combines the Priestley-Taylor model with deep learning methodologies. Shang et al. (2023) extended this approach by coupling the LightGBM model with the Priestley-Taylor model for *ET* estimation over the Tibetan Plateau. Furthermore, Chen et al. (2023b) developed a deep learning model informed by atmospheric turbulence processes and subsequently employed this hybrid model to assess long-term trends in global land surface *ET*. While previous studies have examined the integration of machine learning with physical process models for regional *ET* simulation, research on the development of hybrid models to estimate individual *ET* components remains limited (Rong et al., 2024).

Prominent dual-source models include those based on temperature decomposition processes, such as the Two-Source Energy Balance (TSEB) model (Kustas et al., 2012) and the feature space model integrating land surface temperature with vegetation coverage (*LST-fc*) (Chen et al., 2020, 2023), as well as models founded on surface

conductance processes, including the MOD16 algorithm derived from the Penman-Monteith (P-M) equation (Mu et al., 2007, 2011). Another category comprises dual-source *ET* models operating on semi-empirical principles, notably represented by the Priestley-Taylor Jet Propulsion Laboratory (PT-JPL) model (Fisher et al., 2008). Within this classification, both the TSEB and feature space models utilize *LST* as an indicator of surface moisture conditions, facilitating the disaggregation of *ET* components by distinguishing between vegetation surface temperature (T_v) and soil surface temperature (T_s). The dual-source model based on the P-M equation achieves component separation by integrating the energy balances of vegetation and soil with turbulent diffusion processes occurring within the vegetation-soil-atmosphere continuum (Mu et al., 2007, 2011). Conversely, the PT-JPL model employs empirical PT coefficients to quantify the surface moisture of both vegetation and soil surfaces, thereby achieving effective *ET* partitioning (Marshall et al., 2020). Efforts have been made to evaluate the simulation performance of existing dual-source models for *ET* and its components (e.g., Colaizzi et al., 2014; Guzinski et al., 2015; Yang et al., 2018; Chen et al., 2021). However, comprehensive comparative analyses across diverse vegetation and climatic conditions remain limited. Consequently, this gap impedes the establishment of a scientifically robust foundation necessary for the informed selection of dual-source models in practical applications.

To address these challenges, this study aims to: 1) develop and validate a hybrid dual-source *ET* model that integrates machine learning with physical process representations to enable robust partitioning of *ET* components, namely the Residual Neural Network-Penman-Monteith (RNN-PM) model; 2) compare the performance of this newly developed dual-source hybrid *ET* model with several traditional dual-source *ET* models in simulating *ET* and its components under varying vegetation and climatic conditions; and 3) elucidate the physical mechanisms underlying performance differences among the various dual-source models.

2. Materials and methods

2.1. Formulation of hybrid two source energy balance model

The proposed hybrid dual-source model employs the P-M equation as the framework for its physical processes. The classic P-M model accounts for the influences of surface energy balance and atmospheric turbulence processes in *ET* simulation (Penman, 1948; Monteith, 1965). The expressions for estimating *ET* and components using the P-M model are shown in Eqs. (1)–(3) (Mu et al., 2007; 2011).

$$ET = E + T \quad (1)$$

$$T = \frac{\Delta R_{nc} f_c + \rho C_p VPD G_a}{\Delta + \gamma(1 + G_a/G_c)} \quad (2)$$

$$E = \frac{\Delta(R_{ns} - G)(1 - f_c) + \rho C_p VPD G_s}{\Delta + \gamma G_a/G_s} \quad (3)$$

In Eqs. (1)–(3), *ET*, *E*, and *T* denote total evapotranspiration, soil evaporation, and vegetation transpiration (W/m^2), respectively. f_c is the vegetation coverage (unitless). The energy balance includes the ground heat flux (*G*, W/m^2) and net radiation flux (R_n , W/m^2). Key thermodynamic variables are the air density (ρ), the specific heat of air (C_p , 1004 J °C/kg), the psychrometric constant (γ), the slope of the saturated vapor pressure curve (Δ , Pa/K), and the vapor pressure deficit (VPD, Pa). R_{nc} and R_{ns} denote the net radiation flux for canopy and soil surface (W/m^2), respectively. R_{nc} is estimated using the formula: $R_{nc} = R_n(1 - \exp(-k_a LAI))$, while R_{ns} is determined by $R_{ns} = R_n \exp(-k_a LAI)$ (Norman et al., 1995). Here, R_n represents the net radiant flux (W/m^2) and *LAI* represents the leaf area index (m^2/m^2). The extinction coefficient (k_a) is assumed to be 0.8. This value is widely accepted in literature as a representative mean for the diverse vegetation types, approximating the

typical light interception efficiency of moderate-to-dense canopies (e.g., Campbell, 1986; Norman et al., 1995; Leuning et al., 2008; Chen et al., 2022). Furthermore, to verify the rationality of this setting, we conducted a sensitivity analysis (Fig. S2), which demonstrated that the RNN-PM model performance remains stable and robust around $k_a = 0.8$. Finally, G_a , G_c , and G_s define the aerodynamic, canopy, and soil surface conductances ($\text{mol}/\text{m}^2/\text{s}$), respectively. To ensure consistency with hydrological applications, the calculated instantaneous latent heat fluxes (W/m^2) at the half-hourly time step were converted to equivalent water depths (mm) using Eq. (4):

$$ET_{mm} = \frac{\lambda ET_W/m^2}{\lambda \rho_w} \times t_{step} \quad (4)$$

where λ is the latent heat of vaporization (J/kg); ρ_w is the density of water ($1000 \text{ kg}/\text{m}^3$) and t_{step} is the time interval seconds (1800 s for half-hourly data). These half-hourly values were then aggregated to daily totals (mm/day).

In applying the P-M model to estimate ET , accurately determining the three conductance parameters (G_a , G_s , and G_c) is both critically important and challenging. Precise estimation of these parameters necessitates the calculation of complex ABL variables, such as momentum roughness length and thermal roughness length (Penman, 1948; Monteith, 1965; Mu et al., 2007, 2011). This complexity introduces significant challenges and uncertainties into the estimation of conductance parameters and, consequently, the subsequent estimation of ET and its components. To tackle this challenge, the hybrid model framework employs three sub-residual neural networks (RNNs) to estimate the three conductance parameters (G_a , G_s , and G_c) in the P-M model. The estimated conductance parameters are then integrated into the P-M model for calculating ET and its components. The rationale for adopting this tripartite architecture instead of a single unified network lies in the distinct physical mechanisms and driving variables associated with each conductance parameter. By employing three specialized sub-networks, the hybrid model ensures that each conductance is estimated solely based on its physically relevant predictors, as detailed in Table 1. The designed process decoupling strategy minimizes the interference of irrelevant features, such as the potential for LAI to erroneously influence G_a , thereby enhancing the physical interpretability of the hybrid model.

The optimization of the hybrid RNN-PM model is governed by observational constraints from an eddy covariance (EC) system, specifically data for total ET and the transpiration fraction (T/ET). The model's calibration objective is to concurrently minimize simulation errors for both ET and T/ET . This is achieved by minimizing a composite loss function, defined as the aggregated deviation between predicted and measured values of these two variables. These predicted G_a , G_s and G_c serve as inputs for Eqs. (2)–(3) to compute the E and T fluxes, which are then summed via Eq. (1) to yield the total ET . During training, the internal weights (w) and biases (b) of the three sub-networks are iteratively adjusted to refine these conductance estimates. Convergence is achieved when the composite loss function is minimized. The model outputs for ET and its components are considered final only upon reaching this minimum; otherwise, the iterative adjustment of parameters continues. To balance the trade-off between the ET and T/ET simulations without expensive manual tuning, we adopted the multi-task learning loss function based on homoscedastic uncertainty (Kendall & Gal, 2018). The composite loss function is defined as Eq. (5).

$$\eta = \frac{1}{2\sigma_1^2}\eta_{ET} + \frac{1}{2\sigma_2^2}\eta_{T/ET} + \log(\sigma_1) + \log(\sigma_2) \quad (5)$$

where η_{ET} and $\eta_{T/ET}$ are the root-mean-square-error (RMSE) for the ET and T/ET predictions, respectively. The parameters σ_1 and σ_2 represent the task-dependent observation noise for each task. During the training process, σ_1 and σ_2 are treated as learnable parameters and optimized simultaneously with the network weights. This formulation automatically decreases the contribution of the task with higher uncertainty to

Table 1

Summary of input datasets of the three sub-RNNs (Input variables, Abbreviation, Unit, and Data source).

Subnetwork	Input variables	Abbreviation	Unit	Data source
RNN _a	Air temperature	T _a	K	In-situ observations
	Wind speed	u	m/s	In-situ observations
	Vapour pressure deficit	VPD	Pa	In-situ observations
	Land surface temperature	LST	K	Landsat OLI
	Air temperature	T _a	K	In-situ observations
	Solar radiation	R _a	W/m ²	In-situ observations
RNN _c	Wind speed	u	m/s	In-situ observations
	Vapour pressure deficit	VPD	hpa	In-situ observations
	Leaf area index	LAI	m ² /m ²	Landsat OLI
	Soil moisture	SM	%	Landsat OLI
	Land surface temperature	LST	K	Landsat OLI
	Air temperature	T _a	K	In-situ observations
RNN _s	Wind speed	u	m/s	In-situ observations
	Vapour pressure deficit	VPD	hpa	In-situ observations
	Solar radiation	R _a	W/m ²	In-situ observations
	Soil moisture	SM	%	Landsat OLI
	Land surface temperature	LST	K	Landsat OLI

the total gradient, while the log terms serve as regularizers to prevent the uncertainty from growing to infinity. Fig. 1 presents a schematic diagram of the hybrid RNN-PM model structure.

2.2. Input data set of RNN-PM model and hyperparameter setting

The RNN-PM model recognizes the distinct physical processes driving each of the three conductance parameters and accordingly employs different environmental variables to estimate G_a , G_s and G_c . Specifically, G_a is primarily driven by atmospheric turbulence factors such as wind speed and roughness, whereas G_c is regulated by plant physiological responses including LAI and stomatal sensitivity to light. The RNN_a utilizes wind speed (u), VPD, air temperature (T_a), and Land Surface Temperature (LST) (Liu et al., 2007). Although structural parameters like roughness length are not explicitly included, the combination of u and the temperature gradient ($LST - T_a$) allows the network to implicitly capture atmospheric stability regimes and mechanical turbulence intensity. The RNN_c inputs include LAI , soil moisture (SM), solar radiation (R_a), T_a , u , LST and VPD . This selection mirrors the physiological constraints found in the classic Jarvis-Stewart model (Parlange & Waggoner, 1970). R_a accounts for the light response of stomata; T_a , u and VPD capture atmospheric stress; and SM and LST represent root-zone water availability. Crucially, LAI is included to scale the conductance from leaf-level to canopy-level. In contrast, G_s is governed by soil physical processes like soil moisture diffusion (van de Griend et al., 1994). The RNN_s incorporates SM , LST , T_a , u , R_a , and VPD . Here, SM and LST serve as direct proxies for the soil water supply constraint on evaporation. Concurrently, R_a , T_a , u , VPD represent the atmospheric demand, enabling the network to learn the diffusion-limited phase of soil evaporation. Table 1 lists the specific information of the driving datasets of the three sub-RNN networks. The separation of predictors across RNN_a, RNN_s, and RNN_c is therefore a deliberate, streamlined design choice. Rather than exposing all three sub-networks

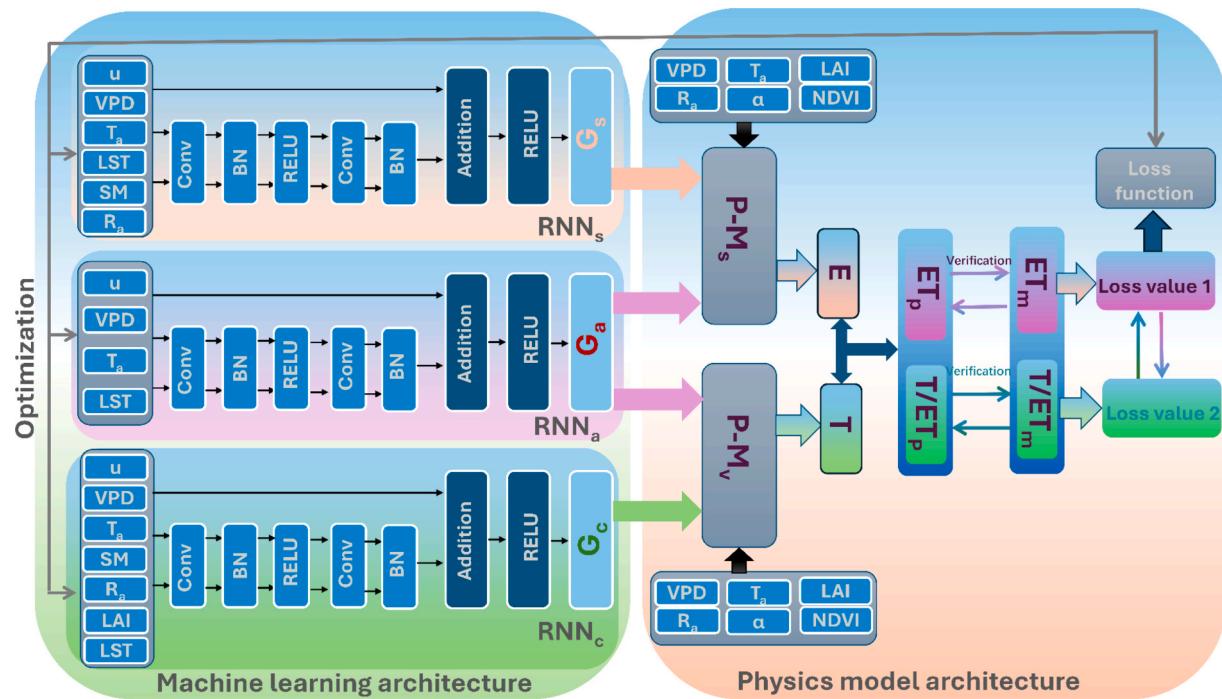


Fig. 1. Schematic of the RNN-PM Model structure. This diagram illustrates the integration of three sub-RNN networks within the Penman-Monteith (P-M) model to predict the conductance parameters: G_a , G_s , and G_c . These predicted parameters are subsequently input into the P-M model to simulate E , T , and ET . The model's simulation accuracy for ET , as well as the T/ET , is validated against observed data to ensure error minimization.

to the full set of available variables, each network is restricted to those predictors that are most directly relevant to its underlying process. This decoupling reduces the risk of overfitting and spurious correlations and enhances interpretability by ensuring that each conductance is inferred from variables with a clear physical justification. While additional structural predictors could be included (i.e. vegetation height), they are often difficult to obtain consistently at regional scales, and their marginal benefit must be weighed against increased model complexity and reduced scalability.

The hyperparameter optimization process in the RNN-PM model aims to find the most effective combination of hyperparameters to maximize model performance (Yang and Shami, 2020). Key hyperparameters in the RNN model include the number of layers (NL), learning rate (LR), batch size (BS), number of epochs (NE), momentum (MO), and dropout rates (DP). The genetic algorithm is applied to fine-tune each hyperparameter within the three sub-RNN frameworks (Aszemi and Dominic, 2019). The specific settings of various hyperparameter in the three sub-RNN frameworks are detailed in Table 2. Additionally, all three sub-RNN networks utilize the ReLU activation function (Agarap, 2018). To prevent overfitting, an early stopping strategy (ES) is implemented (Yao et al., 2007). The ES strategy halts training immediately when performance on the validation set starts to deteriorate, thus safeguarding against the model overfitting the training data.

Table 2

Detailed settings of the various hyperparameters within the three sub-RNNs.

Number of layers	Learning rate	Batch size	Number of epochs	Dropout rates	Momentum
RNN _a	0.010	32	100	0.15	0.9
RNN _c	0.015	32	100	0.20	0.9
RNN _s	0.015	32	100	0.20	0.9

2.3. Comparison with seven other dual-source ET models

To comprehensively evaluate the performance of the RNN-PM model, we conducted a comparative analysis involving seven classical models to simulate ET and its components. This intercomparison aimed to assess the accuracy of the RNN-PM model relative to traditional dual-source ET models, thereby providing a thorough understanding of the characteristics and uncertainties associated with various dual-source modeling approaches. Among the seven classical models, two are based on surface temperature decomposition processes: the Two-Source Energy Balance (TSEB) model (Norman et al., 1995) and the Pixel Component Arranging Comparing and Layered Energy Partition (PCA-LEP) model (Chen et al., 2020). The TSEB model assumes that the soil and vegetation layers each have independent energy balance processes and employs an iterative algorithm to partition LST into T_s and T_c . Following this decomposition, the model calculates the sensible heat fluxes for vegetation (H_c) and soil (H_s), subsequently determining E and T using the respective energy balance equations for the soil and vegetation layers. Contrasting with the TSEB model, the PCA-LEP employs the trapezoidal feature space theory for estimating E and T . Specifically, the PCA-LEP model initially constructs two-dimensional (2D) scatter plots based on $T_s f_c$ and $T_c f_c$. The model then identifies the theoretical dry/wet boundaries within the two scatterplots. The evaporation fractions for soil (EF_s) and vegetation (EF_v) for any given pixel in the two scatterplots are determined by its relative position to the theoretical dry/wet boundaries. Once EF_s and EF_v are established, E and T are subsequently calculated using the R_{ns} and R_{nv} (Chen et al., 2020).

Additionally, we conducted a comparison of the RNN-PM model with the original P-M model (Mu et al., 2007), which represents a purely physically based model grounded in surface conductance processes. We also compared it with the PT-JPL model, a two-source model that combines semi-empirical and semi-physical mechanisms (Fisher et al., 2008). The original P-M model employs the same parameterization scheme as the RNN-PM for E and T , as outlined in Eqs. (1)–(3) for all variables except the conductance values. In P-M model, G_c is estimated using an enhanced conductance model that accounts for leaf

photosynthetic efficiency (Leuning et al., 2008). G_s is computed as a function of soil moisture content (van de Griend et al., 1994), and G_a is determined in relation to wind speed (Liu et al., 2007). In the PT-JPL model, the Priestley–Taylor (PT) coefficients for both vegetation and soil are set to 1.26 (Priestley & Taylor, 1972). To simulate T , the model considers constraints like the LAI, VPD, T_a and SM, all of which influence potential vegetation transpiration. Similarly, for the simulation of E , the model incorporates the effects of SM and VPD on potential soil evaporation. Furthermore, we also compared the RNN-PM model with two other types of conductance-based models (PML-type model, and ET Monitor-type model). Please refer to the [supplementary materials](#) for details.

Finally, the performance of the RNN-PM model was compared with three pure machine learning models: a RNN model (Sherstinsky, 2020), a CatBoost model (Prokhorenkova et al., 2018), and a support vector regression (SVR) model (Drucker et al., 1996). These three models respectively represent regression approaches based on neural networks, decision trees, and kernel functions. To discern the components of ET , each machine learning model employed two sub-models dedicated to simulating E and T independently. The output nodes of these two sub-models were interconnected to facilitate the estimation of ET . All three machine learning models utilized the same loss function as the RNN-PM model, aimed at constraining the simulation errors of ET and T/ET . Additionally, all models employed the same driving data set. Specifically, the dataset for the sub-model simulating E comprised variables such as R_a , T_a , u , VPD, LST, SM, and u , while the data set for the sub-model simulating T included R_a , T_a , u , VPD, LST, SM, u , and LAI.

2.4. Study site and data collection

2.4.1. Study site

Data were sourced from 47 terrestrial flux sites in the National Ecological Observatory Network (NEON), a National Science Foundation-funded initiative (Keller et al., 2008). This network spans

diverse ecosystems across the United States, designed to capture a wide array of environmental conditions. The sites collectively represent a broad spectrum of ET characteristics by encompassing significant variability in vegetation, soil moisture, climate, surface roughness, and atmospheric turbulence. Fig. 2 presents the spatial distribution and classification of 47 selected EC sites utilized for model training and validation.

2.4.2. Eddy covariance observations

The RNN-PM model was trained and validated using data collected from 47 NEON flux sites between 2019 and 2021. Each flux tower is equipped with a three-dimensional (3D) ultrasonic anemometer and a gas analyzer. At the top of flux towers, a sonic anemometer (CSAT-3, Campbell Scientific Inc., Logan, Utah, USA) records 3D wind speed at a high frequency of 20 Hz. Concentrations of H_2O and CO_2 were monitored using an enclosed-path infrared gas analyzer (Li-Cor, Inc., model LI-840A and LI-850; Lincoln, Nebraska, USA), paired with a mass flow controller to maintain a steady flow rate through the sampling cell. T_a was recorded using triple-redundant aspirated platinum resistance thermometers (MetOne Instruments, Inc., model 076B-7388; Grant Pass, Oregon, USA) at a frequency of 1 Hz. The final ET flux calculation is based on the covariance between H_2O concentration and wind speed, assessed over a half-hourly period (NEON, 2024j, k). The flux correction process involves a series of adjustments, including density and sonic temperature corrections, coordinate rotation to account for terrain-induced tilts, time lag compensation, and correction for high-frequency signal losses (Massman & Lee, 2002). Following these corrections, quality control procedures are implemented, such as stationarity and integral turbulence tests, along with the identification and exclusion of low-quality data collected during periods of weak turbulence (Mauder et al., 2013). Finally, to preserve the continuity and integrity of the measured flux time series, the marginal distribution sampling (MDS) gap-filling algorithm was employed to address instances of missing flux data (Moffat et al., 2007).

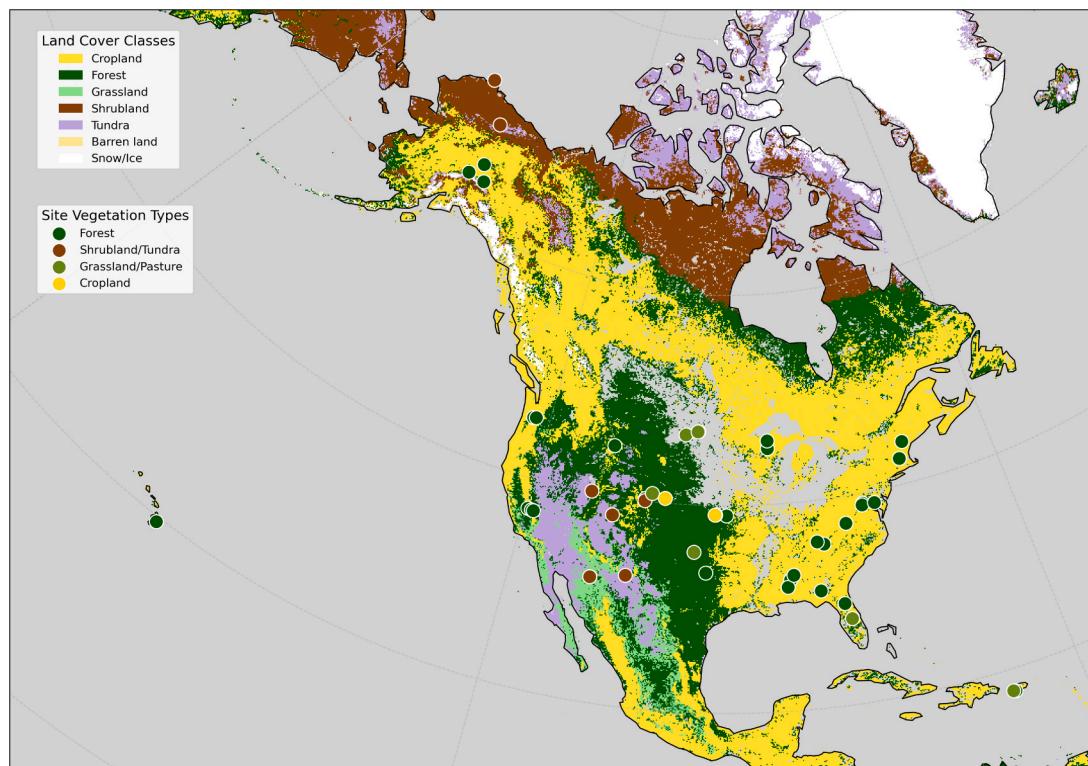


Fig. 2. Spatial distribution and classification of NEON sites utilized for model training and validation. Land use data are derived from the GLASS-GLC dataset (Liu et al. 2020).

A critical challenge in model validation is that EC systems provide only the total ET flux, lacking direct measurements of its constituents, E and T . To assess the RNN-PM model's component-level accuracy, we therefore partitioned the EC-observed ET using high-frequency turbulence (HFT) methods. The HFT partitioning approach is predicated on the theoretical assumption (Scanlon & Sahu, 2008) that T and photosynthesis are co-located, sharing common sources and sinks. This principle allows ET to be partitioned by analyzing high-frequency correlations between H_2O and CO_2 concentrations, which are concurrently measured by the EC system. A key benefit of HFT approaches is the ability to generate continuous, half-hourly T/ET ratios without reliance on ancillary meteorological data. In this study, we applied four distinct HFT methods across 47 NEON sites: Flux-variance Similarity (FVS) (Scanlon & Kustas, 2010), Conditional Eddy Covariance (CEC) (Zahn et al., 2022), a combined CEC and Water-Use Efficiency approach (Zahn et al., 2022; Zahn & Bou-Zeid, 2024), and Conditional Eddy Accumulation (CEA) (Zahn et al., 2022; Zahn & Bou-Zeid, 2024). The definitive T/ET ratio used for validation was derived from the ensemble mean of these four methods. The ensemble mean was selected to minimize the potential bias and random errors associated with any single algorithm, providing a more robust benchmark in the absence of a priori knowledge regarding the superiority of a specific method across diverse ecosystems. Since the ET components derived from the high-frequency turbulence method are directly calculated from the original H_2O and CO_2 concentration data collected by the EC system without any auxiliary data, the HFT approach is generally regarded as an observation-based method for decomposing ET components. Its results are widely used as a standard value for validating various models of E and T .

To further verify the reliability of the HFT-derived T/ET benchmark, we conducted a cross-validation using stable isotope data, which is widely regarded as an independent reference standard. We selected 13 NEON sites where sufficient isotopic measurements of precipitation, soil water, and xylem water were available during the study period. The T/ET ratios derived from the isotope mass balance method were compared against the HFT ensemble averages. The results demonstrate a robust consistency between the two independent methods, yielding a RMSE of 0.16 and a coefficient of determination (R^2) of 0.58 (see Fig. 3). This strong agreement at the 13 validation sites provides high confidence in using the HFT ensemble as the ground truth for evaluating the RNN-PM model across the entire network.

2.4.3. Environmental variable observations

In addition to flux measurements, NEON sites provide a

comprehensive suite of vegetation, soil, and meteorological observations. These include T_a (NEON, 2024a, b), R_a (NEON, 2024e, f), VPD (NEON, 2024c, d), and u (NEON, 2024g, h), all recorded at a frequency of 1 Hz. SM is measured at multiple depths, from 5 to 200 cm, using an automated soil moisture monitoring system (1 Hz) (NEON, 2024i). The raw 1 Hz data are aggregated into half-hourly and daily averages to align with the temporal resolution of EC flux observations.

2.4.4. Satellite data collection

The vegetation and soil variables at regional scale were derived from Landsat Operational Land Imager (OLI) imagery. Specifically, the $NDVI$ and f_c were extracted using the visible Near-Infrared (NIR) and Red bands (0.433–0.680 μm) bands (Chen et al., 2021). LST was obtained from the thermal infrared band (10.6–12.5 μm) (Jimenez-Munoz et al., 2014). Surface albedo (α) was derived from the visible, near-infrared, and shortwave infrared bands (0.433–2.300 μm) (Liang et al., 1999). LAI was extracted using the visible light band, based on the Soil-Adjusted Vegetation Index (SAVI) (0.433–0.680 μm) (Fang et al., 2019). Surface SM was calculated through the triangular feature space method, which determines SM by locating the dry/wet boundaries in $LST-f_c$ scatter plots (Carlson, 2007). Ground validation results indicate that the regionally scaled SM retrieval based on Landsat OLI imagery across 43 EC sites achieves an average Mean Absolute Percentage Difference (MAPD) of 3.5% and a coefficient of determination (R^2) of 0.95. These results demonstrate the effectiveness of the triangular feature space method in estimating SM at the regional scale. Finally, all Landsat images were resampled to a spatial resolution of 30 m. To rigorously match the gridded satellite data with the dynamic site-scale EC observations, we employed a two-dimensional flux footprint model (Kljun et al., 2015). For each satellite overpass time, we calculated the footprint source area function based on the concurrent aerodynamic conditions (wind speed, wind direction, and atmospheric stability). The site-scale values for all satellite-derived variables (e.g., $NDVI$, LST , and f_c) were then computed as the footprint-weighted averages of the underlying 30 m pixels. This approach ensures that the input features for the RNN-PM model accurately represent the heterogeneous land surface actually captured by the flux sensors, rather than a simple geometric average around the tower.

2.5. Model operation

The evaluation of the RNN-PM model's performance encompasses three aspects: 1) Assessing its ability to simulate site-scale ET and its

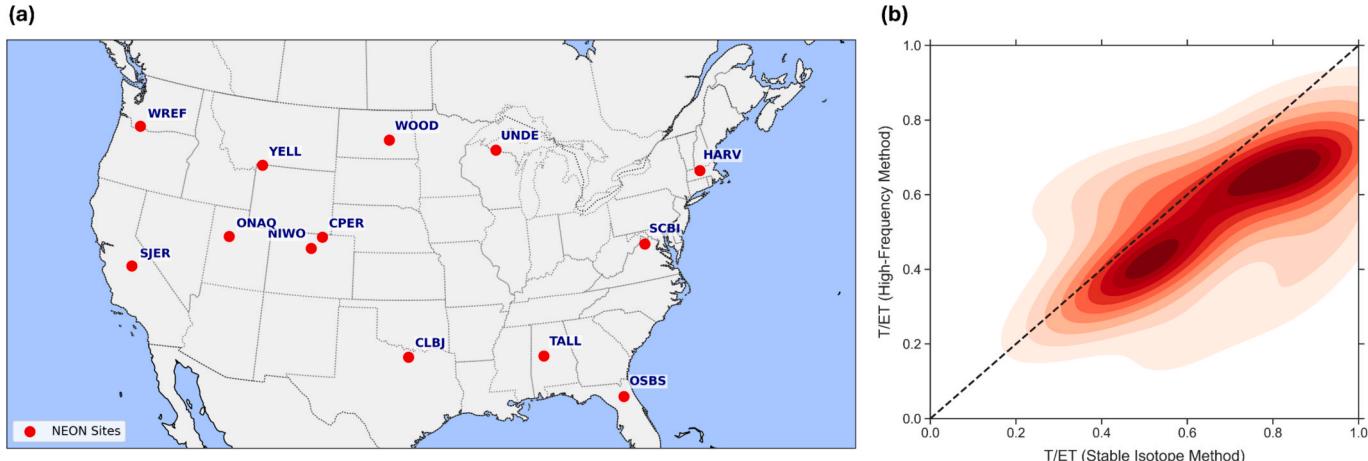


Fig. 3. (a) Geographic distribution of the 13 NEON sites used for model validation. (b) Comparison of transpiration fractions (T/ET) derived from the stable isotope method and the high-frequency temperature (HFT) method. The plot utilizes Kernel Density Estimation (KDE) to visualize the distribution of paired observations, with darker red regions indicating higher data density. The black dashed line represents the 1:1 identity line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

components, along with a comparative analysis against seven other classic dual-source *ET* models; 2) Investigating the model's capability to simulate *ET* and its components outside the training set, which sheds light on its generalization potential; 3) Examining the model's performance in the context of extreme climatic events.

Initially, the RNN-PM model's capability to simulate site-scale *ET* was evaluated. The RNN-PM model performed the simulation of site-scale *ET* on a daily scale. To enhance the efficiency of the training process, all meteorological, vegetation, and soil driving variables were normalized, standardizing their values within a range of 0–1 (Singh & Singh, 2020). Subsequently, the RNN-PM model underwent 150 rounds of cross-validation to evaluate its robustness and reliability. In each iteration, both the training and testing datasets were resampled. The observational dataset, comprising both *ET* and *T/ET* data, was randomly partitioned. A majority subset (80%) was allocated for model calibration, while the remaining 20% was reserved as a hold-out set for validation. For transparency, we additionally evaluated RNN-PM using a single 80%/20% train-validation split without repeated cross-validation. The resulting performance statistics are reported in Table S3 and closely match those obtained under the 150-fold cross-validation protocol. Furthermore, to benchmark the *ET* simulation accuracy of the RNN-PM model, its performance was contrasted against a suite of seven conventional dual-source *ET* models. The simulation performance of these eight models for *ET* and its components was evaluated using four statistical metrics: root-mean-square-error (RMSE), R², bias (defined as mean bias error), and Kling–Gupta efficiency (KGE). The KGE is calculated as follows (Gupta et al., 2009):

$$[MICK]GE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (6)$$

where r represents the Pearson correlation coefficient between simulated and observed values, α represents the variability ratio ($\sigma_{\text{sim}}/\sigma_{\text{obs}}$), and β represents the bias ratio ($\mu_{\text{sim}}/\mu_{\text{obs}}$). Here, σ and μ denote the standard deviation and mean of the simulated and observed data, respectively. Besides, to assess the relative importance of individual predictors, we employed a leave-one-variable-out approach. Specifically, each variable was sequentially removed from the model, after which the model was retrained and validated. The resulting changes in model performance metrics were then used to quantify the contribution of each variable to the overall model skill.

Secondly, the generalization capability of the RNN-PM model was evaluated using *ET* monitoring data from four additional flux sites obtained from the NEON network, which were not included in the model's training set. The eight dual-source *ET* models, previously trained and calibrated using data from the initial 43 NEON flux sites, were directly applied to predict *ET* and *T/ET* at the four additional sites. The predicted values were subsequently compared with the observed *ET* and *T/ET* measurements to evaluate the models' generalization capabilities and their accuracy in predicting *ET* beyond the original training domain.

Finally, the capability of the RNN-PM model in simulating *ET* and components under extreme climate conditions was evaluated. Eight representative datasets exemplifying extreme climate scenarios were selected from the original vegetation, soil and meteorological datasets. These included extreme hot (97–100th percentile of T_a), extreme cold (0–3rd percentile of T_a), extreme wet (97–100th percentile of SM), extreme drought (0–3rd percentile of SM), extreme radiation (97–100th percentile of R_a), cloudy conditions (0–3rd percentile of R_a), sparse vegetation (0–3rd percentile of f_c) and dense vegetation (97–100th percentile of f_c). The selected eight extreme climate datasets were input into the trained RNN-PM model and seven other models to predict *ET* and its components. The predicted values were then compared with corresponding observed measurements to assess each model's performance in simulating *ET* and the *T/ET* ratio under extreme climatic conditions.

3. Results

3.1. Performance of hybrid RNN-PM model

The RNN-PM model reliably estimates *ET* and *T/ET* across NEON sites. Fig. 4 presents a scatter density plot for the model's verification of *ET* and *T/ET*. Additionally, Tables 3 and 4 details the specific statistical parameters. Notably, the statistics for *ET* and *T/ET* simulations at different sites demonstrated minimal variation, indicating that the RNN-PM model consistently performs well across sites with diverse vegetation types. Specifically, the KGE of *ET* as simulated by the RNN-PM model across NEON sites ranged from 0.87 to 0.92, with an average of 0.89. The RMSE spans from 0.52 to 0.58 mm/day, averaging 0.55 mm/day across NEON sites. The accuracy of *T/ET* simulation mirrors that of *ET*, with an average KGE of 0.87 (ranging from 0.85 to 0.89) and an RMSE of 0.06 (0.05–0.07). However, the bias results reveal a slight underestimation in the RNN-PM model for both *ET* and *T/ET*, exhibiting average biases of −0.18 mm/day (ranging from −0.35 to −0.02 mm/day) and −0.010 (−0.020 ~ 0.003) respectively.

The performance of the RNN-PM model was further compared with seven other dual-source models. As depicted in Fig. 4c–d and Tables 3 and 4, the RNN-PM model demonstrates superior accuracy for *ET* and *T/ET* simulations among the eight models. The ranking of model performance, from best to worst, is as follows: RNN-PM, RNN, CatBoost, TSEB, PCALEP, P-M, SVR, and PT-JPL. Notably, the performance of the three pure physics process models varied significantly across different EC sites. Specifically, the two dual-source models (TSEB and PCALEP), which are based on the temperature decomposition process, demonstrated higher accuracy in simulating *ET* and *T/ET* at sites with lower *LAI*. For example, at EC sites where *LAI* was below 1.5 m²/m², the average KGE for *ET* and *T/ET* simulations reached 0.93 and 0.90, respectively, for the TSEB and PCALEP models (Figs. 5 and 6). In these cases, the simulation performance of these two thermal infrared-based models even surpassed that of the RNN-PM model. Conversely, at EC sites with higher *LAI*, the performance of both models declined significantly. When *LAI* exceeded 4.0 m²/m², the average KGE for *ET* and *T/ET* simulations using the TSEB and PCALEP models dropped to 0.70 and 0.72, respectively. In contrast to the two thermal infrared-based models, the P-M model, which is based on the conductance process, exhibits a distinctly different performance pattern. It exhibits excellent performance at sites with higher *LAI*, but its effectiveness decreases at sites with lower *LAI*. Notably, at the EC sites with *LAI* larger than 4.0 m²/m², the performance of the P-M model surpasses that of the RNN-PM model, achieving an average KGE greater than 0.90. However, at sites where *LAI* is below 1.5 m²/m², the performance of the P-M model is inferior to that of the PT-JPL model, with an average KGE at these sites falling below 0.70 (Figs. 5 and 6).

Unlike models based solely on physical processes, the three models based entirely on machine learning exhibit more consistent performance across diverse EC sites. The discrepancy in average RMSE for *ET* simulations between the best and worst-performing EC sites among three machine learning models (RNN, CatBoost and SVR) is just 0.12 mm/day, with average KGE differences being only 0.09. For *T/ET* simulation, the average RMSE difference is merely 0.05, with KGE showing a minimal variation of 0.06. However, a notable disparity in performance is observed among the three machine learning models. The CatBoost and RNN significantly outperform the SVR in simulations of both *ET* and *T/ET* (refer to Fig. 4c–d). In contrast, the PT-JPL model, which combines semi-empirical and semi-physical processes, exhibited the poorest performance among the eight models. This is evidenced by its average KGE values of only 0.71 for *ET* simulations and 0.63 for *T/ET* simulations.

To further evaluate the robustness and generalization capability of the RNN-PM model, we conducted a granular comparative analysis against eight models across six major vegetation types. As depicted in Fig. 7, the RNN-PM model demonstrates superior overall accuracy among the eight models, yet the performance ranking varies

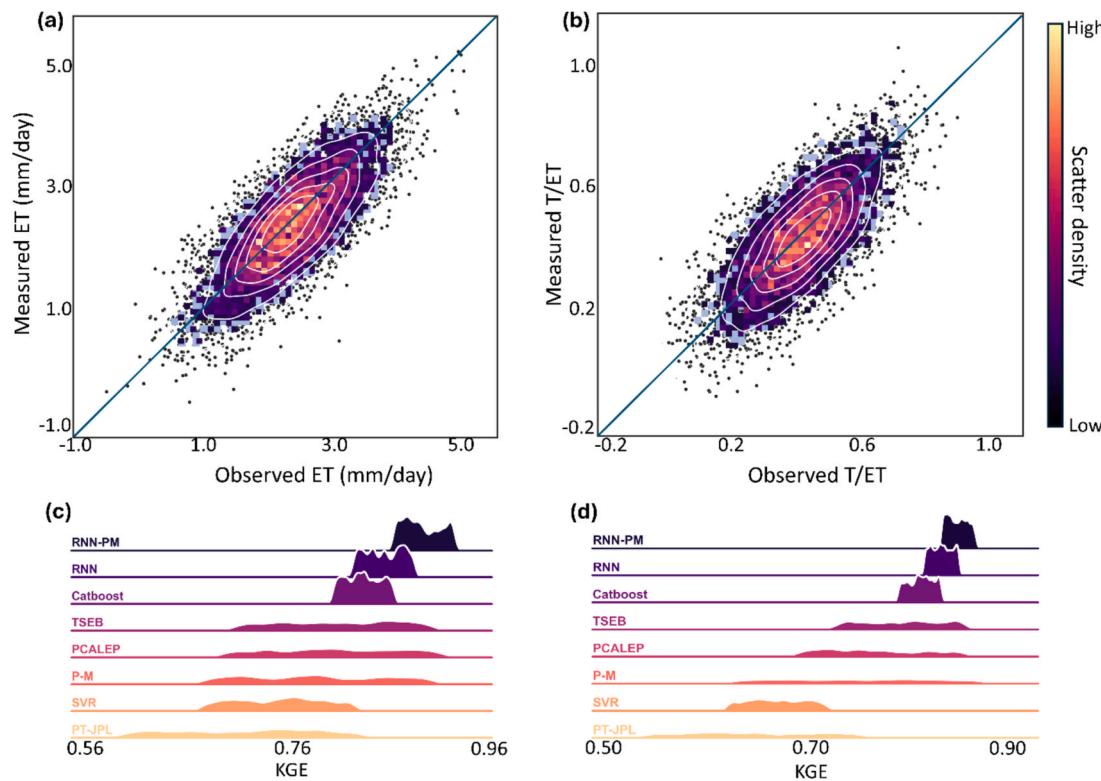


Fig. 4. (a) Scatter density plot comparing observed and RNN-PM model-simulated ET during the model verification period. (b) Similar comparison for the T/ET. (c) Distribution of KGE scores for the performance of eight dual-source models in ET simulations across the NEON sites during the model verification period; (d) Similar analysis for T/ET.

Table 3

Statistical parameters of accuracy of RNN-PM model for ET simulation across the NEON sites.

	KGE	R ²	RMSE (mm/day)	Bias (mm/day)
RNN-PM	0.89	0.83	0.55	-0.18
RNN	0.85	0.80	0.62	0.23
Catboost	0.83	0.79	0.64	-0.25
TSEB	0.80	0.77	0.70	0.29
PCALEP	0.79	0.76	0.72	0.30
P-M	0.77	0.74	0.78	-0.35
SVR	0.75	0.71	0.85	-0.43
PT-JPL	0.71	0.68	0.95	0.51

Table 4

Statistical parameters of accuracy of RNN-PM model for T/ET simulation across the NEON sites.

	KGE	R ²	RMSE	bias
RNN-PM	0.87	0.82	0.06	-0.01
RNN	0.85	0.80	0.07	0.02
Catboost	0.82	0.78	0.09	0.04
TSEB	0.78	0.74	0.13	0.08
PCALEP	0.76	0.72	0.16	0.09
P-M	0.72	0.68	0.20	0.13
SVR	0.66	0.61	0.25	-0.19
PT-JPL	0.63	0.59	0.28	-0.23

significantly across biomes, revealing distinct structural characteristics of the physical baselines. In sparsely vegetated ecosystems such as Shrublands and Grasslands, where *LAI* is typically low (<1.5 m²/m²), the two thermal infrared-based models (TSEB and PCALEP) exhibited remarkable performance, achieving KGE values exceeding 0.90. In these specific conditions, their accuracy was comparable to, and occasionally

surpassed, that of the RNN-PM model. Conversely, the conductance-based P-M model showed its lowest performance in these regions, likely due to the difficulty in parameterizing soil evaporation resistance without explicit soil surface temperature constraints. A distinct reversal in performance was observed in dense forest ecosystems (e.g. Deciduous Broadleaf Forest, Evergreen Broadleaf Forest and Evergreen Needleleaf Forest), where *LAI* often exceeds 4.0 m²/m². Here, the P-M model demonstrated excellent accuracy (KGE > 0.90), rivaling the performance of the RNN-PM model. However, the performance of TSEB and PCALEP declined significantly, highlighting the limitations of temperature decomposition methods in closed canopies where soil signal is masked. Among the machine learning models, RNN and CatBoost showed consistent performance across all vegetation types, significantly outperforming SVR, though they generally lagged slightly behind the hybrid RNN-PM. The semi-empirical PT-JPL model exhibited the poorest performance across most biomes. Ultimately, the RNN-PM model stands out not by being the absolute best in every single biome, but by its unique stability. It effectively avoids the low-LAI failure of the P-M model and the high-LAI failure of TSEB/PCALEP, securing the highest overall accuracy.

Overall, according to the verification results, the RNN-PM model exhibits the most robust and consistent performance in simulating *ET* and *T/ET* across the NEON sites. The two models based on temperature decomposition perform most effectively under low *LAI* conditions, but their accuracy declines significantly in high *LAI* environments. In contrast, the P-M model, which is grounded in surface conductance processes, exhibits limited effectiveness in low *LAI* scenarios but demonstrates superior performance under high *LAI* conditions. The performances of the three purely machine learning-based models can provide consistent performance across different EC sites, with the RNN and CatBoost models outshining the SVR model. The PT-JPL model, hindered by its reliance on empirical parameter settings, demonstrates the weakest performance among the evaluated models.

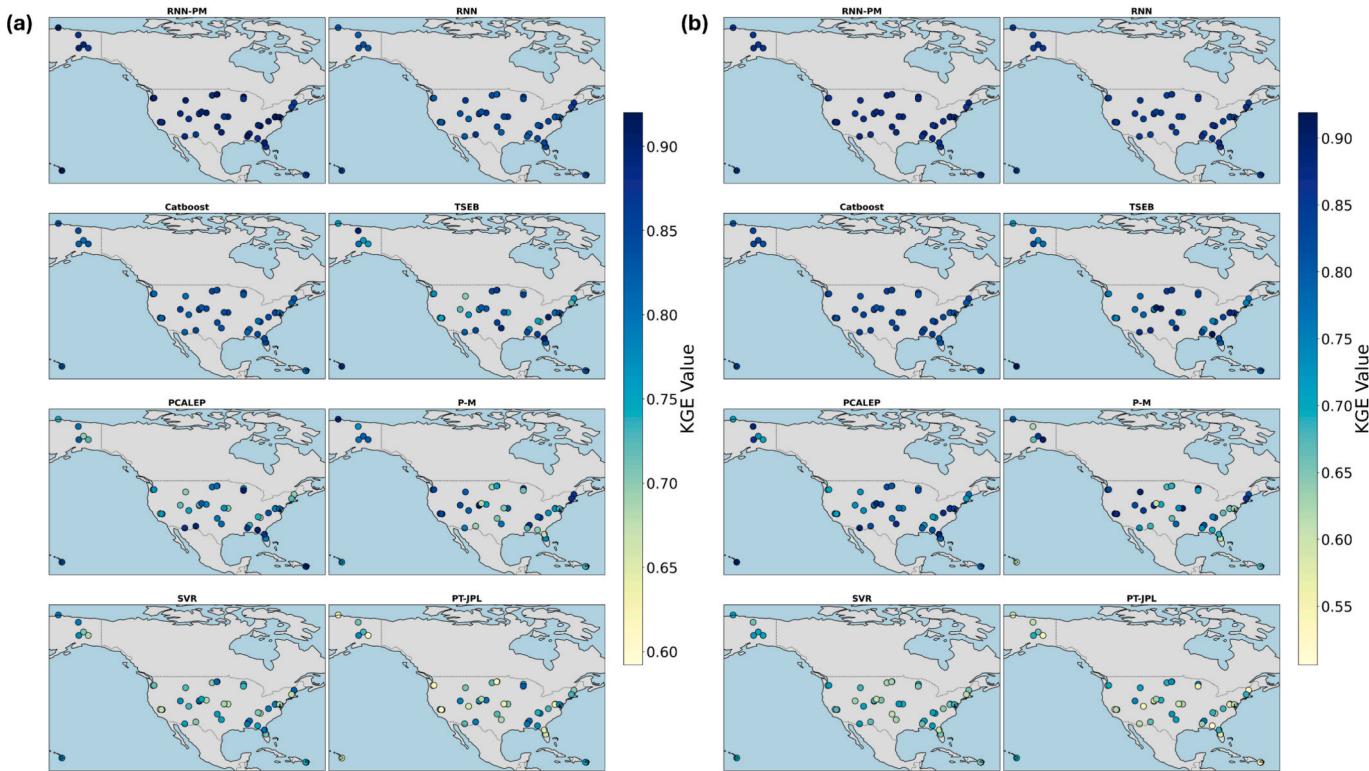


Fig. 5. (a) Spatial distribution of KGE for ET simulations conducted by eight dual-source models across NEON sites during the model verification period. (b) Spatial distribution of KGE for the T/ET simulations by the same models across the NEON sites during the model verification period.

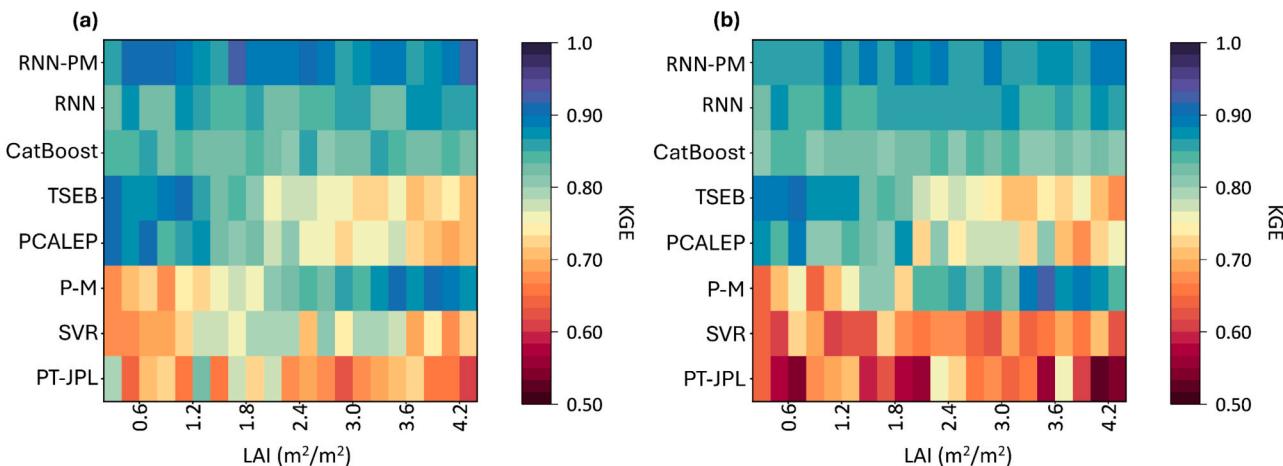


Fig. 6. (a) Distribution of KGE for ET simulated by eight dual-source models under varying LAI conditions. (b) Distribution of KGE for the T/ET simulated by the eight dual-source models under varying LAI conditions. The results depict the mean KGE across different EC sites with identical LAI conditions.

3.2. Generalization potential testing and variable importance

To explicitly test the model's extrapolation performance, its generalization capacity was assessed at four EC sites not included in the original calibration dataset. At these external locations (Fig. 8), the RNN-PM model's *ET* predictions achieved robust metrics: an average RMSE of 0.62 mm/day, and a KGE of 0.85. The *T/ET* predictions were similarly strong, yielding an RMSE of 0.07, and a KGE of 0.83. While these statistics reflect a slight reduction in accuracy compared to the training set, all error metrics remained well within acceptable ranges, confirming the model's strong generalization capabilities. This stability starkly contrasted with the three models based purely on machine learning (CatBoost, RNN, SVR). When applied to four new EC sites not

included in the original 43-site training set, these pure machine learning models exhibited a severe degradation in accuracy. Specifically, the three models based purely on physical processes (TSEB, PCALEP, and P-M), along with the PT-JPL model, exhibited performance at the four additional sites that were consistent with their performance at the initial 43 EC sites. In contrast, the three models based solely on machine learning (CatBoost, RNN, and SVR) exhibited a significant decline in performance at the four additional sites compared to their performance at the initial 43 EC sites used for training. In *ET* simulation, the average RMSE for these purely machine learning models exceeded 1.0 mm/day, with both R^2 and KGE falling below 0.70. Similarly, in simulating *T/ET*, their average RMSE exceeded 0.25, with R^2 and KGE both below 0.60 (Fig. 8). Additionally, it was observed that the three purely machine

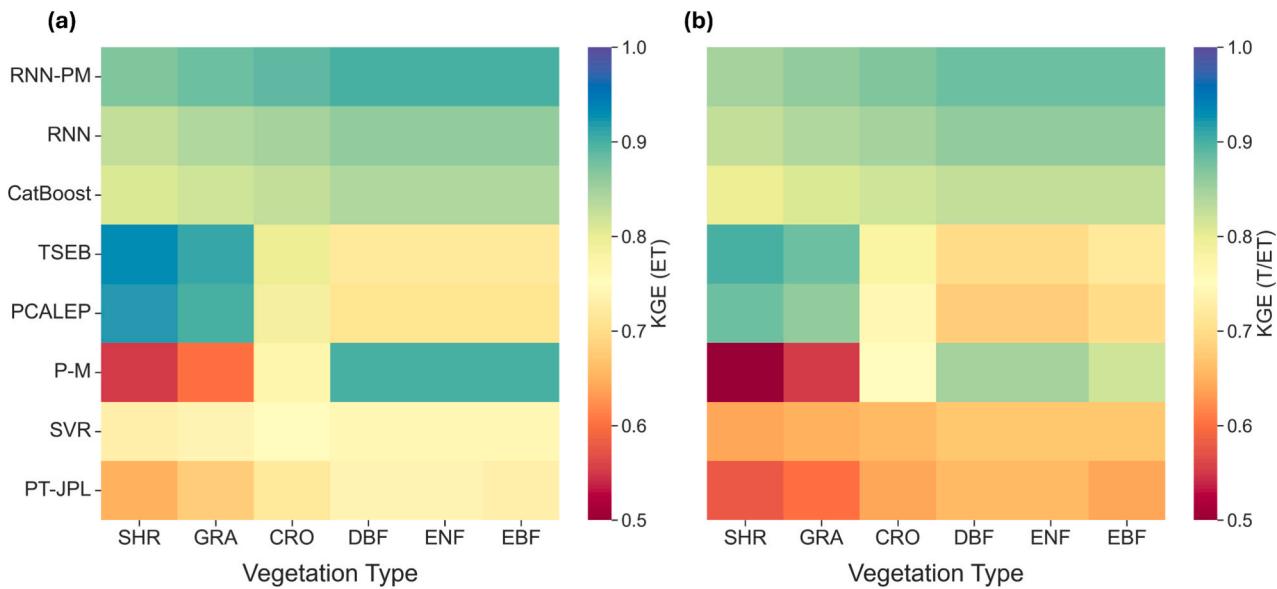


Fig. 7. Comparative analysis of model performance across diverse vegetation types. The heatmaps display the Kling-Gupta Efficiency (KGE) for (a) daily evapotranspiration (ET) and (b) the transpiration fraction (T/ET). Vegetation types on the x-axis are generally arranged from sparse (low LAI) to dense canopies (high LAI). Note the contrasting performance patterns: thermal-based models (TSEB, PCALEP) excel in sparse vegetation but decline in forests, whereas the conductance-based P-M model shows the opposite trend. The RNN-PM model demonstrates superior robustness across the entire gradient. Abbreviations: SHR: Shrublands; GRA: Grasslands; CRO: Croplands; DBF: Deciduous Broadleaf Forests; ENF: Evergreen Needleleaf Forests; EBF: Evergreen Broadleaf Forests.

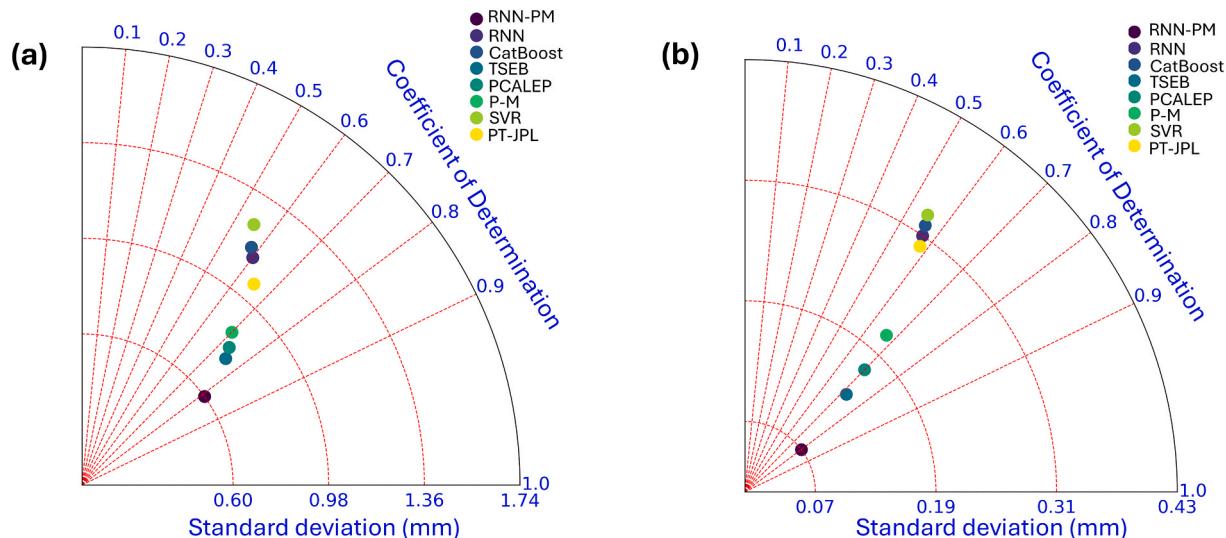


Fig. 8. (a) Taylor plots showing statistical performance of ET simulations by eight dual-source models at 4 EC sites outside the training regime. (b) Similar plots for T/ET.

learning-based models frequently generated *ET* and *T/ET* estimates lacking physical plausibility at the four additional sites. Examples include cases in which the estimated *ET* exceeded the total net radiation flux or the *T/ET* ratio surpassed 1. These findings strongly suggest that the absence of physical constraints renders the pure machine learning models incapable of reliable extrapolation beyond their training domain.

The performance of various dual-source models was further evaluated under eight extreme climate scenarios. As envisaged in Fig. 9a–b, the RNN-PM model and the three models based purely on physical processes produced reliable estimates of *ET* and *T/ET* under extreme climate conditions. The performance of these four models (RNN-PM, TSEB, PCALEP and PM) in extreme scenarios was relatively comparable to their performance in regular climate scenarios. However, the performance of the three models based solely on machine learning

(CatBoost, RNN, and SVR) deteriorated markedly under extreme climate conditions compared to datasets from regular climate scenarios. More precisely, the three purely machine learning-based models exhibited a significant tendency to overestimate *ET* and *T/ET* under extreme conditions of cold, drought, cloudy, and sparse vegetation. In these four scenarios, their average bias exceeded 1.0 mm/day for *ET* simulations and was greater than 0.25 for *T/ET* simulations. Conversely, under extreme conditions of hot, wet, intense radiation, and dense vegetation, the three machine learning models significantly underestimated *ET* and *T/ET*, with average biases for *ET* simulations lower than -1.0 mm/day and less than -0.25 for *T/ET* simulations (Fig. 9a–b). In short, the overall performance ranking of the eight models under extreme climate conditions, from highest to lowest, was as follows: RNN-PM, TSEB, PCALEP, P-M, PT-JPL, RNN, CatBoost, and SVR.

Identifying the most influential variables by ranking feature

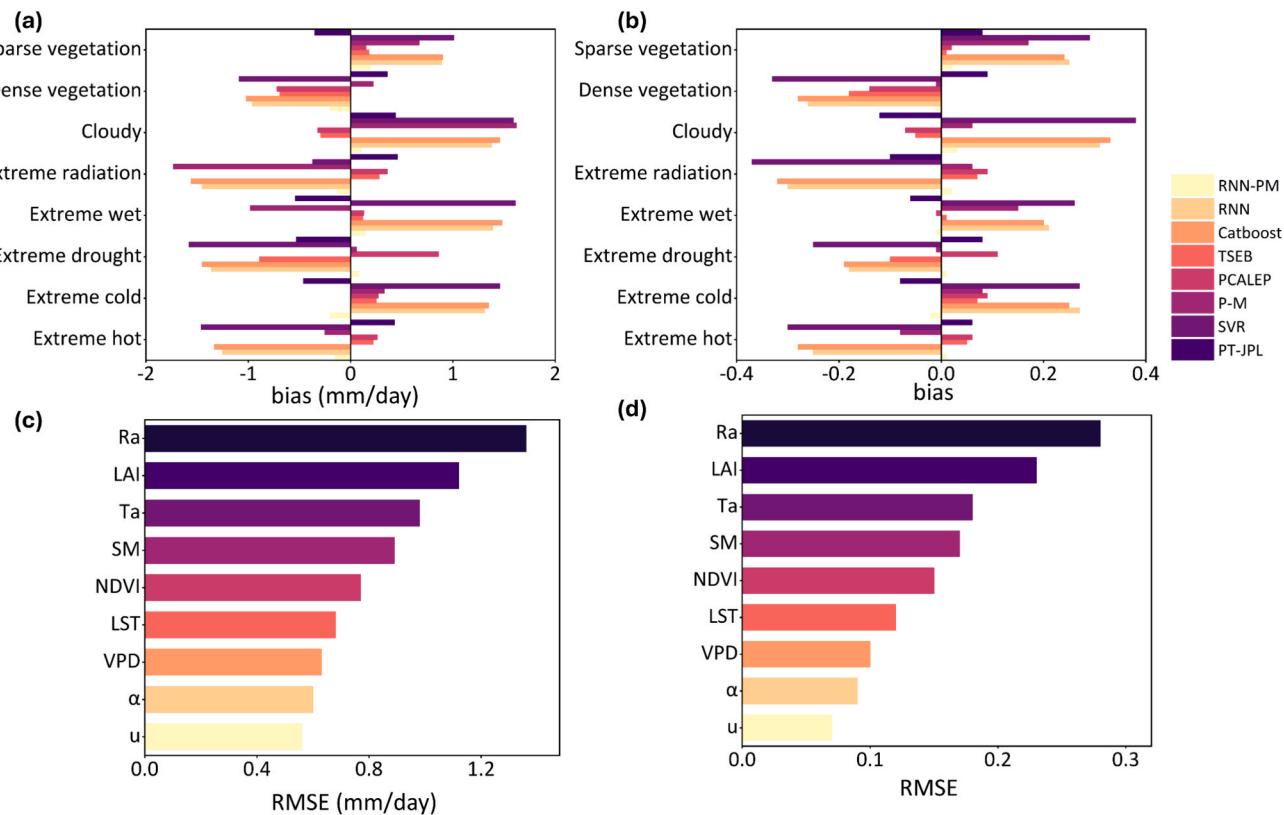


Fig. 9. (a) Bias distribution in ET simulations conducted by eight dual-source models under eight extreme climate scenarios. (b) Similar distribution for the T/ET. (c) Variable importance ranking in the RNN-PM model, indicated by changes in the RMSE for ET simulations when specific variables are excluded from the input dataset. (d) Similar analysis for T/ET.

importance is essential in machine learning models, as it helps optimize data collection, refine model performance, and enhance interpretability. In the RNN-PM model, R_a emerges as the most critical input variable, followed by LAI, T_a , SM, NDVI, LST, VPD, α , and u . Excluding R_a from the

dataset results in a substantial increase in the RMSE for ET and T/ET simulations by 0.45 mm/day and 0.49, respectively. Conversely, removing u leads to only a minor increase in RMSE of 0.02 mm/day for ET and 0.01 for T/ET simulations. Previous global-scale studies have

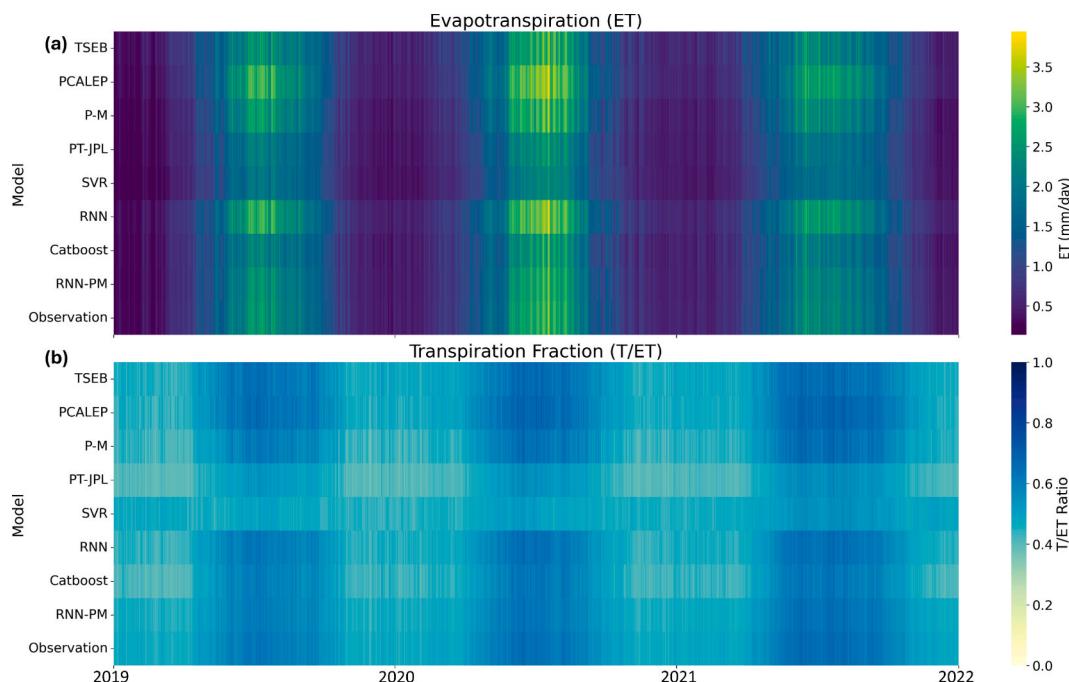


Fig. 10. (a) Temporal variations in measured daily ET simulated by eight dual-source models during the experimental period. (b) Similar analysis for the T/ET (The results represent the average daily values of ET and T/ET across NEON sites).

also confirmed the dominant role of R_a in ET simulations, while highlighting the relatively minor importance of u (e.g. Chen et al., 2023 a, b).

3.3. Spatial and temporal patterns of ET and its components

Fig. 10 illustrates the seasonal variations of simulated ET and T/ET by eight models compared with corresponding measured values from 2019 to 2021. Generally, all models successfully captured the seasonal dynamics of ET and T/ET , with peaks occurring in summer due to higher R_a , T_a and LAI , and lower values in winter with reduced R_a , T_a and LAI . However, there are substantial discrepancies in the ET and T/ET time series produced by the eight models. The three purely machine learning models and P-M model considerably underestimated ET and T/ET in summer, with average biases reaching -0.92 mm/day and -0.13 , respectively, during the summer season. The two thermal infrared-based models (TSEB and PCALEP) and the PT-JPL model, conversely, significantly overestimated ET and T/ET during summer (average biases being 0.86 mm/day and 0.12). In contrast, throughout the entire experiment period, the RNN-PM model did not exhibit any significant overestimation or underestimation phenomena.

To demonstrate the scalability of the RNN-PM model from site-level flux towers to landscape scales, we applied the trained model to Landsat imagery to generate regional maps of ET and its components at four selected NEON sites (**Fig. 11**). Generally, the inferred ET and its components exhibit a wide range of variation, reflecting the heterogeneity of land use surrounding the NEON flux sites. For instance, at the DEJU site, ET and T reach as high as 3.1 mm/day and 2.2 mm/day, respectively, in the southwestern region, compared to just 1.3 mm/day for ET and 0.9 mm/day for T in the southeastern region. At the SJER site, the ET varies spatially between 0.3 and 4.8 mm/day, averaging 3.2 mm/day, while T ranges from 0.2 to 3.7 mm/day, with an average of 2.1 mm/day. Moreover, the regional patterns of ET and T are broadly similar, typically higher in areas with greater vegetation cover, while E displays a contrasting spatial pattern, being higher in areas with lower vegetation cover. It is important to note that these spatial patterns represent an extrapolation of the site-level training results to the surrounding landscape. While the spatial distribution shows a logical positive correlation with vegetation cover (f_c) ($r = 0.85$, $P < 0.01$) and distinct partitioning between E and T , these regional maps have not been directly validated

against independent large-scale observations (e.g., basin-scale water balance) due to data limitations. Therefore, the spatial pattern of ET and components should be interpreted as a demonstration of the model's potential for high-resolution mapping, rather than a fully validated regional product. Overall, the observed seasonal dynamics and spatial patterns of ET and its components confirm that the RNN-PM model effectively captures both temporal and spatial variations across different NEON sites.

4. Discussion

4.1. Mechanisms causing the varying performance of eight dual-source ET models

The verification results reveal significant variations in the performance of the eight models for simulating ET and its components. These performance disparities among the eight models primarily stem from their differing model mechanisms. The performance of the two models based on temperature decomposition process is notably better in areas with lower LAI , primarily due to the sparse vegetation characteristic of these regions. In these environments, the thermal infrared remote sensing is particularly effective at capturing rapid surface temperature changes, with minimal obstruction from the vegetation canopy (Kustas & Anderson, 2009). Conversely, in high LAI areas, the denser vegetation hinders the ability of the thermal infrared remote sensing to detect surface temperature changes accurately, leading to a noticeable drop in model performance. The uncertainty in the TSEB model primarily stems from its iterative procedure, which repeatedly decreases T until E becomes positive. Previous studies indicated that the iterative process often fails to converge, frequently resulting in an overestimation of T and an underestimation of E , particularly in environments where T dominates over E , such as regions characterized by dense vegetation (Chen et al., 2021). Conversely, in regions with sparse vegetation, E tends to exceed T , thereby significantly reducing the uncertainty associated with the TSEB model. Consequently, the accuracy of the TSEB model improves substantially in low LAI areas compared to high LAI regions.

Unlike the TSEB model, the uncertainty in the PCALEP model primarily stems from determining the extreme wet and dry boundaries

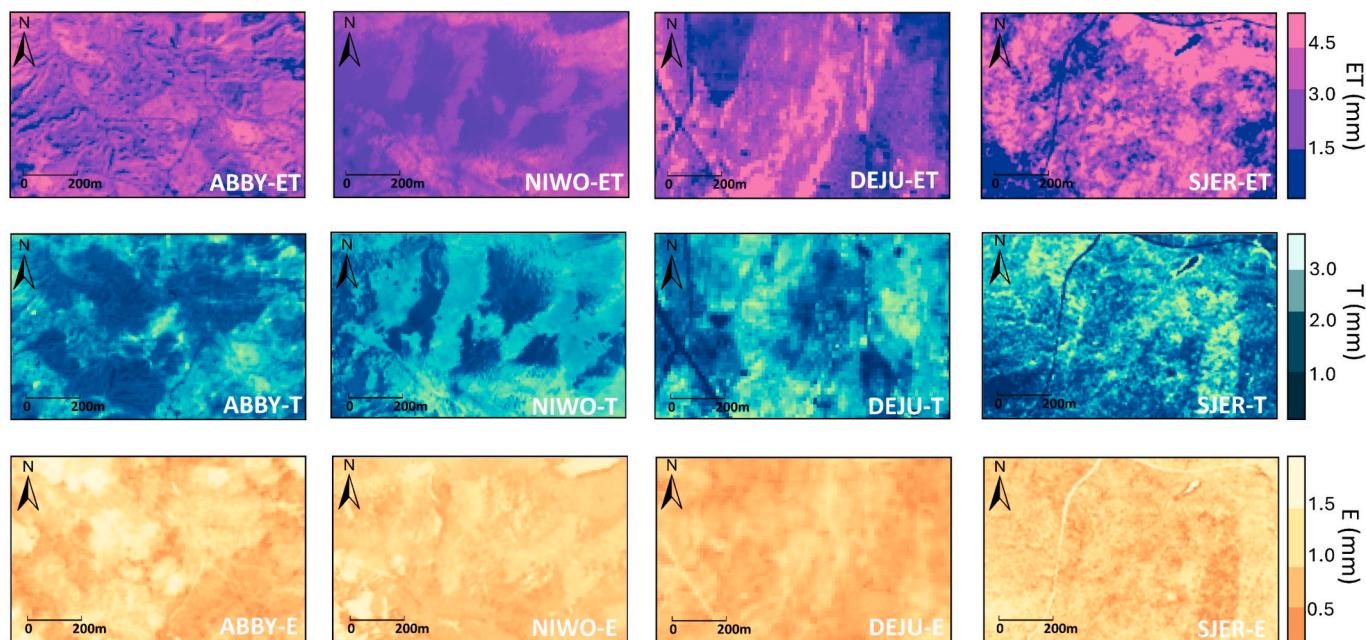


Fig. 11. Spatial distribution of daily ET , E , and T as mapped by the RNN-PM model at four selected EC sites. The spatial resolution is $30 \text{ m} \times 30 \text{ m}$. The results represent the long-term mean values of ET , E , and T from 2019 to 2021.

within the model framework. The developers of the trapezoid framework-based model have acknowledged the significant challenges and subjective nature involved in identifying end-members within the trapezoidal feature space (Moran et al., 1994). Variations in the domain size of the study region and the spatial resolution of remote sensing images affect shape of the feature space, which, in turn, leads to shifting boundaries for extremely dry and wet conditions. Additionally, cloud contamination in satellite imagery often results in anomalous scatter points, complicating end-member selection and adding considerable uncertainty. Lastly, the heterogeneity of the study area also influences the reliability of pinpointing extreme condition boundaries. For example, in arid regions, identifying the coldest points may be unrealistic due to the absence of such conditions, while in humid agricultural areas, determining the driest points poses a challenge because of the perpetual moisture presence. Regions with higher vegetation cover tend to exhibit more homogeneous surface characteristics, which partly explains why the PCALEP model performs poorly in areas with high *LAI* but demonstrates improved performance under low *LAI* conditions.

The P-M model exhibits better performance in high *LAI* regions but is less effective in low *LAI* areas. The model is characterized by its comprehensive physical description of the turbulent diffusion conductance process across surface-soil-vegetation nexus (Mu et al., 2007, 2011). However, a notable limitation is the complexity of its conductance parameterization schemes, which introduces greater uncertainty, particularly in areas with sparse vegetation where the energy and water exchange between soil and vegetation layer interplay significantly. Numerous studies have also demonstrated that P-M type models perform more reliably in areas with dense vegetation but exhibit poorer performance in regions with sparse vegetation (Stannard, 1993; Ershadi et al., 2015; Nyolei et al., 2021). In regions with high *LAI*, where vegetation biomass and coverage are relatively abundant, the vegetation can essentially be treated as a 'big leaf'. In this case, surface conductance is predominantly influenced by vegetation physiological attributes, with minimal impact from soil variables due to the shading of the soil surface by the canopy, thus leading to reduced uncertainty in conductance estimation. Conversely, in low-*LAI* regions, where vegetation biomass and coverage are limited, surface conductance is joint regulated by both vegetation physiological and soil moisture dynamics, often resulting in a coupling effect between the two. This coupling effect complicates accurate estimation of surface conductance in these areas, thereby diminishing the performance of the P-M model.

The PT-JPL model demonstrates the weakest performance among eight models. Its primary source of uncertainty lies in the estimation of empirical parameters. In this study, a Priestley-Taylor coefficient of 1.26 was adopted for both soil and vegetation, a value consistent with the foundational recommendations of the model developers (Priestley & Taylor, 1972; Fisher et al., 2008). However, whether this value accurately represents potential *ET* under varied environmental conditions remains questionable. Previous research indicates that the PT coefficient could significantly exceed 1.26 in certain contexts (Lhomme, 1997; Crago et al., 2023). Moreover, the process of downscaling potential evaporation and transpiration to actual values involves various environmental variables, introducing further uncertainties. For instance, in simulating vegetation transpiration, factors like the *LAI*, *VPD*, *T_a* and *SM* are incorporated in the downscaling of α_v . However, other potential constraints that influence *T*, such as atmospheric CO₂ concentration, are overlooked in the PT-JPL model. Additionally, the environmental variables incorporated into the PT-JPL model exhibit notable interdependencies. For example, *VPD* is strongly correlated with *T_a* and *SM*. When both variables are used simultaneously as constraint factors for transpiration, this overlap may introduce redundancy in the input information, potentially affecting model performance. On the other hand, neglecting the advection effect in PT-JPL model can lead to significant errors, particularly in areas where there are substantial temperature and humidity gradients between irrigated fields and surrounding dry lands (Dhungel et al., 2021; Bambach et al., 2022).

Therefore, the PT-JPL model is better suited for application in large-scale areas rather than at the site scale since the advection effect can be largely disregarded across extensive regions.

Finally, the three purely machine learning-based models consistently performed well across the initial 43 EC sites but exhibited poor results at 4 additional EC sites outside their training set and under extreme climate conditions. This is primarily because these models lack inherent constraints based on physical processes, leading to mechanistic speculations devoid of a physical basis when encountering new datasets (Chen et al., 2023). This issue is particularly evident in *ET* simulations under extreme climate scenarios, where these models often output predictions that defy physical logic. For instance, under cloudy conditions, the latent heat flux simulated by the SVR and CatBoost models exceeds the available net radiation flux, which is physically unrealistic. Similarly, under extreme drought conditions, the RNN and SVR models frequently produce *T/ET* values greater than 1, which is also implausible. Furthermore, it's noteworthy that even under conventional *ET* datasets, the SVR model significantly underperforms compared to the RNN and CatBoost models. Previous research has demonstrated that models based on kernel functions are better suited for classification tasks than regression (Cervantes et al., 2020). In summary, although the pure machine learning models demonstrate relatively high performance in simulating *ET* and *T/ET* among the eight models, caution should be exercised regarding the reliability of their outputs when applied beyond the training dataset or under extreme conditions in practical applications.

4.2. Advantage of the proposed RNN-PM model

The proposed RNN-PM model, when compared to existing hybrid models that blend physical processes with machine learning, holds a distinct advantage in its ability to simulate the *ET* components (Zhao et al., 2019; Shang et al., 2023; Chen et al., 2023). Pure machine learning models operate as black boxes, mapping inputs directly to *ET* outputs without adhering to conservation laws, occasionally leading to biophysically implausible predictions. By embedding the neural networks into the P-M equation, the RNN-PM model enforces energy balance constraints and explicitly resolves intermediate conductance parameters (G_a , G_c , G_s). Additionally, the RNN-PM model strategically balances the simulation errors of *ET* and the *T/ET* within its loss function. Verification results demonstrate that the RNN-PM model maintains comparable error rates for both *ET* and *T/ET* simulations, suggesting that the loss function is effective in simultaneously constraining errors in the simulation of both variables.

The RNN-PM model also boasts a strong generalization potential and robust performance under extreme climate datasets, which is another notable advantage. A significant drawback of classic machine learning models is their poor predictive accuracy for data outside the training set, often yielding *ET* and *T/ET* simulation results that defy physical logic under extreme climate conditions. This issue stems from their reliance solely on data-driven processes to form nonlinear relationships between driving and predictor variables, while disregarding the physical principles underlying *ET*. When training datasets are limited in sample size or suffer from poor data quality, the performance of purely machine learning models declines significantly. In contrast, by integrating the energy balance and turbulent exchange processes of *ET*, the RNN-PM model moves beyond mere mechanical fitting. This enhancement not only improves the model's performance outside the training set and in extreme climatic scenarios but also effectively prevents the generation of physically implausible *ET* simulation results. Additionally, the proposed RNN-PM model offers the advantage of consistent performance across diverse vegetation conditions compared to models based solely on physical processes. Although its performance is slightly inferior to that of thermal infrared-based models under low *LAI* conditions and marginally less effective than conductance-based models under high *LAI* conditions, the RNN-PM model consistently ranks among the top-performing models and exhibits notable stability across a wide range

of surface conditions. This consistent performance enables the RNN-PM model to be effectively applied in ET simulations across a wide range of climates and vegetation types.

The last advantage of the RNN-PM model lies in its requirement for fewer, easily accessible driving variables. Specifically, the RNN-PM model employs commonly available meteorological variables (e.g., T_a , LST , R_a , VPD , and u), in conjunction with vegetation variables ($NDVI$ and LAI) and a soil variable (SM). These inputs are readily accessible: T_a , R_a , VPD , and u can be obtained from surface meteorological stations or various reanalysis datasets, while LST , SM , $NDVI$, and LAI are easily derived from satellite observations. This reduced reliance on hard-to-obtain input data enhances the model's applicability at both site and regional scales, making it less constrained by the availability of driving variables.

4.3. Model uncertainties

Despite its robust performance, the RNN-PM model is subject to several uncertainties that may impact the reliability of ET and T/ET simulations. The primary source of uncertainty arises from the assumptions underlying the dual-source P-M model. The classic P-M model presupposes that regional ET is a linear combination of evaporation and transpiration, without any interaction between the two (as per Eq. (1)). In reality, the coupling effect between evaporation and transpiration is significant (Dong et al., 2020; Xue et al., 2023). For example, vegetation can modify local microclimatic conditions through transpiration, leading to a reduction in local VPD , which subsequently decreases soil evaporation rates. While fully coupled schemes, such as the Shuttleworth-Wallace (S-W) model, theoretically resolve this interaction by introducing additional aerodynamic resistances between the soil-canopy-atmosphere nexus, we retained the uncoupled approach in this study. The S-W framework requires the simultaneous estimation of multiple interacting resistances between the soil-vegetation-atmosphere nexus. Without vertical temperature and humidity profile observations to constrain these internal variables, training a neural network to predict erroneous results. Therefore, the uncoupled P-M framework was selected as a strategic trade-off, offering sufficient physical constraints while maintaining a parsimonious structure suitable for data-driven parameterization. Another important limitation is that the RNN-PM model does not account for the contribution of canopy interception (I) to ET . In the current implementation, ET is represented as the sum of transpiration and soil evaporation only, and evaporation of intercepted water on wet canopies during and shortly after rainfall events is not treated as a separate term. As a result, the measured ET during wet periods includes an additional I that is either implicitly lumped into E and T or effectively omitted in the model. This mismatch can lead to a systematic underestimation of total ET and potential distortions in the inferred partitioning between T and E under wet-canopy conditions, particularly at densely vegetated sites.

To assess the potential impact of I on ET simulation, we conducted a simple sensitivity analysis in which model performance was re-evaluated after excluding time steps associated with precipitation and immediate post-rainfall periods. Removing these wet-canopy conditions led to a modest improvement in KGE and a reduction in bias for ET and T/ET , while leaving the overall performance ranking among models unchanged (Table S2). These results suggest that, although interception-related errors can be substantial at specific times, their influence on aggregate network-wide statistics is limited by the relatively small fraction of wet-canopy events in the three-year record. Looking ahead, a natural extension of the RNN-PM framework is to evolve from a dual-source to a three-source scheme in which canopy interception is represented explicitly as an additional component. This could be achieved by coupling RNN-PM with a simple interception module based on a bucket- or Gash-type model, in which a canopy water storage variable is driven by precipitation and depleted by wet-canopy evaporation. The latent heat flux associated with interception

evaporation would then be diagnosed separately from transpiration and soil evaporation, allowing the model to partition ET into T , E , and I and to reduce structural bias during rainfall events. Developing and testing such a three-source RNN-PM-I configuration is an important direction for future regional and global applications. Another limitation stems from the simplified representation of vegetation structure. While the RNN-PM model utilizes LAI as a primary structural descriptor, it omits other detailed canopy parameters such as vegetation height, displacement height, and roughness length. These parameters play a theoretical role in modulating aerodynamic resistance, particularly over rough canopies (e.g., forests). Their omission was a strategic choice to enhance data availability and generalization. However, this simplification may limit the model's ability to fully resolve G_a under complex phenological conditions or in ecosystems with distinct vertical structures. Future work should investigate whether integrating vertical structure data from spaceborne LiDAR (e.g., GEDI or ICESat-2) can further refine the estimation of aerodynamic conductance without compromising the model's broad applicability.

Besides, uncertainty also arises from the estimation of key parameters and the setting of empirical ones. Specifically, the model employs Beer's Law to partition net radiation (R_n) into soil (R_{ns}) and vegetation (R_{nv}) layers. This approach relies on the exponential relationship between R_n and LAI to allocate R_n (Norman et al., 1995). However, the assumption that R_{ns} and R_{nv} exhibit a simple exponential relationship with LAI across various vegetation types requires further validation. Chen et al. (2022) demonstrated that Beer's Law might systematically overestimate R_{nv} and underestimate R_{ns} in areas with higher LAI . Consequently, utilizing Beer's Law to decompose R_n introduces uncertainties in the simulation of ET and its components. Furthermore, while the RNN-PM model significantly reduces the number of empirical parameters, the remaining empirical ones still introduce errors into the model's results. For instance, we directly set the vegetation extinction coefficient (K_a) to 0.8, as recommended by the model developer (Norman et al., 1995; Leuning et al., 2008). However, K_a can vary substantially across different vegetation types (Campbell, 1986). To quantify the impact of this simplification, we conducted a sensitivity analysis by varying K_a from 0.2 to 1.6 (see Fig. S2 in Supplementary Material). The results indicate that the model output is moderately sensitive to K_a , with a $\pm 20\%$ variation in K_a leads to a mean deviation of 1.8% in total ET and 2.1% in the T/ET ratio. Interestingly, the model demonstrates a degree of resilience; the data-driven nature of the conductance sub-networks allows them to partially adapt to variations in the radiation partitioning scheme during the training process, thereby mitigating the error propagation compared to rigid physical models. Nevertheless, using a fixed K_a remains a source of uncertainty, particularly for ecosystems with extreme canopy structures.

In addition to uncertainties inherent in its mechanisms, the RNN-PM model also faces uncertainties arising from ground observations. The model utilizes three sub-RNN architectures to simulate the conductance parameters (G_a , G_s and G_c), which play a critical role in bridging the physical process model with the machine learning framework. However, the absence of ground-based observations for these conductance parameters limits the ability to effectively evaluate the accuracy of the three sub-RNN models. Moreover, errors are present in ET and T/ET observations using EC systems. A limitation of the EC system is its inability to effectively capture large eddies, primarily detecting small eddies near the surface, leading to energy non-closure issues, especially pronounced in heterogeneous underlying surfaces and nighttime (Liu et al., 2024). In this study, the average energy closure rate at 47 EC sites was 0.82. Although we applied a statistical model-based correction for the energy closure ratio, the error introduced by energy non-closure in the ET observation is still significant (Foken, 2008). On the other hand, the separation of ET components using high-frequency turbulence methods also entails uncertainties. The high-frequency turbulence method is predicated on the foundational premise that transpiration and photosynthesis are perfectly coupled, originating from coincident

sources and sinks (Scanlon & Sahu, 2008). This assumption of co-location, however, may become invalid in ecosystems with high vegetation heterogeneity. In such complex canopies, the source-sink processes for water and carbon fluxes can diverge, which in turn introduces systematic biases into the ET partitioning results. Consequently, future research is expected to incorporate multiple ET partitioning methods to develop a benchmark dataset for site-scale T/ET, thereby enhancing the reliability of RNN-PM model evaluation.

Lastly, uncertainty in the RNN-PM model also arises from the retrieval of several surface parameters based on satellite observations (e.g. SM, LAI, α , and LST). For instance, site-scale validation conducted at 47 NEON flux sites demonstrates that the retrieval accuracy of LAI, α , and LST from Landsat OLI imagery ranges between 0.80 and 0.95. Similar issues are present in SM estimation using the triangle feature space method, where errors due to inaccurately positioned extreme wet and dry boundaries are significant. Undoubtedly, these surface parameter inaccuracies from satellite observations will accumulate into the final ET and T/ET simulations through error propagation. Furthermore, although we applied a dynamic footprint model to match satellite inputs with EC observations, uncertainties associated with the footprint estimation itself remain a source of error. The accuracy of the footprint model relies on precise parameterizations of surface roughness and zero-plane displacement, which are often estimated empirically. Nevertheless, the use of footprint-weighted inputs significantly minimizes the spatial representation error compared to fixed-window approaches.

A further consideration is that the eight dual-source models differ not only in their underlying physical mechanisms but also in how they are parameterized and evaluated. The three purely physical models and the semi-empirical PT-JPL model rely on fixed parameterizations developed and tested in previous studies; in our experiments, these models were implemented following their original formulations and applied directly to all NEON sites without local calibration. In contrast, the RNN-PM model is explicitly trained against ET and T/ET observations and therefore requires an internal optimization procedure. To avoid overfitting and to obtain robust estimates of model skill, we employed repeated cross-validation, in which performance metrics are computed only on validation subsets that are not used for training. Additional sensitivity tests using a single train-validation split (Table S3) yielded similar performance metrics and did not alter the relative ranking among models. Thus, the comparison presented here should be interpreted as an intercomparison of practically deployable model configurations under a common forcing and evaluation dataset, rather than as an experiment in which all models share identical calibration machinery.

5. Conclusions

The present research constructed a novel hybrid framework (RNN-PM) that couples a physics (Penman-Monteith) model with machine learning for the explicit simulation of evapotranspiration (ET) and its components (E and T). The architecture leverages three RNN sub-networks to dynamically estimate key conductances (G_a , G_c , G_s), which are then supplied to the P-M model to calculate the resultant fluxes. Optimization and constraint of the model were performed using ET and T/ET observations from NEON sites. Its efficacy was quantified via a comprehensive benchmark against seven classic dual-source models, which were grouped as either purely process-based or data-driven. This comparative analysis was designed to investigate the physical mechanisms and uncertainties driving model performance discrepancies. The principal conclusions are:

Across the NEON flux sites, the RNN-PM model's simulations were highly consistent and accurate. For total ET, average performance

metrics were strong (KGE = 0.89, RMSE = 0.55 mm/day). The T/ET ratio was also reliably simulated, achieving an average KGE of 0.87 and an RMSE of 0.06.

The model's strong generalization capacity was confirmed by its accurate simulation of ET and T/ET outside the training dataset and during extreme climatic events. Furthermore, the model adeptly captured seasonal dynamics of ET and components. An input variable analysis identified the main drivers, ranked by importance: R_a , LAI, T_a , SM, NDVI, LST, VPD, α , and u .

The performance of eight dual-source models varies significantly across sites with diverse vegetation types and climatic conditions. Models based on temperature decomposition processes perform better in simulating ET and T/ET in sparsely vegetated areas but exhibit reduced accuracy in regions with dense vegetation. In contrast, models based on the conductance process perform more effectively in densely vegetated areas but are less suitable for sparsely vegetated regions. Although the three purely machine learning models provide relatively reliable simulations of ET and its components, they exhibit limited generalization capability and reduced accuracy under extreme climatic conditions. Among the eight models evaluated, the PT-JPL model demonstrated the poorest performance in simulating both ET and T/ET.

CRediT authorship contribution statement

Han Chen: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stephen Good:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Kelly Caylor:** Supervision, Resources, Project administration, Conceptualization. **Richard P. Fiorella:** Writing – review & editing, Writing – original draft, Conceptualization. **Lixin Wang:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The 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.

Acknowledgments

The authors acknowledge the financial support from the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Environmental Systems Science program under grant number DE-SC0024297. This work was supported by the U.S. Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). K.K.C. was additionally supported by a grant from the Zegar Family Foundation (SB200109). L.W. acknowledges funding support from the National Science Foundation grants (DEB-2307257 and DEB-2406931).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2026.134985>.

Data availability

All data to support the analysis can be accessed from the websites as follows: Partitioned ET components data of NEON stations based on high-frequency turbulence method: Zahn and Bou-Zeid, (2024) (<https://doi.org/10.5281/zenodo.12191876>); Environment variable data for NEON sites (NEON, 2024 a, b, c, d, e, f, g, h, i, j, k) (<https://data.neonscience.org/data-products/explore>); Landsat images (<https://ear.thexplorer.usgs.gov/>). The ET and T/ET datasets from NEON sites, generated by the eight dual-source models, are available on the Zenodo repository (<https://zenodo.org/records/17204669>).

References

- Anderegg, W.R., Trugman, A.T., Bowling, D.R., Salvucci, G., Tuttle, S.E., 2019. Plant functional traits and climate influence drought intensification and land-atmosphere feedbacks. Proc. Natl. Acad. Sci. 116 (28), 14071–14076. <https://doi.org/10.1073/pnas.1904747116>.
- Aouade, G., Jarlan, L., Ezzahar, J., Er-Raki, S., Napolý, A., Benkaddour, A., Boone, A., 2019. Evapotranspiration partition using the multiple energy balance version of the ISBA-Ag's land surface model over two irrigated crops in a semi-arid Mediterranean region (Marrakech, Morocco). Hydrol. Earth Syst. Sci. Discuss. 2019, 1–44. <https://doi.org/10.5194/hess-24-3789-2020>.
- Agarap, A.F., 2018. Deep learning using rectified linear units (relu). arXiv preprint arXiv: 1803.08375. <https://doi.org/10.48550/arXiv.1803.08375>.
- Azemi, N.M., Dominic, P.D.D., 2019. Hyperparameter optimization in convolutional neural network using genetic algorithms. Int. J. Adv. Comput. Sci. Appl. 10 (6). <https://doi.org/10.14569/IJACSA.2019.0100638>.
- Bambach, N., Kustas, W., Alfieri, J., Prueger, J., Hippis, L., McKee, L., McElrone, A.J., 2022. Evapotranspiration uncertainty at micrometeorological scales: the impact of the eddy covariance energy imbalance and correction methods. Irrig. Sci. 40 (4), 445–461. <https://doi.org/10.1007/s00271-022-00783-1>.
- Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions. In: Biggins, J. (eds) Progress in Photosynthesis Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-0519-6_48.
- Campbell, G.S., 1986. Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. Agric. For. Meteorol. 36 (4), 317–321. [https://doi.org/10.1016/0168-1923\(86\)90010-9](https://doi.org/10.1016/0168-1923(86)90010-9).
- Carlson, T., 2007. An overview of the "triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. Sensors 7 (8), 1612–1629. <https://doi.org/10.3390/s7081612>.
- Cervantes, J., García-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408, 189–215. <https://doi.org/10.1016/j.neucom.2019.10.118>.
- Chappell, A., Van Pelt, S., Zobeck, T., Dong, Z., 2010. Estimating aerodynamic resistance of rough surfaces using angular reflectance. Remote Sens. Environ. 114 (7), 1462–1470. <https://doi.org/10.1016/j.rse.2010.01.025>.
- Chen, H., Huang, J.J., Berg, A., McBean, E., 2020. Development of a trapezoidal framework-based model (PCALEP) for partition of land evapotranspiration. J. Hydrol. 589, 124994. <https://doi.org/10.1016/j.jhydrol.2020.124994>.
- Chen, H., Wei, Y., 2025. Declining soil evaporation on a drying earth. Agric. For. Meteorol. 368, 110550. <https://doi.org/10.1016/j.agrformet.2025.110550>.
- Chen, H., Huang, J.J., McBean, E., Singh, V.P., 2021a. Evaluation of alternative two-source remote sensing models in partitioning of land evapotranspiration. J. Hydrol. 597, 126029. <https://doi.org/10.1016/j.jhydrol.2021.126029>.
- Chen, H., Jiang, A.Z., Huang, J.J., Li, H., McBean, E., Singh, V.P., Zhou, Z., 2022. An enhanced shuttleworth-wallace model for simulation of evapotranspiration and its components. Agric. For. Meteorol. 313, 108769. <https://doi.org/10.1016/j.agrformet.2021.108769>.
- Chen, Y., Cao, R., Chen, J., Liu, L., Matsushita, B., 2021b. A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky–Golay filter. ISPRS J. Photogramm. Remote Sens. 180, 174–190. <https://doi.org/10.1016/j.isprsjprs.2021.08.015>.
- Colaizzi, P.D., Agam, N., Tolk, J.A., Evett, S.R., Howell, T.A., Gowda, P.H., Anderson, M.C., 2014. Two-source energy balance model to calculate E, T, and ET: Comparison of Priestley-Taylor and Penman-Monteith formulations and two time scaling methods. Trans. ASABE 57(2), 10, 479–498. <https://doi.org/10.13031/trans.57.10423>.
- Crago, R.D., Szilagyi, J., Qualls, R.J., 2023. What is the Priestley-Taylor wet-surface evaporation parameter? Testing four hypotheses. Hydrol. Earth Syst. Sci. 27 (17), 3205–3220. <https://doi.org/10.5194/hess-27-3205-2023>.
- Dhungel, R., Aiken, R., Evett, S.R., Colaizzi, P.D., Marek, G., Moorhead, J.E., Lin, X., 2021. Energy imbalance and evapotranspiration hysteresis under an advective environment: evidence from lysimeter, eddy covariance, and energy balance modeling. Geophys. Res. Lett. 48 (1), e2020GL091203. <https://doi.org/10.1029/2020GL091203>.
- Dong, J., Dirmeyer, P.A., Lei, F., Anderson, M.C., Holmes, T.R., Hain, C., Crow, W.T., 2020. Soil evaporation stress determines soil moisture-evapotranspiration coupling strength in land surface modeling. Geophys. Res. Lett. 47 (21), e2020GL090391. <https://doi.org/10.1029/2020GL090391>.
- Drucker, H., Burges, C.J., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. Adv. Neural Inf. Proces. Syst. 9, 155–161.
- Ershadi, A., McCabe, M.F., Evans, J.P., Wood, E.F., 2015. Impact of model structure and parameterization on Penman-Monteith type evaporation models. J. Hydrol. 525, 521–535. <https://doi.org/10.1016/j.jhydrol.2015.04.008>.
- Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An overview of global leaf area index (LAI): methods, products, validation, and applications. Rev. Geophys. 57 (3), 739–799. <https://doi.org/10.1029/2018RG000608>.
- Farquhar, G.D., Sharkey, T.D., 1982. Stomatal conductance and photosynthesis. Annu. Rev. Plant Physiol. 33 (1), 317–345.
- Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112 (3), 901–919. <https://doi.org/10.1016/j.rse.2007.06.025>.
- Foken, T., 2008. The energy balance closure problem: an overview. Ecol. Appl. 18 (6), 1351–1367. <https://doi.org/10.1890/06-0922.1>.
- Guzinski, R., Nieto, H., Stisen, S., Fensholt, R., 2015. Inter-comparison of energy balance and hydrological models for land surface energy flux estimation over a whole river catchment. Hydrol. Earth Syst. Sci. 19 (4), 2017–2036. <https://doi.org/10.5194/hess-19-2017-2015>.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Jaafer, H.H., Mourad, R.M., Kustas, W.P., Anderson, M.C., 2022. A global implementation of single-and dual-source surface energy balance models for estimating actual evapotranspiration at 30-m resolution using google earth engine. Water Resour. Res. 58 (11), e2022WR032800. <https://doi.org/10.1029/2022WR032800>.
- Jarvis, P., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. Philos. Trans. R. Soc. London. B, Biol. Sci. 273 (927), 593–610. <https://doi.org/10.1098/rstb.1976.0035>.
- Jimenez-Munoz, J.C., Sobrino, J.A., Skoković, D., Mattar, C., Cristobal, J., 2014. Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. IEEE Geosci. Remote Sens. Lett. 11 (10), 1840–1843.
- Keller, M., Schimel, D.S., Hargrove, W.W., Hoffman, F.M., 2008. A continental strategy for the national ecological observatory network. Ecol. Soc. Am. 282–284.
- Kendall, A., Gal, Y., Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7482–7491.
- Kool, D., Agam, N., Lazarovitch, N., Heitman, J.L., Sauer, T.J., Ben-Gal, A., 2014. A review of approaches for evapotranspiration partitioning. Agric. For. Meteorol. 184, 56–70. <https://doi.org/10.1016/j.agrformet.2013.09.003>.
- Kool, D., Kustas, W.P., Ben-Gal, A., Agam, N., 2021. Energy partitioning between plant canopy and soil, performance of the two-source energy balance model in a vineyard. Agric. For. Meteorol. 300, 108328. <https://doi.org/10.1016/j.agrformet.2021.108328>.
- Koppa, A., Rains, D., Hulsman, P., Poyatos, R., Miralles, D.G., 2022. A deep learning-based hybrid model of global terrestrial evaporation. Nat. Commun. 13 (1), 1912. <https://doi.org/10.1038/s41467-022-29543-7>.
- Kustas, W.P., Alfieri, J.G., Anderson, M.C., Colaizzi, P.D., Prueger, J.H., Evett, S.R., Howell, T.A., 2012. Evaluating the two-source energy balance model using local thermal and surface flux observations in a strongly advective irrigated agricultural area. Adv. Water Resour. 50, 120–133. <https://doi.org/10.1016/j.advwatres.2012.07.005>.
- Kustas, W., Anderson, M., 2009. Advances in thermal infrared remote sensing for land surface modeling. Agric. For. Meteorol. 149 (12), 2071–2081. <https://doi.org/10.1016/j.agrformet.2009.05.016>.
- Kljun, N., Calanca, P., Rotach, M.W., Schmid, H.P., 2015. A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). Geosci. Model Dev. 8 (11), 3695–3713. <https://doi.org/10.5194/gmd-8-3695-2015>.
- Leuning, R., Zhang, Y.Q., Rajaud, A., Cleugh, H., Tu, K., 2008. A simple surface conductance model to estimate regional evaporation using modis leaf area index and the penman-monteith equation. Water Resour. Res. 44, W10419. <https://doi.org/10.1029/2007wr006562>.
- Lee, J.E., Lintner, B.R., Neelin, J.D., Jiang, X., Gentine, P., Boyce, C.K., Worden, J., 2012. Reduction of tropical land region precipitation variability via transpiration. Geophys. Res. Lett. 39 (19). <https://doi.org/10.1029/2012GL053417>.
- Lhomme, J.P., 1997. A theoretical basis for the Priestley-Taylor coefficient. Bound.-Lay. Meteorol. 82 (2), 179–191. <https://doi.org/10.1023/A:1000281114105>.
- Liang, S., Strahler, A.H., Walther, C., 1999. Retrieval of land surface albedo from satellite observations: a simulation study. J. Appl. Meteorol. 38 (6), 712–725. [https://doi.org/10.1175/1520-0450\(1999\)038<0712:ROLSAF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<0712:ROLSAF>2.0.CO;2).
- Liu, H., Liu, C., Huang, J., Desai, A.R., Zhang, Q., Ghannam, K., Katul, G.G., 2024. Scalar flux profiles in the unstable atmospheric surface layer under the influence of large eddies: implications for eddy covariance flux measurements and the non-closure problem. Geophys. Res. Lett. 51 (1), e2023GL106649. <https://doi.org/10.1029/2023GL106649>.
- Liu, H., Gong, P., Wang, J., Clinton, N., Bai, Y., Liang, S., 2020. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. Earth Syst. Sci. Data 12 (2), 1217–1243. <https://doi.org/10.5194/essd-12-1217-2020>.
- Liu, S., Lu, L., Mao, D., Jia, L., 2007. Evaluating parameterizations of aerodynamic resistance to heat transfer using field measurements. Hydrol. Earth Syst. Sci. 11 (2), 769–783. <https://doi.org/10.5194/hess-11-769-2007>.

- Marshall, M., Tu, K., Andreo, V., 2020. On parameterizing soil evaporation in a direct remote sensing model of ET: PT-JPL. *Water Resour. Res.* 56 (5), e2019WR026290. <https://doi.org/10.1029/2019WR026290>.
- Massari, C., Avanzi, F., Bruno, G., Gabellani, S., Penna, D., Camici, S., 2022. Evaporation enhancement drives the European water-budget deficit during multi-year droughts. *Hydrol. Earth Syst. Sci.* 26 (6), 1527–1543. <https://doi.org/10.5194/hess-24-2687-2020>.
- Massari, C., Modanesi, S., Dari, J., Gruber, A., De Lannoy, G.J., Girotto, M., Brocca, L., 2021. A review of irrigation information retrievals from space and their utility for users. *Remote Sens. (Basel)* 13 (20), 4112. <https://doi.org/10.3390/rs13204112>.
- Massman, W.J., Lee, X., 2002. Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. *Agric. For. Meteorol.* 113 (1–4), 121–144. [https://doi.org/10.1016/S0168-1923\(02\)00105-3](https://doi.org/10.1016/S0168-1923(02)00105-3).
- Mauder, M., Cuntz, M., Drije, C., Graf, A., Rebmann, C., Schmid, H.P., Steinbrecher, R., 2013. A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements. *Agric. For. Meteorol.* 169, 122–135. <https://doi.org/10.1016/j.agrformet.2012.09.006>.
- Medlyn, B.E., Duursma, R.A., Eamus, D., Ellsworth, D.S., Prentice, I.C., Barton, C.V., Wingate, L., 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Glob. Chang. Biol.* 17 (6), 2134–2144. <https://doi.org/10.1111/j.1365-2486.2010.02375.x>.
- Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., Stauch, V.J., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. For. Meteorol.* 147 (3–4), 209–232. <https://doi.org/10.1016/j.agrformet.2007.08.011>.
- Monteith, J.L., 1965. Evaporation and environment. In: *Symposia of the society for experimental biology* (Vol. 19, pp. 205–234). Cambridge University Press (CUP) Cambridge.
- Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* 111 (4), 519–536. <https://doi.org/10.1016/j.rse.2007.04.015>.
- Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* 115 (8), 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>.
- Moran, M.S., Clarke, T.R., Inoue, Y., Vidal, A., 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sens. Environ.* 49 (3), 246–263. [https://doi.org/10.1016/0034-4257\(94\)90020-5](https://doi.org/10.1016/0034-4257(94)90020-5).
- Nourani, V., Elkirian, G., Abdullahi, J., 2019. Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J. Hydrol.* 577, 123958. <https://doi.org/10.1016/j.jhydrol.2019.123958>.
- Nourani, V., Ahmad, R., Zhang, Y., Dąbrowska, D., 2025. Ensemble machine learning-based extrapolation of Penman-Monteith-Leuning evapotranspiration data. *Ecol. Ind.* 170, 113012. <https://doi.org/10.1016/j.ecolind.2024.113012>.
- National Ecological Observatory Network (NEON): Triple aspirated air temperature (DP1.00003.001), 2024a. <https://data.neonscience.org/> /data-products/DP1.00003.001.
- National Ecological Observatory Network (NEON): Triple aspirated air temperature (DP1.00003.001), 2024b. <https://doi.org/10.48443/PDE7-K607>.
- National Ecological Observatory Network (NEON): Relative humidity (DP1.00098.001), 2024c. <https://data.neonscience.org/data-products/DP1.00098.001>.
- National Ecological Observatory Network (NEON): Relative humidity (DP1.00098.001), 2024d. <https://doi.org/10.48443/K9VK-5K27>.
- National Ecological Observatory Network (NEON): Shortwave and longwave radiation (net radiometer) (DP1.00023.001), 2024e. <https://data.neonscience.org/data-products/DP1.00023.001>.
- National Ecological Observatory Network (NEON): Shortwave and longwave radiation (net radiometer) (DP1.00023.001), 530.2024f. <https://doi.org/10.48443/9QPC-5V70>.
- National Ecological Observatory Network (NEON): 2D wind speed and direction (DP1.00001.001), 2024g. <https://doi.org/10.48443/YEX7-7Z81>.
- National Ecological Observatory Network (NEON): 2D wind speed and direction (DP1.00001.001), 2024h. <https://data.neonscience.org/> /data-products/DP1.00001.001.
- National Ecological Observatory Network (NEON): Soil water content and water salinity (DP1.00094.001), provisional data. 2024i. <https://data.neonscience.org/data-products/DP1.00094.001>.
- National Ecological Observatory Network (NEON): Bundled data products – eddy covariance (DP4.00200.001), 2024j. <https://doi.org/10.48443/J9PT-M241>.
- National Ecological Observatory Network (NEON): Bundled data products – eddy covariance (DP4.00200.001), 2024k. <https://data.neonscience.org/data-products/DP4.00200.001>.
- Nelson, J.A., Pérez-Priego, O., Zhou, S., Poyatos, R., Zhang, Y., Blanken, P.D., Jung, M., 2020. Ecosystem transpiration and evaporation: Insights from three water flux partitioning methods across FLUXNET sites. *Glob. Chang. Biol.* 26 (12), 6916–6930. <https://doi.org/10.1111/gcb.15314>.
- Norman, J.M., Kustas, W.P., Humes, K.S., 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agric. For. Meteorol.* 77 (3–4), 263–293. [https://doi.org/10.1016/0168-1923\(95\)02265-Y](https://doi.org/10.1016/0168-1923(95)02265-Y).
- Nyolei, D., Diels, J., Mbilinyi, B., Mbungu, W., van Griensven, A., 2021. Evapotranspiration simulation from a sparsely vegetated agricultural field in a semi-arid agro-ecosystem using Penman-Monteith models. *Agric. For. Meteorol.* 303, 108370. <https://doi.org/10.1016/j.agrformet.2021.108370>.
- Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Running, S.W., 2020. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrol. Earth Syst. Sci.* 24 (3), 1485–1509. <https://doi.org/10.5194/hess-24-1485-2020>.
- Parlange, J.Y., Waggoner, P.E., 1970. Stomatal dimensions and resistance to diffusion. *Plant Physiol.* Aug 1;46(2):337-42. <https://doi.org/10.1104/pp.46.2.337>.
- Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. *Proc. R. Soc. Lond. A* 193 (1032), 120–145. <https://doi.org/10.1098/rspa.1948.0037>.
- Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* 100 (2), 81–92. [https://doi.org/10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2).
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. In: *Advances in Neural Information Processing Systems*, p. 31.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566 (7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>.
- Rong, Y., Wang, W., Wu, P., Wang, P., Zhang, C., Wang, C., Huo, Z., 2024. A novel hybrid deep learning framework for evaluating field evapotranspiration considering the impact of soil salinity. *Water Resour. Res.* 60 (9), e2023WR036809. <https://doi.org/10.1029/2023WR036809>.
- Scanlon, T.M., Kustas, W.P., 2010. Partitioning carbon dioxide and water vapor fluxes using correlation analysis. *Jan* 15;150(1):89–99 *Agric. Forest Meteorol.* <https://doi.org/10.1016/j.agrformet.2009.09.005>.
- Scanlon, T.M., Sahu, P., 2008. On the correlation structure of water vapor and carbon dioxide in the atmospheric surface layer: a basis for flux partitioning. *Water Resour. Res.* 44 (10). <https://doi.org/10.1029/2008WR006932>.
- Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J.B., Zhang, L., 2023. Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau. *Remote Sens. Environ.* 289, 113519. <https://doi.org/10.1016/j.rse.2023.113519>.
- Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D* 404, 132306.
- Singh, D., Singh, B., 2020. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* 97, 105524. <https://doi.org/10.1016/j.asoc.2019.105524>.
- Stannard, D.I., 1993. Comparison of Penman-Monteith, Shuttleworth-Wallace, and modified Priestley-Taylor evapotranspiration models for wildland vegetation in semiarid rangeland. *Water Resour. Res.* 29 (5), 1379–1392. <https://doi.org/10.1029/93WR00333>.
- Tran, B.N., Van Der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G., Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. *Hydrol. Earth Syst. Sci.* 27 (24), 4505–4528. <https://doi.org/10.5194/hess-27-4505-2023>.
- Timmermans, W.J., Kustas, W.P., Anderson, M.C., French, A.N., 2007. An intercomparison of the surface energy balance algorithm for land (SEBAL) and the two-source energy balance (TSEB) modeling schemes. *Remote Sens. Environ.* 108 (4), 369–384. <https://doi.org/10.1016/j.rse.2006.11.028>.
- Torres, A.F., Walker, W.R., McKee, M., 2011. Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agric. Water Manage.* 98 (4), 553–562. <https://doi.org/10.1016/j.agwat.2010.10.012>.
- van de Griend, A.A., Owe, M., 1994. Bare soil surface resistance to evaporation by vapor diffusion under semiarid conditions. *Water Resour. Res.* Feb;30(2):181-8. <https://doi.org/10.1029/93WR02747>.
- Van Olden, A.P., Holtlag, A.A.M., 1985. Estimation of atmospheric boundary layer parameters for diffusion applications. *J. Appl. Meteorol. Climatol.* 24 (11), 1196–1207. [https://doi.org/10.1175/1520-0450\(1985\)024<1196:EOABLP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1985)024<1196:EOABLP>2.0.CO;2).
- Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. *Rev. Geophys.* 50 (2). <https://doi.org/10.1029/2011RG000373>.
- Wei, Z., Yoshimura, K., Wang, L., Miralles, D.G., Jasechko, S., Lee, X., 2017. Revisiting the contribution of transpiration to global terrestrial evapotranspiration. *Geophys. Res. Lett.* 44 (6), 2792–2801. <https://doi.org/10.1002/2016GL072235>.
- Xue, K., Song, L., Xu, Y., Liu, S., Zhao, G., Tao, S., Zhao, L., 2023. Estimating ecosystem evaporation and transpiration using a soil moisture coupled two-source energy balance model across FLUXNET sites. *Agric. For. Meteorol.* 337, 109513. <https://doi.org/10.1016/j.agrformet.2023.109513>.
- Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* 415, 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>.
- Yang, Y., Roderick, M.L., Guo, H., Miralles, D.G., Zhang, L., Fatichi, S., Yang, D., 2023. Evapotranspiration on a greening Earth. *Nat. Rev. Earth & Environ.* 4 (9), 626–641. <https://doi.org/10.1038/s43017-023-00464-3>.
- Yang, Y., Long, D., Guan, H., Liang, W., Simmons, C., Batelaan, O., 2015a. Comparison of three dual-source remote sensing evapotranspiration models during the MUSOEXE-12 campaign: Revisit of model physics. *Water Resour. Res.* 51 (5), 3145–3165. <https://doi.org/10.1002/2014WR015619>.
- Yang, Y., Qiu, J., Zhang, R., Huang, S., Chen, S., Wang, H., Fan, Y., 2018. Intercomparison of three two-source energy balance models for partitioning evaporation and transpiration in semiarid climates. *Remote Sens. (Basel)* 10 (7), 1149. <https://doi.org/10.3390/rs10071149>.
- Yang, Y., Su, H., Zhang, R., Tian, J., Li, L., 2015b. An enhanced two-source evapotranspiration model for land (ETEML): Algorithm and evaluation. *Remote Sens. Environ.* 168, 54–65. <https://doi.org/10.1016/j.rse.2015.06.020>.

- Yao, Y., Rosasco, L., Caponnetto, A., 2007. On early stopping in gradient descent learning. *Constr. Approx.* 26, 289–315. <https://doi.org/10.1007/s00365-006-0663-2>.
- Zahn, E., Bou-Zeid, E., 2024. Partitioning of water and CO₂ fluxes at NEON sites into soil and plant components: a five-year dataset for spatial and temporal analysis. *Earth Syst. Sci. Data*. <https://doi.org/10.5194/essd-2024-272>.
- Zahn, E., Bou-Zeid, E., Good, S.P., Katul, G.G., Thomas, C.K., Ghannam, K., Kustas, W.P., 2022. Direct partitioning of eddy-covariance water and carbon dioxide fluxes into ground and plant components. *Agric. For. Meteorol.* 315, 108790. <https://doi.org/10.1016/j.agrformet.2021.108790>.
- Zhao, W.L., Gentile, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Qiu, G.Y., 2019. Physics-constrained machine learning of evapotranspiration. *Geophys. Res. Lett.* 46 (24), 14496–14507. <https://doi.org/10.1029/2019GL085291>.