Journal of Environmental Management 416 (2026) 130780

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

A hybrid deep learning framework for estimating urban tree transpiration

Xiang Zhang [a,*], Xue Zhong [b], Jun-Ru Yan [c], Qi Li [a,d,e], Kai-Xin Liu [f], Ling-Ye Yao [g], Long-Hao Zhang [h], Astrid Moser-Reischl [a], Thomas Rötzer [a], Stephan Pauleit [a], Mohammad A. Rahman [i]

[a] Chair for Strategic Landscape Planning and Management, Technical University of Munich, Emil-Ramann-Str. 6, Freising, 85354, Germany
[b] School of Architecture, Southwest Jiaotong University, Chengdu, 611756, China
[c] College of Horticulture and Landscape Architecture, Southwest University, Chongqing, 400715, China
[d] State Key Laboratory of Subtropical Building and Urban Science, Guangzhou, 510641, China
[e] School of Architecture, South China University of Technology, Guangzhou, 510641, China
[f] Institute of Regional Science, Karlsruhe Institute of Technology, Karlsruhe, 76131, Germany
[g] School of Architecture, The Chinese University of Hong Kong, New Territories, Hong Kong, China
[h] Department of Architecture, Tianjin University, Tianjin, 300072, China
[i] School of Agriculture, Food and Ecosystem Sciences, The University of Melbourne, Burnley, Victoria, 3121, Australia

ARTICLE INFO

Keywords:
Urban tree transpiration
Physics-guided deep learning
Explainable machine learning
Species-specific response
Jarvis-type model

ABSTRACT

Urban trees regulate urban water-energy exchange via transpiration, yet accurate estimates in urban settings remain challenging. Traditional Jarvis-type models often require frequent recalibration and can produce biased estimates when parameters vary or structural assumptions are violated. Alternatively, hybrid deep learning (DL) models that integrate data-driven approaches with process knowledge represent a promising alternative, merging the flexibility of DL with prior information from process models. Nonetheless, the effectiveness of these hybrids for estimating urban tree transpiration has rarely been evaluated. This study investigates three physics-guided DL strategies that incorporate the Jarvis model: residual learning (M2), weighted fusion (M3), and feature learning (M4), against three baselines: Jarvis, pure DL, and Random Forest (RF). We also analyze SHAP-based explainability to assess feature contributions and interactions. Utilizing observed hourly sap flow and environmental data from Munich and Würzburg (2015-2021), we estimate hourly latent heat flux (*LE*) for two physiologically contrasting species, *Tilia cordata* and *Robinia pseudoacacia*, from shortwave radiation (R_s), air temperature (T_a), vapor pressure deficit (*VPD*), soil moisture content (θ), and day of year (*DOY*). All DL models outperform the Jarvis baseline, with M4 showing the best overall performance: test NSE 0.86 (0.90) and RMSE 11.43 (1.93) W/m^2 for two species, while maintaining strong performance with limited training data. SHAP identifies model-derived species-specific response patterns consistent with known differences in hydraulic architecture, and diagnoses regimes where Jarvis assumptions underperform. This framework offers a practical pathway for embedding process knowledge into broader urban evapotranspiration and microclimate models.

1. Introduction

Latent heat flux (*LE*) from plant transpiration is a key component of the terrestrial water and surface energy balance (Jasechko et al., 2013; Evaristo et al., 2015). By consuming available energy through phase change, *LE* cools leaves and the near-surface air and can influence regional and global climate (Monteith and Unsworth, 2013). Accurate estimation of *LE* supports improvements in plant water-use efficiency, reliable long-term estimates of coupled carbon, water and energy fluxes, and precise control of urban microclimates (Liu et al., 2023). In urban areas, *LE* shows significant spatiotemporal variability due to heterogeneous hydrological conditions, including impervious surfaces, restricted rooting volumes, episodic irrigation, and diverse species and canopy structures (Rahman et al., 2018, 2024). Intensifying heat waves and the urban heat island (UHI) further elevate the role of natural cooling strategies, with urban trees as the core (Zhang et al., 2022; Zhang et al., 2025). Estimating tree-level *LE* in urban environments is therefore both scientifically important and practically urgent.

* Corresponding author.
E-mail address: xxiang.zhang@tum.de (X. Zhang).

https://doi.org/10.1016/j.jenvman.2026.130780
Received 11 February 2026; Received in revised form 10 August 2026; Accepted 21 August 2026
Available online 28 August 2026
0301-4797/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Direct observation of urban-tree *LE* remains challenging. Thermal sap flow sensors and gas-exchange chambers require careful installation, calibration, and long-term maintenance (Flo et al., 2019; de Aguiar-Campos et al., 2025). These methods can be costly and prone to interruptions, especially in urban settings (Wang et al., 2021). Therefore, reliable estimation models are essential. Tree-level *LE* is mainly driven by shortwave radiation (R_s), air temperature (T_a), vapor pressure deficit (*VPD*), and soil moisture content (θ), along with aerodynamic coupling (Mitchell et al., 2009; Verhoef and Egea, 2014; Wang et al., 2020). A common approach is the energy-resistance framework (Monteith, 1965). The Penman-Monteith (PM) equation relates *LE* to available energy, aerodynamic resistance, and canopy resistance. When *LE* observations such as sap flow or eddy covariance are available, canopy resistance or conductance can be inferred by inversion (Allen et al., 2006). Otherwise, these quantities must be parameterized to calculate *LE*. A widely used phenomenological choice is the Jarvis-type model (Jarvis, 1976), which multiplies bounded response functions (range 0-1) for R_s, T_a, *VPD* and θ by a maximum conductance, assuming these factors act independently. Variants of this structure can be parameterized more conveniently to estimate *LE* without explicitly modeling aerodynamic resistance (Wang et al., 2016, 2020), although its influence is not eliminated (Whitley et al., 2009). Despite the interpretability and perceived extrapolation capability of these process models (Houshmandfar et al., 2021), they rely on empirical parameters that can vary among species and may fluctuate throughout a growing season, requiring regular re-estimation (Gaur and Drewry, 2024). The independence assumption may be violated by nonlinear interactions, lags, and hysteresis (Whitley et al., 2009; Houshmandfar et al., 2021; Gaur and Drewry, 2024). Furthermore, physiological responses are not strictly linear (Cai et al., 2019), and important drivers may be overlooked. Accordingly, there is significant potential to enhance predictive accuracy and stability, particularly under changing conditions, by integrating data-driven components into these models when appropriate.

Unlike process-based models, statistical models learn the relationships between observed *LE* and predictor variables, whether linear or nonlinear (Ali et al., 2018; Aghelpour et al., 2020). Machine learning (ML) methods such as support vector machines (SVM), random forests (RF), gradient boosted trees (XGBoost), and artificial neural networks (ANN) are widely used and often outperform process models in modelling vegetation *LE* or stomatal conductance (Whitley et al., 2009; Cai et al., 2019; He et al., 2022; Wu et al., 2022; Gaur and Drewry, 2024). Several studies on crops like wheat and maize have reported that SVM or RF models show higher accuracy compared to Jarvis-type and PM models (Houshmandfar et al., 2021; Zheng et al., 2021). However, conventional ML still relies on manual feature design and can struggle to represent complex data distributions (Liu et al., 2023, 2025). Deep learning (DL), a subfield of ML, offers greater representational capacity: it enables end-to-end representation learning from raw inputs, captures strong nonlinearities and higher-order interactions, and can be combined with process consistency constraints (LeCun et al., 2015). Recent applications in hydrology have demonstrated that deep neural networks can outperform SVM, XGBoost, and shallow ANNs for estimating crop *LE*, tree transpiration, and evapotranspiration (Li et al., 2022; Xing et al., 2022; Qian et al., 2024). Despite these advantages, relying solely on ML or DL has its limitations (Reichstein et al., 2019). These models do not explicitly enforce physical or plant physiological constraints, which can lead to physically unrealistic estimations (Reichstein et al., 2019). Additionally, compared to process models, they are more challenging to interpret, complicating error attribution and sensitivity analysis (Houshmandfar et al., 2021).

A promising direction in modeling is to combine process knowledge with ML in physics-guided hybrids (Chen et al., 2020;Chen et al., 2022; Liu et al., 2023, 2025). These models can be designed to adhere to established theories and physical laws while also learning relations that traditional process formulations cannot capture when there is informative data (Shen et al., 2023). The two approaches are complementary rather than substitutes (Reichstein et al., 2019): process models bring mechanistic interpretability and the potential for extrapolation beyond observed conditions, while data-driven methods offer flexibility and adaptability in the face of changing regimes and unobserved heterogeneity (Chen et al., 2022; Willard et al., 2022). Hybrids have been successfully applied in various modeling applications, such as lake temperature (Read et al., 2019), vegetation *LE* (Liu et al., 2023), terrestrial evapotranspiration (Chen et al., 2022; Liu et al., 2025), crop phenology (Ou et al., 2025), and urban thermal environment (Zhong et al., 2026a, 2026b), often achieving higher accuracy than either process models or standalone ML methods. Common strategies include using the process model's outputs as additional inputs to the DL model (Reichstein et al., 2019; Willard et al., 2022), allowing DL to emulate uncertain subcomponents within a process framework, and incorporating physical constraints through penalties or regularization in the loss function to guide neural networks. Recent research has demonstrated accurate tree-level *LE* estimation, with additional improvements gained by combining the process estimate with a learned residual (Liu et al., 2023). Notably, the performance of these models depends on the accuracy of the embedded process representation; if key mechanisms are incorrectly specified or omitted, the hybrid model may inherit or even amplify biases (Zhao et al., 2019; Bennett and Nijssen, 2021).

Recent studies highlight the potential of DL and physics-guided hybrid models for estimating plant *LE*. However, few studies focus on interpretability, and even fewer address the interpretability of *LE* estimations across different tree species. Species exhibit significant differences in water use and stomatal regulation. For example, *Acacia aptaneura* can maintain relatively high transpiration rates under arid conditions (Nolan et al., 2018), while drought-tolerant *Robinia pseudoacacia* typically reduces transpiration by closing its stomata and sometimes losing leaves (Hanley et al., 2021; Rahman et al., 2024). Without interpretable analyses of species-specific transpiration responses, models remain predictive but not explanatory. This limits the understanding of the mechanisms that govern transpiration and offers insufficient guidance for experimental design, model refinement, and management strategies. Explainable ML can help characterize model-derived associations and response patterns, although these patterns should not be interpreted as causal mechanisms (Gaur and Drewry, 2024; Gaur et al., 2025; Zhang et al., 2025). Techniques such as SHAP (Shapley Additive Explanations) offer both local and global attributions (Lundberg et al., 2020), marginal response profiles, and interaction diagnostics. These methods have proven useful in plant physiology and hydrologic estimation. For instance, SHAP has been used both to quantify how spectral and meteorological predictors influence stomatal conductance in pepper (Zhang et al., 2025), and to identify key parameters and interactions affecting stomatal conductance across different plant functional types, aiding the interpretation of large-scale evapotranspiration estimates (Gaur and Drewry, 2024; Gaur et al., 2025).

More importantly, since both the embedded process model and the coupling strategy differ across applications, the effectiveness of hybrid approaches depends on the specific domain. This is especially relevant for urban trees, for which these methods have rarely been evaluated for species-level *LE* estimations, and physiological drivers are not well understood. To fill these gaps, we developed a physics-guided DL frame-

work that includes three hybrid models incorporating the Jarvis model. We compare these hybrids with the Jarvis, pure DL, and RF baselines for hourly *LE* estimations at the species-level. Additionally, we employ the SHAP method to quantify feature contributions and characterize model-derived associations between environmental and phenological factors and estimated *LE* across species. We hypothesize that: i) compared to the Jarvis model, DL models can better capture nonlinearities and interactions among environmental, phenological and physiological controls, leading to improved out-of-sample accuracy while preserving physiologically interpretable signals; ii) hybrid schemes that combine process knowledge with data-driven learning can achieve higher accuracy than pure DL and greater robustness across complementary validation settings; and iii) when the Jarvis model introduces systematic bias due to issues of separability, independence, or fixed subfunction parameters, DL can learn structured, species-specific residual corrections.

2. Materials and methods

2.1. Data collection and processing

We conducted observations at nine urban sites in Würzburg and Munich (2015-2021), Bavaria, Germany. Table 1 summarizes tree attributes, while Table A1 and Figs. A1-A2 in Appendix A provide background climate information and photographs of the monitored trees, respectively. Two dominant street tree species were selected for their contrasting hydraulic strategies: *Tilia cordata* is diffuse porous, anisohydric, and shade tolerant with relatively low water use efficiency; in contrast, *Robinia pseudoacacia* is ring porous, isohydric, and light demanding with a high water use efficiency (Moser-Reischl et al., 2019). This selection allowed for species-specific tests of transpiration responses.

The environmental variables included R_s (W/m^2), T_a (°C), *VPD* (kPa), and θ (%) at 30 cm depth, which were recorded at 10-min intervals. Since the absolute values of θ are not directly comparable across sites, we normalized it to the soil moisture deficit $\theta_{def} = \frac{\theta_{max}-\theta}{\theta_{max}-\theta_{min}}$, which was used as input for both the Jarvis and DL models. Here, θ_{max} and θ_{min} are the maximum and minimum θ observed at each site. Sap flow density (SFD, ml/cm^2/h) was measured at 5-min intervals using Granier-type thermal dissipation probes (TDPs; Ecomatik, Dachau, Germany) following standardized protocols applied across all sites (Granier, 1987). To reduce individual variability and ensure consistency, we averaged SFD by species at each site. All variables were aggregated to hourly values, obvious outliers were removed, and we retained only daytime records with $R_s > 0$. After preprocessing, the final datasets contained 25,273 and 12,671 hourly observations for *T. cordata* and *R. pseudoacacia*, respectively. The target was the *LE* of trees, computed from SFD:

$$LE = L_v \cdot SFD \cdot \frac{A_{sap}}{A_c} \tag{1}$$

Where L_v is the latent heat of vaporization, 2.45 MJ/kg; A_{sap} (cm^2) is sapwood area; A_c (m^2) is canopy projected area. Further details on the measurement campaigns can be found in the references listed in Table 1.

2.2. Model development

The primary objective of this study is to compare different physics-guided strategies within a well-established hybrid DL framework, rather than to exhaustively evaluate multiple ML algorithms. Here, "physics-guided" is used in a broad, process-informed sense, referring to the incorporation of Jarvis-derived process estimates into DL models through residual correction, adaptive fusion, or feature learning. We therefore used neural networks as the foundation for the hybrid models, following previous studies that combine process knowledge with data-driven learning (Reichstein et al., 2019; Chadalawada et al., 2020; Willard et al., 2022). Existing work has developed several

Table 1
Structural attributes and monitoring periods of sampled trees, including sampling locations for *T. cordata* and *R. pseudoacacia*, sapwood area (A_{sap}, cm^2), leaf area index (LAI), tree age (Age, years), tree height (H, m), crown projection area (A_c, m^2), diameter at breast height (DBH, cm), number of instrumented trees (n), and measurement periods (YYYY.MM.DD). Morphological parameters represent site-level means.

Cities	Locations	A_{sap}	LAI	Age	H	A_c	DBH	n	Measurement period
T. cordata									
Munich	Bordeaux Platz[a]	528.60	2.39	40.0	15.1	67.1	29.2	3	2015.08.03-10.21
	Pariser Platz[a]	923.80	2.55	89.0	16.8	81.7	44.7	3	2015.08.03-10.15
	Messestadt Riem[b]	364.00	3.64	36.0	10.6	35.0	23.7	10	2016.06.06-10.12
	Nordbad[c]	751.19	2.54	51.0	16.2	72.0	37.9	6	2020.07.08-10.07 2021.05.18-09.30
Würzburg	Zu-Rhein-Straße[d]	821.00	3.51	54.7	13.8	51.1	42.8	4	2018.04.19-10.28 2019.04.10-10.17
	Rottendorfer Straße[d]	508.77	3.44	41.6	14.8	37.7	29.7	5	2018.04.19-10.28 2019.04.10-10.17
	Rennweg[d]	634.56	3.06	48.0	15.2	79.7	35.4	3	2018.04.18-11.07 2019.07.01-10.10
	Parade Platz[d]	572.12	3.23	43.0	13.3	76.9	32.1	5	2018.04.18-11.13 2019.04.14-10.17
	Ludwigkai[d]	1065.39	3.28	67.4	15.7	88.4	52.6	5	2018.04.19-11.19
R. pseudoacacia									
Munich	Messestadt Riem[b]	155.00	2.61	32.0	12.0	52.0	27.4	10	2016.06.06-10.12
Würzburg	Rottendorfer Straße[d]	209.15	2.12	38.3	11.7	56.5	38.4	3	2018.04.19-10.08 2019.04.10-10.17
	Rennweg[d]	313.22	2.35	55.3	17.7	65.0	55.7	3	2018.04.18-11.19 2019.04.10-10.17
	Zu-Rhein-Straße[d]	202.86	2.06	37.0	12.8	56.50	37.3	2	2019.06.27-10.17

[a] : (Rahman et al., 2017a).
[b] : (Rahman et al., 2018).
[c] : (Shashua-Bar et al., 2023).
[d] : (Rahman et al., 2020; Rötzer et al., 2021).

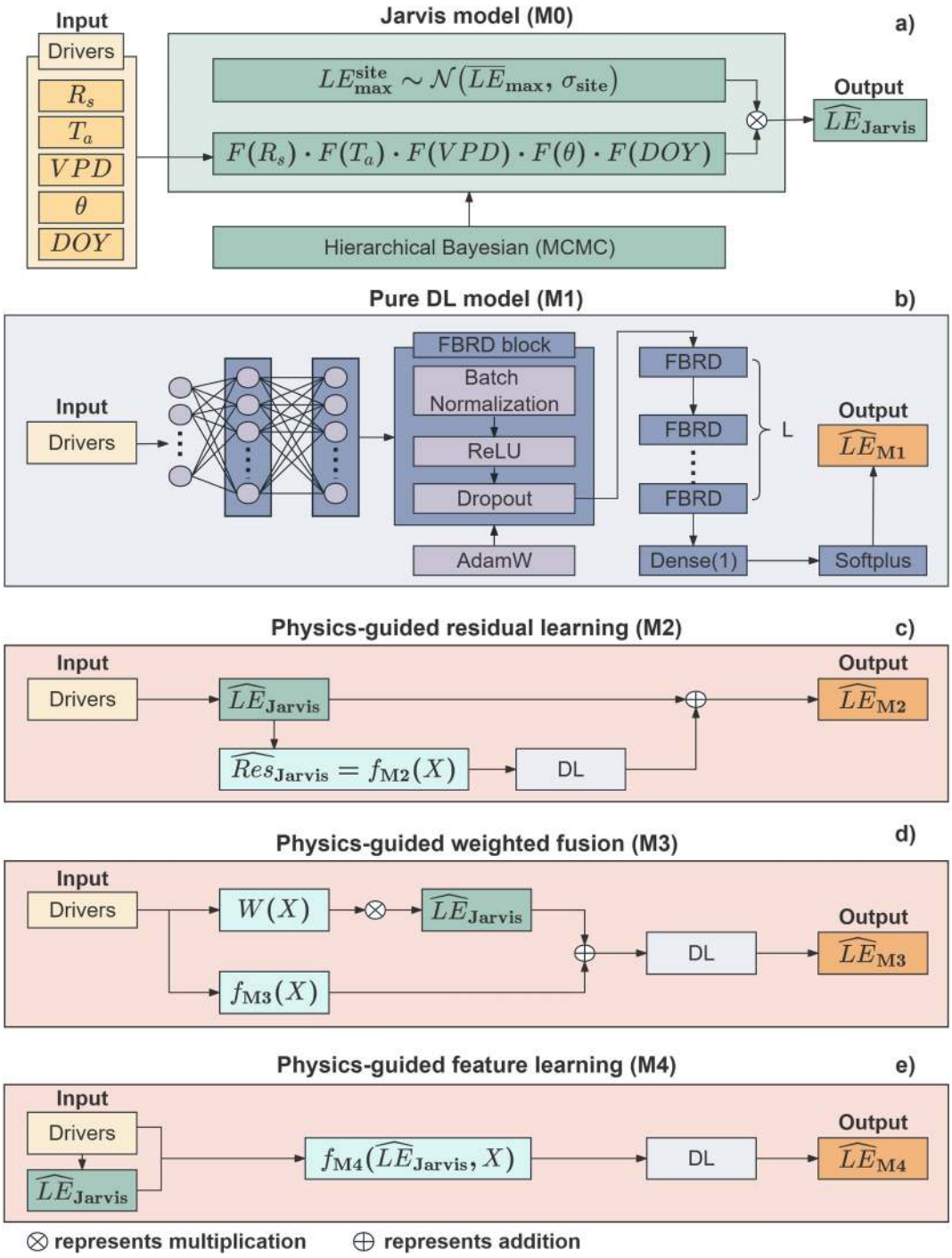

Fig. 1. Hybrid DL Modeling framework. **a)** Physical Jarvis model (M0): species-level LE_{max} is scaled by multiplicative response functions, and parameters are inferred using hierarchical Bayesian modelling with Markov chain Monte Carlo (MCMC) to yield $\widehat{LE}_{Jarvis}$. **b)** Pure DL model (M1): inputs are mapped to LE by a stack of Fully connected-BatchNorm-ReLU-Dropout (FBRD) blocks followed by a Dense (1) output. **c)** Physics-guided residual learning (M2): a DL subnetwork learns the Jarvis residual (Res_{Jarvis}). **d)** Physics-guided weighted fusion (M3): the Jarvis term is adaptively controlled by a learnable weight $W(X)$ and combined with a data-driven correction. **e)** Physics-guided feature learning (M4): $\widehat{LE}_{Jarvis}$ is used as an additional input feature.

physics-neural-network hybrids, in which process information or physical relationships are embedded into the network architecture, training objectives, or input representations (Herath et al., 2021; Chen et al., 2022; Cai et al., 2022; Liu et al., 2023, 2025). While other learners, such as RF or XGBoost, could potentially be combined with process-based models, achieving a comparable architecture-level and differentiable integration would require substantially different algorithmic developments (Wang et al., 2025), which is beyond the scope of this study. Within this DL-based framework, we focus on six models. Two baselines were defined first: the Jarvis model (M0) and a pure DL model (M1). Subsequently, we designed three hybrid strategies (M2-M4) that progressively enhance the flexibility of the DL component, as shown in Fig. 1. In addition, to provide a non-DL benchmark, we trained an RF model using the same input variables and data-splitting procedure as M1, because RF is widely used and well suited for nonlinear tabular regression (Breiman, 2001).

2.2.1. Jarvis model (M0)

The Jarvis-type stomatal conductance model is defined as:

$$g_s = g_{max} \cdot F(R_s) \cdot F(T_a) \cdot F(VPD) \cdot F(\theta) \tag{2}$$

Where g_{max} (m/s) denotes the maximum g_s. The four dimensionless response functions (from 0 to 1) represent the constraint imposed by R_s, T_a, VPD, and θ. These functions follow the Stewart-modified Jarvis formulation (Stewart, 1988), replacing difficult physiological states with observable environmental drivers. This structure can likewise be used to directly infer SFD or transpiration (Whitley et al., 2013; Wang et al., 2020):

$$\widehat{LE}_{Jarvis} = LE_{max} \cdot F(R_s) \cdot F(T_a) \cdot F(VPD) \cdot F(\theta) \cdot F(DOY) \tag{3}$$

$$F(R_s) = \frac{R_s}{k_1 + R_s} \tag{3a}$$

$$F(T_a) = \left(\frac{T_a - T_{a,min}}{T_{a,opt} - T_{a,min}}\right) \cdot \left(\frac{T_{a,max} - T_a}{T_{a,max} - T_{a,opt}}\right)^{\frac{T_{a,max} - T_{a,opt}}{T_{a,opt} - T_{a,min}}} \tag{3b}$$

$$F(VPD) = \exp\left(\frac{-k_2 (VPD - k_3)^2}{VPD}\right) \tag{3c}$$

$$F(\theta) = \begin{cases} 1, \theta_{def} < k_4 \\ \left(\frac{k_5 - 1}{1 - k_4}\right)\theta_{def} + \left(\frac{1 - k_4 \cdot k_5}{1 - k_4}\right) \end{cases} \tag{3d}$$

Here, $\widehat{LE}_{Jarvis}$ denotes the LE estimated by the Jarvis model, and LE_{max} is the maximum LE; the subfunctions (Eqs. 3a, 3b, 3c, and 3d) are adopted from previous studies (Stewart, 1988; Poyatos et al., 2007; Samanta et al., 2008; Wang et al., 2016). Where $T_{a,min}$, $T_{a,opt}$, and $T_{a,max}$ represent the minimum, optimal, and maximum T_a for transpiration, and k_1 to k_5 are parameters to be calibrated. Notably, the specific subfunctions may vary across studies. In our previous study, we compared alternative subfunction formulations and selected the present set based on its predictive performance and physiological plausibility (Zhang et al., 2026). We further include an independent seasonal term to capture phenology-related variation (Zhang et al., 2026):

$$F(DOY) = \begin{cases} \exp\left(-\frac{(DOY - DOY_{max})^2}{2DOY_{pre}^{\ 2}}\right), DOY \le DOY_{max} \\ \exp\left(-\frac{(DOY - DOY_{max})^2}{2DOY_{post}^{\ 2}}\right), DOY > DOY_{max} \end{cases} \tag{3e}$$

Where DOY_{max} is the day-of-year (DOY) of peak LE, the DOY_{pre} and DOY_{post} are pre- and post-peak width parameters allowing an asymmetric seasonal peak. Because seasonal processes do not operate at hourly or daily scales and to improve parameter identifiability, we suppressed high-frequency variability by aggregating observations into nonoverlapping 15-day windows.

Given that even within a species the site-specific LE_{max} can differ substantially due to soil conditions (soil water potential) and osmotic pressure (Buckley et al., 2012), we estimated a species-specific set of Jarvis parameters (constant across sites within each species) and modelled the site-specific LE_{max} (i.e. LE_{max}^{site}) as a random effect:

$$LE_{max}^{site} \sim \mathcal{N}(\overline{LE}_{max}, \sigma_{site}) \tag{3f}$$

Where $\overline{LE}_{max}$ is the species-level mean and σ_{site} is the between-site standard deviation. This hierarchical specification captures between-site heterogeneity in $\overline{LE}_{max}$ while avoiding the need to fix a single value for all locations.

We calibrated the Jarvis parameters using a Hierarchical Bayesian approach with MCMC (Clark, 2005). Specifically, we ran four parallel chains with 8000 total iterations and discarded the first half as warm-up. Convergence was evaluated following Vehtari's method (Vehtari et al., 2021) using the potential scale reduction statistic ($\widehat{R} < 1.01$), and effective sample sizes ($ESS > 400$). Details of the Bayesian implementation and prior specifications are provided in Appendix B.

2.2.2. Deep learning model (M1)

Here, we use a multilayer perceptron (MLP) as the DL architecture (Chen et al., 2022; Liu et al., 2023). Deep MLPs often outperform shallow neural networks, though at higher computational cost. We adopt the same environmental inputs as in the Jarvis model and encode seasonal cycles with Fourier features of DOY (Tancik et al., 2020). The model estimates LE and is trained by minimizing mean squared error (MSE) (Gaur and Drewry, 2024):

$$X = [R_s, T_a, VPD, \theta_{def}, DOY] \tag{4}$$

$$\widehat{LE}_{M1} = f_{M1}(X) \tag{5}$$

$$LOSS_{M1} = \frac{1}{N}\sum_{i=1}^{N}\left(LE_{true,i} - f_{M1}(X_i)\right)^2 \tag{6}$$

Where $LE_{true,i}$ is the measured LE. Following a previous study (Liu et al., 2023), we adopt a similar feedforward deep network configuration (Fig. 1b). To improve training stability and mitigate overfitting, each hidden layer utilizes both Batch Normalization (BN) and Dropout: BN stabilizes the distributions of activations within layers and accelerates convergence (Santurkar et al., 2019), while Dropout randomly deactivates units to reduce co-adaptation. The hidden layers employ ReLU nonlinearities to capture strong nonlinearity and alleviate the vanishing gradient problem (Kulathunga et al., 2020). The DL model consists of stacked hidden-layer blocks, each following the order: Fully Connected-BN-ReLU-Dropout (FBRD), and the network stacks L such FBRD blocks. The output head is a single scalar regression unit with a Dense (1) configuration and uses Softplus activation to ensure the LE is non-negative. To enhance generalization and stabilize optimization, we use AdamW with decoupled weight decay (Loshchilov and Hutter, 2019), meaning that L2 weight decay is applied separately from the gradient update. Since hyperparameters critically affect DL performance, during the initial training phase, we perform a joint Bayesian optimization with Keras Tuner (Koppa et al., 2022), then based on validation performance select a shared configuration across two species and the DL models (M1-M4) to ensure fair and reproducible comparisons (Table C1, Appendix C).

2.2.3. Hybrid DL models (M2-M4)

To incorporate process knowledge into the data-driven framework, we develop three strategies that are linked to Jarvis:

i) Physics-guided residual learning (M2)

The core concept of M2 (Fig. 1c) is to take the $\widehat{LE}_{Jarvis}$ as a baseline and to explicitly incorporate its residuals during training. This approach incorporates process-derived information and directs the learning capacity toward the aspects that the Jarvis model fails to explain. The network predicts the residual from the inputs and reconstructs LE. In this way, it acts as a residual correction to the Jarvis estimate, formally:

$$Res_{Jarvis} = LE_{true} - \widehat{LE}_{Jarvis} \tag{7}$$

$$\widehat{Res}_{Jarvis} = f_{M2}(X) \tag{8}$$

$$\widehat{LE}_{M2} = \widehat{LE}_{Jarvis} + f_{M2}(X) \tag{9}$$

$$LOSS_{M2} = \frac{1}{N}\sum_{i=1}^{N}\left(LE_{true,i} - \widehat{LE}_{Jarvis,i} - f_{M2}(X_i)\right)^2 \tag{10}$$

Where Res_{Jarvis} is the observed residual relative to $\widehat{LE}_{Jarvis}$, and $\widehat{Res}_{Jarvis}$ is the DL-estimated residual.

ii) Physics-guided weighted fusion (M3)

Building on M2, M3 (Fig. 1d) considers the $\widehat{LE}_{Jarvis}$ as an additive process-derived component that has a learnable weight $W(X) \in [0, 1]$, and allows the network backbone to learn data-driven compensation $f_{M3}(X)$. We assume the Jarvis estimate may underperform in certain extreme environmental or physiological conditions, so $W(X)$ adapts with the inputs. The model is:

$$\widehat{LE}_{M3} = \widehat{LE}_{Jarvis} \cdot W(X) + f_{M3}(X) \tag{11}$$

$$LOSS_{M3} = \frac{1}{N}\sum_{i=1}^{N}\left(LE_{true,i} - \widehat{LE}_{Jarvis,i} \cdot W(X_i) - f_{M3}(X_i)\right)^2 \tag{12}$$

(iii) Physics-guided feature learning (M4)

M4 (Fig. 1e) operates independently from the other hybrids, providing the DL model with enhanced flexibility. It inputs the $\widehat{LE}_{Jarvis}$ directly as an additional feature to test if the process prior improves DL performance:

$$\widehat{LE}_{M4} = f_{M4}(X, \widehat{LE}_{Jarvis}) \tag{13}$$

$$LOSS_{M4} = \frac{1}{N}\sum_{i=1}^{N}\left(LE_{true,i} - f_{M4}\left(X_i, \widehat{LE}_{Jarvis,i}\right)\right)^2 \tag{14}$$

The three hybrids are designed to be mutually independent and are organized in a way that progressively enhances modeling flexibility from M2 to M4. This structured approach allows for a clear evaluation of how various coupling strategies impact the accuracy of species-level *LE* estimations and supports a transparent comparison with the Jarvis and pure DL baselines.

2.3. Model performance evaluation

To assess how modeling choices affect *LE* estimations, we report three standard statistics: NSE (Nash-Sutcliffe efficiency), RMSE (root mean squared error), and MAE (mean absolute error). Lower RMSE and MAE, and NSE closer to 1 indicate better performance:

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(\widehat{LE}_i - LE_{true,i}\right)^2}{\sum_{i=1}^{N}\left(LE_{true,i} - \overline{LE_{true}}\right)^2} \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{LE}_i - LE_{true,i}\right)^2} \tag{16}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{LE}_i - LE_{true,i}\right| \tag{17}$$

For each species, the primary model assessment was based on a 70/30 random hold-out split. All retained hourly observations were temporally shuffled and randomly split at the observation level using a fixed random seed, with 70% used for training and 30% used as the held-out test set, consistent with the Jarvis calibration stage. To avoid direct data leakage between model development and evaluation, the held-out test set was not used for model training, hyperparameter tuning, or model selection, and model performance on this 30% test set was used as the main evaluation result. Accordingly, this observation-level random split evaluates within-domain predictive performance and should not be interpreted as a test of spatial extrapolation to independent sites or trees.

In addition, we used three complementary procedures to further assess model performance and robustness. i) Five-fold cross-validation on the training set: the training data were partitioned into five folds. Each fold was used once for validation, while the remaining four folds were used for training; ii) leave-one-year-out validation: considering the strong association between site and year and the unequal site-year sample sizes, we reconstructed a balanced validation dataset by restricting records to *DOY* 180-290, a period covering most monitoring campaigns, and excluding site-year combinations with small sample sizes, including Munich-Pariser Platz-2015 and Würzburg-Rennweg-2019 for *T. cordata*. For each species, one year from 2015 to 2021 was held out in turn for testing, while the remaining years were used for training. This procedure was applied to RF and M1-M4. iii) training-size sensitivity: to assess the small-sample performance of the hybrid DL models, we trained M1-M4 on varying subsamples of the original training data, including 20%, 30%, 40%, 50%, and 60%. Each model was then evaluated on the same held-out test set. To account for sampling variability, each subsample was trained with five independent repeats using different random draws.

All model training and statistical analyses were conducted using Python 3.12. Hierarchical Bayesian inference (MCMC) was performed with PyMC v4, and DL models were implemented with TensorFlow v2.20. The RF baseline was implemented using scikit-learn v1.7.2, and its hyperparameters were optimized by joint Bayesian optimization using Optuna v4.9.0. The experiments were run on a Windows 11 system with an Intel Xeon Gold 6258R (2.70 GHz) CPU and two NVIDIA RTX A6000 GPUs.

2.4. Interpretation of DL model outputs

Although DL models often demonstrate higher accuracy than process models, interpretability remains challenging (Reichstein et al., 2019; Houshmandfar et al., 2021). To address this, we used the SHapley Additive exPlanations (SHAP) method, which is based on cooperative game theory. This method assigns each predictor an expected marginal contribution to the model output (Lundberg et al., 2020), averaged across all possible combinations of predictors.

In this study, SHAP values were computed with DeepExplainer (a specialized interpreter for DL) from shap v0.48 and interpreted from two complementary perspectives. At the local level, SHAP quantifies how much each predictor increases or decreases the estimated *LE* relative to a baseline for each individual sample. At the global level, it ranks predictors by their overall importance, calculated as the average of their absolute SHAP contributions across all samples (Jiang et al., 2022). Importantly, we employ SHAP to characterize species-specific model-response patterns associated with environmental and phenological drivers in *T. cordata* and *R. pseudoacacia*, allowing for direct comparisons between the species. We also use SHAP dependence plots to display

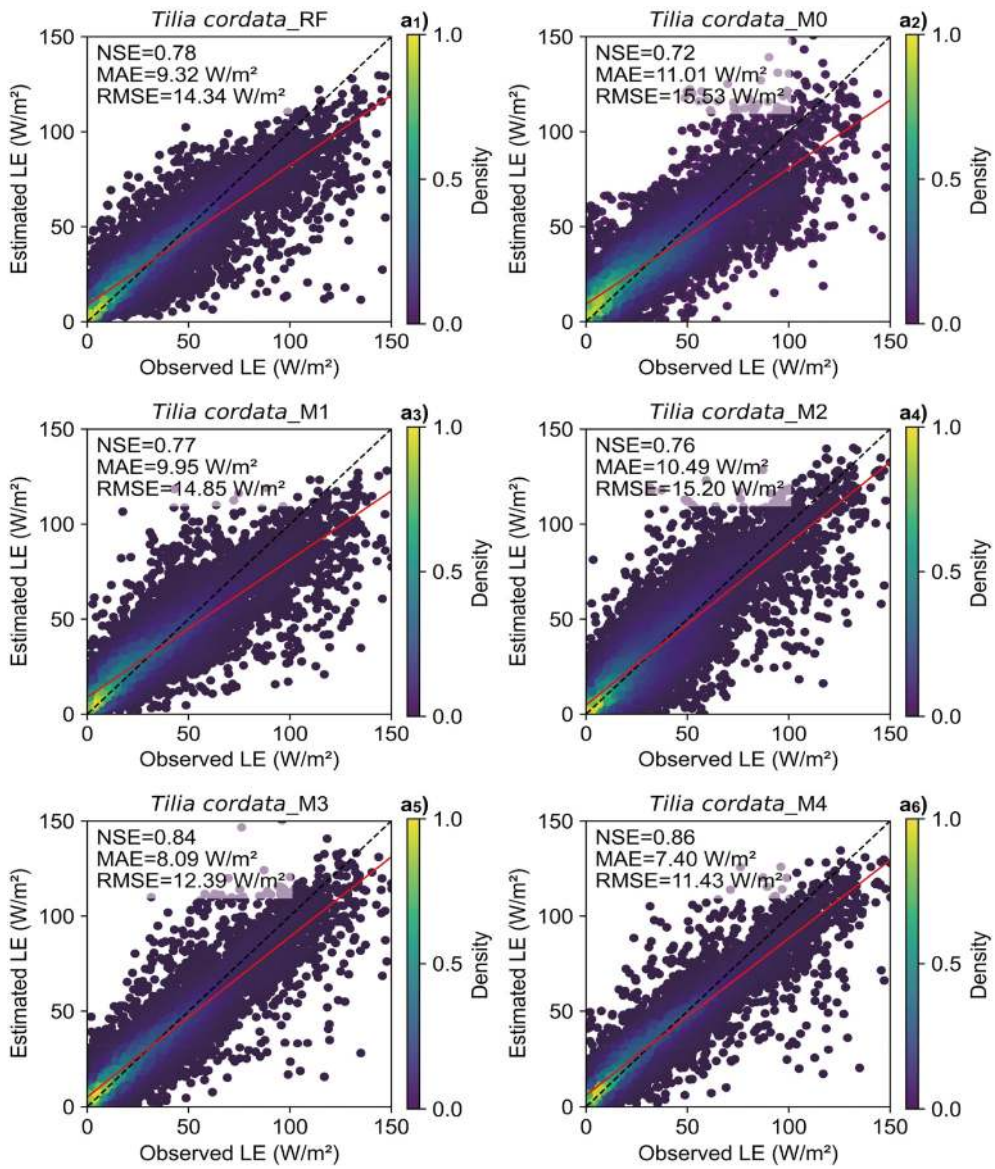

Fig. 2. Estimated versus observed *LE* for **(a1-a6)** *T. cordata* and **(b1-b6)** *R. pseudoacacia* across six models. Subplots **(1-6)** correspond to RF, M0, M1, M2, M3, and M4, respectively. The black dotted line represents the 1:1 reference line, and the red solid line represents the least-squares regression. All performance metrics were calculated on the 30% held-out test set, including NSE, MAE (W/m^2), and RMSE (W/m^2). The color scale denotes normalized point density (0-1), with warmer colors indicating denser regions.

marginal model-response curves and to indicate potential interactions among drivers. For more detailed information about SHAP calculation, see the reference (Shrikumar et al., 2019).

3. Results

3.1. Performance of all models

On the 30% held-out test set, we evaluated all models for *T. cordata* and *R. pseudoacacia* (Fig. 2). Compared to the Jarvis baseline (M0), all RF and DL models improved accuracy. In particular, the hybrid strategies, especially models M3 and M4, achieved further gains relative to the pure DL (M1). For *T. cordata*, M3 increased NSE by 0.07 and reduced the MAE (RMSE) by 1.86 (2.46) W/m^2; M4 increased NSE by 0.09 and reduced MAE (RMSE) by 2.55 (3.42) W/m^2. For *R. pseudoacacia*, M3 increased NSE by 0.11 and reduced MAE (RMSE) by 0.55 (0.84) W/m^2; M4 increased NSE by 0.11 and reduced MAE (RMSE) by 0.60 (0.84) W/m^2. M2 demonstrated that DL can learn structured residuals of Jarvis. However, it performed slightly worse than M1 for *T. cordata*, as its focus on the Jarvis residual increases sensitivity to residual learnability and data noise. M3 and M4 achieved further improvements, with M4 performing best in the random hold-out evaluation. The RF baseline also provided a competitive nonlinear tabular ML benchmark, outperforming the Jarvis baseline, M1, and M2 in this learning task, particularly for *R. pseudoacacia*, which had a relatively small sample size in the present training dataset.

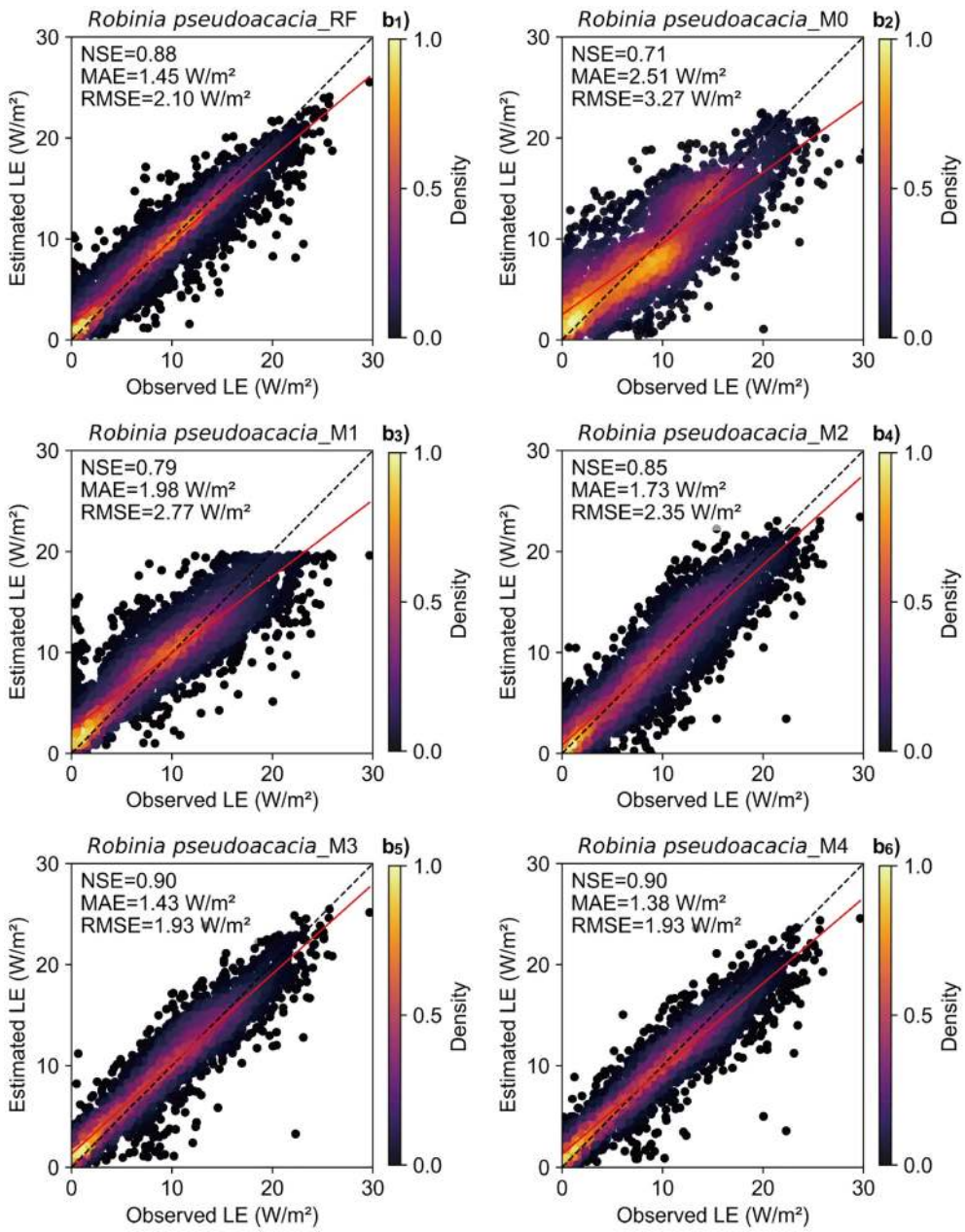

Fig. 2. (*continued*).

Table 2
Leave-one-year-out validation performance for each species. Values are reported as mean ± standard deviation across held-out years. Performance metrics include NSE, MAE (W/m^2), and RMSE (W/m^2).

Species	Models	NSE	MAE	RMSE
T. cordata	RF	−0.32 ± 0.97	23.09 ± 10.01	30.21 ± 12.61
	M1	−0.12 ± 0.41	21.71 ± 7.00	28.95 ± 8.84
	M2	0.53 ± 0.22	13.96 ± 6.22	18.73 ± 8.22
	M3	0.52 ± 0.27	14.11 ± 6.91	18.87 ± 9.36
	M4	0.50 ± 0.26	14.93 ± 7.19	20.00 ± 9.85
R. pseudoacacia	RF	0.11 ± 0.14	4.21 ± 0.85	5.27 ± 1.10
	M1	0.18 ± 0.04	3.99 ± 0.72	5.04 ± 0.86
	M2	0.64 ± 0.04	2.60 ± 0.27	3.31 ± 0.40
	M3	0.64 ± 0.05	2.69 ± 0.43	3.34 ± 0.48
	M4	0.60 ± 0.01	2.83 ± 0.51	3.55 ± 0.63

3.2. Generalization and robustness of hybrid models

To further assess model robustness beyond the random hold-out test, we tested the models using two additional validation procedures and one training-size sensitivity experiment: i) Five-fold cross-validation on the training set (Table C2, Appendix C), ii) leave-one-year-out validation (Table 2; detailed year-specific results in Table C3, Appendix C), and iii) learning curves using a fixed test set while varying the training fraction to 20%, 30%, 40%, 50%, and 60% (Fig. 3).

The results from five-fold cross-validation are consistent with those from the held-out test set, and the leave-one-year-out validation provides a stricter site-year temporal robustness test. Compared with the random hold-out evaluation, model accuracy decreases, indicating that transfer across held-out site-year conditions is more challenging than

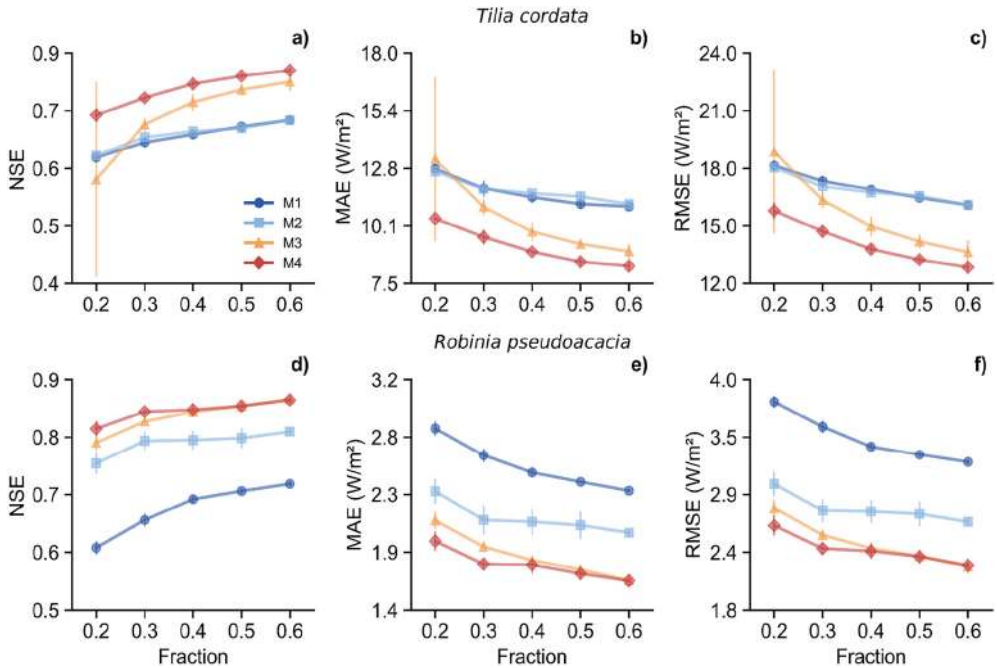

Fig. 3. Scaling experiment for the DL models using different training-data fractions (20-60% of the original training set). Results are shown for *T. cordata* **(a-c)** and *R. pseudoacacia* **(d-f)**. Performance is plotted against the training fraction: **(a, d)** NSE, **(b, e)** MAE, and **(c, f)** RMSE. Error bars indicate variability across repeated runs.

within-domain prediction. Nevertheless, the hybrid models remain more stable than the pure DL model and RF baseline. M2 and M3 performed slightly better than M4 for both species in the leave-one-year-out validation, whereas the RF baseline showed the largest decline relative to the random hold-out evaluation.

The training-size sensitivity experiment was conducted only for the DL-based models (M1-M4), because its purpose is to assess the small-sample behavior of the hybrid DL strategies. M2 shows limited additional improvement over M1 for *T. cordata*, but is more stable for *R. pseudoacacia.* In the small-sample regime, M3 shows large variance at the 20% setting for *T. cordata*, indicating that the learned weight on the Jarvis component is unstable when data are scarce. By contrast, M4 is the most robust and accurate model for both species below 40%. As the sample size increases, the results of M3 and M4 gradually converge, and for *R. pseudoacacia*, their performances are nearly identical at larger training fractions. Overall, M4 showed the most stable performance at smaller training fractions, whereas M3 and M4 converged as the available training data increased.

3.3. Global importance of DL model inputs

Because M1 is a pure DL model without process priors, its SHAP attributions provide a baseline understanding of how predictors relate to model-predicted *LE*. In contrast, M4, which achieves the best predictive performance in the random hold-out evaluation, incorporates additional $\widehat{LE}_{Jarvis}$ inputs to retain Jarvis-derived process information while allowing the DL component to maintain maximal flexibility. Comparing M4 with M1 (Fig. 4) helps distinguish information already represented by the Jarvis prior from residual patterns that remain to be learned from the data, thereby informing potential refinement of the Jarvis formulation. In M1 (Fig. 4a), the beeswarm plots and the global importance ranking agree that *DOY* is the dominant predictor, indicating a primary seasonal modulation of hourly *LE* within the model. For *T. cordata*, *DOY* shows a clear transition from positive SHAP values early in the season to negative values later, consistent with a peak in late spring to early summer followed by a systematic decline. For *R. pseudoacacia*, the pattern is nearly opposite: SHAP values are negative early and become increasingly positive later, reflecting a later phenology and a delayed maximum transpiration. *VPD* ranks second and behaves bidirectionally: SHAP values tend to be positive at moderate *VPD*, consistent with demand-driven enhancement, and increasingly negative at high *VPD*, consistent with stronger stomatal regulation under high atmospheric demand. The θ_{def} shows marked species contrasts: for *T. cordata,* high θ_{def} is associated with strongly negative SHAP values, indicating reduced model-predicted *LE* under drier soil conditions; for *R. pseudoacacia*, the responses are mixed and not monotonic. Other environmental variables showed lower global importance for both species.

In M4 (Fig. 4b), the top-ranked predictor is $\widehat{LE}_{Jarvis}$. The prior accounts for most of the model-attributed variation, while the DL primarily makes minor, state-dependent corrections rather than reconstructing the full environmental and phenological controls. The *DOY* remains important, highlighting a residual seasonal gap in the Jarvis formulation that warrants further adjustment. For the other environmental variables, the beeswarm clouds narrow and center more symmetrically around zero, showing that their role in M4 is mainly to provide minor, conditional corrections. Notably, the importance of θ_{def} decreases, especially for *R. pseudoacacia* (mean SHAP = 0.28), indicating that the prior has captured much of their shared variance.

3.4. Marginal responses of individual predictors

To characterize the marginal effect of each predictor on model-predicted *LE*, we plot SHAP dependence for M1 (Fig. 5) and M4 (Fig. 6) to visualize how predictor values relate to SHAP attributions. For *T. cordata* (Fig. 5a1), the SHAP values for *DOY* peak around 140, then decline and turn negative near 230. For *R. pseudoacacia* (Fig. 5b1), the

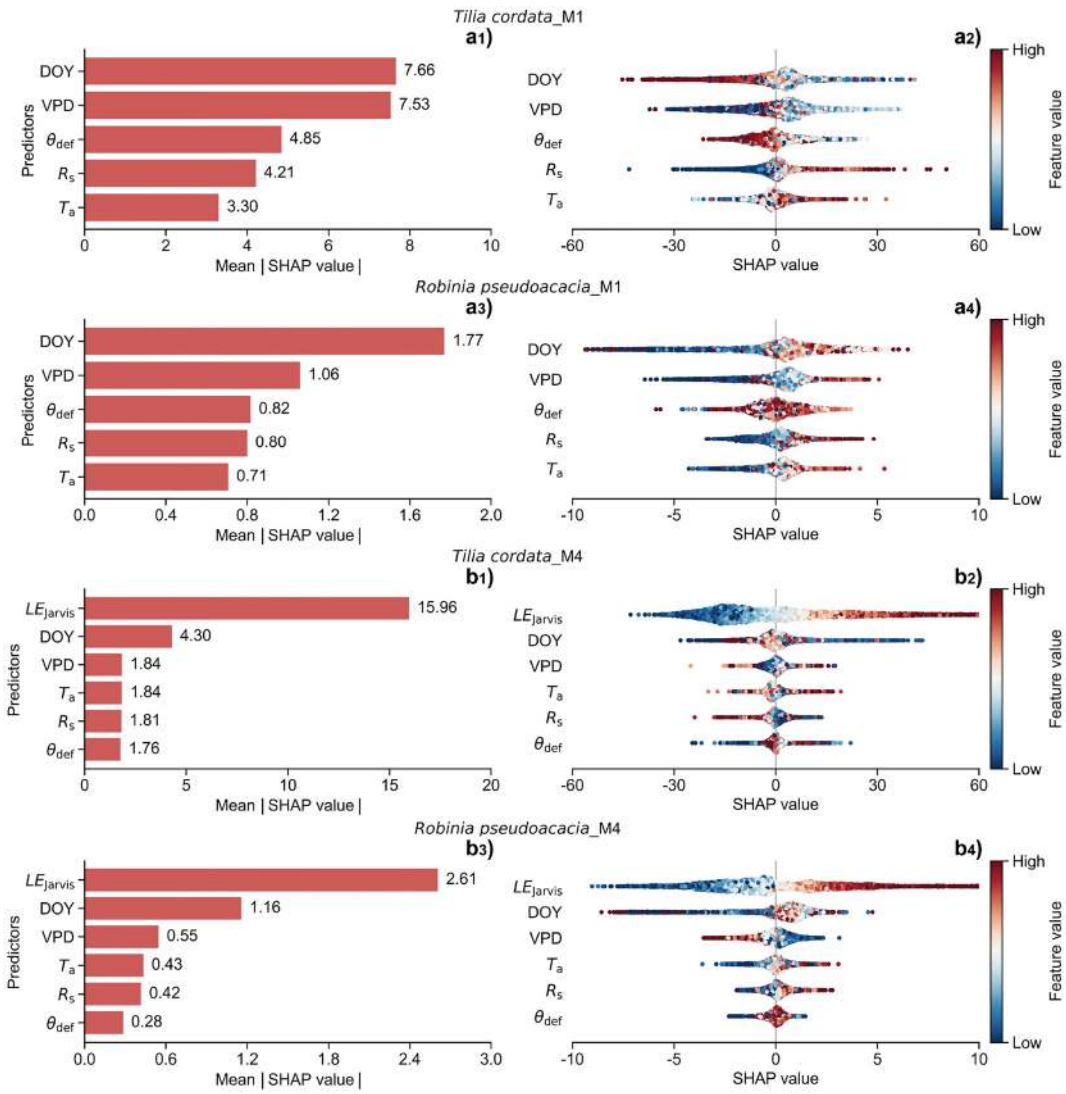

Fig. 4. SHAP explanations for the pure DL model M1 **(a1-a4)** and the process-informed feature learning model M4 **(b1-b4)**. For each model, results are shown for *T. cordata* **(a1-a2, b1-b2)** and *R. pseudoacacia* **(a3-a4, b3-b4)**. Global feature importance plots **(a1, a3, b1, b3)** show mean absolute SHAP values (|SHAP|) for the input predictors, including day of year (*DOY*), vapor pressure deficit (*VPD*), shortwave radiation (R_s), soil moisture deficit (θ_{def}), and air temperature (T_a); for M4, the Jarvis-derived estimate $\widehat{LE}_{Jarvis}$ is also included. Longer bars indicate higher overall contribution to model predictions. Beeswarm plots **(a2, a4, b2, b4)** show individual samples positioned by their SHAP values, indicating their contribution to the predicted *LE*, and colored by the corresponding feature values from low to high. Denser point clouds indicate more observations.

SHAP values for *DOY* become positive later (around 170) and show two peaks (around 190 and 260), turning negative after 275, consistent with a later phenology. R_s and *VPD* show similar patterns for both species (Fig. 5a2-a3 and b2-b3): SHAP values are negative under low light when energy is limited, increase with R_s, and saturate beyond 500 W/m^2 (diminishing returns), and are positive at low *VPD* and increasingly negative at high *VPD*. However, *R. pseudoacacia* experiences an earlier downturn (by ~0.3 kPa) and a quicker transition to negative SHAP values, which is consistent with an earlier reduction in transpiration under increasing atmospheric demand. For T_a, *T. cordata* (Fig. 5a4) shows a generally increasing response, whereas *R. pseudoacacia* (Fig. 5b4) is more parabolic, with a peak between 20 and 30 °C; because values above 30 °C are infrequent, caution is warranted when interpreting this tail of the distribution. Regarding θ_{def}, *T. cordata* (Fig. 5a5) declines after 0.6 and becomes strongly negative beyond 0.8; *R. pseudoacacia* shows smaller, non-monotonic changes (Fig. 5b5), suggesting either weaker direct sensitivity or stronger dependence on multi-variable interactions.

For M4, since $\widehat{LE}_{Jarvis}$ absorbs much of the explainable variance, the SHAP clouds for all predictors contract, especially for *T. cordata*, where most environmental drivers fluctuate only slightly around zero. Visible corrections are mainly observed during the early and late *DOY* periods (Fig. 6a1), as well as under low-demand conditions: slight positive corrections at low R_s and *VPD* indicate that Jarvis tends to underestimate *LE* when energy or atmospheric demand is limited (Fig. 6a2-a3); conversely, more negative corrections at high θ_{def} imply a stronger downward adjustment for *T. cordata* under dry-soil conditions than captured by the Jarvis response (Fig. 6a5). For *R. pseudoacacia*, the residual structure shows more distinct trends: i) R_s maintains persistent positive corrections above 300 W/m^2 (Fig. 6b2), suggesting Jarvis may

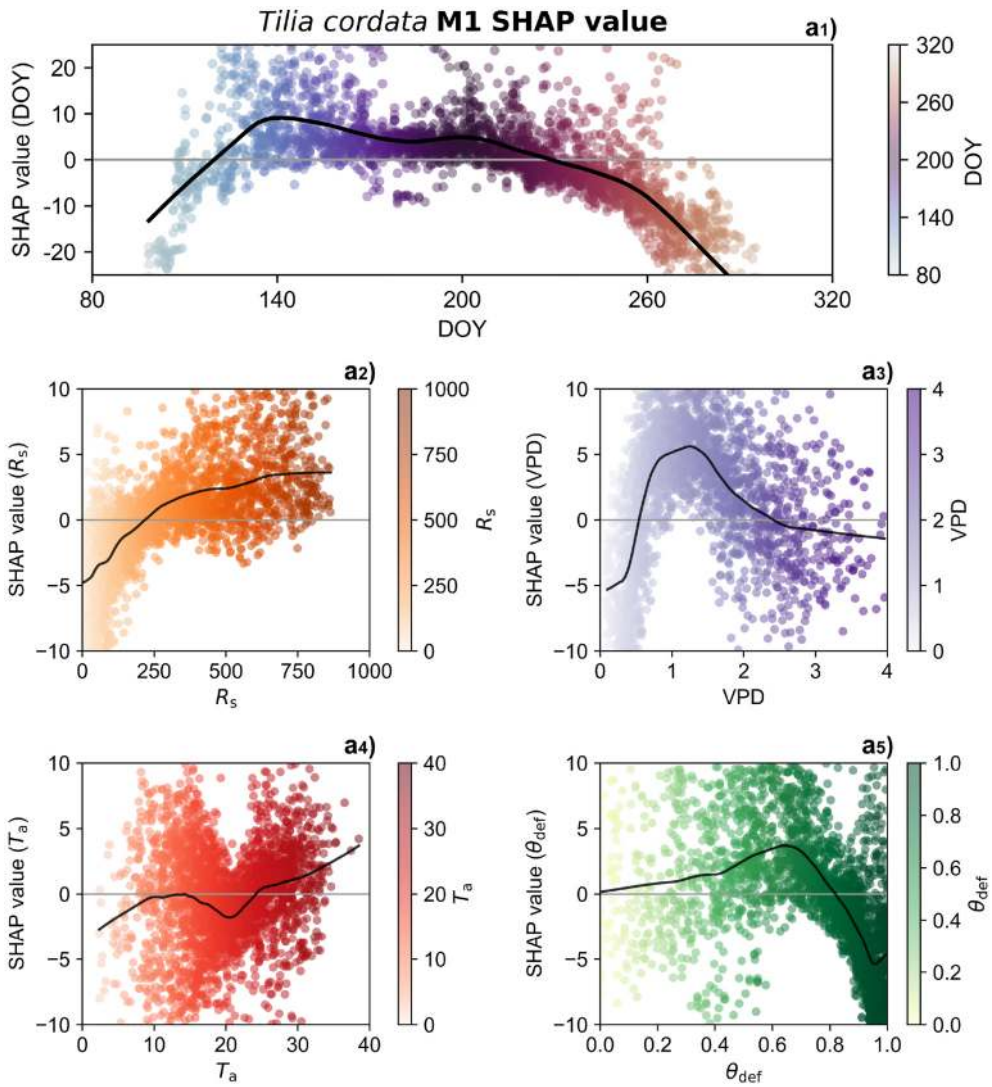

Fig. 5. SHAP dependence plots of predictors in M1 for *T. cordata* **(a1-a5)** and *R. pseudoacacia* **(b1-b5)**, including day of year (*DOY*), vapor pressure deficit (*VPD*), shortwave radiation (R_s), soil moisture deficit (θ_{def}), and air temperature (T_a). Points are colored by the corresponding predictor value. The black curve is a LOWESS smoother highlighting the marginal model response.

underestimate the radiation response for this species; ii) *VPD* shows stable negative corrections beyond 1 kPa, indicating a stronger downward model adjustment under atmospheric dryness (Fig. 6b3); iii) T_a yields nearly monotonic positive corrections (Fig. 6b4), showing that higher temperatures are associated with positive model corrections within the observed range.

4. Discussion

4.1. Performance, robustness, and validation boundaries of hybrid strategies

Our comparison shows that the hybrid strategies generally improve hourly *LE* estimation by combining the Jarvis-type process model with data-driven learning. The RF baseline is competitive in the random hold-out evaluation but declines substantially in the leave-one-year-out validation, even below M1. This suggests that DL itself may provide greater robustness than RF under shifted conditions, while the additional incorporation of Jarvis-derived process information further stabilizes the hybrid models (Fan et al., 2021; Liu et al., 2023). M4 performs best in the random hold-out and limited-data evaluations, whereas the more Jarvis-constrained M2 and M3 are slightly more stable in the leave-one-year-out validation, possibly because stronger process guidance further reduces the learning burden under site-year shifts (Liu et al., 2023, 2025). Therefore, no single hybrid strategy should be interpreted as universally superior across all validation settings.

The leave-one-year-out validation also defines the generalization boundary of this study. Because site and year are closely linked, it should be interpreted as a site-year temporal robustness test rather than strict temporal or spatial extrapolation. In this test, the Jarvis formulation and site-level prior information are treated as available. A fully blind setting would mix the extrapolation error of the Jarvis model with the transferability of the DL correction. Thus, the hybrid framework can stabilize learning when the prior is informative, but biased or poorly calibrated

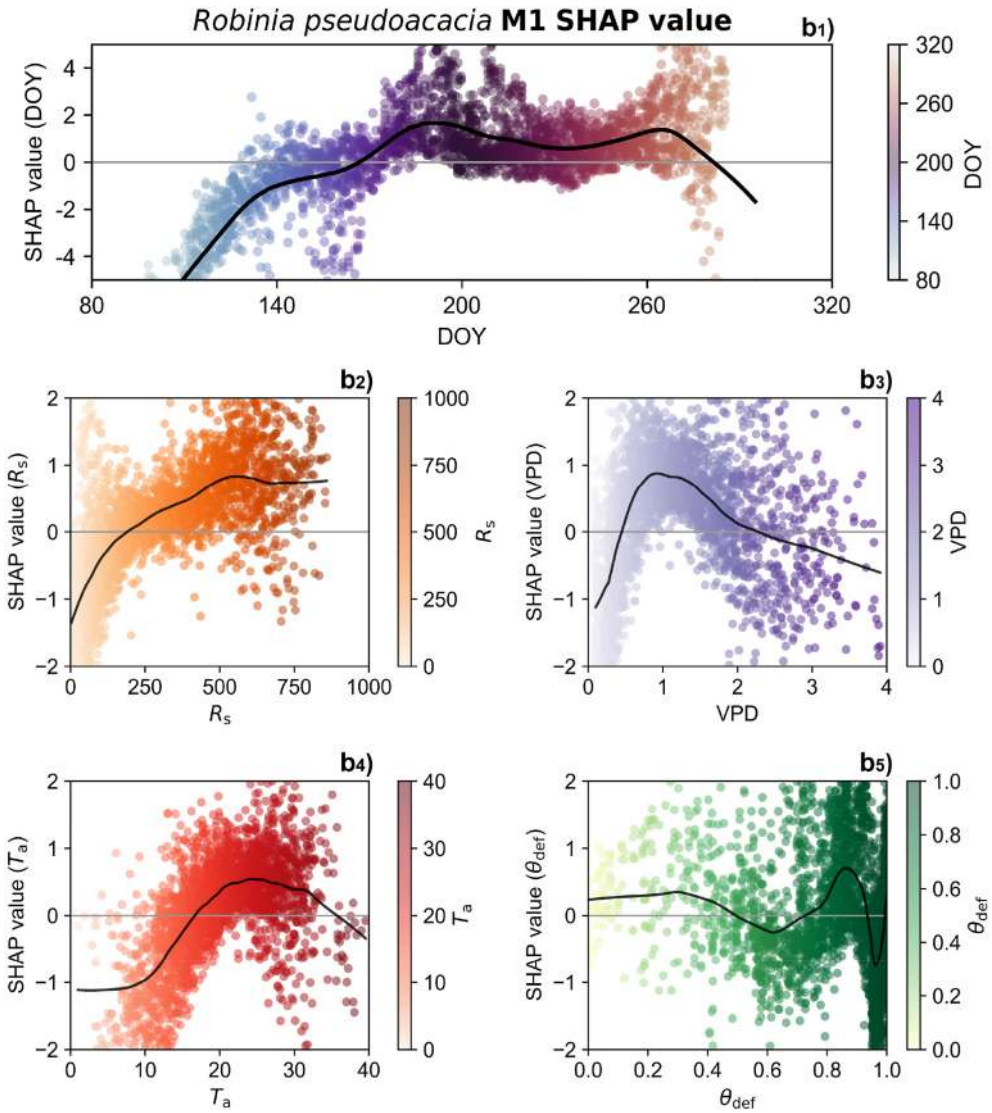

Fig. 5. (*continued*).

Jarvis inputs may propagate into the hybrid predictions, meaning that the present validation supports conditional robustness rather than strict extrapolation to entirely new sites, years, species, or management regimes.

Because prediction under heat and drought stress is important for urban tree cooling and irrigation management, we further evaluated the DL-based models under high-stress subsets defined by the 95th percentiles of T_a, *VPD*, and θ_{def} (Table C4, Appendix C). M3 and M4 remain robust under these less frequent conditions, with M4 generally showing the strongest performance. This indicates that the hybrid models can still capture useful response patterns under high-stress conditions, despite the limited number of such observations. However, performance under rare heatwave or drought events should still be interpreted cautiously and tested with longer-term datasets.

Although M4 performed well overall, it should be interpreted as a pragmatic process-informed feature-learning strategy rather than as a more physically faithful or physics-constrained model (Zhan et al., 2024). Consistent with previous studies, process-model estimates can provide informative priors, while DL can correct systematic biases and nonlinearities that are difficult to parameterize (Reichstein et al., 2019; Read et al., 2019; Irrgang et al., 2021), supporting a practical strategy that combines a streamlined process core with data-driven corrections.

4.2. Model-derived response patterns and physiological interpretation

Plant responses to environmental and physiological drivers are highly nonlinear and strongly interactive (Cai et al., 2019; Gaur and Drewry, 2024). Jarvis-type and stomatal conductance models, such as Ball-Berry and Leuning (Ball et al., 1987; Leuning et al., 1995), are transparent and effective in many settings. However, their empirical assumptions and fixed functional forms make it difficult to fully represent multi-factor interactions among meteorology, phenology, and biochemical processes (Houshmandfar et al., 2021; Gaur and Drewry, 2024), especially at the species-level. Even with hierarchical Bayesian (MCMC) calibration, parameter estimates remain constrained by model structure and data coverage, leading to expected residual bias.

Although black-box ML models are often criticized for their limited interpretability, especially when post-hoc explanations are used instead

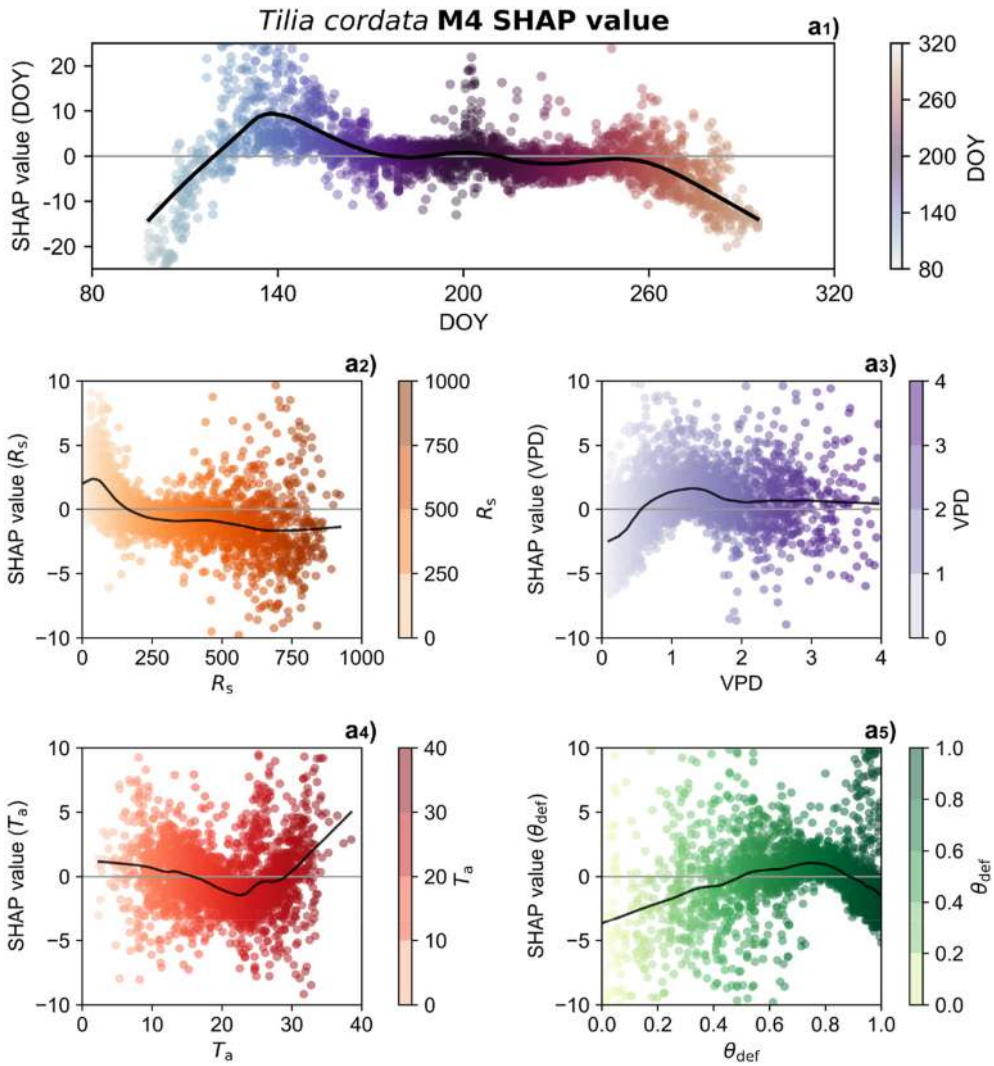

Fig. 6. SHAP dependence of predictors in M4 for *T. cordata* **(a1-a5)** and *R. pseudoacacia* **(b1-b5)**.

of inherently interpretable model designs (Rudin, 2019), explainability tools can still provide useful diagnostic insights when used cautiously. In this study, SHAP analysis was used to identify model-derived response patterns and compare them with known physiological evidence. These patterns should therefore be interpreted as associations learned by the model rather than as direct causal validation of physiological mechanisms. This interpretation is consistent with previous studies showing that explainable, data-driven methods can reveal response patterns that correspond to underlying physiological principles (Gaur and Drewry, 2024; Zhang et al., 2025).

In the pure DL model, the most influential predictor, *DOY*, shows patterns consistent with known differences in wood anatomy and phenology. Specifically, *T. cordata* (diffuse-porous) can resume water transport in previous-year xylem and typically senesces earlier, showing a mid-season decline in SHAP effects that turns negative earlier in the year (Moser-Reischl et al., 2019). In contrast, *R. pseudoacacia* (ring-porous) must produce new earlywood to replace winter-embolized vessels (Panchen et al., 2014; Moser-Reischl et al., 2019); its canopy activity extends later, and SHAP values become positive later with later seasonal peaks. Among environmental drivers, *VPD* ranks next and is pivotal for both species. As atmospheric dryness increases, the model attribution shifts from positive to negative, a pattern consistent with a transition from demand-driven enhancement to stronger stomatal limitation. Relative to *T. cordata*, *R. pseudoacacia* shows an earlier downturn and negative attributions, which is consistent with its more isohydric and stomata-conservative strategy (Rötzer et al., 2021). Meanwhile, θ_{def} shows strong effects for *T. cordata* (suppression strengthens beyond 0.8), while responses for *R. pseudoacacia* are weaker and non-monotonic. The fitted $F(\theta)$ subfunction shows a similar pattern: the onset threshold occurs only at high deficit (parameter k_5 close to 1, shown in Appendix B, Table B3), consistent with deeper water access and relative insensitivity to shallow drying (Zhou et al., 2022). The R_s exhibits increasing and then saturating marginal gains for both species, with no marked interspecific difference. T_a is the weakest driver overall; *R. pseudoacacia* shows slightly stronger temperature sensitivity, which may relate to its small and pinnate leaflets (compound leaf) (Leigh et al., 2017), though this interpretation warrants further investigation.

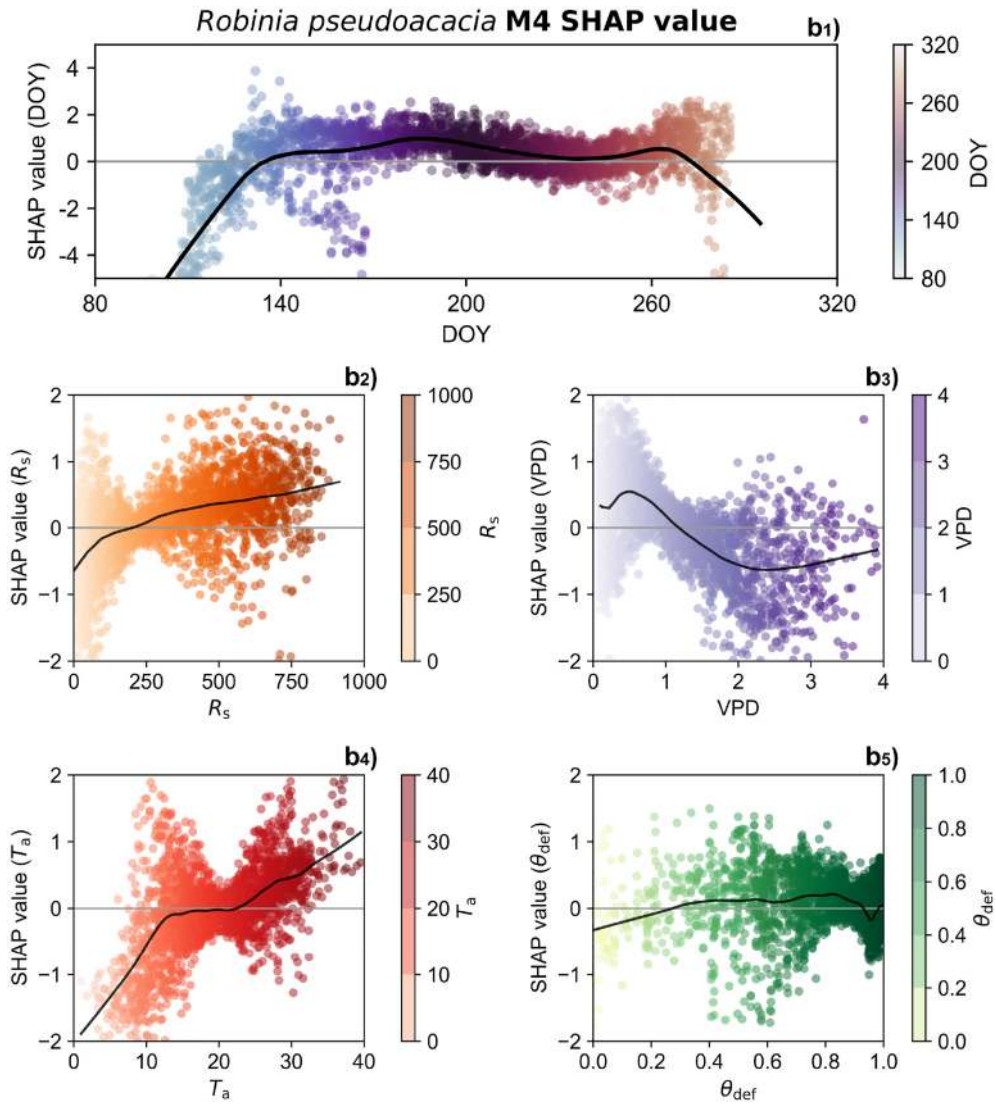

Fig. 6. (*continued*).

4.3. Using hybrid models to identify structural limits of the Jarvis formulation

Our results reveal a second use case for hybrids: using the data-driven component to monitor and diagnose systematic biases of process models like Jarvis at the species-level. Because the Jarvis prior absorbs most of the shared variance among drivers, M4 reduces confusion from collinearity and highlights areas where corrections to the process formulation are needed. In short, M1 SHAP illustrates the overall drivers of *LE*, whereas M4 SHAP pinpoints the locations of Jarvis misfit. Although Jarvis with *DOY*, R_s, T_a, *VPD*, and θ_{def} explains much of the variance for both *T. cordata* and *R. pseudoacacia*, M4 still exposes structured residuals.

To clarify the sources of these corrections, we plot M4 SHAP dependence using color mapping based on a second predictor to visualize interactions (Appendix C). For *T. cordata* (Fig. C1), we observe that under higher R_s with elevated *VPD*, the positive effect of radiation weakens (Fig. C1b). Negative corrections at low θ_{def} suggest the Jarvis model tends to overestimate *LE* around rainfall and other low-demand situations. For *R. pseudoacacia* (Fig. C2), deviations are more pronounced: persistent positive corrections occur under medium to high radiation, while strong negative corrections appear at higher *VPD*; temperature corrections are nearly linear and positive. The parabolic pattern in M1 likely reflects collinearity between high T_a and *VPD*. These patterns point to structural limits in the Jarvis formulation: i) multiplicative separability and factor independence compress conditional dependencies among environmental drivers into single-factor curves, which can lead to premature saturation of R_s. However, under low *VPD*, elevated R_s can still enhance the transpiration potential of light-demanding *R. pseudoacacia* (Fig. C2b); ii) approximating leaf-level *VPD* by air *VPD* neglects boundary-layer and leaf-temperature effects (Novick et al., 2024), which may be particularly relevant for the stomata-conservative *R. pseudoacacia* (Fig. C2f); iii) strong collinearity (where high R_s often co-occurs with elevated T_a and *VPD*) renders single-factor response functions inadequate for disentangling key drivers; iv) fixed empirical parameters represent averages over the sample period and may not capture within-species divergence driven by hydroclimatic background and microclimate heterogeneity (Anys and

Weiler, 2025; Zahradníková et al., 2025); for example, street-canyon geometry can amplify transpiration differences for *T. cordata* (Rahman et al., 2017a, 2017b), and surface sealing (impervious cover) alters soil water supply, rooting volume, and heat loading, further diversifying within-species responses (Rahman et al., 2019, 2024).

Thus, future improvements may relax separability by allowing interactions among drivers rather than assuming strictly independent Jarvis subfunctions and could also introduce dynamic parameter responses that allow parameters currently treated as constants in Jarvis to adapt across seasons and urban settings. The effect of θ_{def} is likely conditional: sensitivity may increase under high-demand transpiration (high R_s) and decrease otherwise, reflecting greater water limitation. Implementing these extensions will require longer-term, multi-species testing.

4.4. Limitations, uncertainty, and future directions

Despite these advances, limitations remain. Because this study used sap flow and environmental data for two tree species from urban sites in Munich and Würzburg during 2015-2021, the results should be interpreted as species- and context-specific evidence, rather than as universally transferable conclusions. From a broader perspective, the same framework could be extended to urban evapotranspiration or microclimate modelling, but this requires further testing across more species, climates, urban forms, irrigation regimes, and soil conditions.

We distinguished between species but did not examine individual trees in detail. Sap flow was averaged at the species-site level, which helps to mitigate sensor noise but also obscures within-species variability. Trees can exhibit different physiological strategies even within the same species in urban environments (Anys and Weiler, 2025; Zahradníková et al., 2025), and capturing this diversity may require adding mechanistic drivers and trait information, such as wood density and leaf turgor-loss point, as well as longer observation periods.

The conversion from sap flow to *LE* relies on several assumptions, including sensor calibration, sapwood area estimates, radial sap-flow profiles, canopy projected area, and upscaling from point-scale sensor measurements to whole-tree fluxes. These sources of uncertainty may propagate into the derived *LE* target variable. Random measurement noise may reduce apparent model performance, while systematic scaling errors in sapwood area or canopy projected area may shift the magnitude of *LE* for a given species or site. Such uncertainty may also influence the magnitude of SHAP attributions. Nevertheless, because all models were trained and evaluated using the same *LE* derivation workflow, the relative comparison among modelling strategies remains meaningful, while fine-scale species differences or predictor attributions should be interpreted with this uncertainty in mind.

The DL component is not fixed to a single architecture; alternative neural designs can be integrated with process knowledge. Currently, physics guidance in our framework is implemented in a soft form: the Jarvis model enters through its calculated values, rather than through strict physical constraints. Future work could explore stronger forms of physics-guided learning, such as physics-guided loss functions or network structures, energy-balance constraints, PM formulations, stomatal-conductance bounds, or water-availability constraints, and test whether these approaches improve robustness across different urban contexts.

5. Conclusions

We integrated the Jarvis model into DL in three ways and systematically evaluated their performance for species-level hourly *LE* estimation in *T. cordata* and *R. pseudoacacia*. Rather than identifying a single universally superior model, our results highlight how different coupling strategies trade off robustness, flexibility and physical interpretability. In particular, including $\widehat{LE}_{Jarvis}$ as a feature improved performance in the random hold-out and limited-training-data evaluations. This process-informed feature-based coupling, which does not force a strict decomposition of the learning target, provides a practical and reproducible approach for hybrid urban tree *LE* modelling when data are limited.

SHAP analyses further identify species-specific model response patterns associated with environmental and phenological drivers. Additionally, the species-specific biases identified in the Jarvis model highlight the value of DL for diagnosing process-model limitations. The framework could be extended to broader evapotranspiration systems, for example, as a hybrid module or data-driven correction for Jarvis-based stomatal or canopy resistance terms in Shuttleworth-Wallace or PM models (Shuttleworth and Wallace, 1985). These findings emphasize the complementarity of the two approaches: process models provide prior structure and interpretability, while DL offers flexible corrections. This interplay is important for future model development and diagnosis.

CRediT authorship contribution statement

Xiang Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Xue Zhong:** Software, Methodology, Conceptualization. **Jun-Ru Yan:** Methodology, Formal analysis, Conceptualization. **Qi Li:** Software, Methodology, Conceptualization. **Kai-Xin Liu:** Visualization, Validation, Software. **Ling-Ye Yao:** Writing – review & editing, Conceptualization. **Long-Hao Zhang:** Software, Resources. **Astrid Moser-Reischl:** Writing – review & editing, Investigation, Funding acquisition, Data curation. **Thomas Rötzer:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Data curation. **Stephan Pauleit:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Data curation. **Mohammad A. Rahman:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgements

This work was supported by the Alexander von Humboldt Fellowship; Climate Experience Würzburg: Influence of Trees on the Microclimate of the City of Würzburg (TKP01KPB-71924) within the Center for Urban Ecology and Climate Adaptation funded by the Bavarian State Ministry of the Environment and Consumer Protection; German Research Foundation through the project CUT-Climate and Urban Trees: Effects of Trees on Urban Climates during Climate Change (PA2626/3-1, PR 292/21-1); the National Natural Science Foundation of China (No. 32501726); the China Postdoctoral Science Foundation Postdoctoral Fellowship Program (No. GZC20251115); the China Scholarship Council (Nos. 202306150006, 202506250066); and the Research Impact Fund of the Hong Kong Research Grants Council (No. CUHK R4040-22). We thank the Department for Municipal Green Areas of Munich and the relevant municipal authorities in Würzburg for their support, and we are grateful to all colleagues and field assistants for their dedicated efforts in instrumentation, maintenance, and data collection.

Appendix A

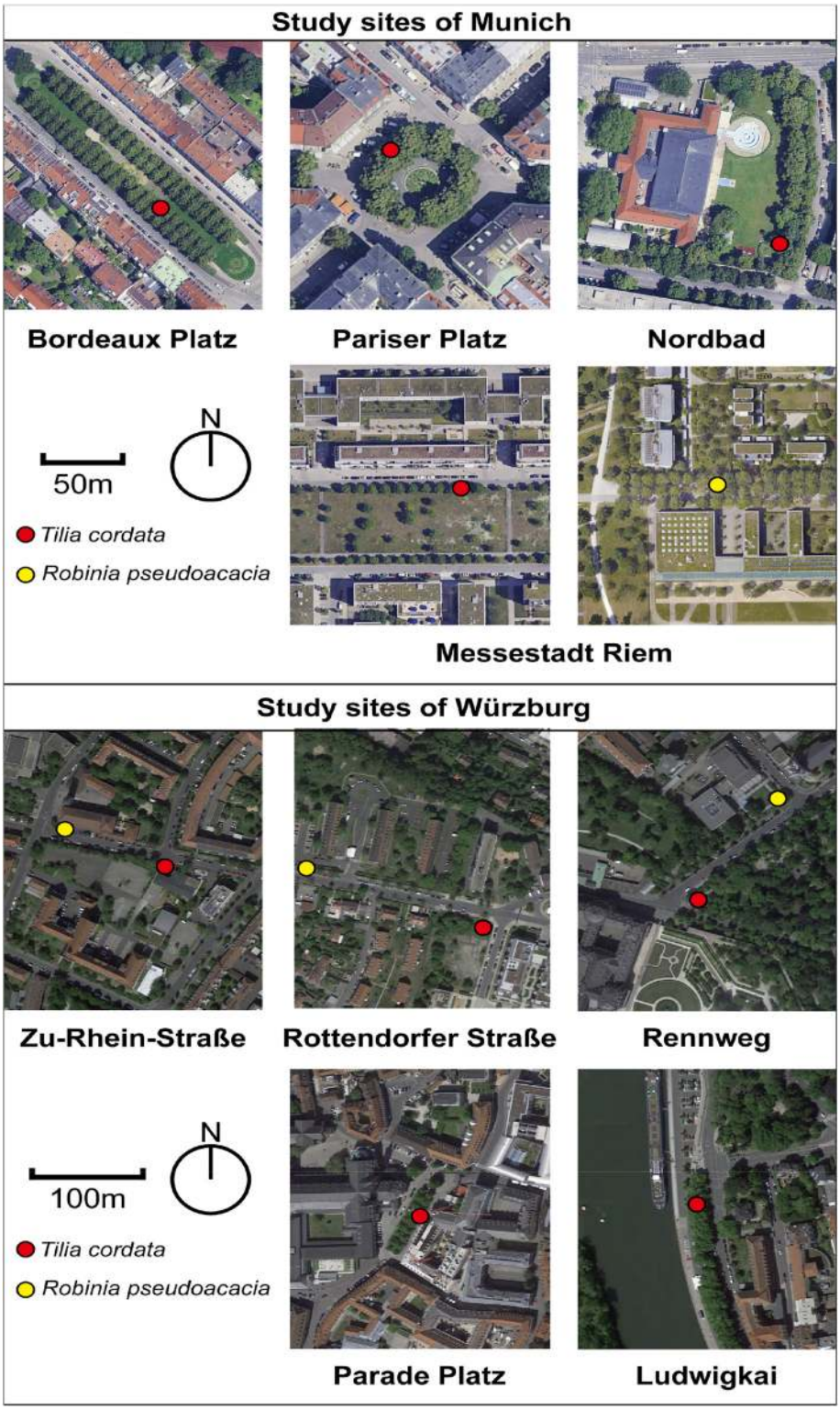

Fig. A1. Aerial views of the study sites in Munich and Würzburg. Base images were obtained from Google Earth. Red and yellow points indicate monitored *Tilia cordata* and *Robinia pseudoacacia* plots, respectively.

Fig. A2. Field photographs of monitored trees in the study plots. Further details on the monitoring campaigns can be found in the references listed in Table 1.

Table A1
Background information on Munich and Würzburg.

Attribute	Munich	Würzburg
Coordinates	48°08′ N, 11°35′ E	49°48′ N, 9°56′ E
Elevation	520 m	177 m
Regional setting	Capital of Bavaria	Northern Bavaria
Climate zone (Köppen-Geiger)	Temperate oceanic (Cfb)	
Mean annual temperature	9.1 °C	9.6 °C
Mean January temperature	−4 °C	1 °C
Mean July temperature	24 °C	20 °C
Annual precipitation	959 mm	599 mm

Appendix B

The constant parameters in these functional forms determine species-specific transpiration responses. The Bayesian MCMC framework was employed. Compared to other widely used uncertainty estimation approaches in transpiration modeling (such as the Generalized Likelihood Uncertainty Estimation and the Pareto optimality), the Bayesian approach provides a coherent probabilistic characterization, allows incorporation of prior knowledge, and supports inference in high-dimensional spaces (Hill, 1974; Samanta et al., 2007, 2008). Accordingly, it is well suited and widely

adopted for Jarvis-type and other transpiration modeling (Chen et al., 2020).

The core of Bayesian inference is to update parameter uncertainty by combining priors with data through a likelihood (Zhu et al., 2013). Let $O = \{o_1, o_2, \cdots, o_n\}$ be observed SFD and $\widehat{o}_j(\beta)$ the model prediction. Assuming independent and identically distributed Gaussian errors with variance ${\sigma_{res}}^2$, the likelihood is:

$$p\left(O|\beta, {\sigma_{res}}^2\right) \propto {\sigma_{res}}^{-n} \prod_{j=1}^{n} \exp\left(-\frac{\left(o_j - \widehat{o}_j(\beta)\right)^2}{2{\sigma_{res}}^2}\right) \tag{B1}$$

Priors were uniform for β within literature-based feasible ranges and uniform on log $({\sigma_{res}}^2)$, i.e., $p(\beta, {\sigma_{res}}^2) \propto \frac{1}{{\sigma_{res}}^2}$ (Zhu et al., 2013), yielding:

$$p\left(\beta, {\sigma_{res}}^2|O\right) \propto {\sigma_{res}}^{-(n+2)} \prod_{j=1}^{n} \exp\left(-\frac{\left(o_j - \widehat{o}_j(\beta)\right)^2}{2{\sigma_{res}}^2}\right) \tag{B2}$$

Because the posterior has no closed form, we approximated $p(\beta, {\sigma_{res}}^2|O)$ via MCMC (Gelman et al., 2013). Specifically, MCMC samples the joint posterior by iteratively proposing parameter values and accepting them with probability given by the posterior ratio (Hastings, 1970). We used the No-U-Turn Sampler (NUTS) in PyMC 4.0 (Python 3.12), an adaptive Hamiltonian method widely used in modern Bayesian inference because it automatically tunes key sampler parameters during warm-up and typically converges faster than Metropolis-Hastings (Hoffman and Gelman, 2011). We assigned a HalfNormal(20) prior to the σ_{site}, implying positive variability with a weakly informative scale that favors moderate values while still allowing larger site-to-site differences if supported by the data (the unit is W/m^2). The prior distribution of Jarvis parameters used in this study and the diagnostic summary are shown in Tables B1-B3.

Table B1
The prior distribution of Jarvis parameters used in this study.

Subfunctions	Calibration parameters	Lower bound	Upper bound
$F(R_s)$	k_1	0	500
$F(T_a)$	$T_{a,min}$	0	15
	$T_{a,opt}$	20	30
	$T_{a,max}$	35	50
$F(VPD)$	k_2	0.01	1
	k_3	0.5	4
$F(\theta)$	k_4	0	0.3
	k_5	0	5
$F(DOY)$	DOY_{pre}	50	250
	DOY_{max}	100	270
	DOY_{post}	50	250

Table B2
Posterior summaries of the hierarchical models for *T. cordata*. Mean: posterior mean values; sd: posterior standard deviation; HDI_2.5% and HDI_97.5%: Lower and upper bounds of the 95% highest posterior density interval; ESS: effective sample size; $\widehat{R}$: potential scale reduction factor.

Parameters	mean	sd	HDI_2.5%	HDI_97.5%	ESS	$\widehat{R}$
DOY_{max}	139.76	1.05	137.63	141.75	5129	1.00
DOY_{pre}	50.18	0.19	50.00	50.56	3796	1.00
DOY_{post}	109.87	1.26	107.49	112.40	5455	1.00
k_1	40.76	1.15	38.54	43.03	5906	1.00
$T_{a,min}$	0.06	0.06	0.00	0.18	3821	1.00
$T_{a,opt}$	26.37	0.56	25.31	27.49	3095	1.00
$T_{a,max}$	49.68	0.31	49.06	50.00	4406	1.00
k_2	0.09	0.01	0.06	0.11	3192	1.00
k_3	2.01	0.13	1.78	2.26	2741	1.00
k_4	0.58	0.01	0.56	0.60	3894	1.00
k_5	0.71	0.01	0.70	0.72	5200	1.00

Table B3
Posterior summaries of the hierarchical models for *R. pseudoacacia*.

Parameters	mean	sd	HDI_2.5%	HDI_97.5%	ESS	$\widehat{R}$
DOY_{max}	267.60	2.30	262.91	270.00	3449	1.00
DOY_{pre}	132.15	2.65	126.84	137.05	3473	1.00
DOY_{post}	190.50	41.09	115.17	249.96	3832	1.00
k_1	20.04	0.97	18.23	22.00	5876	1.00
$T_{a,min}$	0.11	0.11	0.00	0.34	2753	1.00
$T_{a,opt}$	24.29	0.31	23.67	24.91	4810	1.00
$T_{a,max}$	49.48	0.49	48.50	50.00	2419	1.00
k_2	0.03	0.00	0.02	0.03	2600	1.00
k_3	2.82	0.15	2.51	3.00	2623	1.00

(*continued on next page*)

Table B3 (*continued*)

Parameters	mean	sd	HDI_2.5%	HDI_97.5%	ESS	$\hat{R}$
k_4	0.60	0.23	0.20	0.95	2791	1.00
k_5	1.00	0.00	1.00	1.00	4336	1.00

Appendix C

Table C1
The hyperparameter search space used in this study and the optimized results.

Hyperparameters	Search space	Optimized results
Number of hidden layers	[1-20]	18
Number of hidden layer neurons	[8,16,32,64,128,256]	128
Dropout	[0-1]	0.05
Learning rate	[0.00001-0.01]	0.00005
Batch size	[16,32,64,128,256,512]	128
Weight decay coefficient	[0.00001-0.01]	0.00005

Table C2
Five-fold cross-validation results (with standard deviation) of RF and DL models for the two tree species. Performance metrics include NSE, MAE (W/m^2), and RMSE (W/m^2).

Models	*Tilia cordata*			*Robinia pseudoacacia*		
	NSE	MAE	RMSE	NSE	MAE	RMSE
RF	0.80 ± 0.01	8.97 ± 0.05	13.80 ± 0.28	0.88 ± 0.01	1.38 ± 0.06	2.04 ± 0.08
M1	0.75 ± 0.01	10.28 ± 0.22	15.46 ± 0.20	0.75 ± 0.01	2.13 ± 0.05	2.98 ± 0.06
M2	0.75 ± 0.01	10.75 ± 0.13	15.47 ± 0.28	0.81 ± 0.02	1.94 ± 0.11	2.63 ± 0.16
M3	0.84 ± 0.01	8.17 ± 0.12	12.24 ± 0.21	0.87 ± 0.02	1.59 ± 0.06	2.17 ± 0.12
M4	0.86 ± 0.01	7.60 ± 0.24	11.65 ± 0.42	0.88 ± 0.01	1.50 ± 0.07	2.08 ± 0.12

Table C3
Leave-one-year-out validation performance across held-out years for the two tree species. n denotes the number of hourly observations in each held-out test year. Performance metrics include NSE, MAE (W/m^2), and RMSE (W/m^2).

Species	Model	Held-out year	n	NSE	MAE	RMSE
T. cordata	RF	2015	903	0.44	14.15	18.45
		2016	949	−0.09	34.02	43.26
		2018	4643	−2.13	27.75	35.34
		2019	2781	−0.45	14.59	20.14
		2020	976	0.55	13.89	18.76
		2021	951	−0.22	34.17	45.29
	M1	2015	903	0.16	17.28	22.67
		2016	949	0.15	30.09	38.33
		2018	4643	−0.72	20.30	26.20
		2019	2781	−0.55	13.79	20.83
		2020	976	0.28	18.23	23.89
		2021	951	−0.03	30.56	41.76
	M2	2015	903	0.74	9.57	12.49
		2016	949	0.82	13.58	17.54
		2018	4643	0.30	12.11	16.74
		2019	2781	0.48	8.74	12.07
		2020	976	0.54	13.79	19.03
		2021	951	0.29	25.93	34.50
	M3	2015	903	0.76	9.13	12.18
		2016	949	0.85	12.32	16.01
		2018	4643	0.53	10.68	13.70
		2019	2781	0.58	7.77	10.85
		2020	976	0.13	19.07	26.10
		2021	951	0.30	25.70	34.40
	M4	2015	903	0.75	9.27	12.28
		2016	949	0.73	16.54	21.38
		2018	4643	0.52	10.40	13.83
		2019	2781	0.62	7.82	10.38
		2020	976	0.13	18.95	26.21
		2021	951	0.23	26.62	35.94

(*continued on next page*)

Table C3 (*continued*)

Species	Model	Held-out year	n	NSE	MAE	RMSE
R. pseudoacacia	RF	2016	946	−0.03	4.23	5.45
		2018	1871	0.25	3.35	4.09
		2019	2617	0.11	5.04	6.27
	M1	2016	946	0.22	3.74	4.75
		2018	1871	0.15	3.43	4.36
		2019	2617	0.18	4.80	6.00
	M2	2016	946	0.61	2.57	3.36
		2018	1871	0.63	2.35	2.88
		2019	2617	0.69	2.89	3.69
	M3	2016	946	0.68	2.33	3.05
		2018	1871	0.58	2.59	3.08
		2019	2617	0.65	3.16	3.90
	M4	2016	946	0.59	2.64	3.46
		2018	1871	0.61	2.43	2.97
		2019	2617	0.60	3.40	4.21

Table C4
Model performance under high-stress environmental subsets for M1-M4. Performance metrics include NSE, MAE (W/m^2), and RMSE (W/m^2). High-stress subsets were identified from the full dataset using the 95th percentile of soil moisture deficit (θ_{def}), air temperature (T_a), and vapor pressure deficit (*VPD*), and then mapped to the 30% held-out test set for model evaluation. Other two-variable combinations and the three-variable combination contained fewer than 10 records and were therefore excluded.

Species	Model	Extreme condition	n	NSE	MAE	RMSE
T. cordata	M1	High θ_{def}	362	0.72	7.78	11.98
		High T_a	442	0.77	8.13	11.16
		High *VPD*	429	0.75	8.62	11.69
		High $T_a + VPD$	353	0.76	8.50	11.66
		Mean ± SD	-	0.75 ± 0.02	8.25 ± 0.38	11.62 ± 0.34
	M2	High θ_{def}	362	0.66	8.75	13.29
		High T_a	442	0.57	11.28	15.15
		High *VPD*	429	0.60	11.25	14.83
		High $T_a + VPD$	353	0.60	11.29	14.96
		Mean ± SD	-	0.61 ± 0.04	10.64 ± 1.26	14.56 ± 0.85
	M3	High θ_{def}	362	0.85	6.31	8.78
		High T_a	442	0.84	6.86	9.31
		High *VPD*	429	0.85	6.84	9.21
		High $T_a + VPD$	353	0.85	6.89	9.24
		Mean ± SD	-	0.85 ± 0.01	6.72 ± 0.28	9.14 ± 0.24
	M4	High θ_{def}	362	0.84	6.61	9.25
		High T_a	442	0.90	5.47	7.51
		High *VPD*	429	0.88	5.84	8.01
		High $T_a + VPD$	353	0.90	5.81	7.91
		Mean ± SD	-	0.88 ± 0.03	5.93 ± 0.48	8.17 ± 0.75
R. pseudoacacia	M1	High θ_{def}	154	0.76	1.99	2.59
		High T_a	220	0.84	1.50	1.93
		High *VPD*	232	0.85	1.49	1.92
		High $T_a + VPD$	192	0.84	1.59	2.02
		Mean ± SD	-	0.82 ± 0.04	1.64 ± 0.23	2.12 ± 0.32
	M2	High θ_{def}	154	0.81	1.76	2.33
		High T_a	220	0.82	1.59	2.08
		High *VPD*	232	0.85	1.48	1.92
		High $T_a + VPD$	192	0.85	1.55	1.98
		Mean ± SD	-	0.83 ± 0.02	1.60 ± 0.12	2.08 ± 0.18
	M3	High θ_{def}	154	0.87	1.56	1.94
		High T_a	220	0.91	1.14	1.48
		High *VPD*	232	0.91	1.16	1.52
		High $T_a + VPD$	192	0.91	1.16	1.53
		Mean ± SD	-	0.90 ± 0.02	1.26 ± 0.20	1.61 ± 0.22
	M4	High θ_{def}	154	0.84	1.64	2.12
		High T_a	220	0.92	0.99	1.40
		High *VPD*	232	0.92	1.00	1.40
		High $T_a + VPD$	192	0.92	0.99	1.42
		Mean ± SD	-	0.90 ± 0.04	1.15 ± 0.32	1.58 ± 0.36

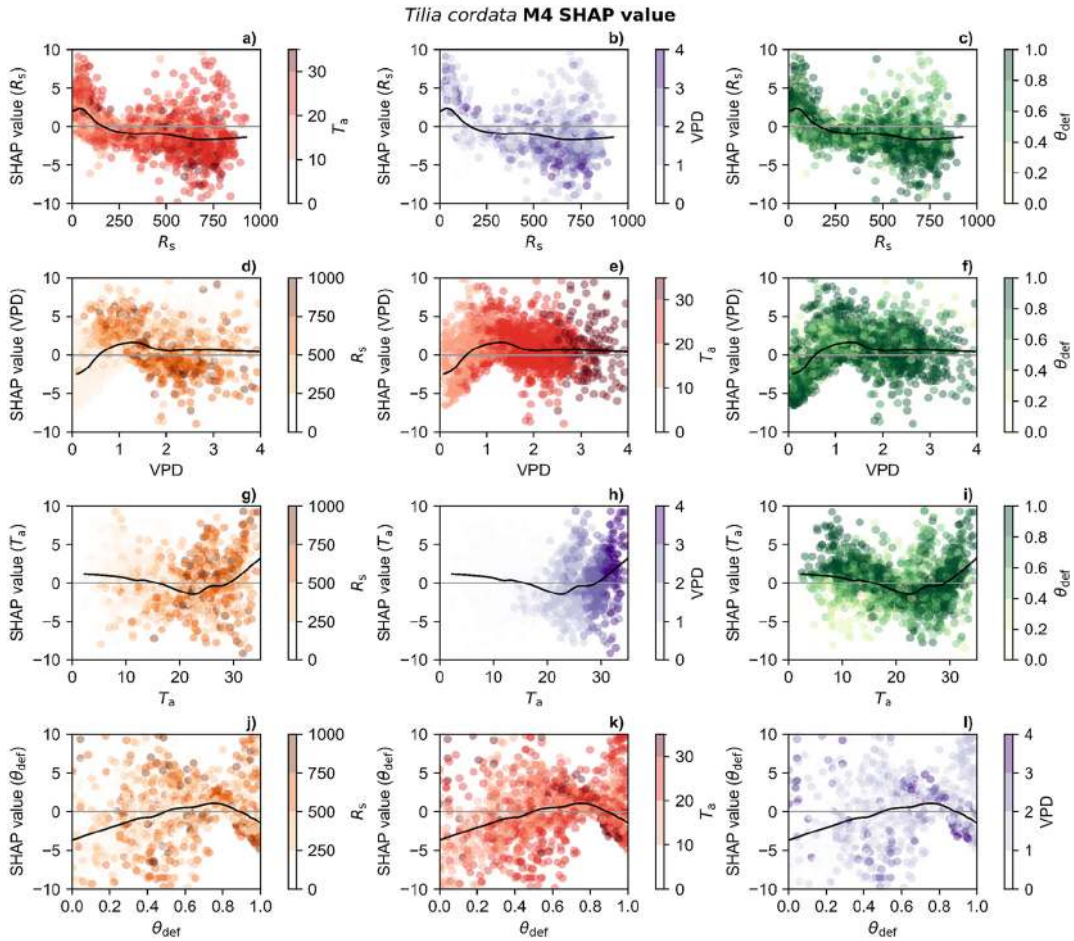

Fig. C1. SHAP interaction dependence for *T. cordata* in the physics-guided feature model (M4). Each panel plots the SHAP value of a focal predictor (y-axis) against its observed value (x-axis), with points colored by a second predictor (color bar at right). Rows correspond to focal predictors from top to bottom: shortwave radiation (R_s), vapor pressure deficit (*VPD*), air temperature (T_a), and soil moisture deficit (θ_{def}).

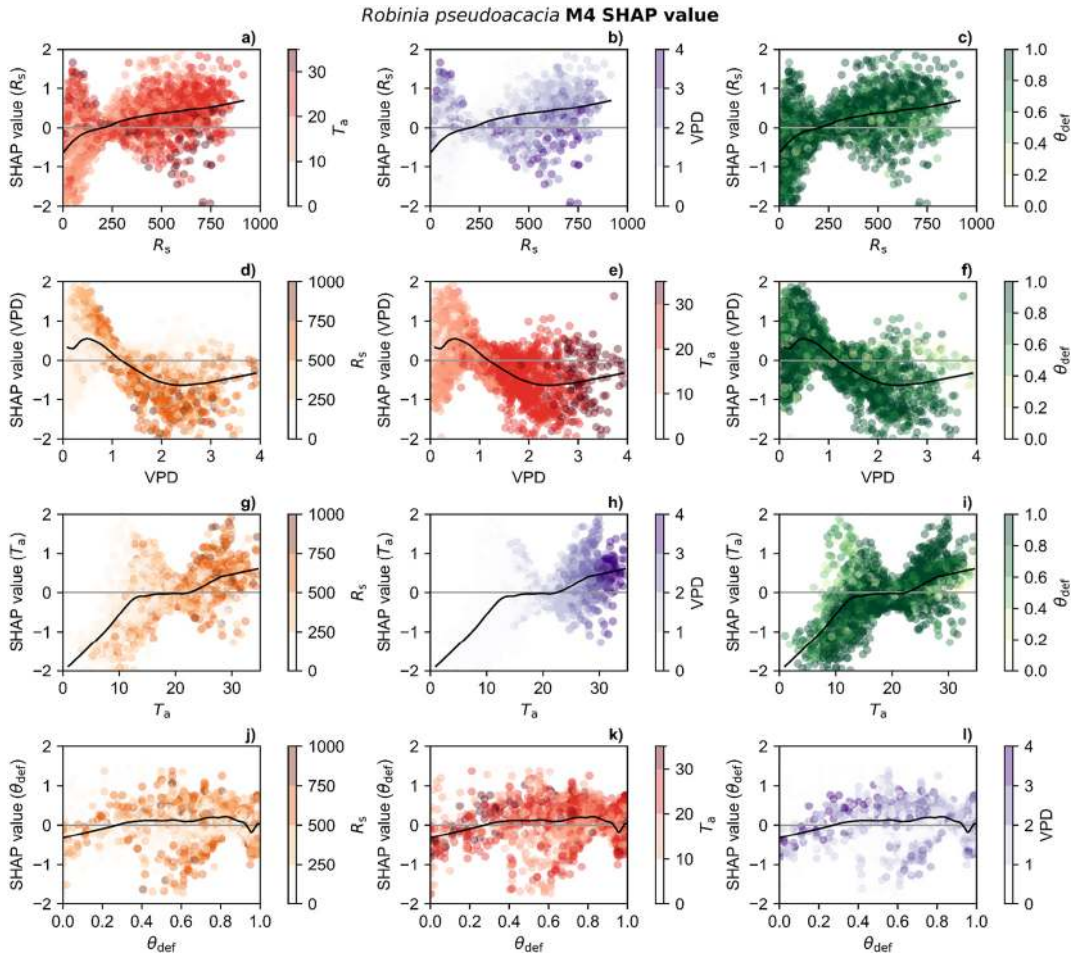

Fig. C2. SHAP interaction dependence for *R. pseudoacacia* in the physics-guided feature model (M4).

Data availability

Data will be made available on request.

References

Aghelpour, P., Bahrami-Pichaghchi, H., Kisi, O., 2020. Comparison of three different bio-inspired algorithms to improve ability of neuro fuzzy approach in prediction of agricultural drought, based on three different indexes. Comput. Electron. Agric. 170, 105279. https://doi.org/10.1016/j.compag.2020.105279.

Ali, M., Deo, R.C., Downs, N.J., Maraseni, T., 2018. Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting. Comput. Electron. Agric. 152, 149–165. https://doi.org/10.1016/j.compag.2018.07.013.

Allen, R.G., Pruitt, W.O., Wright, J.L., Howell, T.A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J.B., Smith, M., Pereira, L.S., Raes, D., Perrier, A., Alves, I., Walter, I., Elliott, R., 2006. A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 penman-monteith method. Agric. Water Manag. 81, 1–22. https://doi.org/10.1016/j.agwat.2005.03.007.

Anys, M., Weiler, M., 2025. Controls on tree transpiration dynamics in an urban environment. Agric. For. Meteorol. 372, 110726. https://doi.org/10.1016/j.agrformet.2025.110726.

Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In: Biggins, J. (Ed.), Progress in Photosynthesis Research: Volume 4 Proceedings of the Viith International Congress on Photosynthesis Providence, Rhode Island, USA, August 10–15, 1986. Springer, Netherlands, Dordrecht, pp. 221–224. https://doi.org/10.1007/978-94-017-0519-6_48.

Bennett, A., Nijssen, B., 2021. Deep learned process parameterizations provide better representations of turbulent heat fluxes in hydrologic models. Water Resour. Res. 57. https://doi.org/10.1029/2020WR029328 e2020WR029328.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Buckley, T.N., Turnbull, T.L., Adams, M.A., 2012. Simple models for stomatal conductance derived from a process model: cross-validation against sap flux data. Plant Cell Environ. 35, 1647–1662. https://doi.org/10.1111/j.1365-3040.2012.02515.x.

Cai, H., Liu, S., Shi, H., Zhou, Z., Jiang, S., Babovic, V., 2022. Toward improved lumped groundwater level predictions at catchment scale: mutual integration of water balance mechanism and deep learning method. J. Hydrol. 613, 128495. https://doi.org/10.1016/j.jhydrol.2022.128495.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010.

Chadalawada, J., Herath, H.M.V.V., Babovic, V., 2020. Hydrologically informed machine learning for rainfall-runoff modeling: a genetic programming-based toolkit for automatic model induction. Water Resour. Res. 56. https://doi.org/10.1029/2019WR026933 e2019WR026933.

Chen, D., Hsu, K., Duan, X., Wang, Y., Wei, X., Muhammad, S., 2020. Bayesian analysis of jujube canopy transpiration models: does embedding the key environmental factor in jarvis canopy resistance sub-model always associate with improving transpiration modeling? Agric. Water Manag. 234, 106112. https://doi.org/10.1016/j.agwat.2020.106112.

Chen, H., Huang, J.J., Dash, S.S., Wei, Y., Li, H., 2022. A hybrid deep learning framework with physical process description for simulation of evapotranspiration. J. Hydrol. 606, 127422. https://doi.org/10.1016/j.jhydrol.2021.127422.

Chen, H., Huang, J.J., McBean, E., 2020. Partitioning of daily evapotranspiration using a modified shuttleworth-wallace model, random forest and support vector regression, for a cabbage farmland. Agric. Water Manag. 228, 105923. https://doi.org/10.1016/j.agwat.2019.105923.

Clark, J.S., 2005. Why environmental scientists are becoming bayesians. Ecol. Lett. 8, 2–14. https://doi.org/10.1111/j.1461-0248.2004.00702.x.

de Aguiar-Campos, N., Edwards, W., Laurance, S.G.W., 2025. Tropical forest transpiration estimates are geographically, ecologically and methodologically

biased: a systematic review of sap flow research. Agric. For. Meteorol. 373, 110738. https://doi.org/10.1016/j.agrformet.2025.110738.
Evaristo, J., Jasechko, S., McDonnell, J.J., 2015. Global separation of plant transpiration from groundwater and streamflow. Nature 525, 91–94. https://doi.org/10.1038/nature14983.
Fan, J., Zheng, J., Wu, L., Zhang, F., 2021. Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. Agric. Water Manag. 245, 106547. https://doi.org/10.1016/j.agwat.2020.106547.
Flo, V., Martinez-Vilalta, J., Steppe, K., Schuldt, B., Poyatos, R., 2019. A synthesis of bias and uncertainty in sap flow methods. Agric. For. Meteorol. 271, 362–374. https://doi.org/10.1016/j.agrformet.2019.03.012.
Gaur, S., Aslan-Sungur (Rojda), G., VanLoocke, A., Drewry, D.T., 2025. Explainable machine learning to quantify the value of proximal remote sensing in latent energy flux estimation. Agric. Water Manag. 317, 109643. https://doi.org/10.1016/j.agwat.2025.109643.
Gaur, S., Drewry, D.T., 2024. Explainable machine learning for predicting stomatal conductance across multiple plant functional types. Agric. For. Meteorol. 350, 109955. https://doi.org/10.1016/j.agrformet.2024.109955.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, third ed. Chapman and Hall/CRC, New York. https://doi.org/10.1201/b16018.
Granier, A., 1987. Evaluation of transpiration in a douglas-fir stand by means of sap flow measurements. Tree Physiol. 3, 309–320. https://doi.org/10.1093/treephys/3.4.309.
Hanley, P.A., Arndt, S.K., Livesley, S.J., Szota, C., 2021. Relating the climate envelopes of urban tree species to their drought and thermal tolerance. Sci. Total Environ. 753, 142012. https://doi.org/10.1016/j.scitotenv.2020.142012.
Hastings, W.K., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57, 97–109. https://doi.org/10.1093/biomet/57.1.97.
He, B., Jia, B., Zhao, Y., Wang, X., Wei, M., Dietzel, R., 2022. Estimate soil moisture of maize by combining support vector machine and chaotic whale optimization algorithm. Agric. Water Manag. 267, 107618. https://doi.org/10.1016/j.agwat.2022.107618.
Herath, H.M.V.V., Chadalawada, J., Babovic, V., 2021. Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. Hydrol. Earth Syst. Sci. 25, 4373–4401. https://doi.org/10.5194/hess-25-4373-2021.
Hill, B.M., 1974. Bayesian inference in statistical analysis. Technometrics 16, 478–479. https://doi.org/10.1080/00401706.1974.10489222.
Hoffman, M.D., Gelman, A., 2011. The No-U-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. https://doi.org/10.48550/arXiv.1111.4246.
Houshmandfar, A., O'Leary, G., Fitzgerald, G.J., Chen, Y., Tausz-Posch, S., Benke, K., Uddin, S., Tausz, M., 2021. Machine learning produces higher prediction accuracy than the jarvis-type model of climatic control on stomatal conductance in a dryland wheat agro-ecosystem. Agric. For. Meteorol. 304–305, 108423. https://doi.org/10.1016/j.agrformet.2021.108423.
Irrgang, C., Boers, N., Sonnewald, M., Barnes, E.A., Kadow, C., Staneva, J., Saynisch-Wagner, J., 2021. Towards neural earth system modelling by integrating artificial intelligence in Earth system science. Nat. Mach. Intell. 3, 667–674. https://doi.org/10.1038/s42256-021-00374-3.
Jarvis, P.G., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. Philos. Trans. R. Soc. Lond. B Biol. Sci. 273, 593–610. https://doi.org/10.1098/rstb.1976.0035.
Jasechko, S., Sharp, Z.D., Gibson, J.J., Birks, S.J., Yi, Y., Fawcett, P.J., 2013. Terrestrial water fluxes dominated by transpiration. Nature 496, 347–350. https://doi.org/10.1038/nature11983.
Jiang, S., Zheng, Y., Wang, C., Babovic, V., 2022. Uncovering flooding mechanisms across the contiguous United States through interpretive deep learning on representative catchments. Water Resour. Res. 58. https://doi.org/10.1029/2021WR030185 e2021WR030185.
Koppa, A., Rains, D., Hulsman, P., Poyatos, R., Miralles, D.G., 2022. A deep learning-based hybrid model of global terrestrial evaporation. Nat. Commun. 13, 1912. https://doi.org/10.1038/s41467-022-29543-7.
Kulathunga, N., Ranasinghe, N.R., Vrinceanu, D., Kinsman, Z., Huang, L., Wang, Y., 2020. Effects of the Nonlinearity in Activation Functions on the Performance of Deep Learning Models. https://doi.org/10.48550/arXiv.2010.07359.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Leigh, A., Sevanto, S., Close, J.d., Nicotra, A.b., 2017. The influence of leaf size and shape on leaf thermal dynamics: does theory hold up under natural conditions? Plant Cell Environ. 40, 237–248. https://doi.org/10.1111/pce.12857.
Leuning, R., Kelliher, F.M., De Pury, D.G.G., Schulze, E.-D., 1995. Leaf nitrogen, photosynthesis, conductance and transpiration: scaling from leaves to canopies. Plant Cell Environ. 18, 1183–1200. https://doi.org/10.1111/j.1365-3040.1995.tb00628.x.
Li, Y., Ye, J., Xu, D., Zhou, G., Feng, H., 2022. Prediction of sap flow with historical environmental factors based on deep learning technology. Comput. Electron. Agric. 202, 107400. https://doi.org/10.1016/j.compag.2022.107400.
Liu, B., He, X., Liu, N., 2023. Hybrid residual deep learning models with physical knowledge for improving plant transpiration estimation. Comput. Electron. Agric. 212, 108135. https://doi.org/10.1016/j.compag.2023.108135.
Liu, B., He, X., Lyu, W., Tao, L., 2025. Physics-augmented deep learning models for improving evapotranspiration estimation in global land regions. Agric. Water Manag. 317, 109634. https://doi.org/10.1016/j.agwat.2025.109634.
Loshchilov, I., Hutter, F., 2019. Decoupled Weight Decay Regularization. https://doi.org/10.48550/arXiv.1711.05101.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.
Mitchell, P.J., Veneklaas, E., Lambers, H., Burgess, S.S.O., 2009. Partitioning of evapotranspiration in a semi-arid eucalypt woodland in south-western Australia. Agric. For. Meteorol. 149, 25–37. https://doi.org/10.1016/j.agrformet.2008.07.008.
Monteith, J., Unsworth, M., 2013. Principles of Environmental Physics: Plants, Animals, and the Atmosphere. Academic Press.
Monteith, J.L., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205–234.
Moser-Reischl, A., Rahman, M.A., Pauleit, S., Pretzsch, H., Rötzer, T., 2019. Growth patterns and effects of urban micro-climate on two physiologically contrasting urban tree species. Landsc. Urban Plan. 183, 88–99. https://doi.org/10.1016/j.landurbplan.2018.11.004.
Nolan, R.H., Tarin, T., Rumman, R., Cleverly, J., Fairweather, K.A., Zolfaghar, S., Santini, N.S., O'Grady, A.P., Eamus, D., 2018. Contrasting ecophysiology of two widespread arid zone tree species with differing access to water resources. J. Arid Environ. 153, 1–10. https://doi.org/10.1016/j.jaridenv.2018.01.003.
Novick, K.A., Ficklin, D.L., Grossiord, C., Konings, A.G., Martínez-Vilalta, J., Sadok, W., Trugman, A.T., Williams, A.P., Wright, A.J., Abatzoglou, J.T., Dannenberg, M.P., Gentine, P., Guan, K., Johnston, M.R., Lowman, L.E.L., Moore, D.J.P., McDowell, N. G., 2024. The impacts of rising vapour pressure deficit in natural and managed ecosystems. Plant Cell Environ. 47, 3561–3589. https://doi.org/10.1111/pce.14846.
Ou, J., Chen, F., Zhang, M., Batchelor, W.D., Wang, B., Wu, D., Ma, X., Zhang, Z., Hu, K., Feng, P., 2025. A continuous learning framework based on physics-guided deep learning for crop phenology simulation. Agric. For. Meteorol. 368, 110562. https://doi.org/10.1016/j.agrformet.2025.110562.
Panchen, Z.A., Primack, R.B., Nordt, B., Ellwood, E.R., Stevens, A.-D., Renner, S.S., Willis, C.G., Fahey, R., Whittemore, A., Du, Y., Davis, C.C., 2014. Leaf out times of temperate woody plants are related to phylogeny, deciduousness, growth habit and wood anatomy. New Phytol. 203, 1208–1219. https://doi.org/10.1111/nph.12892.
Poyatos, R., Villagarcía, L., Domingo, F., Piñol, J., Llorens, P., 2007. Modelling evapotranspiration in a scots pine stand under mediterranean mountain climate using the GLUE methodology. Agric. For. Meteorol. 146, 13–28. https://doi.org/10.1016/j.agrformet.2007.05.003.
Qian, H., Wang, W., Chen, G., 2024. Assessing forecast performance of daily reference evapotranspiration: a comparison of equations, machine and deep learning using weather forecasts. J. Hydrol. 644, 132101. https://doi.org/10.1016/j.jhydrol.2024.132101.
Rahman, M.A., Arndt, S., Bravo, F., Cheung, P.K., van Doorn, N., Franceschi, E., del Río, M., Livesley, S.J., Moser-Reischl, A., Pattnaik, N., Rötzer, T., Paeth, H., Pauleit, S., Preisler, Y., Pretzsch, H., Tan, P.Y., Cohen, S., Szota, C., Torquato, P.R., 2024. More than a canopy cover metric: influence of canopy quality, water-use strategies and site climate on urban forest cooling potential. Landsc. Urban Plan. 248, 105089. https://doi.org/10.1016/j.landurbplan.2024.105089.
Rahman, M.A., Hartmann, C., Moser-Reischl, A., von Strachwitz, M.F., Paeth, H., Pretzsch, H., Pauleit, S., Rötzer, T., 2020. Tree cooling effects and human thermal comfort under contrasting species and sites. Agric. For. Meteorol. 287, 107947. https://doi.org/10.1016/j.agrformet.2020.107947.
Rahman, M.A., Moser, A., Anderson, M., Zhang, C., Rötzer, T., Pauleit, S., 2019. Comparing the infiltration potentials of soils beneath the canopies of two contrasting urban tree species. Urban For. Urban Green. 38, 22–32. https://doi.org/10.1016/j.ufug.2018.11.002.
Rahman, M.A., Moser, A., Gold, A., Rötzer, T., Pauleit, S., 2018. Vertical air temperature gradients under the shade of two contrasting urban tree species during different types of summer days. Sci. Total Environ. 633, 100–111. https://doi.org/10.1016/j.scitotenv.2018.03.168.
Rahman, M.A., Moser, A., Rötzer, T., Pauleit, S., 2017a. Within canopy temperature differences and cooling ability of *tilia cordata* trees grown in urban conditions. Build. Environ. 114, 118–128. https://doi.org/10.1016/j.buildenv.2016.12.013.
Rahman, M.A., Moser, A., Rötzer, T., Pauleit, S., 2017b. Microclimatic differences and their influence on transpirational cooling of *tilia cordata* in two contrasting street canyons in munich, Germany. Agric. For. Meteorol. 232, 443–456. https://doi.org/10.1016/j.agrformet.2016.10.006.
Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-guided deep learning predictions of lake water temperature. Water Resour. Res. 55, 9173–9190. https://doi.org/10.1029/2019WR024922.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.
Rötzer, T., Moser-Reischl, A., Rahman, M.A., Hartmann, C., Paeth, H., Pauleit, S., Pretzsch, H., 2021. Urban tree growth and ecosystem services under extreme drought. Agric. For. Meteorol. 308–309, 108532. https://doi.org/10.1016/j.agrformet.2021.108532.
Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x.
Samanta, S., Clayton, M.K., Mackay, D.S., Kruger, E.L., Ewers, B.E., 2008. Quantitative comparison of canopy conductance models using a bayesian approach. Water Resour. Res. 44. https://doi.org/10.1029/2007WR006761.
Samanta, S., Mackay, D.S., Clayton, M.K., Kruger, E.L., Ewers, B.E., 2007. Bayesian analysis for uncertainty estimation of a canopy transpiration model. Water Resour. Res. 43. https://doi.org/10.1029/2006WR005028.

Santurkar, S., Tsipras, D., Ilyas, A., Madry, A., 2019. How does batch normalization help optimization? https://doi.org/10.48550/arXiv.1805.11604.

Shashua-Bar, L., Rahman, M.A., Moser-Reischl, A., Peeters, A., Franceschi, E., Pretzsch, H., Rötzer, T., Pauleit, S., Winters, G., Groner, E., Cohen, S., 2023. Do urban tree hydraulics limit their transpirational cooling? A comparison between temperate and hot arid climates. Urban Clim. 49, 101554. https://doi.org/10.1016/j.uclim.2023.101554.

Shen, C., Appling, A.P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C.J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Song, Y., Beck, H.E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C., Mohanty, B., Roy, T., Xu, C., Lawson, K., 2023. Differentiable modelling to unify machine learning and physical models for geosciences. Nat. Rev. Earth Environ. 4, 552–567. https://doi.org/10.1038/s43017-023-00450-9.

Shrikumar, A., Greenside, P., Kundaje, A., 2019. Learning important features through propagating activation differences. https://doi.org/10.48550/arXiv.1704.02685.

Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops-an energy combination theory. Q. J. R. Meteorol. Soc. 111, 839–855. https://doi.org/10.1002/qj.49711146910.

Stewart, J.B., 1988. Modelling surface conductance of pine forest. Agric. For. Meteorol. 43, 19–35. https://doi.org/10.1016/0168-1923(88)90003-2.

Tancik, M., Srinivasan, P.P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J.T., Ng, R., 2020. Fourier features let networks learn high frequency functions in low dimensional domains. https://doi.org/10.48550/arXiv.2006.10739.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C., 2021. Rank-normalization, folding, and localization: an improved r^ for assessing convergence of MCMC (with discussion). Bayesian Anal. 16, 667–718. https://doi.org/10.1214/20-BA1221.

Verhoef, A., Egea, G., 2014. Modeling plant transpiration under limited soil water: comparison of different plant and soil hydraulic parameterizations and preliminary implications for their use in land surface models. Agric. For. Meteorol. 191, 22–32. https://doi.org/10.1016/j.agrformet.2014.02.009.

Wang, H., Guan, H., Liu, N., Soulsby, C., Tetzlaff, D., Zhang, X., 2020. Improving the jarvis-type model with modified temperature and radiation functions for sap flow simulations. J. Hydrol. 587, 124981. https://doi.org/10.1016/j.jhydrol.2020.124981.

Wang, H., Guan, H., Simmons, C.T., 2016. Modeling the environmental controls on tree water use at different temporal scales. Agric. For. Meteorol. 225, 24–35. https://doi.org/10.1016/j.agrformet.2016.04.016.

Wang, R., Qin, X., Du, Z., Liu, Y., Zhao, Q., Jin, Z., Qiang, D., 2025. Improving terrestrial evapotranspiration estimation using physics-guided machine learning model driven by solar-induced chlorophyll fluorescence. J. Hydrol. 661, 133468. https://doi.org/10.1016/j.jhydrol.2025.133468.

Wang, Y., Horton, R., Xue, X., Ren, T., 2021. Partitioning evapotranspiration by measuring soil water evaporation with heat-pulse sensors and plant transpiration with sap flow gauges. Agric. Water Manag. 252, 106883. https://doi.org/10.1016/j.agwat.2021.106883.

Whitley, R., Medlyn, B., Zeppel, M., Macinnis-Ng, C., Eamus, D., 2009. Comparing the penman–monteith equation and a modified jarvis–stewart model with an artificial neural network to estimate stand-scale transpiration and canopy conductance. J. Hydrol. 373, 256–266. https://doi.org/10.1016/j.jhydrol.2009.04.036.

Whitley, R., Taylor, D., Macinnis-Ng, C., Zeppel, M., Yunusa, I., O'Grady, A., Froend, R., Medlyn, B., Eamus, D., 2013. Developing an empirical model of canopy water flux describing the common response of transpiration to solar radiation and VPD across five contrasting woodlands and forests. Hydrol. Process. 27, 1133–1146. https://doi.org/10.1002/hyp.9280.

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2022. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. https://doi.org/10.48550/arXiv.2003.04919.

Wu, Z., Cui, N., Zhao, L., Han, L., Hu, X., Cai, H., Gong, D., Xing, L., Chen, X., Zhu, B., Lv, M., Zhu, S., Liu, Q., 2022. Estimation of maize evapotranspiration in semi-humid regions of northern China using penman-monteith model and segmentally optimized jarvis model. J. Hydrol. 607, 127483. https://doi.org/10.1016/j.jhydrol.2022.127483.

Xing, L., Cui, N., Liu, C., Zhao, L., Guo, L., Du, T., Zhan, C., Wu, Z., Wen, S., Jiang, S., 2022. Estimation of daily apple tree transpiration in the Loess plateau region of China using deep learning models. Agric. Water Manag. 273, 107889. https://doi.org/10.1016/j.agwat.2022.107889.

Zahradníková, E., Juriš, S., Hrabovský, M., 2025. Growth rate of *betula pendula* roth in urban areas of central Europe: different strategies within a single species. Urban For. Urban Green. 113, 129015. https://doi.org/10.1016/j.ufug.2025.129015.

Zhan, Y., Guo, Z., Yan, B., Chen, K., Chang, Z., Babovic, V., Zheng, C., 2024. Physics-informed identification of PDEs with LASSO regression, examples of groundwater-related equations. J. Hydrol. 638, 131504. https://doi.org/10.1016/j.jhydrol.2024.131504.

Zhang, H., Li, W., Wang, G., Song, F., Wen, Z., Zhang, Hengyuan, Tong, L., Kang, S., 2025. Predicting stomatal conductance of chili peppers using TPE-optimized LightGBM and SHAP feature analysis based on UAVs' hyperspectral, thermal infrared imagery, and meteorological data. Comput. Electron. Agric. 231, 110036. https://doi.org/10.1016/j.compag.2025.110036.

Zhang, X., Ren, P., Yao, L.-Y., Li, Q., Zhong, X., Liu, K.-X., Guo, T.-Y., Yang, J.-H., Zhao, L.-H., Pauleit, S., Rahman, M.A., 2025. Investigating tree shading and thermal variations in a subtropical zone: single tree to urban squares. Build. Environ. 283, 113408. https://doi.org/10.1016/j.buildenv.2025.113408.

Zhang, X., Zhao, L., Yao, L., Zhong, X., Ren, P., 2022. Investigating the micro-scale thermal effects of natural underlying surfaces on adjacent spaces in a subtropical zone with an optimized method. Build. Environ. 222, 109382. https://doi.org/10.1016/j.buildenv.2022.109382.

Zhang, X., Zhong, X., Yan, J.-R., Li, Q., Yao, L.-Y., Liu, K.-X., Moser-Reischl, A., Rötzer, T., Pauleit, S., Rahman, M.A., 2026. Disentangling species-specific controls on urban tree transpiration: a bayesian framework. Agric. For. Meteorol. 389, 111381. https://doi.org/10.1016/j.agrformet.2026.111381.

Zhang, Y., Wang, P., Li, Y., Han, D., Wu, D., Liu, S., Guo, J., 2025. A quantification method for cumulative drought effects on summer maize yield incorporating recovery dynamics based on vegetation water index. Agric. Water Manag. 320, 109865. https://doi.org/10.1016/j.agwat.2025.109865.

Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G.Y., 2019. Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. 46, 14496–14507. https://doi.org/10.1029/2019GL085291.

Zheng, J., Fan, J., Zhang, F., Wu, L., Zou, Y., Zhuang, Q., 2021. Estimation of rainfed maize transpiration under various mulching methods using modified jarvis-stewart model and hybrid support vector machine model with whale optimization algorithm. Agric. Water Manag. 249, 106799. https://doi.org/10.1016/j.agwat.2021.106799.

Zhong, X., Voogt, J., Bailey, B., Zhang, X., Krayenhoff, E.S., 2026a. Evaluating high-resolution mean radiant temperature within an urban street canopy: resolving spatiotemporal variations with LiDAR/thermal infrared scanning and data-driven simulation. Build. Environ. 287, 113819. https://doi.org/10.1016/j.buildenv.2025.113819.

Zhong, X., Voogt, J., Zhan, W., Zhang, X., Shen, Z., Zhao, L., Ren, P., Wang, J., Wu, R., Teng, Y., Zhang, Y., 2026b. UAV-based low-altitude remote sensing in built-environment thermal studies: a review across different spatial scales. Build. Environ. 303, 114937. https://doi.org/10.1016/j.buildenv.2026.114937.

Zhou, Z., Wang, Y., An, Z., Li, R., Xu, Y., Zhang, P., Yang, Y., Wang, T., 2022. Deep root information "hidden in the dark": a case study on the 21-m soil profile of *robinia pseudoacacia* in the critical zone of the chinese Loess plateau. Catena 213, 106121. https://doi.org/10.1016/j.catena.2022.106121.

Zhu, G., Su, Y., Li, X., Zhang, K., Li, C., 2013. Estimating actual evapotranspiration from an alpine grassland on Qinghai-Tibetan plateau using a two-source model and parameter uncertainty analysis by bayesian approach. J. Hydrol. 476, 42–51. https://doi.org/10.1016/j.jhydrol.2012.10.006.