https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Physically constrained multi-task learning for hourly joint estimation of evapotranspiration and transpiration from sparse sap-flow data

Hongyang Fu[1], Yuanyuan Yang[1*], Dengfeng Liu[1*], Qiang Li[2], Zhihua He[3], Huimin Lei[4], Mohd Yawar Ali Khan[5], Fiaz Hussain[6]

[1]State Key Laboratory of Water Engineering Ecology and Environment in Arid Area, Xi'an University of Technology, Xi'an, 710048, China
[2]State Key Laboratory of Soil and Water Conservation and Desertification Control, College of Forestry, Northwest A&F University, Yangling 712100, China
[3]School of Civil Engineering, Sun-Yat-Sen University, Zhuhai 519000, China
[4]State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China
[5]Department of Hydrogeology, Faculty of Earth Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[6]Department of Land and Water Conservation Engineering, Faculty of Agricultural Engineering and Technology, PMAS-Arid Agriculture University Rawalpindi, Rawalpindi 46300, Pakistan

Correspondence to: Yuanyuan Yang (yuanyuanyang@xaut.edu.cn) and Dengfeng Liu (liudf@xaut.edu.cn)

Abstract. Hourly evapotranspiration (*ET*) estimation helps resolve rapid land-surface water and energy responses to radiation, atmospheric dryness, and aerodynamic forcing. However, the same change in total *ET* may arise from different combinations of vegetation transpiration, soil evaporation, and canopy interception evaporation, making it difficult to interpret transpiration and its sub-daily variability when *ET* is modeled only as a single bulk flux. Sap-flow observations provide direct information on transpiration (*T*), but their site coverage is sparse and tree-to-site upscaling remains uncertain. To address this issue, we integrate FLUXNET2015, SAPFLUXNET, and GIMMS LAI4g data to develop a physically constrained multi-task learning framework, termed MLF-ETT, for hourly joint estimation of *ET* and *T*. MLF-ETT estimates total *ET* and the transpiration fraction *T*/*ET*, deriving *T* as a bounded component of total *ET* under $0 \le T \le ET$, so that limited *T* data contribute to joint *ET*-*T* learning rather than serving only as isolated *T* targets. Compared with the baselines of single-task XGBoost and multilayer perceptron models, the multi-task framework primarily improved *T* estimation, particularly under within-site temporal generalization and high evaporative-demand conditions characterized by high vapor pressure deficit, high air temperature, or both. Cross-site spatial generalization showed stronger site dependence, indicating that the transferability of sparse *T* supervision remained limited by cross-site process differences and uncertainty in sap-flow-derived *T* data. Input allocation between shared and task-specific branches strongly affected model performance, whereas increasing the *T*-supervision weight alone did not consistently improve performance. Overall, the framework incorporates limited sap-flow-derived *T* data into hourly joint *ET*-*T* learning and estimates *T* as a physically bounded component of total *ET*, providing a constrained reference for sub-daily *ET* partitioning.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

1 Introduction

Evapotranspiration (*ET*) is a key variable linking the terrestrial water and energy cycles (Oki and Kanae, 2006) and influences regional climate, boundary-layer development, and ecosystem functioning through land-atmosphere feedbacks (Fisher et al., 2017; Liu and Wang, 2025). Compared with daily or coarser estimates, hourly *ET*-related flux estimates can better resolve rapid responses of land-surface water and energy exchanges to radiation, humidity, and aerodynamic forcing (Althoff et al., 2019; Zhong et al., 2019). This sub-daily perspective is particularly important for irrigation scheduling and short-term diagnostics of land-atmosphere water and energy exchanges (Tawegoum et al., 2015).

Existing methods for estimating hourly actual *ET* can be broadly grouped into 2 methodological streams. Hereafter, *ET* refers to actual evapotranspiration. The first includes physically based and process-guided approaches, which estimate latent heat flux through surface-energy-balance closure (Allen et al., 2007; Hashem et al., 2020; Cheng et al., 2026) or derive hourly *ET* using resistance parameterizations and turbulent-flux frameworks (Gowda et al., 2013; Su, 2002); they can also estimate or simulate *ET* using canopy radiative transfer models (Duarte Rocha et al., 2022), numerical weather prediction or geostationary-satellite forcing (S. Ha et al., 2020), flux-equilibrium constraints (Pan et al., 2024; Zheng et al., 2025), and crop-coefficient schemes (Man et al., 2025). The second includes machine-learning and hybrid data-driven approaches, which learn nonlinear relationships between *ET* and multi-source predictors from site observations, meteorological data, and remote sensing, and have been applied to regional hourly reconstruction (Wang et al., 2023), global flux products (Nelson et al., 2024), cross-site extrapolation (Shi and Cai, 2025), thermal-remote-sensing and machine-learning fusion (Reyes Rojas et al., 2021), physics-constrained machine learning (Zhang et al., 2025), field-scale agricultural *ET* estimation (Zhuang et al., 2025), and multi-source-observation-based *ET* estimation in complex ecosystems (Guo et al., 2023). Despite these advances, most existing hourly *ET* modeling studies still treat *ET* as a single bulk flux, while *ET* partitioning is often conducted as a separate analysis rather than being integrated into the *ET* estimation task.

When *ET* is modeled only as an integrated bulk flux, it is difficult to distinguish vegetation water use from soil evaporation or to explain the mechanisms underlying sub-daily flux variability (Stoy et al., 2019). Partitioning *ET* into evaporation and transpiration (*T*), therefore, provides a more process-relevant basis for characterizing canopy regulation (Kool et al., 2014; Grossiord et al., 2020) and for interpreting transpiration as a key component of vegetation water use (Schlesinger and Jasechko, 2014). However, incorporating *T* information into hourly *ET* modeling remains constrained by the limited availability of independent *T* observations. Sap flow observations have long been used to quantify *T*, and SAPFLUXNET has substantially improved data availability by integrating previously fragmented measurements into a standardized global database (Poyatos et al., 2021; Poyatos et al., 2016). However, existing studies have mainly used *T* information for *ET* partitioning or for analyzing long-term variability in the ratio of *T* to *ET* (*T*/*ET*) (Kar et al., 2025; Yu et al., 2022; Xiang et al., 2025; Zhou et al., 2016). It remains unclear how site-limited sap-flow-derived *T* labels can be incorporated into hourly joint *ET*-*T* modeling, under what conditions such information improves *T* estimation, and whether it can support *T* estimation at sites without local *T* measurements.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Multi-task learning (MTL) provides a suitable framework for hourly joint *ET*-*T* modeling. It can learn nonlinear responses of target variables to environmental drivers while exploiting shared information across related tasks and retaining task-specific differences (Zhang and Yang, 2022). In hydrology and environmental sciences, MTL has been used for joint prediction of related hydrological and environmental variables (Yan et al., 2026), including streamflow and water temperature (Sadler et al., 2022), meteorological variables (Ouyang et al., 2025), reference evapotranspiration (Ferreira et al., 2022), and rainfall-runoff processes (Li et al., 2023). These studies demonstrate the broader value of MTL for joint environmental modeling. Because transpiration is a major but variable component of terrestrial *ET*, it also fits the theoretical basis of applying multi-task models to related estimation targets (Wei et al., 2017; Caruana, 1997).

Here, we develop a physically constrained multi-task learning framework, termed MLF-ETT, for hourly joint *ET*-*T* estimation by integrating eddy-covariance-based *ET* references, sap-flow-derived *T* labels, and GIMMS LAI4g canopy information. The contributions of this study are threefold. First, it incorporates site-limited, sap-flow-derived *T* labels into multi-task learning to test whether *T* information improves *T* estimation and supports *ET* learning across different generalization settings. Second, it uses an *ET* and *T*/*ET* output structure to derive *T* as a bounded component of total *ET* under $0 \le T \le ET$, enabling physically constrained *T* estimates at *ET*-only sites. Third, it evaluates how environmental-condition windows, input feature combinations, and *T*-supervision strength affect model performance under within-site temporal and cross-site spatial generalization. Overall, MLF-ETT extends hourly *ET* modeling from single-output bulk-flux estimation toward physically constrained joint *ET*-*T* estimation when *T* labels are available only at a subset of sites.

2 Data

2.1 Data sources

This study used 3 complementary datasets: FLUXNET2015, SAPFLUXNET, and GIMMS LAI4g. FLUXNET2015 (https://fluxnet.org/data/fluxnet2015-dataset/; accessed on 20 April 2026) provided half-hourly or hourly eddy-covariance fluxes and environmental variables, including air temperature (T_a), vapor pressure deficit (*VPD*), air pressure (P_a), wind speed (*WS*), relative humidity (*RH*), net radiation (R_n), gross primary productivity (*GPP*), and latent heat flux (*LE*) (Pastorello et al., 2020). *GPP* was represented by *GPP*_NT_VUT_REF, derived using the nighttime respiration partitioning method.

SAPFLUXNET (https://sapfluxnet.creaf.cat/; accessed on 20 April 2026) provided hourly sap-flow observations and metadata on site conditions, stand structure, species attributes, measurement methods, and observation coverage (Poyatos et al., 2021; Poyatos et al., 2016). These data were used to derive site-scale *T* reference values from tree-level measurements. Given the limited site coverage and methodological heterogeneity of SAPFLUXNET, the derived *T* values were treated as site-limited *T* reference values, and observation coverage was used for sample weighting during model training.

Canopy structural information was obtained from GIMMS LAI4g (https://zenodo.org/records/8281930; accessed on 20 April 2026), which provides global leaf area index (*LAI*) data for vegetated regions from 1982 to 2020 at a semi-monthly

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

temporal resolution and a 1/12° spatial resolution (Cao et al., 2023). *LAI* was used to characterize seasonal changes in canopy structure.

2.2 Data processing

2.2.1 *ET* reference

Hourly *ET* reference values were derived from the energy-balance-corrected latent heat flux (*LE*_CORR) in FLUXNET2015 (Eqs. (1) and (2)). Quality control was adapted from Rong et al. (2024) and Li et al. (2019). Records with missing variables or low-quality flags were removed, and precipitation periods and the following 6 h were excluded to reduce the effects of rainfall, canopy interception, and post-rain evaporation. The analysis was restricted to daytime conditions. Additional filters were applied to remove periods with poor flux representativeness, including $RH > 95$ %, $R_n < 30$ W m^{-2}, and negative *GPP*, *LE*, or *VPD*. Half-hourly records were then aggregated to hourly resolution.

$$ET = \frac{LE \times 3600}{\lambda}, \quad (1)$$

$$\lambda = 2.500 \times 10^6 - 2.386 \times 10^3 T_a, \quad (2)$$

where *LE* is the hourly mean latent heat flux (W m^{-2}); T_a is the hourly mean air temperature (°C); λ is the latent heat of vaporization (Henderson-Sellers, 1984) (J kg^{-1}); and *ET* is expressed in mm h^{-1}. The constants in Eq. (2) have units of J kg^{-1} and J kg^{-1} °C^{-1}, respectively.

2.2.2 *T* reference

Hourly *T* reference values were derived by upscaling SAPFLUXNET tree-level sap-flow measurements to site-level estimates following Bright et al. (2022). To reduce methodological inconsistency, only sites measured with the heat dissipation (HD) method were retained (De Aguiar-Campos et al., 2025). In the absence of sufficient site-specific calibration information, sap-flow density was adjusted using the default correction factor of 1.405 for HD measurements (Flo et al., 2019).

Because sap-flow-derived *T* estimates may be inconsistent with independent *ET* references at larger spatial scales (Ford et al., 2007). The derived *T* reference values were further screened to reduce the influence of sap-flow upscaling bias on *T* supervision. Sites for which more than 5 % of candidate samples had $T > ET$ were regarded as potentially affected by systematic overestimation. For each affected site, a site-specific multiplicative scaling factor was applied to the *T* reference series to reduce the proportion of samples with $T > ET$ to 5 %. This rescaling was applied uniformly to all *T* reference values at the affected site and did not modify the *ET* reference values. This procedure reduced clear-site-scale inconsistency in sap-flow-derived *T* reference values, and the remaining 5 % of samples with $T > ET$ were retained during training via coverage-based sample weighting.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

2.2.3 *LAI* processing and multi-source alignment

Because GIMMS LAI4g is provided at a semi-monthly resolution, each *LAI* value was assigned to hourly flux and sap-flow records within the corresponding half-month period. Missing or invalid *LAI* records were removed. The processed *ET*, *T*, input features, and *LAI* were then timestamp-aligned to generate the final hourly sample set.

2.3 Study sites

After quality control, hourly aggregation, and *LAI* matching, 112 FLUXNET2015 sites contained at least 2000 hours of valid *ET*. Among sites with co-located sap-flow observations, we retained only HD sites with at least 1000 aligned *T*-labeled samples. This screening yielded 7 *T*-labeled sites, defined here as sites with paired *ET* and sap-flow-derived *T* reference values; sites without usable *T* labels but with *ET* reference values were defined as *ET*-only sites. For model evaluation, 2 experimental datasets were defined: 25 evergreen needleleaf forest (ENF) sites for cross-site spatial generalization and 7 *T*-labeled sites for within-site temporal generalization (Fig. 1). The cross-site spatial experiment was restricted to ENF sites to reduce ecosystem-type heterogeneity and to focus on transferability within a relatively homogeneous plant functional type.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

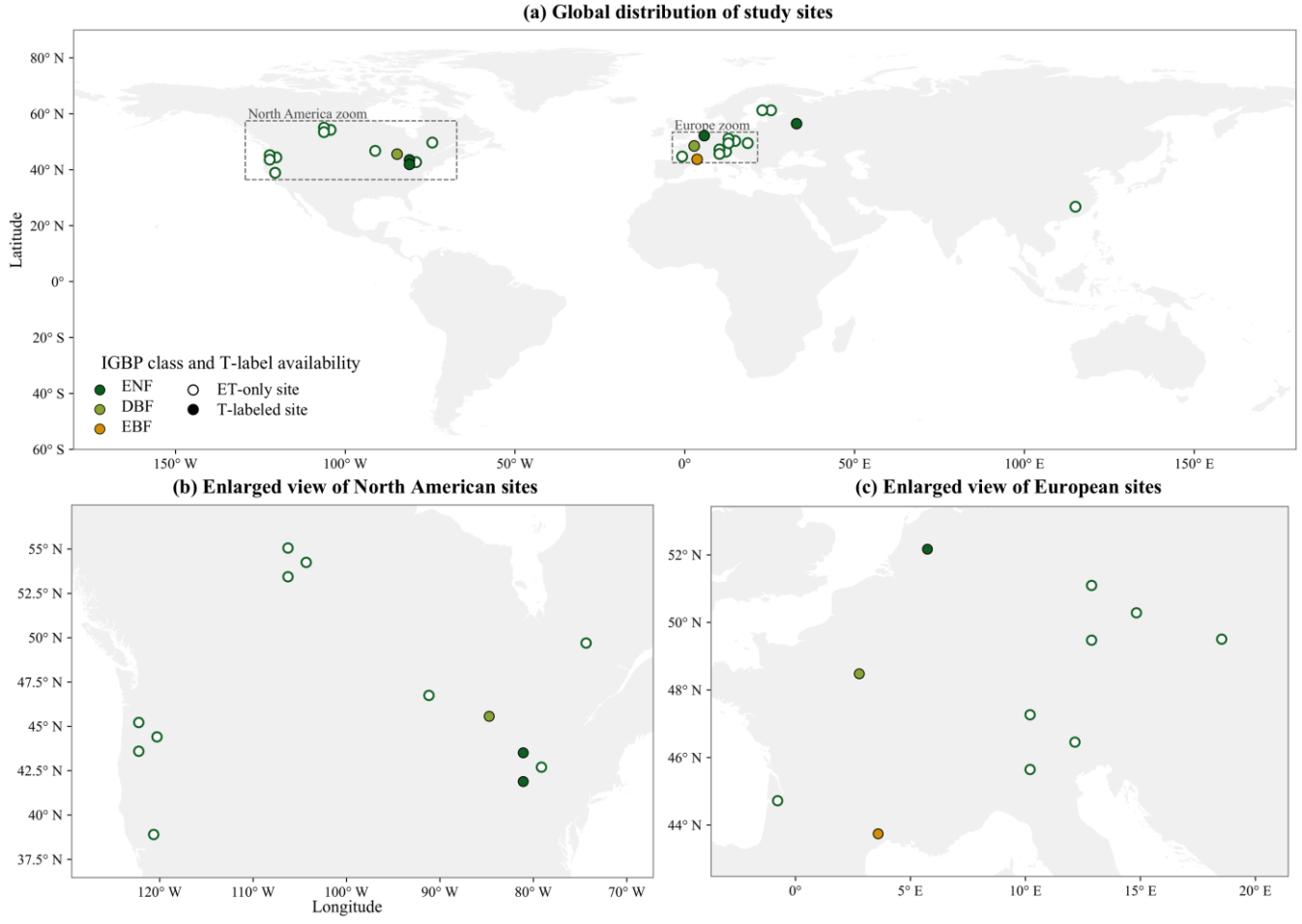


Figure 1. Spatial distribution of the 28 unique study sites used for model training and validation. Note: Colors indicate IGBP classes, and filled symbols indicate the 7 *T*-labeled sites with paired *ET* and sap-flow-derived *T* reference values. Open symbols indicate *ET*-only sites. Map boundaries were generated from Natural Earth data.

The 7 *T*-labeled sites comprised 4 ENF, 2 deciduous broadleaf forest (DBF), and 1 evergreen broadleaf forest (EBF) site according to the International Geosphere-Biosphere Programme (IGBP) classification (Fu et al., 2022; Pastorello et al., 2020). Among the 7 *T*-labeled sites, 4 were also present in the 25-site ENF dataset; therefore, the 2 experimental datasets comprised 28 unique study sites. Table 1 summarizes the observation periods, *ET*-valid hours, *T*-labeled hours, IGBP classes, and geographic coordinates for all sites used for model training and validation.

Table 1. Study site information. Note: *ET*-valid hours indicate the number of valid *ET* samples after quality control, hourly aggregation, and multi-source temporal alignment; *T*-labeled hours indicate the number of samples with *T* reference values; 0 indicates that no usable *T* labels were available at the site. Sites are grouped by IGBP class.

Site code	Data years	*ET*-valid hours	*T*-labeled hours	IGBP	Latitude (°)	Longitude (°)
CA-Obs	1999–2002	3,637	0	ENF	53.9872	-105.1178
CA-Qfo	2004–2010	10,005	0	ENF	49.6925	-74.3421
CA-SF1	2003–2006	5,666	0	ENF	54.4850	-105.8176

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

CA-SF2	2002–2005	5,465	0	ENF	54.2539	-105.8775
CA-TP1	2010–2014	6,618	0	ENF	42.6609	-80.5595
CA-TP3	2008–2014	11,421	4,518	ENF	42.7068	-80.3483
CA-TP4	2008–2014	11,208	7,366	ENF	42.7102	-80.3574
CH-Dav	1997–2011	17,102	0	ENF	46.8153	9.8559
CN-Qia	2003–2005	4,086	0	ENF	26.7414	115.0581
CZ-BK1	2005–2014	9,864	0	ENF	49.5021	18.5369
DE-Lkb	2009–2013	2,244	0	ENF	49.0996	13.3047
DE-Obe	2008–2014	9,328	0	ENF	50.7867	13.7213
DE-Tha	1997–2003, 2009–2014	19,440	0	ENF	50.9626	13.5651
FI-Hyy	2003–2011	10,329	0	ENF	61.8474	24.2948
FI-Let	2009–2012	2,869	0	ENF	60.6418	23.9595
FR-LBr	1996–1998, 2000–2008	13,417	0	ENF	44.7171	-0.7693
IT-Lav	2004, 2007–2014	10,487	0	ENF	45.9562	11.2813
IT-Ren	2010–2012	4,267	0	ENF	46.5869	11.4337
NL-Loo	1996–2012	20,269	1,015	ENF	52.1666	5.7436
RU-Fyo	1998–1999, 2002–2012, 2014	15,860	3,987	ENF	56.4615	32.9221
US-Blo	1997–2007	17,071	0	ENF	38.8953	-120.6328
US-Me2	2005–2014	17,099	0	ENF	44.4526	-121.5589
US-Me3	2004–2009	8,639	0	ENF	44.3154	-121.6078
US-Me5	2000–2002	3,745	0	ENF	44.4372	-121.5668
US-Wi4	2002–2005	3,287	0	ENF	46.7393	-91.1663
FR-Fon	2005–2013	15,849	12,402	DBF	48.4764	2.7801
US-UMB	2000–2014	28,889	3,837	DBF	45.5598	-84.7138
FR-Pue	2004–2014	15,516	15,516	EBF	43.7413	3.5957

3 Methods

3.1 Framework design

MLF-ETT was designed as a physically constrained multi-task learning framework for hourly joint estimation of *ET* and *T* under limited sap-flow-derived *T* data. The framework integrates eddy-covariance-based *ET* references from FLUXNET2015, sap-flow-derived *T* reference values from SAPFLUXNET, and canopy structural information from GIMMS LAI4g. Its core estimation structure consists of an *ET* output branch and a transpiration-ratio output branch. The *ET* branch estimates total *ET*, whereas the ratio branch estimates *T*/*ET*; *T* is then derived as the product of the estimated *ET* and *T*/*ET*, subject to the physical constraint $0 \leq T \leq ET$. This design allows *T* labels to provide paired or auxiliary supervision during training while generating physically bounded *T* estimates for *ET*-only sites.

The workflow of the framework is shown in Fig. 2 and includes data integration, model comparison, training and validation design, and diagnostic analyses. Model comparison included 2 single-task models and 3 multi-task models. The single-task models were XGBoost (XGB) and multilayer perceptron (MLP). In contrast, the multi-task models were Shared-Bottom (SB),

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Multi-gate Mixture-of-Experts (MMOE), and Progressive Layered Extraction (PLE). The validation design included 2 generalization scenarios: within-site temporal generalization at the 7 *T*-labeled sites and cross-site spatial generalization across the 25 ENF sites. Additional diagnostic analyses were conducted for environmental-condition windows, input feature combinations, and *T*-supervision strength.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

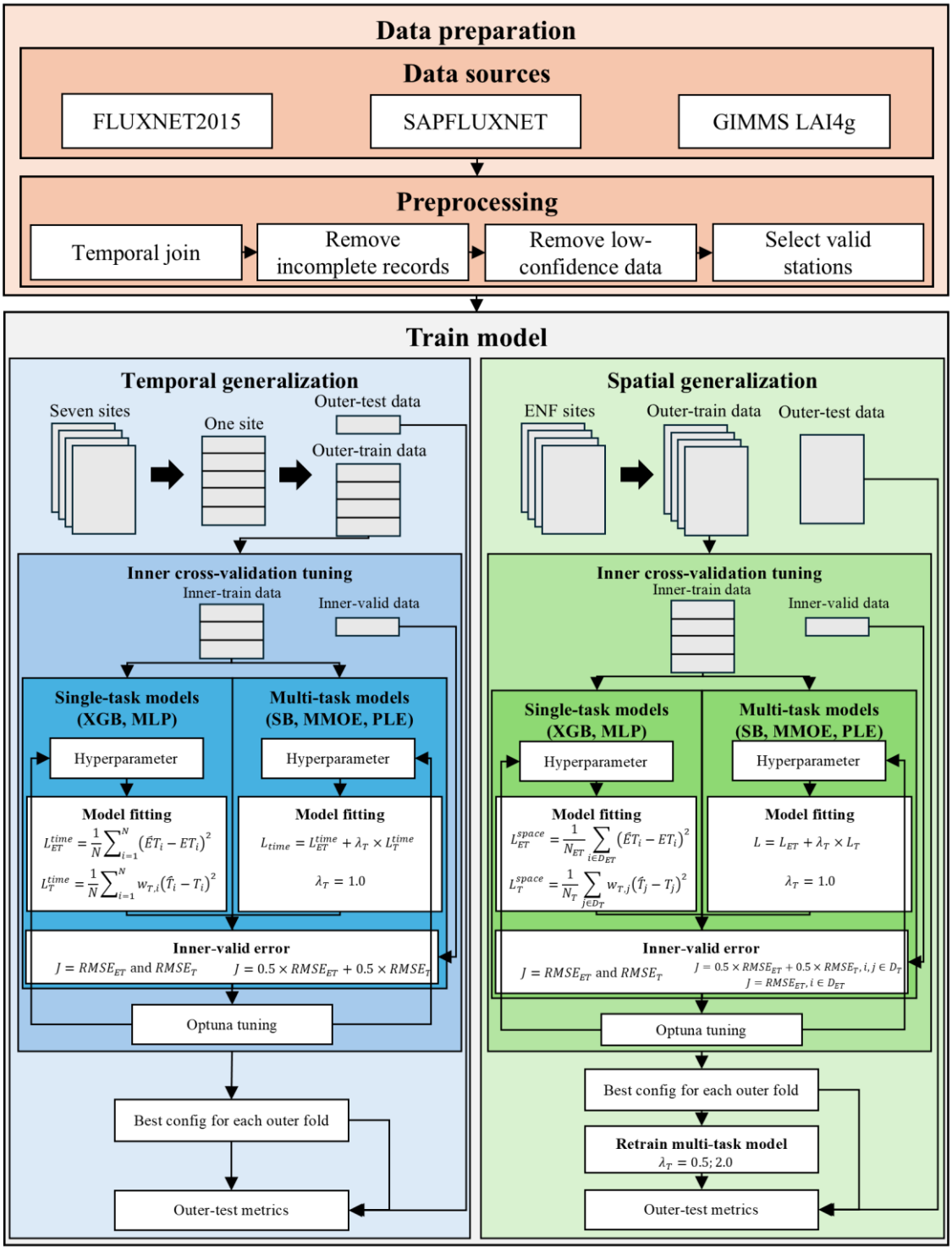


Figure 2. Overall workflow and validation design of the MLF-ETT framework. Note: Within-site temporal generalization was evaluated using block-wise validation at T-labeled sites, whereas cross-site spatial generalization was evaluated using leave-one-site-out validation

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

across ENF sites. Hyperparameters were optimized using Optuna within the inner cross-validation loop for each outer split, and the final models were retrained with the selected hyperparameters. *L* denotes the loss function; $\boldsymbol{\lambda_T}$ denotes the weight of *T* supervision; *w* denotes the sample weight for *T*-labeled samples; *N* denotes the number of samples involved in loss calculation under the corresponding scenario; *D* denotes the set of samples with the corresponding task labels.

3.1.1 Training loss and sample weights

The role of *T* supervision differed between the 2 validation scenarios because *T*-label coverage was uneven. In the within-site temporal scenario, training, validation, and test samples from the 7 *T*-labeled sites all had paired *ET* and *T* reference values. In the cross-site spatial scenario, *ET* reference values were available for all ENF samples, whereas *T* reference values were available only for subsets of samples from 4 sites. Thus, *T* labels provided paired *ET*-*T* supervision in the temporal scenario and sparse auxiliary *T* supervision in the spatial scenario. In the spatial scenario, samples without *T* labels contributed only to the *ET* loss. However, the MLF-ETT output structure still allowed bounded *T* estimates to be generated for these *ET*-only samples.

Accordingly, the single-task *ET* model minimized the *ET* loss over all available samples, whereas the single-task *T* model minimized the *T* loss only over samples with *T* labels:

$$L_{ET} = \frac{1}{N_{ET}} \sum_{i=1}^{N_{ET}} \left(\widehat{ET}_i - ET_i\right)^2, L_T = \frac{1}{N_T} \sum_{j=1}^{N_T} w_{T,j} \left(\widehat{T}_j - T_j\right)^2 , \quad (3)$$

where *N* denotes the number of samples with labels for the corresponding task, and the $\widehat{T}$ was calculated as the product of the *ET* estimate and the estimated *T* fraction. For *T*-labeled hourly samples, *T* supervision was further weighted using sap-flow observation coverage and anomaly flags (Poyatos et al., 2021):

$$w_T = \gamma \times T_{coverage} , \quad (4)$$

where γ is an anomaly penalty, set to 0.1 when *T* > *ET* in the training samples and to 1 otherwise, and $T_{coverage}$ denotes sap-flow observation coverage (%). This weighting reduces the influence of low-coverage or potentially unreliable *T* labels while retaining information on *T* dynamics.

The multi-task training objective was defined as:

$$L = L_{ET} + \lambda_T L_T , \quad (5)$$

where λ_T controls the relative strength of *T* supervision compared with *ET* supervision

3.1.2 Validation schemes

Within-site temporal generalization. Within-site temporal generalization was used to evaluate model performance across time periods within the same site. For each *T*-labeled site, samples were ordered chronologically and divided into 5 contiguous temporal blocks. In each outer-loop split, 1 block was held out as the independent test segment, while the remaining 4 blocks

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

were used for inner cross-validation and model tuning. Final metrics were calculated after concatenating estimates from the 5 held-out blocks.

Cross-site spatial generalization. Cross-site spatial generalization was used to evaluate transferability across ENF sites. Leave-one-site-out (LOSO) validation was implemented across the 25 ENF sites, adapted from Shi et al. (2023) and Tramontana et al. (2016). In each outer-loop split, 1 site was held out for testing, and the remaining sites were used for model training and inner-loop tuning. For the inner 5-fold split, sites with and without *T* labels were assigned to separate inner folds to balance the representation of *T*-labeled and *ET*-only sites.

Sensitivity to T-supervision strength. The multi-task models were first tuned and trained with $\lambda_T = 1.0$, which was used as the reference *T*-supervision setting. To examine the sensitivity to auxiliary *T* supervision under cross-site spatial generalization, the models were retrained with $\lambda_T = 0.5$ and $\lambda_T = 2.0$ under the same training protocol. Their performance was then compared with the $\lambda_T = 1.0$ to isolate the effect of *T*-supervision strength from that of the input-sharing structure.

3.1.3 Hyperparameter optimization and model selection

All candidate models were tuned using inner cross-validation on the outer-loop training set (Table 2), avoiding model selection based on a single fixed training-validation split. Hyperparameter optimization was implemented with Optuna using the tree-structured Parzen estimator (TPE) sampler and 20 trials by default (Akiba et al., 2019). TPE fits parameter distributions from completed trials and preferentially samples candidate parameters from regions more likely to improve performance, thereby increasing search efficiency in high-dimensional mixed hyperparameter spaces (Bergstra et al., 2011). For neural-network models, the maximum number of inner-loop training epochs was set to 100, and early stopping on the validation set was used to record the best epoch. Final outer-loop training did not use early stopping; instead, the median of the best epochs across the inner folds was used as the fixed number of training epochs.

To keep *ET* and *T* on a consistent numerical scale during joint training, the 99th percentile of *ET* was calculated from the outer-loop *ET* training set for each temporal or spatial split, and the same scaling factor was then applied to both *ET* and *T*. Compared with maximum-value scaling, this percentile-based scaling reduces the influence of a small number of extreme samples on training stability.

Hyperparameters for single-task models were selected based on the root mean square error (RMSE) on the inner validation set for the corresponding target. Hyperparameters for multi-task models were selected using the following composite validation objective:

$$J = 0.5RMSE_{ET} + 0.5RMSE_{T}\ , \tag{6}$$

where RMSE_{ET} and RMSE_{T} are the *ET* and *T* root mean square errors on the validation set (mm h^{-1}), respectively. If no *T*-labeled samples were available in the current validation set, the objective was reduced to RMSE_{ET}.

Table 2. Hyperparameter search space of the candidate models.

Model	Hyperparameter	Search range

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

XGB	Number of trees	300–1500
	Tree depth	2–8
	Learning rate	1×10^{-4} to 0.1
	Row/column subsampling ratio	0.6–1.0
	Minimum child weight	1×10^{-2} to 10
	L1 regularization	1×10^{-4} to 0.1
	L2 regularization	1×10^{-3} to 10
	Split penalty	0–5
MLP	Hidden layers	2–4 layers
	Hidden width	(128-64), (256-128), (256-128-64), (512-256-128), (512-256-128-64)
	Dropout	0–0.3
	Learning rate	1×10^{-4} to 3×10^{-3}
	L2 regularization	1×10^{-6} to 1×10^{-3}
SB	Shared-layer depth	3–5 layers
	Shared-layer width	(128-64-32), (256-128-64), (256-128-64-32), (512-256-128-64), (512-256-128-64-32)
	Task-branch depth	2–4 layers
	Task-branch width	(128-64), (128-64-32), (256-128-64), (256-128-64-32), (512-256-128-64)
	Dropout	0–0.5
	Learning rate	1×10^{-4} to 3×10^{-3}
	L2 regularization	1×10^{-6} to 1×10^{-3}
MMOE	Number of experts	4, 6, 8, 10, 12
	Other hyperparameters	Same as SB
PLE	Number of shared experts	2, 3, 4, 6, 8
	Number of task-specific experts	Fixed at 2 per task
	Other hyperparameters	Same as SB

3.2 Machine learning models

We used single-task and multi-task models to estimate hourly *ET* and *T* from environmental, radiative, and vegetation input features, incorporating site-limited sap-flow-derived *T* labels into joint *ET*-*T* learning.

3.2.1 Input features

The 9 input variables were selected to represent atmospheric evaporative demand, surface energy supply, and vegetation physiological and structural conditions (Amani and Shafizadeh-Moghadam, 2023; Zhang et al., 2021). Specifically, T_a, *RH*, P_a, *VPD*, and *WS* characterize near-surface thermodynamic, moisture, and aerodynamic conditions (Feng et al., 2024; Woo, 2026); R_n and solar elevation angle (h_s) represent available energy and radiation geometry (Braghiere et al., 2020); and *GPP* and *LAI* represent vegetation physiological activity and canopy structure (Leng et al., 2024; Shi and Cai, 2025; Medlyn et al., 2011).

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

To examine whether the organization of shared and task-specific information affected joint *ET* and *T* estimation, the 9 input features were grouped into meteorological (M: T_a, *VPD*, P_a, *WS*, and *RH*), radiation (R: R_n and h_s), and vegetation (V: *GPP* and *LAI*) groups. Based on these groups, the multi-task input feature combinations listed in Table 3 were constructed.

Table 3. Input feature combinations for the multi-task models. Note: M, R, and V denote the meteorological, radiation, and vegetation groups, respectively; features to the left of "|" are shared input features, whereas features to the right are branch-specific extra features for the *ET* tower and the transpiration-ratio tower; "—" indicates no branch-specific extra features.

Feature combination	Shared features	*ET* features	Ratio features
All	M, R, V	—	—
M|RV	M	R, V	R, V
MR|V	M, R	V	V
RV|M	R, V	M	M

3.2.2 Single-task models

XGB and MLP were used as single-task models. XGB captures nonlinear relationships and variable interactions through gradient-boosted decision trees (Chen and Guestrin, 2016), whereas MLP learns continuous mappings between inputs and targets using fully connected feed-forward networks (Rumelhart et al., 1986). Both approaches have been widely used to estimate evapotranspiration, latent heat flux, and related environmental variables (Ferreira et al., 2022; Sharafi and Mohammadi Ghaleni, 2024; Feng et al., 2024; Shan et al., 2020). Separate single-task models were trained for *ET* and *T* using the full input feature set (hereafter referred to as "All"). These reference models were used to evaluate whether the multi-task framework yielded gains beyond task-specific learning and to serve as a reference for interpreting model errors.

3.2.3 Multi-task models

The comparison included 3 multi-task models: SB, MMOE, and PLE (Fig. 3). All models used MLP modules and estimated total *ET* and the transpiration fraction *T*/*ET*.

Model inputs and outputs. For each input feature combination, input features were divided into shared input features and branch-specific extra features. Shared features were passed through the shared trunk or expert layers to learn representations common to *ET* and *T*. Branch-specific extra features were then concatenated with these representations before estimation. Consistent with the MLF-ETT output design, each multi-task model used 2 output branches: an *ET* tower and a transpiration-ratio tower. The *ET* tower estimated total *ET*, whereas the ratio tower estimated the transpiration fraction *T*/*ET*; the final *T* estimate was obtained by multiplying the two outputs.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

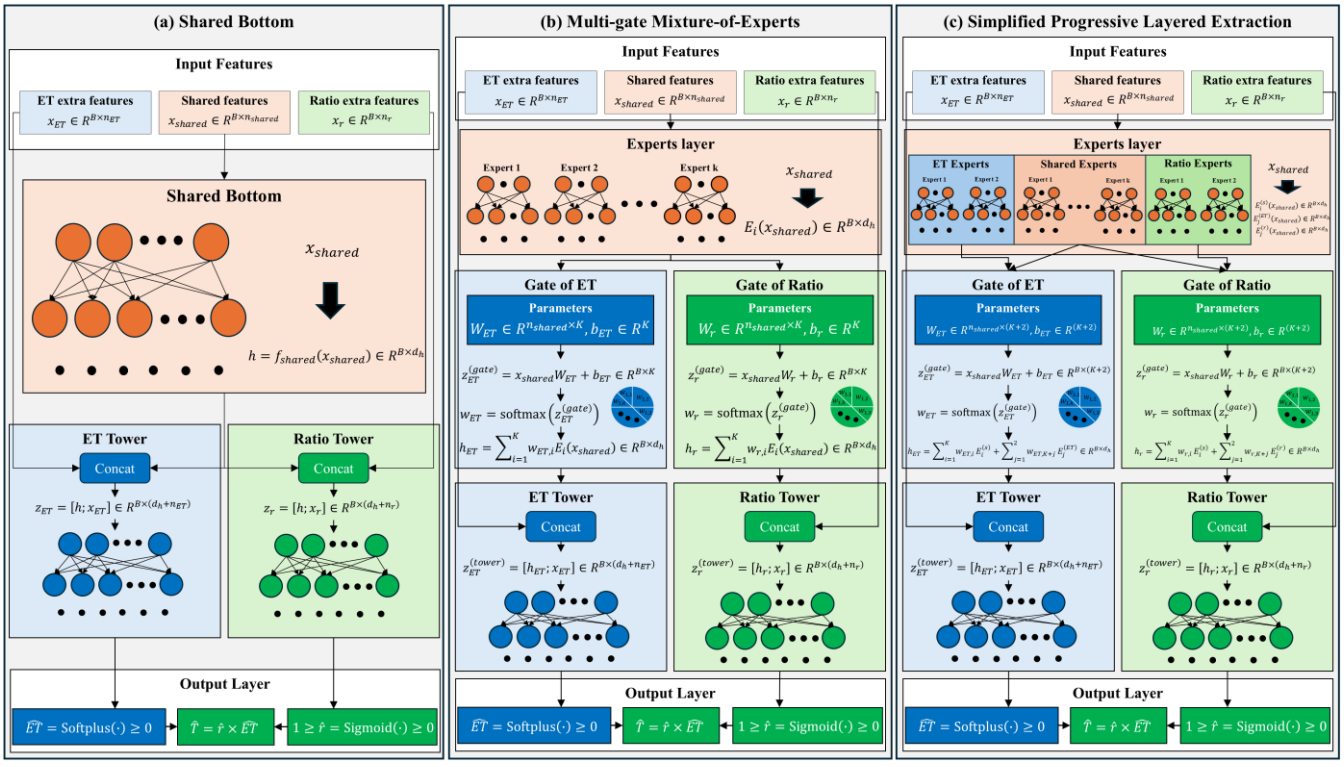


Figure 3. Schematic of the 3 multi-task models. Note: The *ET* tower outputs total evapotranspiration, and the transpiration-ratio tower outputs *T*/*ET*; their product gives the *T* estimate. *B* denotes batch size; *n* denotes the input dimension; d_h denotes the hidden representation dimension; *h* denotes the shared or gate-aggregated hidden representation; *z* denotes the fused task-tower input; *E* denotes an expert network; and *W* and *b* denote the weight matrix and bias term of the gating network, respectively.

SB model. SB is the canonical hard-parameter-sharing multi-task model. It uses a single shared trunk to extract a common representation from the shared input features and then passes this representation to task-specific towers for estimation (Caruana, 1993). In this study, the shared input features were first passed through the shared-bottom trunk:

$$h = f_{shared}(x_{shared}) , \tag{7}$$

where f_{shared} denotes the MLP mapping of the shared-bottom trunk and x_{shared} denotes the shared input features. The shared representation was then concatenated with the extra features of the *ET* tower and the transpiration-ratio tower to form task-specific inputs:

$$z_t = [h; x_t] , \tag{8}$$

where x_t denotes the extra features for task t, with $t \in \{ET, r\}$. SB is parameter-efficient and stable during training, but it may suffer from task conflict or negative transfer when the processes controlling *ET* and *T* differ across sites or environmental conditions.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

MMOE model. MMOE organizes shared information more flexibly through multiple shared expert networks and task-specific gating mechanisms. Instead of forcing both tasks to use a single common representation, MMOE learns task-specific combinations of shared expert outputs (Ma et al., 2018). The shared input features were passed into K parallel expert networks to obtain expert outputs:

$$e_i = E_i(x_{shared}), i = 1,2, \dots, K \,, \tag{9}$$

where $E_i(\cdot)$ denotes the ith expert network. Each task then assigned weights to the expert outputs using its task-specific gating network:

$$w_t = softmax(x_{shared} W_t + b_t) \,, \tag{10}$$

where W_t and b_t denote the weight matrix and bias vector of the corresponding gating network, respectively. The expert outputs were then combined as a weighted sum to form a task-specific hidden representation:

$$h_t = \sum_{i=1}^{K} w_{t,i} e_i(x_{shared}) \,, \tag{11}$$

After obtaining the task-specific hidden representations, each representation was concatenated with the corresponding extra features and passed into the *ET* tower or the transpiration-ratio tower. Compared with SB, MMOE allows *ET* and the transpiration fraction to use the same expert pool with different task-specific weights, although the experts themselves remain fully shared.

PLE model. PLE further separates shared experts from task-specific experts, thereby retaining both transferable information and task-specific information (Tang et al., 2020). This model was included because *ET* and *T* are physically related but not controlled by identical processes, particularly when transferring across sites with different canopy structure and energy partitioning. The shared input features were first passed into shared expert networks to obtain shared expert outputs:

$$e_i^{shared} = E_i^{shared}(x_{shared}), i = 1,2, \dots, K \,, \tag{12}$$

In parallel, the input was passed into task-specific experts, fixed at 2 experts per task, to obtain task-specific representations:

$$e_i^t = E_i^t(x_{shared}), i = 1,2 \,, \tag{13}$$

where E_i^{shared} and E_i^t denote the shared and task-specific expert networks, respectively. The shared and task-specific expert outputs were then weighted and combined to form task-specific hidden representations:

$$h_t = \sum_{i=1}^{K} w_{t,i} E_i^{shared}(x_{shared}) + \sum_{j=1}^{2} w_{t,K+j} E_j^t \,, \tag{14}$$

Compared with SB and MMOE, PLE provides both shared and task-specific pathways, making it more flexible for representing related but non-identical *ET* and *T* processes.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

3.3 Evaluation metrics

Model performance was evaluated using the Kling-Gupta efficiency (KGE, Eqs. (15) and (16)), following Gupta et al. (2009) and RMSE (Eq. (17)). RMSE was used as the optimization criterion during hyperparameter tuning, whereas final results were reported primarily using KGE and secondarily using RMSE. KGE jointly characterizes correlation, variability, and mean bias, making it suitable for evaluating the overall agreement between hourly estimates and reference values. A KGE value closer to 1 indicates stronger agreement.

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}\,, \tag{15}$$

where r is the correlation coefficient between estimated and reference values, α is the standard deviation ratio, and β is the mean ratio.

$$\alpha = \frac{\sigma_{est}}{\sigma_{obs}}, \beta = \frac{\mu_{est}}{\mu_{obs}}\,, \tag{16}$$

where σ_{est} and σ_{obs} are the standard deviations of estimated and reference values, respectively, and μ_{est} and μ_{obs} are their corresponding means. RMSE quantifies the absolute magnitude of estimation errors (mm h^{-1}). A smaller RMSE indicates lower overall model error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{est,i} - y_{obs,i}\right)^2}\,, \tag{17}$$

where n is the sample size, and $y_{est,i}$ and $y_{obs,i}$ denote the estimated and reference values for sample i, respectively.

3.4 Model interpretation

To provide a supplementary interpretation of model dependence on input variables, we applied SHapley Additive exPlanations (SHAP) analysis to the trained single-task MLP and PLE models (Lundberg and Lee, 2017). To reduce computational cost, SHAP values were calculated using selected saved models and sampled training records: the first outer-fold model for each *T*-labeled site for temporal generalisation, and each *T*-labeled-site LOSO model for spatial generalization, with at most 200 sampled records used for interpretation. SHAP values were calculated using DeepExplainer, and variable importance was summarized as the mean absolute SHAP value of each input feature; for the PLE model, SHAP values were calculated for both *ET* and the derived *T* estimate.

4 Results

The multi-task models were evaluated from 4 perspectives: overall site-level performance, representative-site responses, environmental-condition-specific errors, and model-configuration effects. Site-level KGE and ΔKGE relative to the best single-task model were first used to assess overall model performance in terms of temporal and spatial generalization.

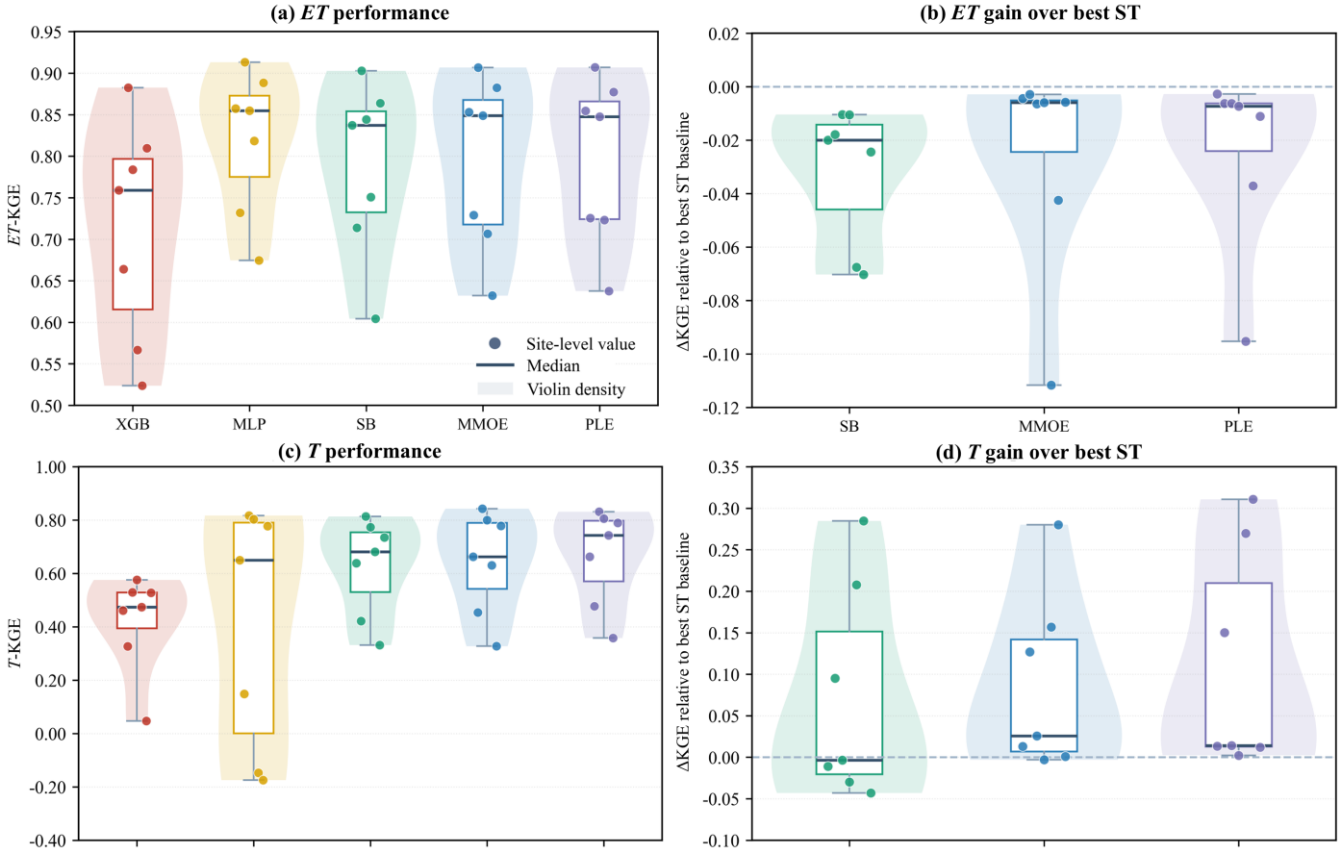
https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Environmental-condition windows, input feature combinations, and *T*-supervision weights were then examined to further evaluate the model's behavior under physically relevant conditions and configurations.

4.1 Within-site temporal generalization

The site-level KGE for *ET* and *T*, and the corresponding ΔKGE relative to the better single-task model, defined as the model with the higher KGE between XGB and MLP for the corresponding target variable, are compared and shown in Fig. 4. Representative *T*-labeled sites were further examined using the scatter-density plots shown in Fig. 5 to determine whether corresponding changes in *ET*-KGE accompanied changes in *T*-KGE.

Within-site temporal generalization, the main performance gain of the multi-task models was observed for *T*-KGE rather than *ET*-KGE (Fig. 4). Relative to the better single-task model at each site, SB, MMOE, and PLE showed positive *ET* gains at 0/7 sites, with median ΔKGE values of −0.0200, −0.0059, and −0.0072, respectively. In contrast, *T* showed clearer positive gains: SB, MMOE, and PLE showed positive *T* gains at 3/7, 6/7, and 7/7 sites, respectively, with median ΔKGE values of −0.0036, +0.0256, and +0.0141. Thus, positive ΔKGE responses under within-site temporal generalization were observed primarily for *T*, particularly for MMOE and PLE.



https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Figure 4. Site-level KGE and ΔKGE for *ET* and T under within-site temporal generalization. Note: For SB, MMOE, and PLE, KGE values were averaged across input feature combinations within each site. ΔKGE was calculated relative to the better single-task model, defined as the model with the higher KGE between XGB and MLP for the corresponding target variable. Dashed horizontal lines indicate ΔKGE = 0.

Across representative *T*-labeled sites, *T*-KGE improvements were also clearer than *ET*-KGE improvements (Fig. 5). The MLP *T*-KGE was 0.650 at CA-TP3 but only 0.148, −0.175, and −0.147 at CA-TP4, RU-Fyo, and NL-Loo, respectively. In comparison, the best multi-task *T*-KGE values at these 4 sites reached 0.688 (PLE-RV|M), 0.491 (PLE-M|RV), 0.471 (PLE-All), and 0.800 (MMOE-RV|M), respectively.

Concurrent *ET*-KGE and *T*-KGE gains occurred at some sites (Fig. 5). At CA-TP3 and CA-TP4, the best multi-task *ET* configurations had slightly higher *ET*-KGE than MLP, with KGE values of 0.864 (PLE-All) versus 0.857 (MLP) and 0.746 (MMOE-M|RV) versus 0.732 (MLP), respectively. At RU-Fyo and NL-Loo, *T*-KGE gains were not accompanied by consistent *ET*-KGE gains, although a few configurations approached or slightly exceeded the MLP result.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

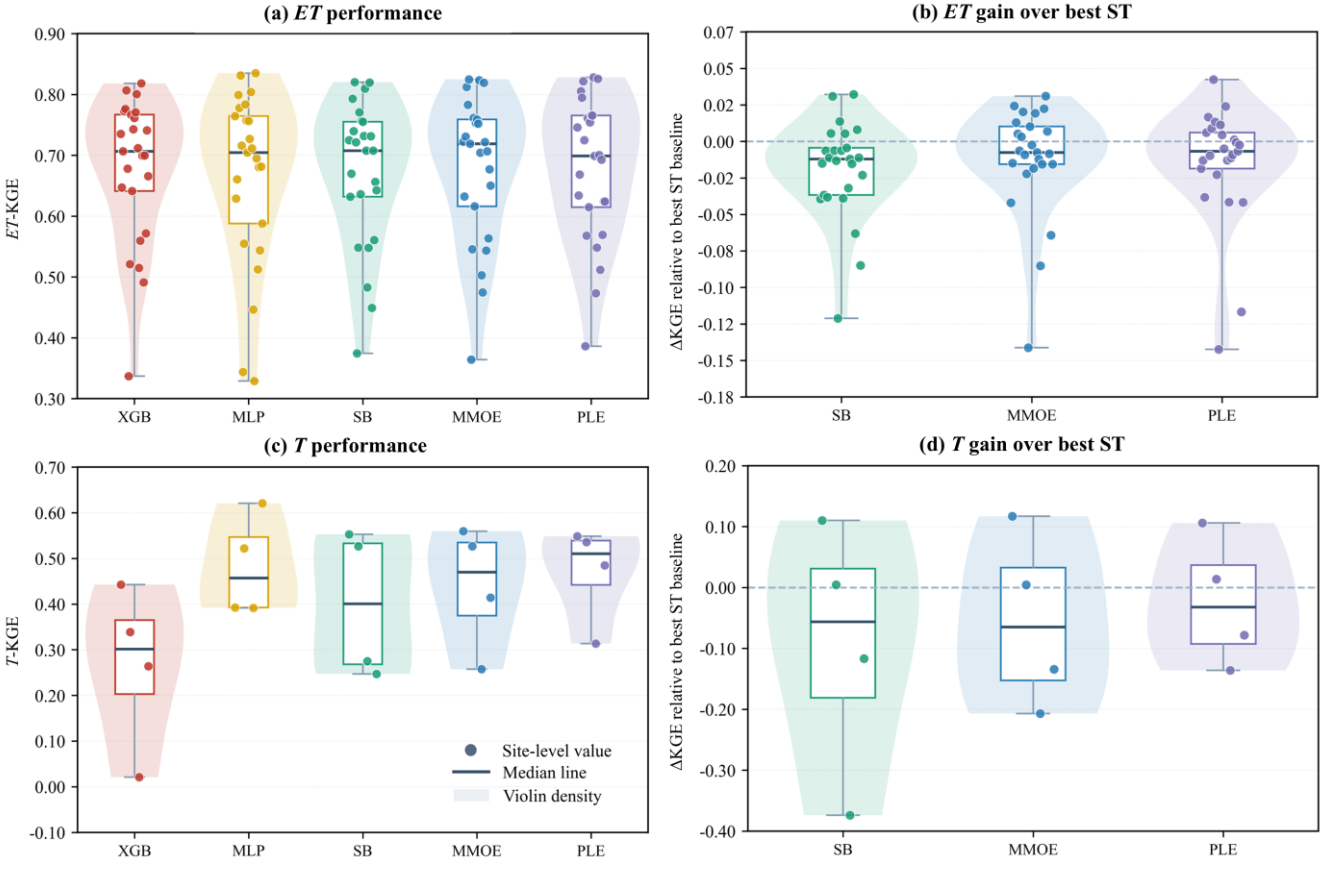
https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Figure 5. Observed versus estimated *ET* and *T* at 4 representative *T*-labeled sites within the ENF dataset under within-site temporal generalization. Note: The left panels show XGB and MLP, whereas the right panels show multi-task models with different input feature combinations. In the multi-task panels, columns indicate input feature combinations; M, R, and V denote meteorological, radiation, and vegetation feature groups, respectively; and features before and after "|" indicate shared and branch-specific inputs. Rows indicate SB, MMOE, and PLE. The grey diagonal line indicates the 1:1 line, and color bars indicate sample counts in the hexbin plots.

4.2 Cross-site spatial generalization

The site-level KGE for *ET* and *T*, and the corresponding ΔKGE relative to the best single-task model, are compared in Fig. 6. Representative *T*-labeled sites were further examined using the scatter-density plots shown in Fig. 7 to assess whether the effects of the multi-task models on *ET*-KGE and *T*-KGE were consistent across sites.

ET-KGE responses were site-dependent under cross-site spatial generalization (Fig. 6). Relative to the better single-task model at each site, SB, MMOE, and PLE showed positive *ET* gains at 6/25, 10/25, and 9/25 sites, respectively, with median ΔKGE values of −0.0120, −0.0076, and −0.0067. *T*-KGE also showed site dependence. SB, MMOE, and PLE showed positive *T* gains at 2/4 *T*-labeled sites, with median ΔKGE values of −0.0561, −0.0649, and −0.0322, respectively. Unlike within-site temporal generalization, the spatial scenario did not show a consistent increase in *T*-KGE, indicating that *T*-related information was less transferable across sites when *T* labels were sparse.

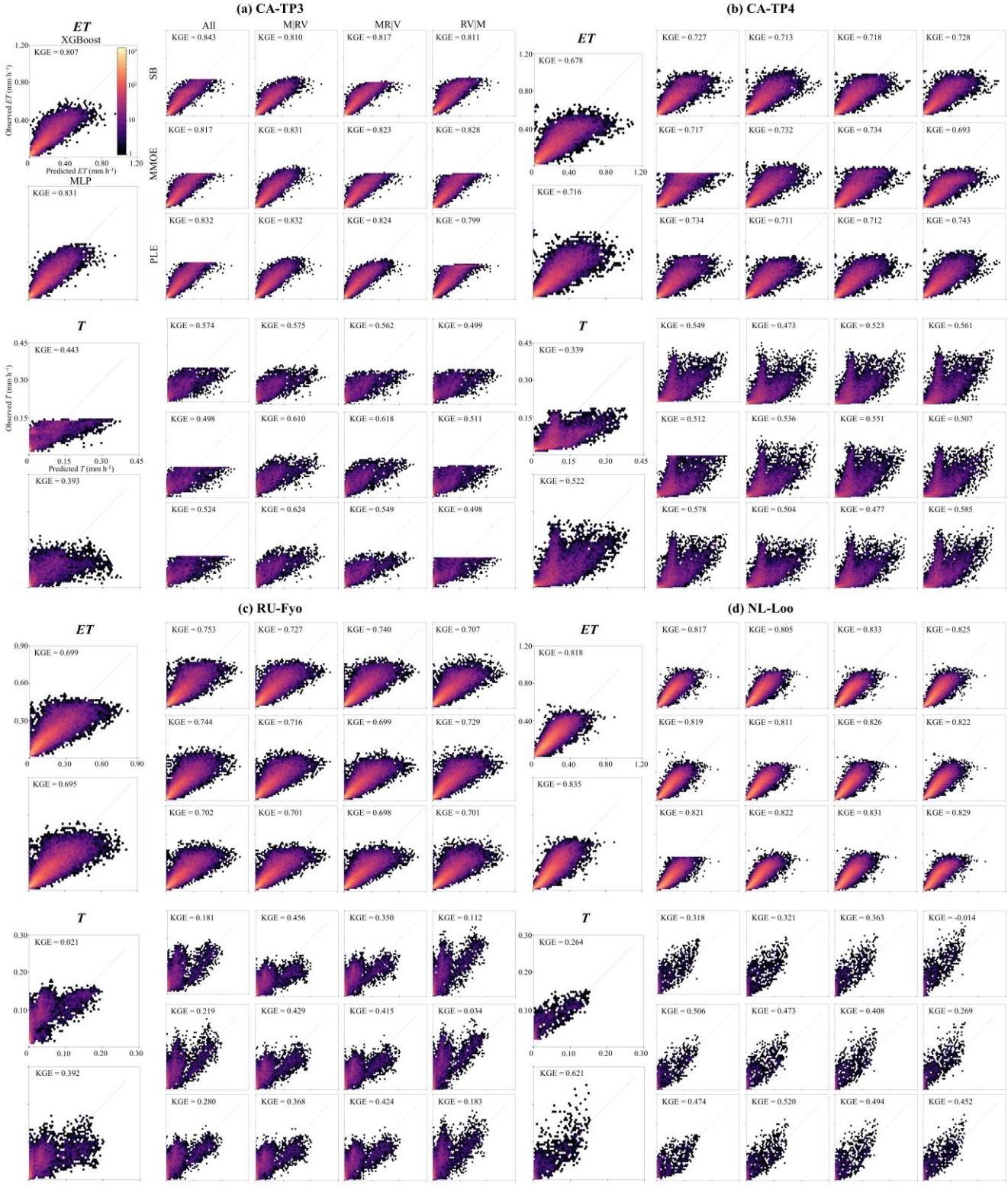


https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Figure 6. Site-level KGE and ΔKGE for *ET* and *T* under cross-site spatial generalization. Note: For SB, MMOE, and PLE, KGE values were averaged across input feature combinations within each test site. ΔKGE was calculated relative to the better single-task model, defined as the model with the higher KGE between XGB and MLP for the corresponding target variable. Dashed horizontal lines indicate ΔKGE = 0.

The alignment between *ET*-KGE and *T*-KGE changes varied across sites (Fig. 7). At RU-Fyo, SB-MR|V reduced *T*-KGE from 0.392 (MLP) to 0.350 but increased *ET*-KGE from 0.695 (MLP) to 0.740. At NL-Loo, PLE-MR|V reduced *T*-KGE from 0.621 (MLP) to 0.494, while *ET*-KGE also decreased slightly from 0.835 (MLP) to 0.831. These representative-site results show that cross-site multi-task responses were site-dependent and that changes in *T*-KGE were not always aligned with changes in *ET*-KGE.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026



https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Figure 7. Observed versus estimated *ET* and *T* at 4 representative *T*-labeled sites within the ENF dataset under cross-site spatial generalization. Note: The left panels show XGB and MLP, whereas the right panels show multi-task models with different input feature combinations. In the multi-task panels, columns indicate input feature combinations; M, R, and V denote meteorological, radiation, and vegetation feature groups, respectively; and features before and after "|" indicate shared and branch-specific inputs. Rows indicate SB, MMOE, and PLE. The grey diagonal line indicates the 1:1 line, and color bars indicate sample counts in the hexbin plots.

4.3 Performance in environmental-condition windows

The RMSE changes across predefined environmental-condition windows are examined. To avoid confounding model and input effects, a single fixed multi-task configuration was selected for each scenario based on the highest combined mean *ET*-KGE and *T*-KGE: PLE-All for within-site temporal generalization and MMOE-M|RV for cross-site spatial generalization. Relative RMSE gains across all samples, high-*RH*, high-*VPD*, high-T_a, and combined high-*VPD*-and-T_a windows are compared in Fig. 8.

PLE-All reduced *T*-RMSE under within-site temporal generalization, and the gain was stronger in high evaporative-demand conditions, including the High *VPD*, High T_a, and combined High *VPD* and T_a windows (Fig. 8). In the all-sample window, the mean relative RMSE gain for *T* was +0.1632, with positive gains at 6/7 sites. In the High *VPD*, High T_a, and High *VPD* and T_a windows, the mean gains increased to +0.2021, +0.2068, and +0.2135, respectively, with positive gains at 6/7 sites in all 3 windows. By contrast, the mean relative RMSE gains for *ET* were negative across the all-sample window and each environmental-condition window, with positive gains at no more than 3 of 7 sites.

Under cross-site spatial generalization, the *T* gain from MMOE-M|RV was also concentrated in high evaporative-demand windows. In the all-sample window, the mean relative RMSE gain for *T* was +0.0115, with positive gains at 2 of 4 sites. The High *RH* window showed a negative gain, with no positive gains at any of the 4 *T*-labeled sites. In contrast, the mean *T* gains in the High *VPD*, High T_a, and High *VPD* and T_a windows were +0.0978, +0.0855, and +0.1021, respectively, with positive gains at 3/4 sites in all 3 windows. The mean relative RMSE gains for *ET* were negative across all 5 environmental-condition windows, indicating no consistent reduction in window-scale *ET*-RMSE.

Overall, the selected multi-task configurations produced greater RMSE reductions for *T* than for *ET* (Fig. 8). The strongest *T*-RMSE reductions occurred under high evaporative demand conditions. In contrast, the High *RH* window under spatial generalization showed a negative *T*-RMSE gain, with no positive gains at any of the 4 *T*-labeled sites. These results indicate that the *T*-related benefit of the multi-task framework was strengthened under high-demand conditions but weakened under humid, low-demand conditions.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

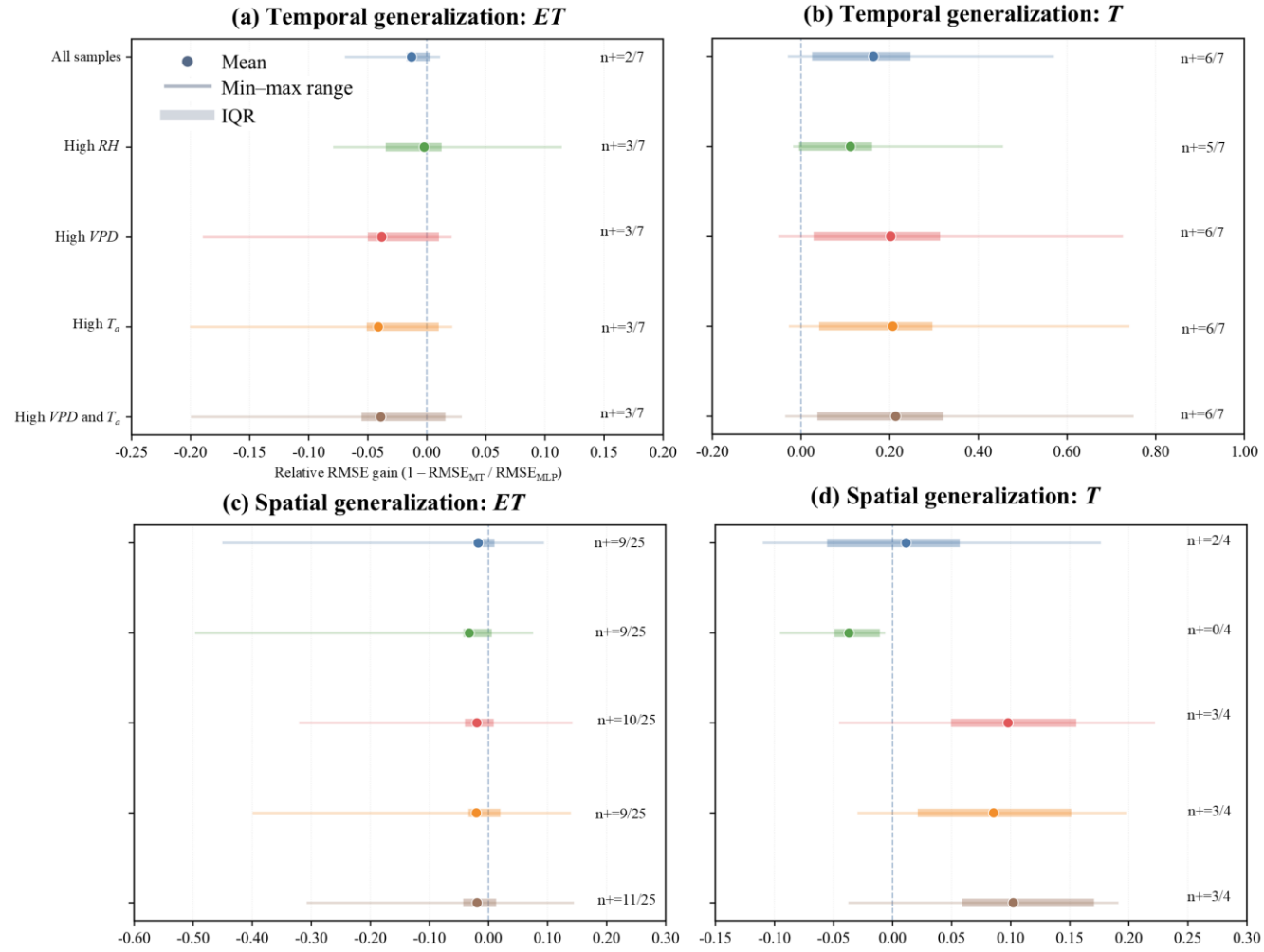


Figure 8. Relative RMSE gain over MLP across environmental-condition windows. Note: Temporal panels use PLE-All, and spatial panels use MMOE-M|RV. Relative RMSE gain was calculated as 1 - $RMSE_{MT}$ / $RMSE_{MLP}$, where MT denotes the selected multi-task model; positive values indicate lower RMSE than MLP. Points, thick bars, and thin lines indicate the mean, interquartile range, and min–max range across sites, respectively. Dashed vertical lines indicate zero gain. The n+ labels indicate the number of sites with positive relative RMSE gain.

4.4 Effects of input feature combinations

The effects of input feature combinations on *ET* and *T* performance were evaluated using mean *ET*-KGE and *T*-KGE across model-input combinations, as well as ΔKGE relative to All. The mean KGE results are summarized in Fig. 9, and the ΔKGE values relative to All are reported in Table 4.

The effects of input feature combinations differed between the 2 generalization scenarios (Fig. 9). Under within-site temporal generalization, variation among multi-task models persisted within a given input feature combination, particularly for *ET*-KGE. However, when summarized by input feature combinations, *ET*-KGE differed only slightly among variants,

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

whereas *T*-KGE showed a clearer separation. RV|M had the highest mean *T*-KGE, reaching 0.6710 compared with 0.6369 for All. Thus, under within-site temporal generalization, input feature combinations mainly affected *T* estimation.

Under cross-site spatial generalization, the separation among input feature combinations was stronger and was observed in both *ET*-KGE and *T*-KGE. All had the highest mean *ET*-KGE (0.6882), whereas M|RV had the highest mean *T*-KGE (0.4906). In contrast, RV|M decreased to 0.3442 for *T*-KGE, markedly lower than its temporal-generalization performance. These results indicate that the effect of input feature combinations was scenario-dependent, with no single input feature combination performing best for both *ET* and *T*.

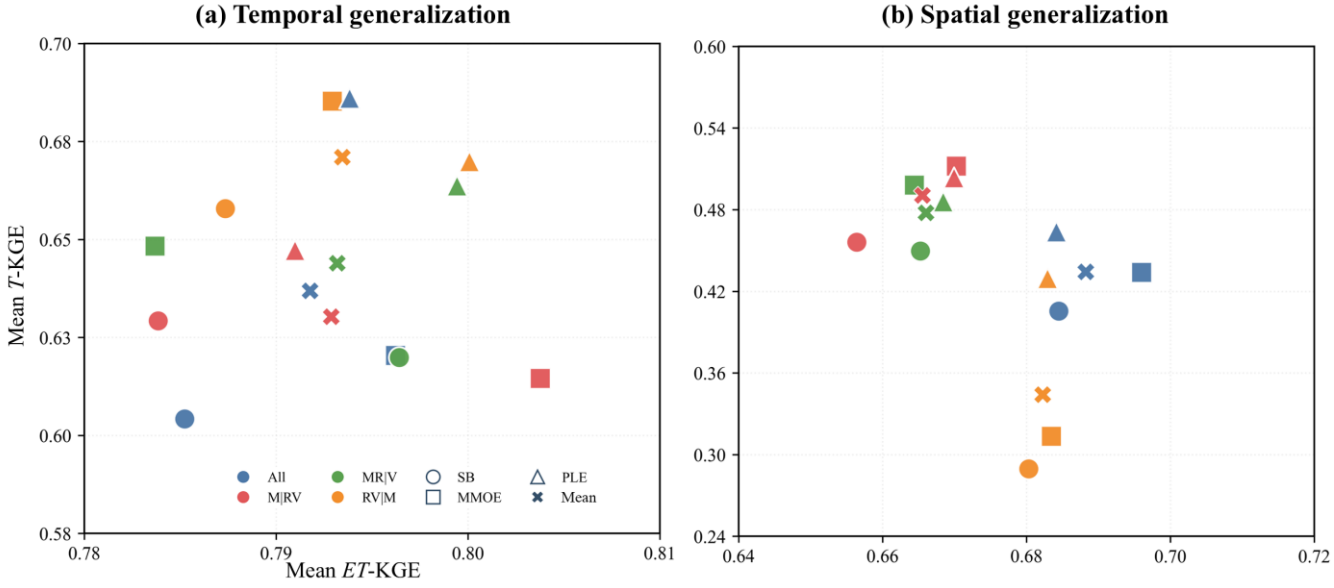


Figure 9. Mean *ET*-KGE and *T*-KGE of different input feature combinations under temporal and spatial generalization. Note: Each point represents a model-input combination averaged across all test sites in the corresponding scenario. Colors indicate input feature combinations, open symbols indicate individual multi-task models, and cross markers indicate the mean across SB, MMOE, and PLE for each input feature combination. M, R, and V denote meteorological, radiation, and vegetation feature groups, respectively; features before and after "|" indicate shared and branch-specific inputs.

The direction of change for each input feature combination relative to All was further quantified using ΔKGE (Table 4). Within-site temporal generalization, input feature combinations had little effect on *ET*. Still, clearer effects on *T*. RV|M were observed, with a mean *T* ΔKGE of +0.034 and positive gains in 15/21 site-model combinations. This indicates that isolated models or sites did not drive the *T* advantage of RV|M.

Under cross-site spatial generalization, the effects of input feature combinations changed direction. M|RV and MR|V had positive mean *T* ΔKGE values of +0.056 and +0.043, respectively, whereas RV|M decreased to −0.090. Thus, the *T* advantage of RV|M under temporal generalization did not carry over to spatial generalization. Overall, input feature combinations affected *ET* and *T* differently across generalization scenarios, indicating that the allocation of meteorological, radiative, and vegetation information between shared and task-specific branches was an important design factor.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Table 4. ΔKGE of input feature combinations relative to All. Note: Each row represents a scenario-input combination. *ET* mean ΔKGE and *T* mean ΔKGE indicate the mean change relative to All across all valid site-model combinations in the corresponding scenario, where $\Delta KGE = KGE_{variant} - KGE_{All}$. Positive n/N indicates the number of positive cases relative to the total number of valid combinations.

Scenario	Input feature combination	*ET* mean ΔKGE	*ET* positive n/N	*T* mean ΔKGE	*T* positive n/N
Temporal	M|RV	+0.001	12/21	-0.007	12/21
	MR|V	+0.001	12/21	+0.007	12/21
	RV|M	+0.002	9/21	+0.034	15/21
Spatial	M|RV	-0.023	29/75	+0.056	9/12
	MR|V	-0.022	33/75	+0.043	8/12
	RV|M	-0.006	33/75	-0.090	3/12

4.5 Sensitivity to *T*-supervision weight

Finally, the sensitivity of *ET* and *T* performance to the *T*-supervision weight was evaluated under cross-site spatial generalization. Mean ΔKGE values at $\lambda_T = 0.5$ and $\lambda_T = 2.0$ relative to $\lambda_T = 1.0$ are compared in Fig. 10.

ET was only weakly affected by changes in the *T*-supervision weight (Fig. 10). For most model-input combinations, *ET* ΔKGE was close to zero, with changes mainly within approximately ±0.01. The largest changes were +0.0119 for SB-M|RV at $\lambda_T = 0.5$ and -0.0086 for PLE-RV|M at $\lambda_T = 2.0$. Thus, changing the *T*-supervision weight alone did not substantially alter overall *ET* performance.

T was more sensitive to the *T*-supervision weight, and the effect was generally stronger at $\lambda_T = 2.0$ than at $\lambda_T = 0.5$. Relative to $\lambda_T = 1.0$, *T* ΔKGE at $\lambda_T = 2.0$ reached +0.0388 for MMOE-All, +0.0370 for MMOE-RV|M, +0.0362 for PLE-All, and +0.0360 for SB-All. By contrast, the largest positive changes at $\lambda_T = 0.5$ were +0.0269 for SB-All and +0.0264 for MMOE-RV|M. Overall, changing the *T*-supervision weight mainly affected *T* estimation rather than *ET*, and its effect depended on the model-input combination.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

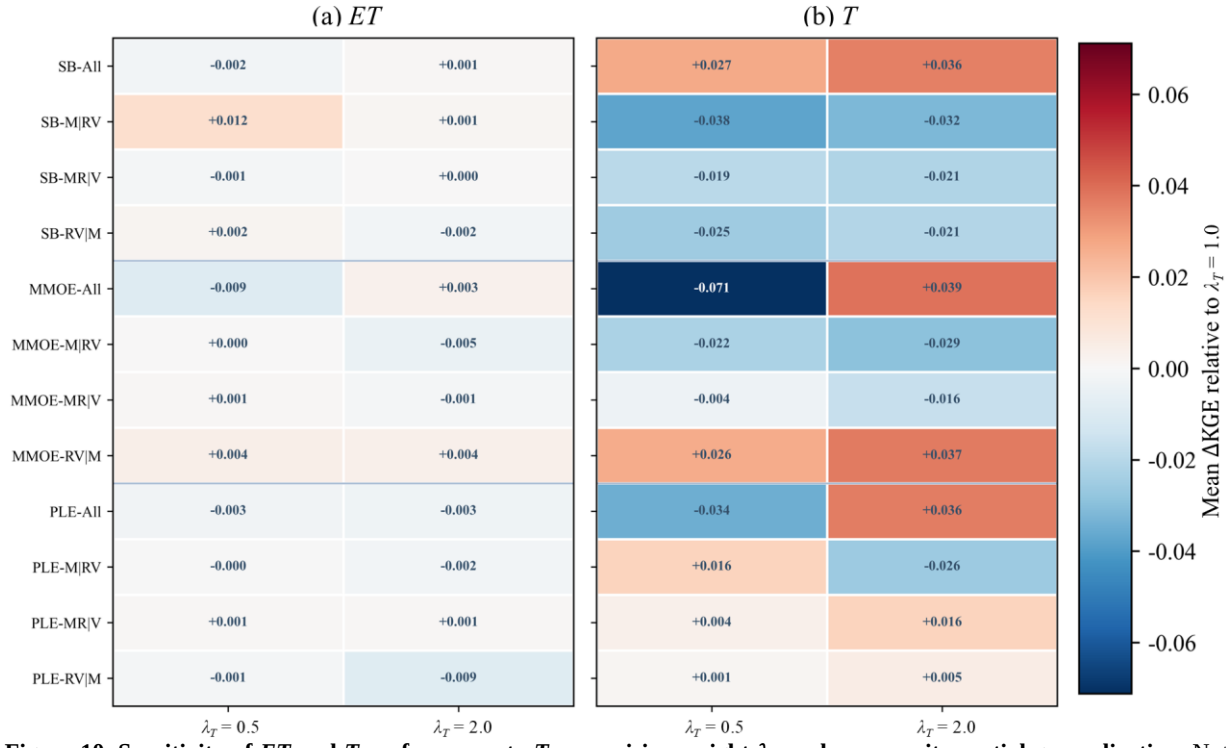


Figure 10. Sensitivity of *ET* and *T* performance to *T*-supervision weight λ_T under cross-site spatial generalization. Note: Each cell shows the mean ΔKGE of a model-input combination relative to $\boldsymbol{\lambda_T = 1.0}$. Columns indicate $\boldsymbol{\lambda_T = 0.5}$ and $\boldsymbol{\lambda_T = 2.0}$, and colors indicate the direction and magnitude of the change.

5 Discussion

5.1 Role of constrained *T* estimates in hourly *ET* modeling

At the hourly scale, total *ET* can change rapidly in response to radiation, atmospheric demand, and aerodynamic forcing, but the same *ET* response may arise from different combinations of canopy transpiration, soil evaporation, and interception evaporation (Kool et al., 2014; Stoy et al., 2019). Estimating *T* together with *ET* therefore provides a way to interpret the vegetation-controlled component of sub-daily *ET* dynamics (Medlyn et al., 2011; Chen and Liu, 2020), rather than treating *ET* only as a bulk water flux.

However, *T* should not be treated simply as an independent numerical target, because auxiliary-task benefits in multi-task learning depend on task-relatedness, shared representation, and supervision quality (Caruana, 1997). A single-task *T* model can produce *T* estimates, but these estimates are not structurally linked to the *ET* estimate for the same sample. In MLF-ETT, the structural coupling between *ET* and *T*/*ET* links *T* estimates directly to total *ET* estimates, preventing unconstrained *T* extrapolation and improving the physical consistency of *T* estimates at *ET*-only sites.

This formulation is particularly useful for *ET*-only sites. Conventional *ET* models provide only total *ET* at these sites, and independent *T* models can only extrapolate unconstrained *T* values. By estimating *T*/*ET* for all samples and multiplying it by

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

estimated *ET*, the proposed framework generates a structurally consistent *T* component even when local *T* labels are unavailable. The machine-learning framework is useful here because it can learn nonlinear driver-response relationships, which are particularly helpful for representing nonlinear transpiration responses to atmospheric demand and canopy activity (Novick et al., 2016; Grossiord et al., 2020).

5.2 *T*-estimation gains and their condition dependence

The role of *T* supervision depended on *T*-label availability and the learnable relationship between *T* labels and input drivers. Within-site temporal generalization, paired *ET* and *T* labels allowed the model to learn site-specific relationships among forcing factors, canopy activity, total *ET*, and sap-flow-derived *T* (Poyatos et al., 2021). At some sites, the single-task MLP was already able to estimate *T* reasonably well, suggesting that the *T* reference values at these sites had a stronger learnable relationship with the input drivers and that *T*-label noise or scale mismatch may have been relatively smaller (Ford et al., 2007). Under such conditions, *T* supervision was more likely to feed back into joint *ET*-*T* learning through shared multi-task representations, rather than only improving an isolated *T* model.

Cross-site spatial generalization was more demanding because *ET* labels were available across all ENF sites, whereas *T* labels were restricted to a small subset of sites. Differences in canopy structure, energy partitioning, soil-water background, and vegetation-atmosphere coupling reduced the transferability of *T*-related representations (Good et al., 2015; Shi et al., 2023; Shi and Cai, 2025). The spatial results, therefore, indicate that sparse auxiliary *T* supervision is more sensitive to cross-site process shifts and observation-scale uncertainty (De Aguiar-Campos et al., 2025).

Environmental conditions were another key factor determining whether multi-task learning improved *T* estimation relative to the single-task model. The *T*-estimation gains were concentrated under high evaporative-demand conditions. Larger *T*-RMSE reductions occurred in the high-*VPD*, high-T_a, and combined high-*VPD*-and-T_a windows, where transpiration is directly related to stomatal regulation, canopy water use, and atmospheric moisture demand (Kar et al., 2025; Nelson et al., 2020; Xing et al., 2024). In contrast, the high-*RH* window under spatial generalization showed no improvement in *T*-RMSE, indicating that *T*-supervision benefits were weaker under humid, low-demand conditions, where *T* signals were less transferable.

Improved *T* estimation should not be interpreted as a guarantee of improved total *ET* estimation. Total *ET* also includes soil evaporation and interception evaporation, and hourly *ET* uncertainty can be affected by energy-balance closure, turbulent-flux representativeness, and timing or scale mismatches between sap-flow-derived *T* and eddy-covariance *ET* (Maltese et al., 2018). Sap-flow-derived *T* labels also contain probe-to-tree-to-site scaling, sapwood-area, sample-tree, and calibration uncertainties (Flo et al., 2019). Thus, *T* supervision is best interpreted as improving the constrained *T* component under favorable label and environmental conditions. At the same time, its feedback to *ET* depends on *T*-label availability and the degree of *T*-label noise or scale mismatch.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

5.3 Input allocation and *T*-supervision strength in multi-task design

Input feature combinations and *T*-supervision strength affected multi-task performance through different pathways. Input feature combinations controlled the allocation of meteorological, radiative, and vegetation information between shared and task-specific branches (Ruder, 2017). Under temporal generalization, RV|M was most favorable for *T* estimation, suggesting that radiation and vegetation variables provided a useful shared canopy-energy background within the same sites (Braghiere et al., 2020; Li et al., 2019; Cheng et al., 2017). Under spatial generalization, M|RV performed better for *T*, indicating that meteorological variables were more transferable as shared features across ENF sites. In contrast, radiative and vegetation-related information was better retained in task-specific branches.

This contrast shows that task-relatedness in *ET*-*T* modeling does not imply that all input features should be shared uniformly. Variables that provide stable, shared information within a site may become more site-specific when transferred across sites with different canopy structures, phenologies, and energy-partitioning regimes. Input feature combinations should therefore be treated as explicit design choices rather than fixed preprocessing options.

T-supervision strength affected the model differently. Changing the *T*-supervision weight altered the contribution of *T* error to the joint training objective, consistent with the role of task weighting in multi-task learning (Wang et al., 2022)and mainly affected *T* performance, whereas *ET* changes remained small across most model-input combinations. Stronger *T* supervision may improve *T* estimation in some configurations, but it can also increase the influence of noisy or weakly transferable *T* labels. Therefore, *T*-supervision strength cannot replace an appropriate input-sharing structure.

Figure 11 provides supplementary evidence for these design implications. *GPP*, R_n, T_a, *VPD*, *LAI*, and solar elevation remained highly important across multiple tasks and scenarios, indicating that carbon-water coupling, energy supply, atmospheric moisture demand, and canopy structure were key information sources for hourly joint *ET*-*T* estimation (Duursma et al., 2013; Han et al., 2024). In contrast, P_a, *RH*, and *WS* contributed less, although their importance varied across scenarios. These results indicate that the input-feature organization changed the variable-importance patterns of key driving factors.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

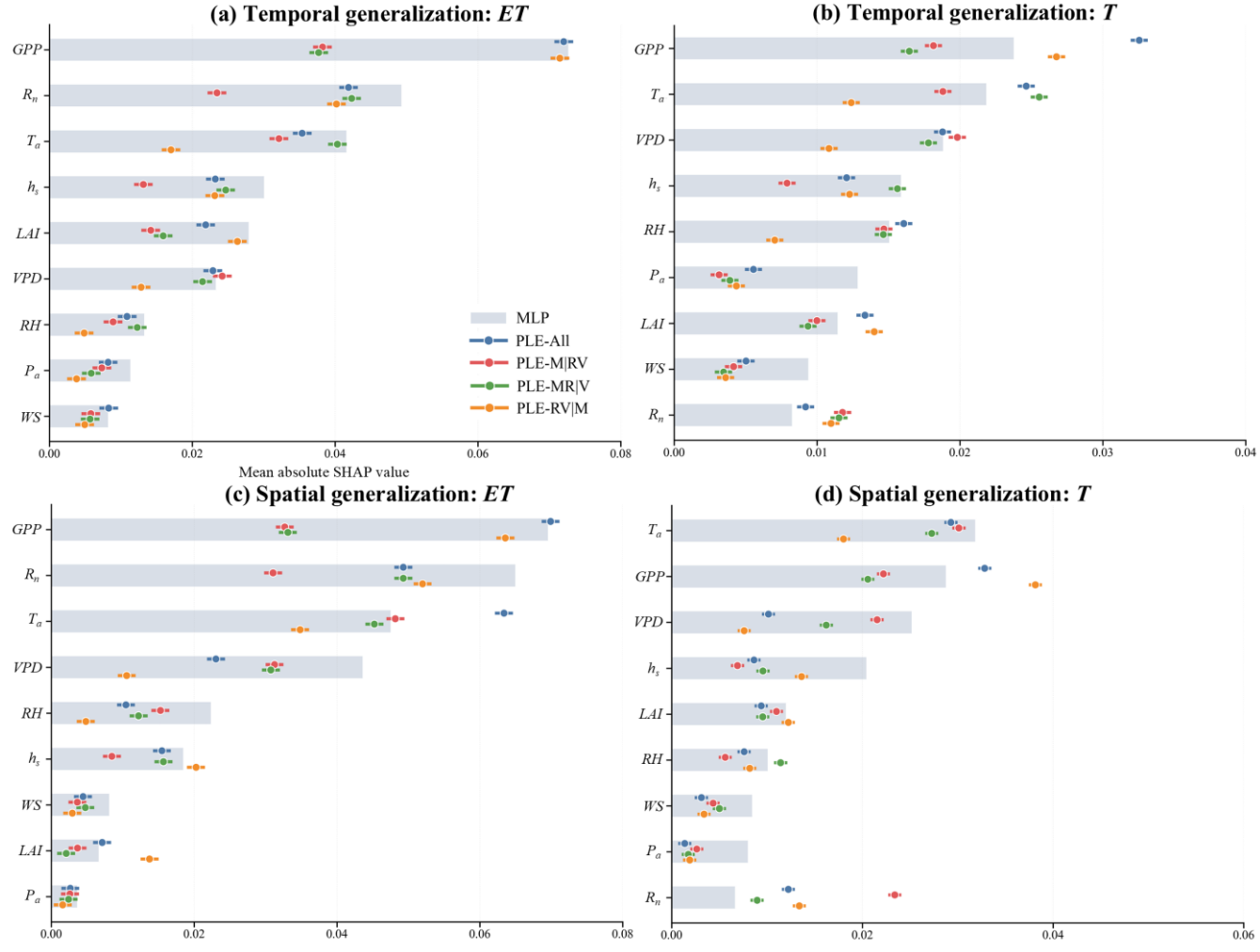


Figure 11. Comparison of mean absolute SHAP values between MLP and PLE input feature combinations. Note: Panels show *ET* and *T* under temporal and spatial generalization. Grey bars indicate the MLP model, and colored points indicate PLE models with different input feature combinations. Variables are sorted by the MLP mean absolute SHAP value within each panel. M, R, and V denote meteorological, radiation, and vegetation feature groups, respectively; features before and after "|" indicate shared and branch-specific inputs.

6 Conclusions

This study developed MLF-ETT, a physically constrained multi-task learning framework for hourly joint estimation of *ET* and *T*, by integrating data from FLUXNET2015, SAPFLUXNET, and GIMMS LAI4g. The framework estimates total *ET* and the transpiration fraction *T*/*ET*, and derives *T* as the product of the estimated *ET* and *T*/*ET*, subject to the physical constraint $0 \leq T \leq ET$. The evaluation was based on 28 unique study sites, including 25 ENF sites for cross-site spatial generalization and 7 *T*-labeled sites for within-site temporal generalization. The evaluation covered model performance across 2 generalization

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

scenarios and the effects of environmental-condition windows, input feature combinations, and *T*-supervision strength. The main conclusions are as follows:

1. The proposed multi-task learning framework enables hourly joint *ET*-*T* estimation through an *ET* and *T*/*ET* output structure, deriving *T* as a bounded component of total *ET* under $0 \leq T \leq ET$. This structure allowed concurrent *ET*-KGE and *T*-KGE gains in some temporally validated configurations. It generated physically constrained *T* estimates for *ET*-only sites under sparse sap-flow-derived *T* data.
2. The multi-task framework reduced *T*-estimation errors relative to the single-task model, with the clearest gains occurring under high evaporative-demand conditions.
3. Input feature combinations and *T*-supervision strength affected the multi-task model performance in different ways. The former determined how meteorological, radiative, and vegetation information was allocated between shared and task-specific branches. In contrast, the latter changed the contribution of *T* supervision to training and affected *T* performance differently across model-input combinations.

Overall, the framework provides a structured way to incorporate limited sap-flow-derived *T* information into hourly *ET* modeling by estimating *T* as a bounded component of total *ET* rather than as an unconstrained extrapolated value.

This study still has several limitations. First, the *T* labels used for supervision were derived from upscaled sap-flow measurements and are affected by sample-tree representativeness, sapwood area estimation, differences in sap-flow methods, and scaling errors. Therefore, the physically constrained output structure ensures bounded *T* estimates but does not eliminate uncertainty in the *T* reference values. Second, the number of sites available for direct *T* validation under spatial generalization was limited, restricting a comprehensive assessment of *T* estimates at sites without local *T* measurements. Third, the spatial generalization experiments were primarily conducted within ENF sites to reduce ecosystem-type heterogeneity, and the framework's transferability across broader ecosystem types has not yet been systematically tested.

Future work could incorporate more independent constraints on *T*, including additional sap-flow measurements, isotope-based partitioning, and remote-sensing *ET*-partitioning products, to better evaluate bounded *T* estimates at *ET*-only sites. Additional input features, such as soil moisture, plant functional type, canopy structure, and site attributes, could help capture site-specific controls on *ET* partitioning and improve cross-site transferability. Future modeling efforts should also combine physically constrained output structures with uncertainty-aware *T* supervision and domain-transfer strategies, thereby extending hourly joint *ET*-*T* estimation from ENF-dominated experiments toward broader multi-ecosystem applications.

Code and data availability

The original data used in this study are publicly available from their respective repositories. FLUXNET2015 data are available from the FLUXNET2015 data portal (https://fluxnet.org/data/fluxnet2015-dataset/). SAPFLUXNET data are available from the SAPFLUXNET project database (https://sapfluxnet.creaf.cat/) and the SAPFLUXNET Zenodo repository (https://zenodo.org/records/2530798). GIMMS LAI4g data are available from Zenodo (https://zenodo.org/records/8281930).

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

The processed hourly datasets and model code generated in this study are archived at Zenodo and are available at https://doi.org/10.5281/zenodo.20095120 (Fu et al., 2026).

Author contribution

Hongyang Fu: conceptualization, data curation, formal analysis, writing–original draft, writing–review and editing. Yuanyuan Yang: conceptualization, supervision, writing–review and editing. Dengfeng Liu: conceptualization, supervision, writing–review and editing, project administration. Qiang Li, Zhihua He, Huimin Lei, Mohd Yawar Ali Khan, Fiaz Hussain: writing–review and editing. All authors have read and agreed to this version of the manuscript.

Competing interests

The authors declare no conflicts of interest.

Financial support

This study was financially supported by the National Natural Science Foundation of China (Grant No. 52279025).

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623-2631, https://doi.org/10.1145/3292500.3330701, 2019.

Allen, R. G., Tasumi, M., and Trezza, R.: Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model, Journal of Irrigation and Drainage Engineering, 133, 380-394, https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380), 2007.

Althoff, D., Filgueiras, R., Dias, S. H. B., and Rodrigues, L. N.: Impact of sum-of-hourly and daily timesteps in the computations of reference evapotranspiration across the Brazilian territory, Agricultural Water Management, 226, 105785, https://doi.org/10.1016/j.agwat.2019.105785, 2019.

Amani, S. and Shafizadeh-Moghadam, H.: A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data, Agricultural Water Management, 284, 108324, https://doi.org/10.1016/j.agwat.2023.108324, 2023.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B.: Algorithms for hyper-parameter optimization, Proceedings of the 25th International Conference on Neural Information Processing Systems, Granada, Spain, 2011.

Braghiere, R. K., Quaife, T., Black, E., Ryu, Y., Chen, Q., De Kauwe, M. G., and Baldocchi, D.: Influence of sun zenith angle on canopy clumping and the resulting impacts on photosynthesis, Agricultural and Forest Meteorology, 291, 108065, https://doi.org/10.1016/j.agrformet.2020.108065, 2020.

Bright, R. M., Miralles, D. G., Poyatos, R., and Eisner, S.: Simple Models Outperform More Complex Big-Leaf Models of Daily Transpiration in Forested Biomes, Geophysical Research Letters, 49, e2022GL100100, https://doi.org/10.1029/2022GL100100, 2022.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Cao, S., Li, M., Zhu, Z., Wang, Z., Zha, J., Zhao, W., Duanmu, Z., Chen, J., Zheng, Y., Chen, Y., Myneni, R. B., and Piao, S.: Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020, Earth Syst. Sci. Data, 15, 4877-4899, https://doi.org/10.5194/essd-15-4877-2023, 2023.

Caruana, R.: Multitask Learning: A Knowledge-Based Source of Inductive Bias, International Conference on Machine Learning, 41-48, https://doi.org/10.1016/b978-1-55860-307-3.50012-5, 1993.

Caruana, R.: Multitask Learning, Machine Learning, 28, 41-75, https://doi.org/10.1023/A:1007379606734, 1997.

Chen, J. M. and Liu, J.: Evolution of evapotranspiration models using thermal and shortwave remote sensing data, Remote Sensing of Environment, 237, 111594, https://doi.org/10.1016/j.rse.2019.111594, 2020.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, https://doi.org/10.1145/2939672.2939785, 2016.

Cheng, H., Liu, D., Ming, G., Han, S., Khan, M. Y. A., Wang, L., and Li, Q.: Estimation of actual evapotranspiration from the SEBAL model and comparison with four datasets in an irrigation district of China, Irrigation Science, 44, 70, https://doi.org/10.1007/s00271-026-01113-5, 2026.

Cheng, L., Zhang, L., Wang, Y.-P., Canadell, J. G., Chiew, F. H. S., Beringer, J., Li, L., Miralles, D. G., Piao, S., and Zhang, Y.: Recent increases in terrestrial carbon uptake at little cost to the water cycle, Nature Communications, 8, 110, https://doi.org/10.1038/s41467-017-00114-5, 2017.

de Aguiar-Campos, N., Edwards, W., and Laurance, S. G. W.: Tropical forest transpiration estimates are geographically, ecologically and methodologically biased: a systematic review of sap flow research, Agricultural and Forest Meteorology, 373, 110738, https://doi.org/10.1016/j.agrformet.2025.110738, 2025.

Duarte Rocha, A., Vulova, S., van der Tol, C., Förster, M., and Kleinschmit, B.: Modelling hourly evapotranspiration in urban environments with SCOPE using open remote sensing and meteorological data, Hydrol. Earth Syst. Sci., 26, 1111-1129, https://doi.org/10.5194/hess-26-1111-2022, 2022.

Duursma, R. A., Payton, P., Bange, M. P., Broughton, K. J., Smith, R. A., Medlyn, B. E., and Tissue, D. T.: Near-optimal response of instantaneous transpiration efficiency to vapour pressure deficit, temperature and [CO2] in cotton (Gossypium hirsutum L.), Agricultural and Forest Meteorology, 168, 168-176, https://doi.org/10.1016/j.agrformet.2012.09.005, 2013.

Feng, J., Wang, W., Xu, F., and Wang, S.: Evaluating the ability of deep learning on actual daily evapotranspiration estimation over the heterogeneous surfaces, Agricultural Water Management, 291, 108627, https://doi.org/10.1016/j.agwat.2023.108627, 2024.

Ferreira, L. B., da Cunha, F. F., and Fernandes Filho, E. I.: Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil, Agricultural Water Management, 259, 107281, https://doi.org/10.1016/j.agwat.2021.107281, 2022.

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., Kilic, A., Tu, K., Miralles, D. D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A. J., French, A., Schimel, D., Famiglietti, J. S., Stephens, G., and Wood, E. F.: The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources, Water Resources Research, 53, 2618-2626, https://doi.org/10.1002/2016WR020175, 2017.

Flo, V., Martinez-Vilalta, J., Steppe, K., Schuldt, B., and Poyatos, R.: A synthesis of bias and uncertainty in sap flow methods, Agricultural and Forest Meteorology, 271, 362-374, https://doi.org/10.1016/j.agrformet.2019.03.012, 2019.

Ford, C. R., Hubbard, R. M., Kloeppel, B. D., and Vose, J. M.: A comparison of sap flux-based evapotranspiration estimates with catchment-scale water balance, Agricultural and Forest Meteorology, 145, 176-185, https://doi.org/10.1016/j.agrformet.2007.04.010, 2007.

Fu, H., Yang, Y., Liu, D., Li, Q., He, Z., Lei, H., Khan, M. Y. A., and Hussain, F.: Physically constrained multi-task learning for hourly joint estimation of evapotranspiration and transpiration from sparse sap-flow data [dataset], https://doi.org/10.5281/zenodo.20095120, 2026.

Fu, J., Wang, W., Shao, Q., Xing, W., Cao, M., Wei, J., Chen, Z., and Nie, W.: Improved global evapotranspiration estimates using proportionality hypothesis-based water balance constraints, Remote Sensing of Environment, 279, 113140, https://doi.org/10.1016/j.rse.2022.113140, 2022.

Good, S. P., Noone, D., and Bowen, G.: Hydrologic connectivity constrains partitioning of global terrestrial water fluxes, Science, 349, 175-177, https://doi.org/10.1126/science.aaa5931, 2015.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Gowda, P. H., Howell, T. A., Paul, G., Colaizzi, P. D., Marek, T. H., Su, B., and Copeland, K. S.: Deriving Hourly Evapotranspiration Rates with SEBS: A Lysimetric Evaluation, Vadose Zone Journal, 12, vzj2012.0110, https://doi.org/10.2136/vzj2012.0110, 2013.

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., Sperry, J. S., and McDowell, N. G.: Plant responses to rising vapor pressure deficit, New Phytologist, 226, 1550-1566, https://doi.org/10.1111/nph.16485, 2020.

Guo, F., Liu, D., Mo, S., Huang, Q., Ma, L., Xie, S., Deng, W., Ming, G., and Fan, J.: Estimation of daily evapotranspiration in gully area scrub ecosystems on Loess Plateau of China based on multisource observation data, Ecological Indicators, 154, 110671, https://doi.org/10.1016/j.ecolind.2023.110671, 2023.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Han, Y., Calabrese, S., Du, H., and Yin, J.: Evaluating biases in Penman and Penman–Monteith evapotranspiration rates at different timescales, Journal of Hydrology, 638, 131534, https://doi.org/10.1016/j.jhydrol.2024.131534, 2024.

Hashem, A. A., Engel, B. A., Bralts, V. F., Marek, G. W., Moorhead, J. E., Rashad, M., Radwan, S., and Gowda, P. H.: Landsat Hourly Evapotranspiration Flux Assessment using Lysimeters for the Texas High Plains, Water, 12, 1192, https://doi.org/10.3390/w12041192, 2020.

Henderson-Sellers, B.: A new formula for latent heat of vaporization of water as a function of temperature, Quarterly Journal of the Royal Meteorological Society, 110, 1186-1190, https://doi.org/10.1002/qj.49711046626, 1984.

Kar, K. K., Haggerty, R., Sharma, H., Dwivedi, D., and Roy, T.: Evapotranspiration Partitioning Using Flux Tower Data in a Semi-Arid Ecosystem, Hydrological Processes, 39, e70083, https://doi.org/10.1002/hyp.70083, 2025.

Kool, D., Agam, N., Lazarovitch, N., Heitman, J. L., Sauer, T. J., and Ben-Gal, A.: A review of approaches for evapotranspiration partitioning, Agricultural and Forest Meteorology, 184, 56-70, https://doi.org/10.1016/j.agrformet.2013.09.003, 2014.

Leng, J., Chen, J. M., Li, W., Luo, X., Xu, M., Liu, J., Wang, R., Rogers, C., Li, B., and Yan, Y.: Global datasets of hourly carbon and water fluxes simulated using a satellite-based process model with dynamic parameterizations, Earth Syst. Sci. Data, 16, 1283-1300, https://doi.org/10.5194/essd-16-1283-2024, 2024.

Li, B., Li, R., Sun, T., Gong, A., Tian, F., Khan, M. Y. A., and Ni, G.: Improving LSTM hydrological modeling with spatiotemporal deep learning and multi-task learning: A case study of three mountainous areas on the Tibetan Plateau, Journal of Hydrology, 620, 129401, https://doi.org/10.1016/j.jhydrol.2023.129401, 2023.

Li, X., Gentine, P., Lin, C., Zhou, S., Sun, Z., Zheng, Y., Liu, J., and Zheng, C.: A simple and objective method to partition evapotranspiration into transpiration and evaporation at eddy-covariance sites, Agricultural and Forest Meteorology, 265, 171-182, https://doi.org/10.1016/j.agrformet.2018.11.017, 2019.

Liu, G. and Wang, W.: Competing Effects of Vegetation Greening-Induced Changes in Summer Evapotranspiration and Precipitation on Water Yield in the Yangtze River Basin Based on WRF Simulations, Water Resources Research, 61, e2024WR038663, https://doi.org/10.1029/2024WR038663, 2025.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, https://doi.org/10.48550/arXiv.1705.07874, 2017.

Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., and Chi, E. H.: Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom, https://doi.org/10.1145/3219819.3220007, 2018.

Maltese, A., Awada, H., Capodici, F., Ciraolo, G., La Loggia, G., and Rallo, G.: On the Use of the Eddy Covariance Latent Heat Flux and Sap Flow Transpiration for the Validation of a Surface Energy Balance Model, Remote Sensing, 10, 195, https://doi.org/10.3390/rs10020195, 2018.

Man, R., Pan, Y., and Lv, Y.: Estimation of Actual Evapotranspiration and Its Components at Hourly and Daily Scales Using Dual Crop Coefficient Method for Water-Saving Irrigated Rice Paddy Field, Agronomy, 15, 2133, https://doi.org/10.3390/agronomy15092133, 2025.

Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M., and Wingate, L.: Reconciling the optimal and empirical approaches to modelling stomatal conductance, Global Change Biology, 17, 2134-2144, https://doi.org/10.1111/j.1365-2486.2010.02375.x, 2011.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Nelson, J. A., Pérez-Priego, O., Zhou, S., Poyatos, R., Zhang, Y., Blanken, P. D., Gimeno, T. E., Wohlfahrt, G., Desai, A. R., Gioli, B., Limousin, J.-M., Bonal, D., Paul-Limoges, E., Scott, R. L., Varlagin, A., Fuchs, K., Montagnani, L., Wolf, S., Delpierre, N., Berveiller, D., Gharun, M., Belelli Marchesini, L., Gianelle, D., Šigut, L., Mammarella, I., Siebicke, L., Andrew Black, T., Knohl, A., Hörtnagl, L., Magliulo, V., Besnard, S., Weber, U., Carvalhais, N., Migliavacca, M., Reichstein, M., and Jung, M.: Ecosystem transpiration and evaporation: Insights from three water flux partitioning methods across FLUXNET sites, Global Change Biology, 26, 6916-6930, https://doi.org/10.1111/gcb.15314, 2020.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Dušek, J., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op de Beeck, M., Billesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G. A., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K. M., Korkiakoski, M., Mammarella, I., Gharun, M., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J. M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, Biogeosciences, 21, 5079-5115, https://doi.org/10.5194/bg-21-5079-2024, 2024.

Novick, K. A., Ficklin, D. L., Stoy, P. C., Williams, C. A., Bohrer, G., Oishi, A. C., Papuga, S. A., Blanken, P. D., Noormets, A., Sulman, B. N., Scott, R. L., Wang, L., and Phillips, R. P.: The increasing importance of atmospheric demand for ecosystem water and carbon fluxes, Nature Climate Change, 6, 1023-1027, https://doi.org/10.1038/nclimate3114, 2016.

Oki, T. and Kanae, S.: Global Hydrological Cycles and World Water Resources, Science, 313, 1068-1072, https://doi.org/10.1126/science.1128845, 2006.

Ouyang, W., Gu, X., Ye, L., Liu, X., and Zhang, C.: Exploring Hydrological Variable Interconnections and Enhancing Predictions for Data-Limited Basins Through Multi-Task Learning, Water Resources Research, 61, e2023WR036593, https://doi.org/10.1029/2023WR036593, 2025.

Pan, X., Yang, Z., Liu, Y., Yuan, J., Wang, Z., Liu, S., and Yang, Y.: A non-parametric method combined with surface flux equilibrium for estimating terrestrial evapotranspiration: Validation at eddy covariance sites, Journal of Hydrology, 631, 130682, https://doi.org/10.1016/j.jhydrol.2024.130682, 2024.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. d., Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. d., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck,

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V. R. d., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., Sánchez-Cañete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Poyatos, R., Granda, V., Molowny-Horas, R., Mencuccini, M., Steppe, K., and Martínez-Vilalta, J.: SAPFLUXNET: towards a global database of sap flow measurements, Tree Physiology, 36, 1449-1455, https://doi.org/10.1093/treephys/tpw110, 2016.

Poyatos, R., Granda, V., Flo, V., Adams, M. A., Adorján, B., Aguadé, D., Aidar, M. P. M., Allen, S., Alvarado-Barrientos, M. S., Anderson-Teixeira, K. J., Aparecido, L. M., Arain, M. A., Aranda, I., Asbjornsen, H., Baxter, R., Beamesderfer, E., Berry, Z. C., Berveiller, D., Blakely, B., Boggs, J., Bohrer, G., Bolstad, P. V., Bonal, D., Bracho, R., Brito, P., Brodeur, J., Casanoves, F., Chave, J., Chen, H., Cisneros, C., Clark, K., Cremonese, E., Dang, H., David, J. S., David, T. S., Delpierre, N., Desai, A. R., Do, F. C., Dohnal, M., Domec, J. C., Dzikiti, S., Edgar, C., Eichstaedt, R., El-Madany, T. S., Elbers, J., Eller, C. B., Euskirchen, E. S., Ewers, B., Fonti, P., Forner, A., Forrester, D. I., Freitas, H. C., Galvagno, M., Garcia-Tejera, O., Ghimire, C. P., Gimeno, T. E., Grace, J., Granier, A., Griebel, A., Guangyu, Y., Gush, M. B., Hanson, P. J., Hasselquist, N. J., Heinrich, I., Hernandez-Santana, V., Herrmann, V., Hölttä, T., Holwerda, F., Irvine, J., Isarangkool Na Ayutthaya, S., Jarvis, P. G., Jochheim, H., Joly, C. A., Kaplick, J., Kim, H. S., Klemedtsson, L., Kropp, H., Lagergren, F., Lane, P., Lang, P., Lapenas, A., Lechuga, V., Lee, M., Leuschner, C., Limousin, J. M., Linares, J. C., Linderson, M. L., Lindroth, A., Llorens, P., López-Bernal, Á., Loranty, M. M., Lüttschwager, D., Macinnis-Ng, C., Maréchaux, I., Martin, T. A., Matheny, A., McDowell, N., McMahon, S., Meir, P., Mészáros, I., Migliavacca, M., Mitchell, P., Mölder, M., Montagnani, L., Moore, G. W., Nakada, R., Niu, F., Nolan, R. H., Norby, R., Novick, K., Oberhuber, W., Obojes, N., Oishi, A. C., Oliveira, R. S., Oren, R., Ourcival, J. M., Paljakka, T., Perez-Priego, O., Peri, P. L., Peters, R. L., Pfautsch, S., Pockman, W. T., Preisler, Y., Rascher, K., Robinson, G., Rocha, H., Rocheteau, A., Röll, A., Rosado, B. H. P., Rowland, L., Rubtsov, A. V., Sabaté, S., Salmon, Y., Salomón, R. L., Sánchez-Costa, E., Schäfer, K. V. R., Schuldt, B., Shashkin, A., Stahl, C., Stojanović, M., Suárez, J. C., Sun, G., Szatniewska, J., Tatarinov, F., Tesař, M., Thomas, F. M., Tor-ngern, P., Urban, J., Valladares, F., van der Tol, C., van Meerveld, I., Varlagin, A., Voigt, H., Warren, J., Werner, C., Werner, W., Wieser, G., Wingate, L., Wullschleger, S., Yi, K., Zweifel, R., Steppe, K., Mencuccini, M., and Martínez-Vilalta, J.: Global transpiration data from sap flow measurements: the SAPFLUXNET database, Earth Syst. Sci. Data, 13, 2607-2649, https://doi.org/10.5194/essd-13-2607-2021, 2021.

Reyes Rojas, L. A., Moletto-Lobos, I., Corradini, F., Mattar, C., Fuster, R., and Escobar-Avaria, C.: Determining Actual Evapotranspiration Based on Machine Learning and Sinusoidal Approaches Applied to Thermal High-Resolution Remote Sensing Imagery in a Semi-Arid Ecosystem, Remote Sensing, 13, 4105, https://doi.org/10.3390/rs13204105, 2021.

Rong, Y., Wang, W., Wu, P., Wang, P., Zhang, C., Wang, C., and Huo, Z.: A Novel Hybrid Deep Learning Framework for Evaluating Field Evapotranspiration Considering the Impact of Soil Salinity, Water Resources Research, 60, e2023WR036809, https://doi.org/10.1029/2023WR036809, 2024.

Ruder, S.: An Overview of Multi-Task Learning in Deep Neural Networks, https://doi.org/10.48550/arXiv.1706.05098, 2017.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533-536, https://doi.org/10.1038/323533a0, 1986.

S. Ha, W., R. Diak, G., and F. Krajewski, W.: Estimating Near Real-Time Hourly Evapotranspiration Using Numerical Weather Prediction Model Output and GOES Remote Sensing Data in Iowa, Remote Sensing, 12, 2337, https://doi.org/10.3390/rs12142337, 2020.

Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J. A., and Kumar, V.: Multi-Task Deep Learning of Daily Streamflow and Water Temperature, Water Resources Research, 58, e2021WR030138, https://doi.org/10.1029/2021WR030138, 2022.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Schlesinger, W. H. and Jasechko, S.: Transpiration in the global water cycle, Agricultural and Forest Meteorology, 189-190, 115-117, https://doi.org/10.1016/j.agrformet.2014.01.011, 2014.
Shan, X., Cui, N., Cai, H., Hu, X., and Zhao, L.: Estimation of summer maize evapotranspiration using MARS model in the semi-arid region of northwest China, Computers and Electronics in Agriculture, 174, 105495, https://doi.org/10.1016/j.compag.2020.105495, 2020.
Sharafi, S. and Mohammadi Ghaleni, M.: Revealing accuracy in climate dynamics: enhancing evapotranspiration estimation using advanced quantile regression and machine learning models, Applied Water Science, 14, 162, https://doi.org/10.1007/s13201-024-02211-5, 2024.
Shi, H. and Cai, X.: Extrapolability improvement of machine learning-based evapotranspiration models via domain-adversarial neural networks, Environmental Modelling & Software, 187, 106383, https://doi.org/10.1016/j.envsoft.2025.106383, 2025.
Shi, H., Luo, G., Hellwich, O., He, X., Xie, M., Zhang, W., Ochege, F. U., Ling, Q., Zhang, Y., Gao, R., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Comparing the use of all data or specific subsets for training machine learning models in hydrology: A case study of evapotranspiration prediction, Journal of Hydrology, 627, 130399, https://doi.org/10.1016/j.jhydrol.2023.130399, 2023.
Stoy, P. C., El-Madany, T. S., Fisher, J. B., Gentine, P., Gerken, T., Good, S. P., Klosterhalfen, A., Liu, S., Miralles, D. G., Perez-Priego, O., Rigden, A. J., Skaggs, T. H., Wohlfahrt, G., Anderson, R. G., Coenders-Gerrits, A. M. J., Jung, M., Maes, W. H., Mammarella, I., Mauder, M., Migliavacca, M., Nelson, J. A., Poyatos, R., Reichstein, M., Scott, R. L., and Wolf, S.: Reviews and syntheses: Turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities, Biogeosciences, 16, 3747-3775, https://doi.org/10.5194/bg-16-3747-2019, 2019.
Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6, 85-100, https://doi.org/10.5194/hess-6-85-2002, 2002.
Tang, H., Liu, J., Zhao, M., and Gong, X.: Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations, Proceedings of the 14th ACM Conference on Recommender Systems, Virtual Event, Brazil, https://doi.org/10.1145/3383313.3412236, 2020.
Tawegoum, R., Leroy, F., Sintes, G., and Chassériaux, G.: Forecasting hourly evapotranspiration for triggering irrigation in nurseries, Biosystems Engineering, 129, 237-247, https://doi.org/10.1016/j.biosystemseng.2014.10.011, 2015.
Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291-4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.
Wang, J., Lin, L., Teng, Z., and Zhang, Y.: Multitask Learning Based on Improved Uncertainty Weighted Loss for Multi-Parameter Meteorological Data Prediction, Atmosphere, 13, 989, https://doi.org/10.3390/atmos13060989, 2022.
Wang, X., Zhong, L., Ma, Y., Fu, Y., Han, C., Li, P., Wang, Z., and Qi, Y.: Estimation of hourly actual evapotranspiration over the Tibetan Plateau from multi-source data, Atmospheric Research, 281, 106475, https://doi.org/10.1016/j.atmosres.2022.106475, 2023.
Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasechko, S., and Lee, X.: Revisiting the contribution of transpiration to global terrestrial evapotranspiration, Geophysical Research Letters, 44, 2792-2801, https://doi.org/10.1002/2016GL072235, 2017.
Woo, D. K.: Estimating actual evapotranspiration from widely available meteorological data with a hybrid CNN–LSTM, Agricultural Water Management, 323, 110078, https://doi.org/10.1016/j.agwat.2025.110078, 2026.
Xiang, Y., Wang, G., Gessler, A., Sun, X., Lin, S., Tang, Z., Sun, S., and Hu, Z.: Long-term variations in the ratio of transpiration to evapotranspiration and their drivers in a humid subalpine forest, Agricultural and Forest Meteorology, 372, 110692, https://doi.org/10.1016/j.agrformet.2025.110692, 2025.
Xing, W., Wang, W., Cai, Y., Yu, Z., Shao, Q., Cao, X., Cao, M., Yang, L., and Yong, B.: Evapotranspiration partitioning through water stable isotopic measurements in a subtropical coniferous forest, Ecohydrology, 17, e2653, https://doi.org/10.1002/eco.2653, 2024.
Yan, Y., Yu, Z., Zhu, J., Wang, J., and Li, Q.: Integrating multi-task learning with a differentiable physics constrained framework for hydrological forecasting, Scientific Reports, 16, 13824, https://doi.org/10.1038/s41598-026-41277-w, 2026.

https://doi.org/10.5194/egusphere-2026-2687
Preprint. Discussion started: 9 July 2026

Yu, L., Zhou, S., Zhao, X., Gao, X., Jiang, K., Zhang, B., Cheng, L., Song, X., and Siddique, K. H. M.: Evapotranspiration Partitioning Based on Leaf and Ecosystem Water Use Efficiency, Water Resources Research, 58, e2021WR030629, https://doi.org/10.1029/2021WR030629, 2022.

Zhang, C., Zhou, C., Luo, G., Ye, S., and Shi, Z.: Physics-constrained machine learning for satellite-derived evapotranspiration in China, Journal of Hydrology, 660, 133512, https://doi.org/10.1016/j.jhydrol.2025.133512, 2025.

Zhang, C., Luo, G., Hellwich, O., Chen, C., Zhang, W., Xie, M., He, H., Shi, H., and Wang, Y.: A framework for estimating actual evapotranspiration at weather stations without flux observations by combining data from MODIS and flux towers through a machine learning approach, Journal of Hydrology, 603, 127047, https://doi.org/10.1016/j.jhydrol.2021.127047, 2021.

Zhang, Y. and Yang, Q.: A Survey on Multi-Task Learning, IEEE Transactions on Knowledge and Data Engineering, 34, 5586-5609, https://doi.org/10.1109/TKDE.2021.3070203, 2022.

Zheng, X., Liu, D., Huang, S., Wang, H., and Meng, X.: Achieving water budget closure through physical hydrological process modelling: insights from a large-sample study, Hydrol. Earth Syst. Sci., 29, 627-653, https://doi.org/10.5194/hess-29-627-2025, 2025.

Zhong, L., Ma, Y., Hu, Z., Fu, Y., Hu, Y., Wang, X., Cheng, M., and Ge, N.: Estimation of hourly land surface heat fluxes over the Tibetan Plateau by the combined use of geostationary and polar-orbiting satellites, Atmos. Chem. Phys., 19, 5529-5541, https://doi.org/10.5194/acp-19-5529-2019, 2019.

Zhou, S., Yu, B., Zhang, Y., Huang, Y., and Wang, G.: Partitioning evapotranspiration based on the concept of underlying water use efficiency, Water Resources Research, 52, 1160-1175, https://doi.org/10.1002/2015WR017766, 2016.

Zhuang, Q., Zhu, W., Yan, N., Faour, G., Ibrahim, M., and Zhu, L.: An Interpretable Machine Learning Approach to Remote Sensing-Based Estimation of Hourly Agricultural Evapotranspiration in Drylands, Agriculture, 15, 2193, https://doi.org/10.3390/agriculture15212193, 2025.