communications earth & environment

A Nature Portfolio journal

Article

https://doi.org/10.1038/s43247-026-03694-7

Three-dimensional canopy morphology and wind dynamics govern global rainfall interception

Check for updates

Qingliang Li[1,2] ✉, Xiaochun Jin[1], Zhongwang Wei[3] ✉, Cheng Zhang[4], Wei Shangguan[3], Jinlong Zhu[1,2], Zijian Zhang[4], Xiaoning Li[1], Yuguang Yan[1,2], Jing Wang[1,2], Xiaoming Shi[5], Yunfeng Lv[6] & Yongjiu Dai[3]

Simulating rainfall interception is essential for understanding the global water cycle. Yet most Earth system models rely on leaf area index-based representations, with limited three-dimensional canopy structure, introducing structural uncertainty. Here, we introduce Feature-Constrained Deep Symbolic Regression, an artificial intelligence approach for explicit canopy interception parameterization. Global analyses show that three-dimensional canopy morphology and wind dynamics dominate interception, while precipitation intensity reduces interception efficiency. Morphology-based parameterizations improve site-level interception loss rate KGESS by 0.27–0.39 over traditional approaches. Budyko-based analyses show improved evapotranspiration at nearly 70% of global flux sites and stronger runoff dynamics across six major river basins. CONUS CoLM simulations show improved runoff over 71.5% of the domain relative to GRDC and streamflow at 55.7% of GRFR stations. Global applications reveal nonlinear interception responses to rainfall intensity and vegetation structure. Together, Feature-Constrained Deep Symbolic Regression enables morphology-based parameterization, reducing rainfall partitioning uncertainty in Earth system models.

Canopy interception is a key regulator of the terrestrial water cycle, controlling the partitioning of precipitation into evaporation, throughfall, and stemflow. This process profoundly shapes runoff generation, soil moisture dynamics, and land–atmosphere energy exchange[1–5]. Interception loss (I) is defined as the fraction of rainfall retained by vegetation and evaporated back to the atmosphere. Globally, it accounts for approximately 10-30% of rainfall, rising to nearly 50% in dense forests[6–8]. Misrepresenting this flux introduces systematic biases that cascade through hydrological models, compromising predictions of water availability, flood risk, and ecosystem responses to climate variability[9–11]. Given its sensitivity to both climate and vegetation change, accurate interception modeling is crucial for reliable Earth system simulations.

Despite its importance, rainfall interception remains poorly constrained in Earth System Models (ESMs). Most ESMs and hydrological models rely on "big-leaf" or "two-big-leaf" simplifications[12–14], reducing complex canopy architecture to simple indices, primarily the Leaf Area Index (LAI) [15-21]. This paradigm fundamentally misrepresents the physical mechanisms governing how three-dimensional (3D) structures modulate rainfall redistribution, temporary storage, and subsequent evaporation. Consequently, I is often calculated using empirical LAI-based rules that lack a robust physical basis and fail to capture the dynamic interplay between canopy structure and meteorological conditions[2,15]. This reliance on LAI introduces significant structural uncertainty into large-scale hydrological and climate predictions (Supplementary Table S1 surveys representative schemes in widely used Land Surface Models[15–21]).

A growing body of evidence underscores the dominance of 3D canopy architecture in controlling interception processes[22,23]. Mechanistic experiments reveal that throughfall partitioning, which includes free throughfall, splash, and drip, depends heavily on voxel-based canopy structure and raindrop microphysics[24]. Microphysical theories further link interception efficiency to explicit structural traits derivable from high-resolution LiDAR[25]. At broader scales, hybrid modeling analyses attribute much of

[1]School of Artificial Intelligence, Changchun Normal University, Changchun, China. [2]Research Institute for Scientific and Technological Innovation, Changchun Normal University, Changchun, China. [3]School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai, China. [4]College of Computer Science and Technology, Jilin University, Changchun, China. [5]Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong, China. [6]College of Geographical Sciences, Changchun Normal University, Changchun, China. ✉e-mail: liqingliang@ccsfu.edu.cn; weizhw6@mail.sysu.edu.cn

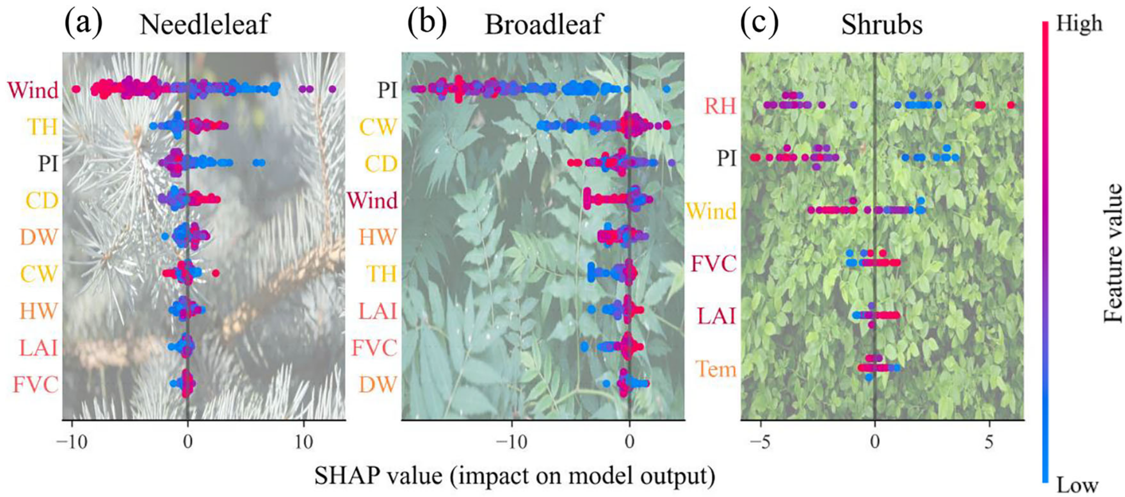

Fig. 1 | SHAP-based feature importance across vegetation types. a For needleleaf forests, variables are ordered by importance, with higher positions indicating stronger influence. Canopy structure is shown in yellow, canopy geometric ratios in orange, vegetation cover in light red, and meteorological factors in dark red. **b** For broadleaf forests, variables are ordered by importance using the same colour categories as in (**a**). **c** For shrubs, variables are ordered by importance and include only meteorological and vegetation-cover variables; wind, temperature, and relative humidity are shown in yellow, orange, and light red, respectively, and vegetation cover is shown in dark red. All variables except precipitation intensity (PI) are included as predictors in the FC-DSR parameterization of $C_{\max}$. PI enters indirectly through $\delta = C_{\max}/PI$ and therefore influences the optimization through the I/P objective.

the variability in global interception to changes in rainfall frequency and intensity rather than LAI alone, emphasizing the importance of weather coupling, which LAI-focused models tend to overlook. Despite these insights, a generalized, physically explicit scheme that translates 3D canopy morphology into scalable interception parameterizations remains elusive.

Machine learning has proven adept at modeling nonlinear dynamics in Earth-system processes[26–28], yet translating these insights into explicit, transferable equations, instead of black-box predictors[29,30] remains a persistent challenge[31]. Symbolic Regression (SR), particularly Deep SR (DSR)[32,33], offers a pathway to discover transparent, physics-consistent formulas directly from data. By integrating domain knowledge and structural priors, such as 3D canopy architecture and meteorological drivers, DSR can be guided toward physically meaningful and interpretable representations of complex processes[34,35]. This approach offers a promising avenue for developing advanced parameterization schemes by bridging the gap between data-driven discovery and process-based modeling.

Here, we introduce a Feature-Constrained Deep Symbolic Regression (FC-DSR) framework to derive explicit, morphology-driven parameterizations for canopy interception. We leverage interpretable artificial intelligence (AI) method, Random Forest (RF) SHapley Additive exPlanations (SHAP), to identify key biophysical drivers, prioritizing 3D canopy structure and wind dynamics to constrain the symbolic search toward physically plausible forms. We compiled a comprehensive multiscale dataset of interception loss ratios (I/P), spanning 1982–2020, where I denotes interception loss and P denotes total precipitation. The resulting scheme is first evaluated within a modified Budyko water-balance framework[14] to assess its impact on evapotranspiration–runoff partitioning. Subsequently, we implement the scheme in the CoLM2024 land surface model (https://github.com/CoLM-SYSU/CoLM) coupled with the CaMa-Flood routing model[36], evaluating improvements in runoff and river discharge from catchment to global scales. Finally, we analyze global interception patterns and long-term trends. Our findings demonstrate that morphology-driven interception enhances hydrological realism and offers a practical, physics-based alternative to conventional LAI-centric approaches in Earth system modeling.

Results

Canopy morphology and wind dynamics dominate interception

To identify the dominant physical controls on I/P and guide the development of FC-DSR, we first used SHAP analysis[37] to quantify variable importance across diverse vegetation types. In this analysis, SHAP values indicate how much each variable increases or decreases the predicted I/P relative to the model's average prediction, with larger absolute values representing stronger influence. The analysis unequivocally demonstrates that 3D canopy morphology and wind dynamics exert substantially stronger control over interception than the traditional LAI. We categorized predictors based on physical traits: 3D structure (Canopy Depth (CD), Canopy Width (CW), Tree Height (TH)), geometric ratios (Height-to-Width (HW), Depth-to-Width (DW)), vegetation cover (LAI, Fractional Vegetation Cover (FVC)), and meteorology (Wind Speed (Wind)) (Fig. 1). By prioritizing variables with high SHAP scores within the FC-DSR framework (see Methods), we guided the discovery process toward the most influential physical drivers, especially CD, CW, and Wind. To improve readability, all symbols and abbreviations used in FC-DSR are summarized in Table 1.

To interpret these drivers within a physically meaningful framework, we build on the Budyko-based formulation described in the "Methods" section and incorporate FVC as a grid-scale adjustment factor. The validity of this modification is supported by the results presented in the "Enhanced Water Balance Partitioning" section. The resulting formulation is written as follows:

$$I/P = (1-\delta)e^{-\delta} * FVC * 100\% \tag{1}$$

Here, δ represents the ratio of canopy storage capacity ($C_{\max}$) to precipitation intensity (PI). $C_{\max}$ is defined for the vegetated fraction of the grid cell, representing canopy storage capacity only where vegetation is present, rather than a value pre-averaged over the full grid cell. Multiplication by FVC therefore converts canopy-level storage into an effective grid-cell mean through area weighting. By contrast, precipitation depth P is defined on a ground-area basis for the entire grid cell and is not multiplied by FVC. Under this formulation, I/P is defined consistently at the grid scale. Both $C_{\max}$ and PI are expressed as equivalent water depths (mm), which makes δ dimensionless. Here, PI is defined as the mean rainfall depth per event and is estimated from monthly precipitation and the number of wet days using the statistical framework of Good et al[14]. As the ratio of $C_{\max}$ to precipitation input, δ governs both the magnitude and the curvature of the interception response. Our SHAP analysis reveals that structural variables (CD, CW) and Wind explain the variability in δ more robustly and consistently across biomes than LAI (Fig. 1). In Fig. 1, variables are ordered by their overall

Table 1 | Symbols used in the interception parameterization scheme

Symbol	Definition	Unit
I	Interception loss	mm
P	Precipitation	mm
I/P	Interception loss ratio	%
C_{max}	Max canopy water storage capacity	mm
PI	Precipitation intensity	mm
δ	Ratio of (C_{max}) to (PI)	-
FVC	Fractional vegetation cover	-
LAI	Leaf area index	-
CD	Canopy depth	m
CW	Canopy width	m
TH	Tree height	m
HW	Height-to-width ratio	-
DW	Depth-to-width ratio	-
Wind	Wind speed	m/s

importance, as measured by the magnitude of their SHAP values, with higher positions indicating stronger influence on the predicted I/P. This finding underscores the necessity of explicitly representing these factors in interception schemes to reduce systematic biases inherent in current models. The attribution patterns align strongly with physical mechanisms (Fig. 2 and Supplementary Fig. S1). In Fig. 2, SHAP dependence plots show how changes in each variable influence the predicted I/P, where positive (negative) values indicate an increase (decrease) in interception. The strength of these relationships is further quantified by the R^2 values shown in Fig. 2 and Supplementary Fig. S1. PI exhibits the strongest overall influence, showing a pronounced negative effect on I/P (Supplementary Fig. S1c; SHAP magnitudes spanning approximately −20 to 5). This confirms that intense rainfall rapidly overwhelms C_{max}, reducing interception efficiency[2,38]. Crucially, Wind and canopy architecture emerge as dominant modulating factors. Wind consistently reduces canopy storage across all vegetation types (Fig. 2d). This aligns with physical processes where wind dislodges intercepted droplets through aerodynamic drag and mechanical shaking, promoting droplet coalescence and fall[39–42].

Table 2 | Explicit canopy interception parameterizations (δ) derived by FC-DSR. CD denotes canopy depth, CW represents canopy width, Wind refers to wind speed, HW is the height-to-width ratio of the canopy, and PI indicates precipitation intensity

Vegetation type	δ parameterizations
Needleleaf	$\frac{CD+CW}{4*(1+Wind)*PI}$
Broadleaf	$\frac{CW}{2*(Wind*HW)*PI}$
Shrubs	$\frac{\frac{1}{1+Wind}+1}{2*PI}$

Structural controls exhibit distinct, vegetation-dependent roles. In needleleaf forests, both CD and CW positively correlate with I/P, with CD showing a stronger contribution (Fig. 2a, b; blue circles). In broadleaf forests, CD shows a positive relationship with interception, whereas the HW ratio is negatively related to it (Fig. 2b, c; yellow symbols).

Discovery of morphology-driven interception parameterizations

Guided by these dominant drivers, the FC-DSR framework derives explicit and compact parameterizations for δ in the canopy interception scheme. The discovered analytical expressions (Table 2) reveal a tight, nonlinear coupling between canopy geometry (CD, CW, HW) and meteorological forcing (Wind).

This interpretation is consistent with recent process-based studies[43] showing that canopy storage capacity depends not only on vegetation amount, such as LAI, but also on canopy structural properties. The functional forms include both main effects and critical interactions (products and ratios) that linear models inherently miss. Physically, these relationships demonstrate that canopy morphology sets the baseline storage capacity, while wind dynamically modulates the realized interception. The inverse relationship with Wind across all types explicitly quantifies the mechanical removal of water from the canopy. The distinct structural terms, such as the additive influence of CD and CW in needleleaf versus the interaction of CW, Wind, and HW in broadleaf, reflect the different ways these canopy architectures interact with rainfall and airflow. Independent validation confirms the robustness and superior performance of the morphology-driven scheme (Fig. 3). Across the three vegetation types, the FC-DSR scheme substantially improves predictive skill, increasing the Kling–Gupta Efficiency Skill Score (KGESS) by 0.27 to 0.39 compared to the conventional empirical baseline[44].

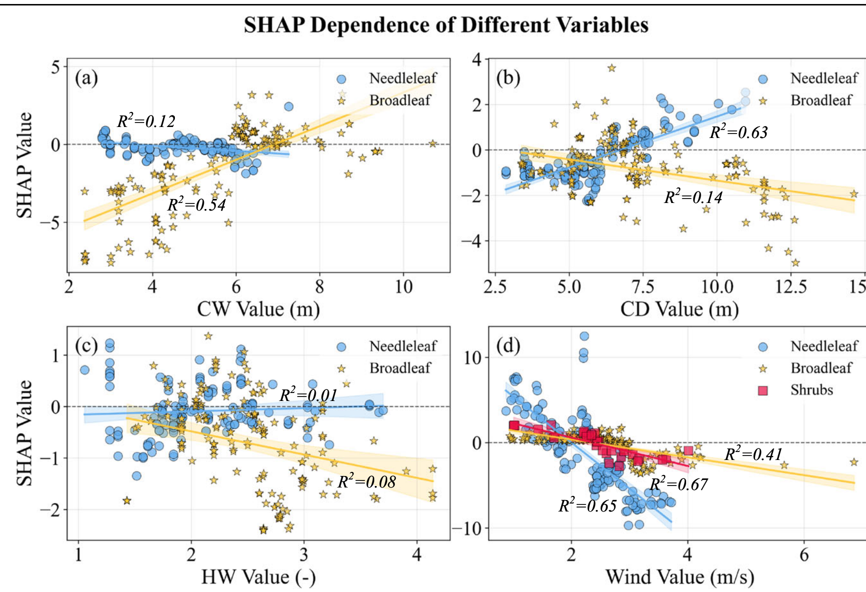

Fig. 2 | SHAP dependence plots for key variables. **a–d** Test-set samples for canopy width (CW), canopy depth (CD), height-to-width ratio (HW), and wind speed (Wind), respectively. Blue circles, yellow stars, and red squares denote needleleaf, broadleaf, and shrub vegetation, respectively. Solid lines indicate linear regression fits for each type, with shaded bands showing 95% confidence intervals. R^2 values are provided for each fitted line to indicate the strength of the linear relationships.

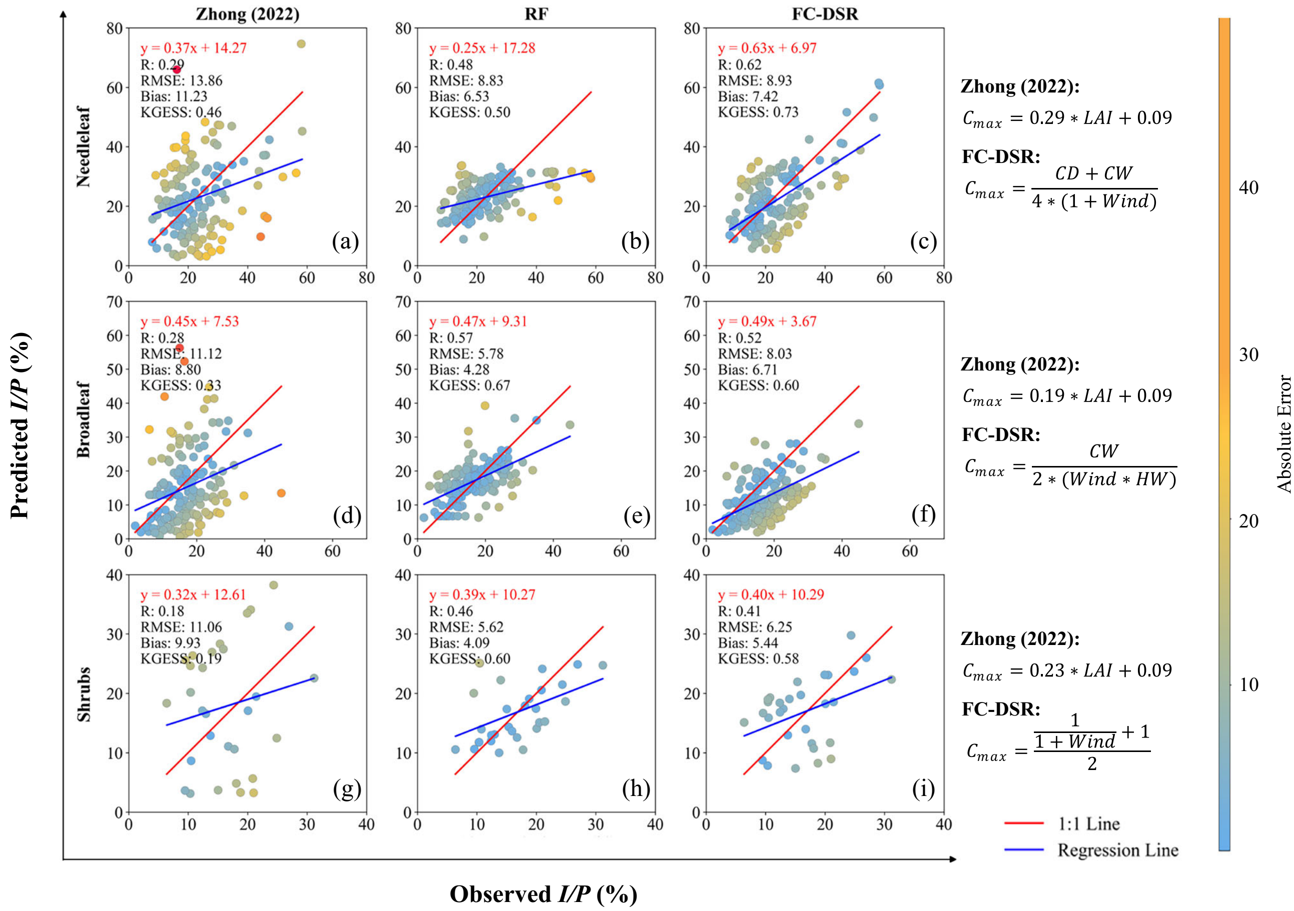

Fig. 3 | Performance of canopy interception parameterizations against observations. a–i Predicted versus observed interception loss rate, *I/P*, for different vegetation types and canopy parameterization schemes. Rows represent vegetation types, including needleleaf, broadleaf, and shrubs, and columns represent parameterization schemes, including Zhong (2022), random forest (RF), and feature-constrained deep symbolic regression (FC-DSR). Point colours indicate absolute error, calculated as |prediction − observation|. Red lines denote the 1:1 line, and blue lines denote regression lines. Formulas on the right show the canopy storage parameterizations from Zhong (2022) and our scheme.

While the RF model achieves high accuracy, FC-DSR provides comparable or superior skill while offering the critical advantage of interpretability. Notably, in needleleaf forests, KGESS reaches 0.73 with FC-DSR, a significant improvement over 0.50 for RF. The larger improvement likely reflects the more coherent structural controls in needleleaf forests. Both CD and CW have positive effects (Fig. 2), with CD playing the dominant role, consistent with conifer canopy structure. The relatively regular and vertically organized architecture of conifer stands likely produces a smooth and physically interpretable response to CD that can be captured effectively by FC-DSR. In contrast, RF relies on recursive local partitions that may fit site-specific variability but do not necessarily recover an equally transferable functional relationship. This may explain why FC-DSR performs especially well when CD is the dominant canopy control. In broadleaf forests, FC-DSR exhibits reduced bias and better captures systematic variation, with fitted relationships ($y = 0.49x + 3.67$) aligning more closely with observations than RF ($y = 0.47x + 9.31$). These results demonstrate that constrained SR successfully distilled complex relationships into transparent, physically interpretable formulations. This morphology-driven scheme provides a practical and robust pathway to incorporate 3D canopy structure and wind effects into large-scale models.

Enhanced water balance partitioning

To assess how the derived parameterizations affect precipitation partitioning at the process level, we incorporated them into a revised Budyko water-balance framework[14]. We tested four configurations (Schemes 1-4; Fig. 4 and Supplementary Tables S2 and S3) that systematically varied the interception scheme (Empirical Baseline vs. FC-DSR) and the scale at which FVC is applied (Process-scale within δ vs. Grid-scale modifier). Validation against evapotranspiration (ET) data from 103 global flux sites (Supplementary Table S4) revealed that Scheme 4, which integrates the FC-DSR parameterization with grid-scale FVC modification, achieved the highest overall performance (Median KGESS = 0.410; Supplementary Fig. S2). Figure 4a, b are based on all available sites and assess the benefit of incorporating FVC as a grid-scale factor. Figure 4c, d are based on a subset of selected vegetation types, including Needleleaf, Broadleaf, Shrubs, and assess the improvement of the new parameterization across different categories of parameterization change. Treating FVC at the grid scale (Schemes 3 and 4) consistently outperformed embedding FVC within the process parameter δ (Schemes 1 and 2). Specifically, Scheme 3 improved upon Scheme 1 at 71.8% of sites, and Scheme 4 improved upon Scheme 2 at 69.9% of sites. This demonstrates that applying FVC as an aerial fraction at the grid scale more effectively captures sub-grid heterogeneity than attempting to modify the microphysical interception process directly. Furthermore, the morphology-driven FC-DSR scheme consistently outperforms the empirical baseline (Fig. 4c, d). Performance gains varied across aggregated vegetation groups (see "Methods"). Notably, in Mixed forests, the interaction between parameterization and scaling was pronounced. The degraded performance when injecting FVC into δ (Schemes 1–2) confirms that a single process-level parameter cannot adequately represent the structural diversity of heterogeneous stands, reinforcing the superiority of the grid-scale treatment of FVC. We further evaluated the impact on global runoff generation by comparing Scheme 4 (FC-DSR) with Scheme 3 (Empirical Baseline) over

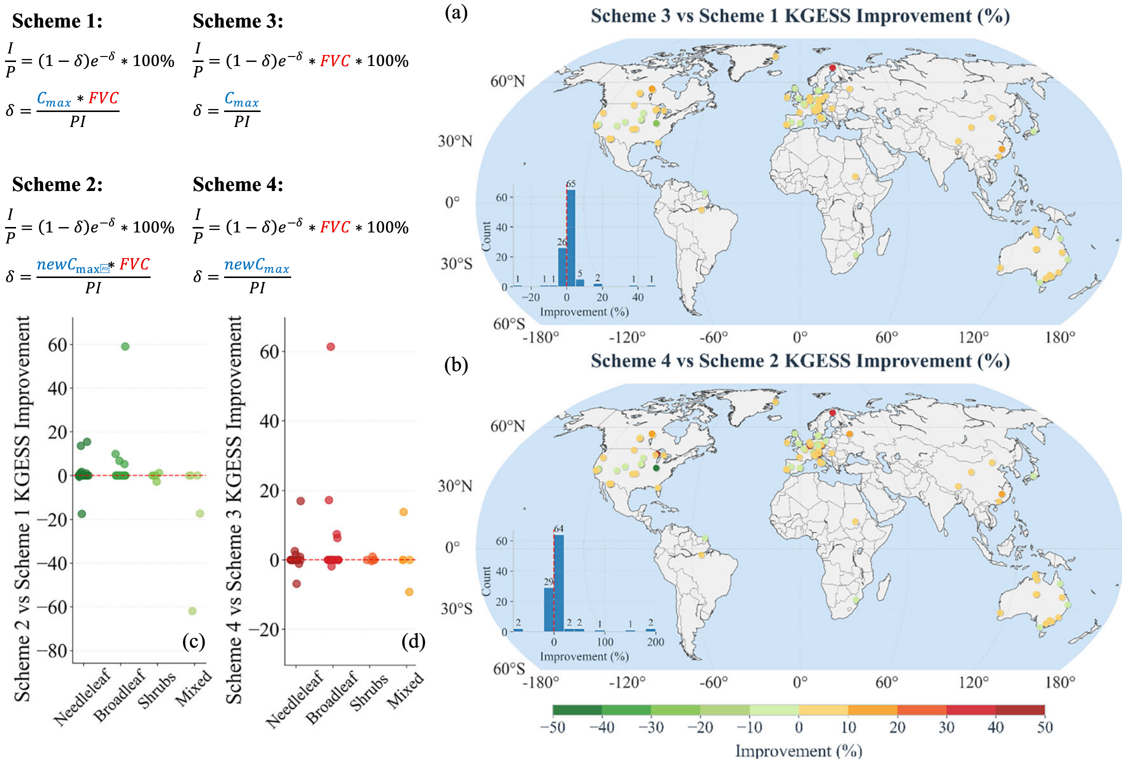

Fig. 4 | Performance comparison of interception parameterizations in the revised Budyko model. **a**, **b** Spatial patterns of KGESS improvements, with Scheme 3 evaluated against Scheme 1 and Scheme 4 against Scheme 2, calculated as (Scheme 3–Scheme 1)/|Scheme 1| and (Scheme 4–Scheme 2)/|Scheme 2|. **c**, **d** Scatterplots of KGESS improvements across vegetation types, showing Scheme 2 relative to Scheme 1 and Scheme 4 relative to Scheme 3, calculated as (Scheme 2–Scheme 1)/|Scheme 1| and (Scheme 4–Scheme 3)/|Scheme 3|.

the period 2000–2019. The morphology-driven scheme induced distinct spatial shifts (Fig. 5a), generally increasing runoff across high-latitude forests and specific tropical regions, while decreasing it elsewhere, reflecting the altered interception and subsequent ET and runoff partitioning.

At the basin scale, we analyzed discharge in six major global river systems (Amazon, Columbia, Danube, Mississippi, Orange, and Paraná) by aggregating modeled runoff upstream of GSHA[45] gauges. These basins encompass diverse hydroclimatic and vegetation regimes, including humid tropical, temperate agricultural, snow-influenced mountainous, temperate mixed, semi-arid, and subtropical seasonal systems. Across these settings, Scheme 4 outperformed Scheme 3 in five of the six basins for both annual time series and cumulative runoff (Fig. 5b–g and Supplementary Fig. S3), demonstrating robust performance and broad transferability. Crucially, Scheme 4 demonstrated enhanced capability in tracking dynamic discharge fluctuations, better capturing peak flows and recession dynamics (Fig. 5b–g, red circles). This improved responsiveness stems from the more realistic, morphology-aware representation of canopy storage. While Scheme 3 showed marginally better performance during some specific low-flow periods (blue circles), Scheme 4 provides greater physical realism and robustness across diverse hydroclimatic regimes.

Impacts on dynamic land surface simulations

To examine whether the proposed scheme improves fully coupled land-surface simulations, we implemented the FC-DSR parameterizations in CoLM2024, hereafter referred to as $CoLM_A$). We specifically replaced the original LAI-based interception rule (Supplementary Text S1) while preserving all other model components and forcings. Simulations were conducted over the contiguous United States (CONUS) and evaluated using the standardized OpenBench platform[46], this region was chosen because it provides extensive, high-quality runoff and streamflow reference data. $CoLM_A$ substantially improved hydrological performance compared to the baseline CoLM.

Analysis of the normalized Root Mean Square Error (RMSE*)* score (nRMSEScore; Fig. 6a, b) shows that $CoLM_A$ enhanced runoff simulation skill relative to GRDC reference data at 71.5% of gauging stations, and improved streamflow skill relative to GRFR[47] records. The most significant runoff improvements were concentrated in the mountainous western U.S. and the semi-arid Southwest (Fig. 6a, blue-circled region). In these complex terrains, characterized by event-driven hydrology and orographic effects, accurate interception partitioning is crucial. By replacing the simplistic LAI rule with the physically grounded FC-DSR scheme, $CoLM_A$ more realistically distributes precipitation among throughfall, canopy evaporation, and soil recharge, enhancing local water-balance fidelity.

The implementation led to physically coherent adjustments throughout the hydrological system. $CoLM_A$ generally increased simulated canopy interception, resulting in a modest reduction in effective surface water input and modeled streamflow across much of the CONUS (Fig. 6c and Supplementary Fig. S4). At representative gauging stations (Fig. 6d–f), the cumulative streamflow simulated by $CoLM_A$ tracks GRDC observations more closely than the original CoLM, particularly in capturing long-term dynamics (Fig. 6d–f, purple boxes). These results indicate that incorporating 3D canopy structure and wind dynamics provides a more realistic representation of land surface hydrological processes.

Global interception patterns driven by 3D structure

After validating the scheme through both site-scale and model-based evaluations, we extended FC-DSR to the global scale to produce new estimates of I and I/P (Fig. 7 and Supplementary Fig. S5). To ensure robustness and maintain physical interpretability across diverse climate zones, input variables (Wind, CD, CW) were harmonized and constrained within biome-specific physical ranges observed during training (see "Methods"). The resulting global patterns reveal pronounced spatial heterogeneity that strongly reflects the underlying 3D canopy morphology. Absolute I peak in humid tropical broadleaf forests (Supplementary Fig. S5a, b). This pattern is

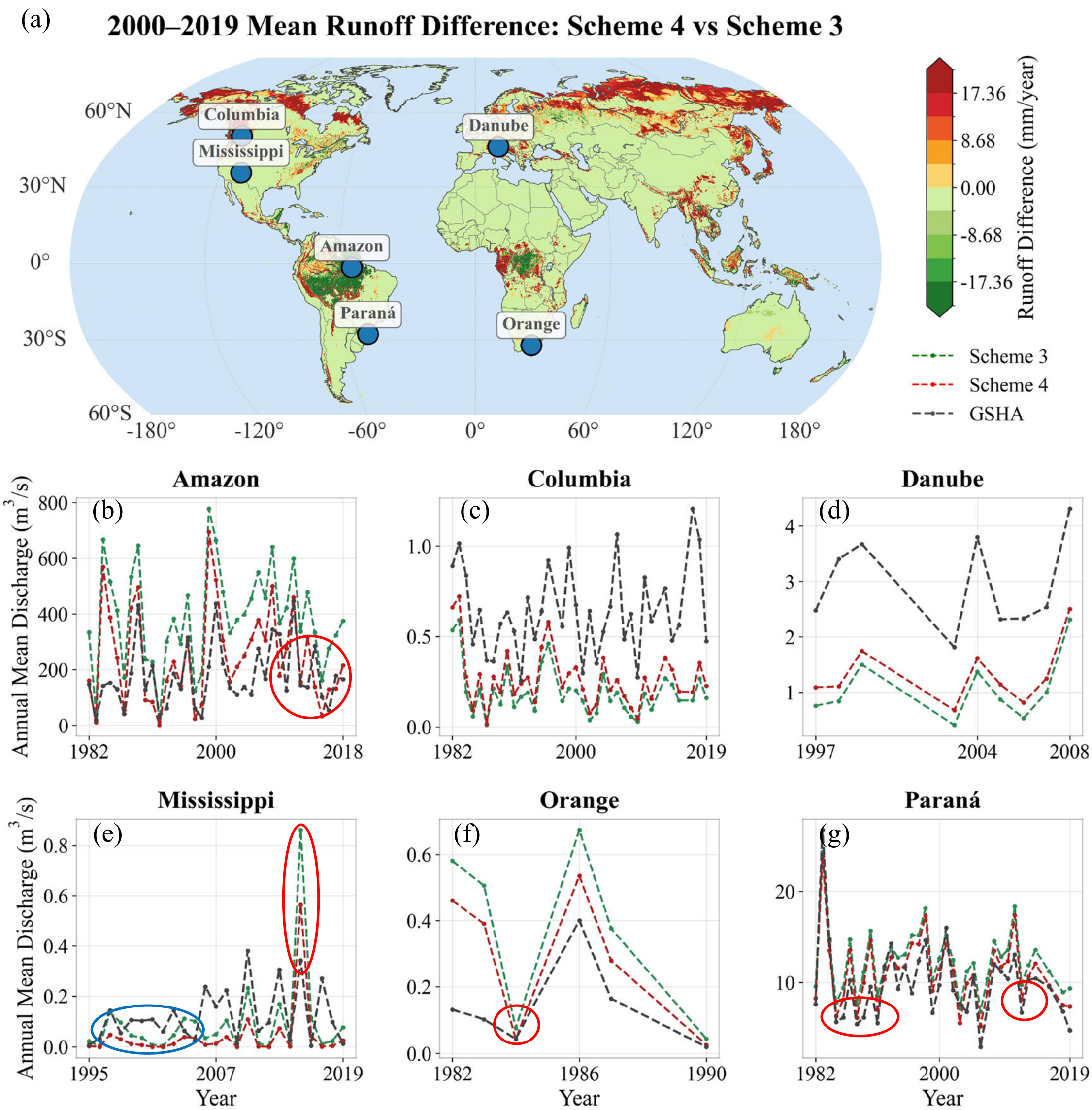

Fig. 5 | Global runoff differences between Scheme 4 and Scheme 3. a Spatial distribution of runoff differences, calculated as the multi-year mean runoff of Scheme 4 minus that of Scheme 3. **b–g** Time-series comparisons of cumulative runoff at representative stations during 2000–2019. Red circles mark runoff stations where Scheme 4 yields cumulative runoff closer to observations than Scheme 3, whereas blue circles indicate stations where Scheme 4 performs worse.

driven by the coincidence of high precipitation and maximal C_{max}, characterized by the largest observed CW and CD globally (Fig. 8b, d).

In contrast, I/P, which indicates the relative importance of interception in the water balance, highlights different structural controls. Temperate and boreal coniferous forests exhibit high I/P values despite lower annual rainfall (Fig. 7a, e). This high efficiency is consistent with their dense, vertically developed canopies (high CD; Fig. 8a, c), which maximize rainfall contact time and storage. The tight spatial covariation between global interception patterns and 3D canopy geometry (CD, CW) confirms that morphology is a first-order control on global water partitioning, surpassing the explanatory power of LAI alone.

Benchmarking and divergence from existing products

To place these estimates in context, compared the morphology-driven product[48] with two established global datasets, GLEAM v4.2a[49] and Lian (2022)[2]. Analysis of the mean differences (2000–2020) reveals significant regional divergences, reflecting fundamental differences in the underlying parameterizations (Fig. 9a–d and Supplementary Fig. S6–S7). Notably, our estimates yield substantially higher I and I/P in the tropics compared to both references. A Köppen climate-zone analysis (Fig. 9e, f) highlights these contrasts. While spatial patterns of I show broad consistency, I/P estimates diverge sharply. In the tropics, our product yields the highest I/P estimates, reaching about 16%, which is 6–8% higher than those of the other products. This difference likely arises from the explicit representation of 3D canopy structure in FC-DSR, particularly CW and CD, which allows the model to better capture C_{max} of complex, multilayered tropical canopies. This interpretation is broadly consistent with the findings reported by Porada et al. in 2018, who showed that non-vascular vegetation can substantially increase C_{max} and enhance rainfall interception at large scales[50]. This aspect is frequently underestimated by models focusing solely on LAI. Conversely, in arid zones, Lian et al. [2] reports the highest I/P (~11%), while our product and GLEAM are lower (~5%). In structure-dominated biomes (needleleaf and broadleaf forests), our morphology-based estimates offer a more physically constrained representation aligned with process understanding[44].

Nonlinear interception responses to global change

We further examined how interception responds to long-term changes in climate and vegetation. We analyzed trends in I and I/P from 1982 to 2020 (Fig. 10a, c) alongside concurrent trends in precipitation sum (PS) and PI (Fig. 10b and Supplementary Table S5).

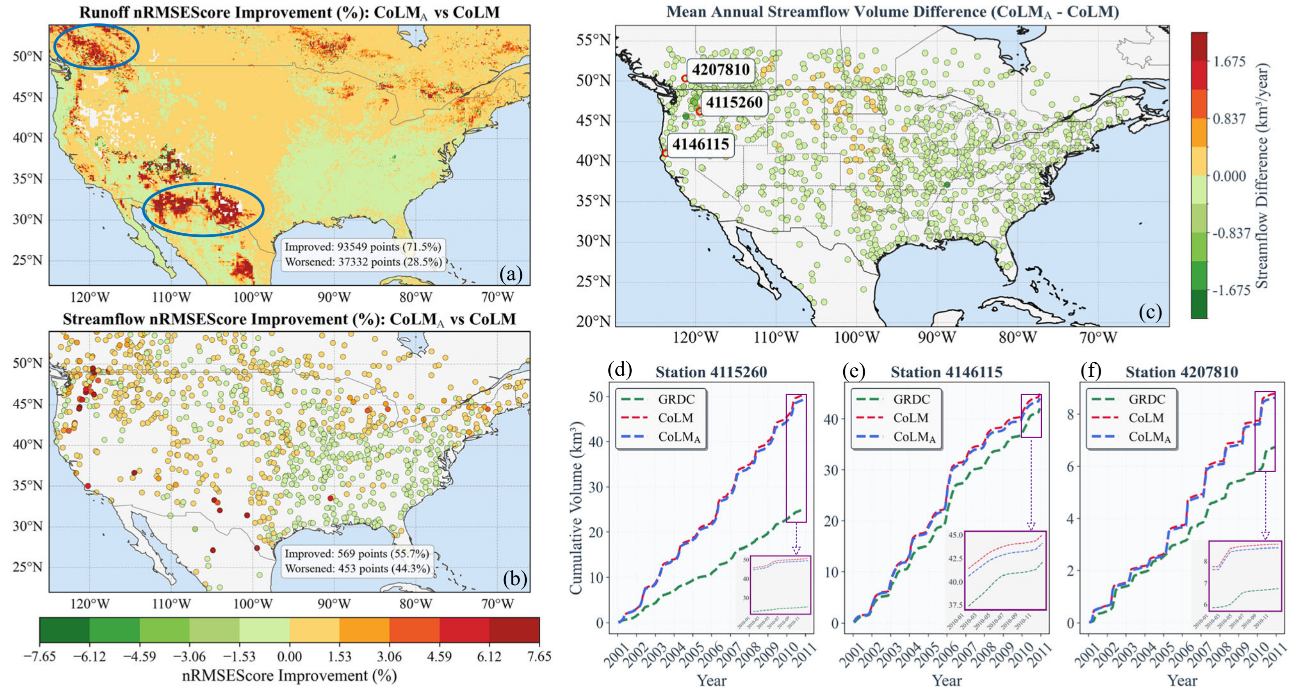

Fig. 6 | Improvements in runoff and streamflow simulations with CoLM$_A$. **a** Relative improvement of CoLM$_A$ over CoLM in nRMSEScore for runoff, using GRFR[47] as the reference dataset. **b** Relative improvement of CoLM$_A$ over CoLM in nRMSEScore for streamflow, using Global Runoff Data Centre discharge data (GRDC) as the reference dataset. Relative improvement is calculated as (CoLM$_A$ − CoLM)/|CoLM| × 100%. **c** Interannual differences in streamflow between CoLM$_A$ and CoLM, calculated as CoLM$_A$ − CoLM. **d–f** Cumulative streamflow at three representative sites. Red lines denote CoLM, blue lines denote CoLM$_A$, and green lines denote GRDC. Purple boxes highlight magnified views of the last 10% of the time series.

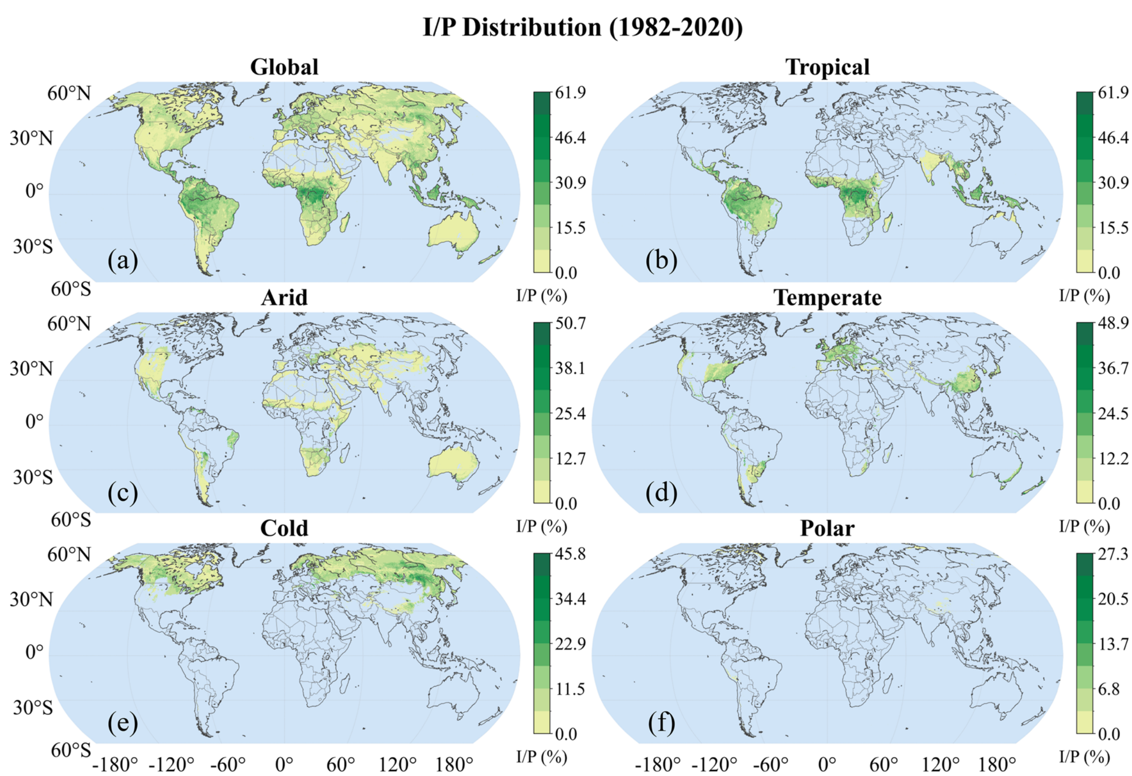

Fig. 7 | Global distribution of canopy interception loss ratio (*I/P*) during 1982–2020. **a** Global spatial pattern of I/P. **b–f** Spatial distributions of I/P in tropical, arid, temperate, cold, and polar climate regions, respectively. Color bars indicate I/P values in percent.

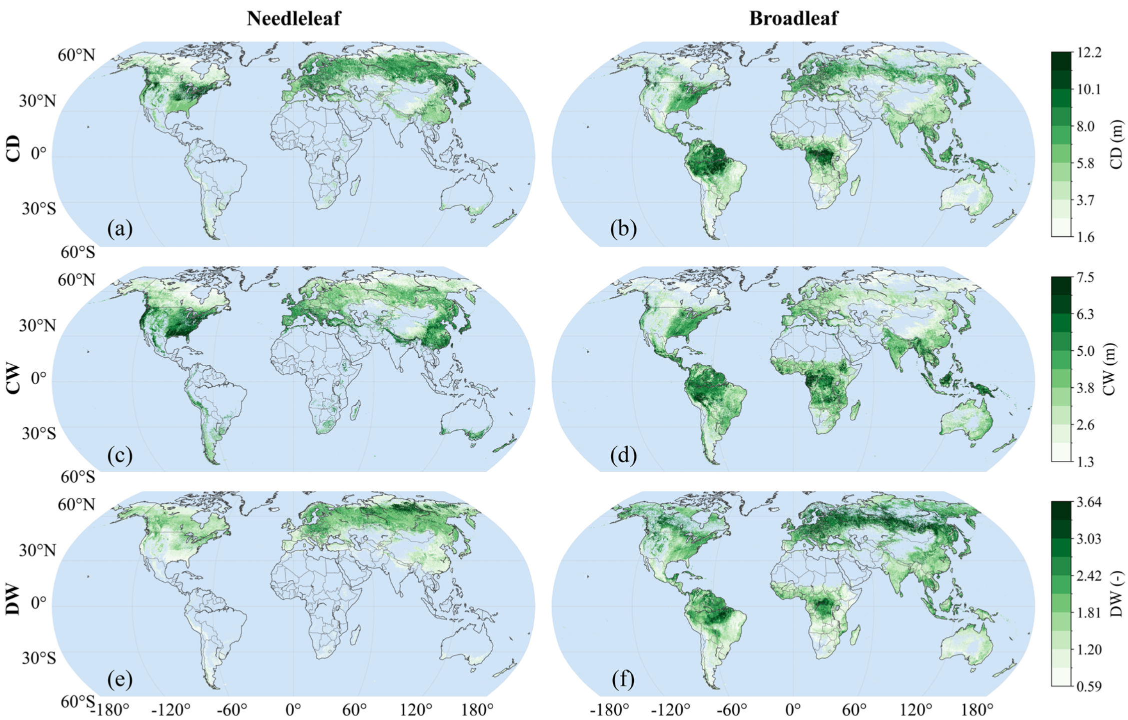

Fig. 8 | Global patterns of canopy structural traits. a, **b** Spatial distributions of canopy depth (CD) for needleleaf and broadleaf vegetation, respectively. **c**, **d** Spatial distributions of canopy width (CW) for needleleaf and broadleaf vegetation, respectively. **e**, **f** Spatial distributions of the canopy depth-to-width ratio (DW), calculated as CD/CW, for needleleaf and broadleaf vegetation, respectively.

The results reveal highly nonlinear and regionally heterogeneous responses. In northern Eurasia (Fig. 10, red box), both I and I/P show clear increases. This trend is consistent with observed increases in both rainfall and LAI (Supplementary Fig. S8a), indicating that vegetation greening and regional wetting have synergistically enhanced canopy interception, despite mixed trends in PI.

Tropical responses were far more complex. In several tropical regions (e.g., purple box), both I and I/P declined, irrespective of local rainfall trends. This decline likely reflects the impacts of forest degradation and climate stress[51], which simplify canopy structure and potentially increase local wind exposure (Supplementary Fig. S8b). These structural changes reduce interception efficiency, thereby overriding the potential effects of changes in precipitation intensity. The Congo basin (blue box) illustrates decoupled responses: decreased overall rainfall reduced total I, while a concurrent decrease in rainfall intensity increased the I/P. Globally, PS, PI, I, and I/P all exhibit positive trends over the past four decades (Fig. 10d–g). Here, each point represents the annual global mean calculated by first aggregating the data to the annual scale and then averaging across all grid cells. Linear trends were estimated using ordinary least squares regression. This underscores a critical dynamic. The climate-driven intensification of the hydrological cycle not only increases water inputs to ecosystems but substantially amplifies the regulatory role of the canopy in partitioning global rainfall. Capturing these complex trends requires models such as the FC-DSR scheme, which explicitly account for interactions among canopy morphology, wind dynamics, and evolving rainfall patterns.

Discussion

Our results show that rainfall interception cannot be adequately represented by LAI-centered formulations alone. Across site observations, water-balance analysis, land-surface simulations, and global application, interception is controlled more strongly by 3D canopy morphology and wind dynamics than by LAI alone. This helps explain why conventional big-leaf interception schemes often perform inconsistently across vegetation types and climatic regimes. Rather than being determined mainly by LAI, interception at the monthly scale emerges from the nonlinear interaction among canopy architecture, aerodynamic forcing, and rainfall characteristics.

A first key result is that 3D canopy structural traits exert distinct and physically meaningful controls across vegetation types. The SHAP rankings in Fig. 1 and the dependence patterns in Fig. 2 show that in needleleaf forests, CD contributes more strongly than CW, suggesting that vertically developed, multilayered canopies enhance rainfall contact time and storage efficiency[38,52]. In broadleaf forests, by contrast, the negative relationship with HW (Fig. 2b, c; yellow symbols) indicates that tall, narrow crowns are less effective at retaining rainfall, likely because they provide less lateral storage area per unit ground area and facilitate more direct drainage and drip pathways. Greater aerodynamic exposure may further accelerate droplet removal and shorten canopy water residence time. This interpretation is consistent with previous studies showing that rainfall partitioning is shaped not only by LAI, but also by crown traits, canopy spatial organization, throughfall pathways, and aerodynamic regulation of wet-canopy evaporation and droplet removal[53]. These contrasting responses show that interception is not controlled by a single universal structural metric. They also help explain why LAI-based approaches, which compress canopy form into a single indicator of LAI, miss important structural heterogeneity. Our findings do not suggest that LAI is unimportant, since LAI remains linked to storage capacity. They do indicate, however, that LAI alone is insufficient to explain interception variability at the monthly scale considered here. Instead, interception is better characterized by the nonlinear interaction between 3D canopy architecture, particularly CD and CW, and near-surface wind dynamics. Together, these factors capture structural heterogeneity and aerodynamic regulation beyond what LAI alone can represent.

A second important result is that FC-DSR improves predictive performance while preserving process interpretability. The strong performance of the explicit FC-DSR parameterizations in Fig. 3 shows that the dominant structure–function relationships can be distilled into compact analytical forms rather than remaining embedded in black-box models. This is particularly clear in needleleaf forests, where FC-DSR outperforms RF and

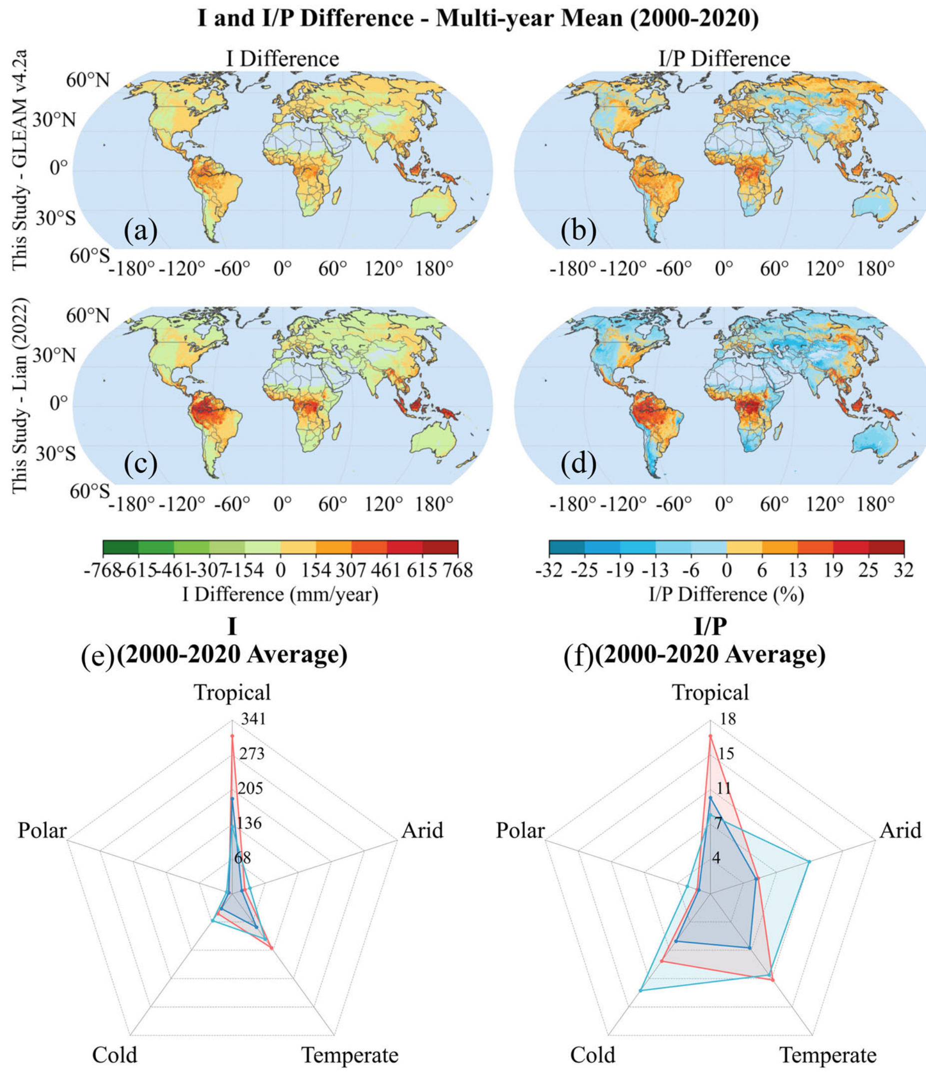

Fig. 9 | Comparison of canopy interception estimates with existing products. **a**, **b** Differences between interception loss (I) and interception loss ratio (I/P) from this study and GLEAM v4.2a for the 2000–2020 mean, calculated as This Study – GLEAM v4.2a. **c**, **d** Differences between I and I/P from this study and Lian (2022) for the 2000–2020 mean, calculated as This Study–Lian (2022). **e**, **f**, Radar plots showing the 2000–2020 mean values of the three products across different climate zones.

yields a much clearer functional relationship between canopy structure and interception. A likely explanation is that structural controls in conifer canopies are more coherent and physically organized, making them more amenable to SR representation. RF can fit local variability effectively, but its recursive partitioning does not necessarily recover transferable functional relationships. The contrast in Fig. 3 therefore suggests that the advantage of FC-DSR is not merely methodological convenience. It directly supports the broader claim that interception can be represented more faithfully when canopy structure is included explicitly and mechanistically.

The Budyko-based experiments further show that improved interception parameterization requires appropriate spatial scaling. The comparison among Schemes 1–4 in Fig. 4 indicates that FVC should act as a grid-scale modifier of interception flux rather than being embedded directly in the process-level storage parameter. This distinction is physically important. Embedding FVC in δ implicitly assumes that sub-grid vegetation cover alters the intrinsic canopy storage process itself, whereas our results indicate that FVC primarily serves as an aerial weighting factor that translates canopy-scale interception to the grid-cell scale. The degradation that occurs when FVC is introduced directly into δ, especially in mixed forests (Fig. 4c, d), highlights the risk of conflating process representation with spatial aggregation. In heterogeneous landscapes, a single process-scale parameter cannot capture the diversity of canopy architectures within a grid cell. Figure 4 therefore provides more than a technical modeling result. It offers broader guidance on how structural information should be incorporated into large-scale land-surface models.

The hydrological analyses provide an additional test of whether the new interception scheme matters beyond local interception estimates. The morphology-driven scheme improves the simulation of discharge variability, especially during hydrological transitions and high-flow periods, indicating that interception errors propagate nonlinearly into runoff generation and streamflow timing. Because interception is the first partitioning step after rainfall reaches the canopy, even moderate biases at this stage can alter effective water input to soils and river networks, thereby affecting peak flows, recession behavior, and annual runoff totals. The stronger performance of Scheme 4 in five of six major river basins (Fig. 5), together with the improvement at 71.5% of CONUS runoff stations (Fig. 6), indicates that morphology-aware interception is not simply improving point-scale fit. It is improving the realism of coupled hydrological behavior.

These hydrological gains help clarify why the proposed scheme matters for Earth system modeling more broadly. The strongest improvements are concentrated in event-responsive regions, especially the mountainous western CONUS and semi-arid transition zones in Fig. 6. This pattern suggests that interception efficiency depends not only on storage capacity, but also on how rapidly canopy storage is overwhelmed by intense rainfall and how quickly stored water is removed by aerodynamic forcing. Traditional LAI-based schemes largely interpret interception as a capacity determined by LAI, but the responses in Figs. 5–6 show that this view is incomplete. In regions where runoff is highly sensitive to rainfall partitioning, a more realistic interception formulation affects not only water balance closure, but also the timing of land atmosphere exchange and runoff

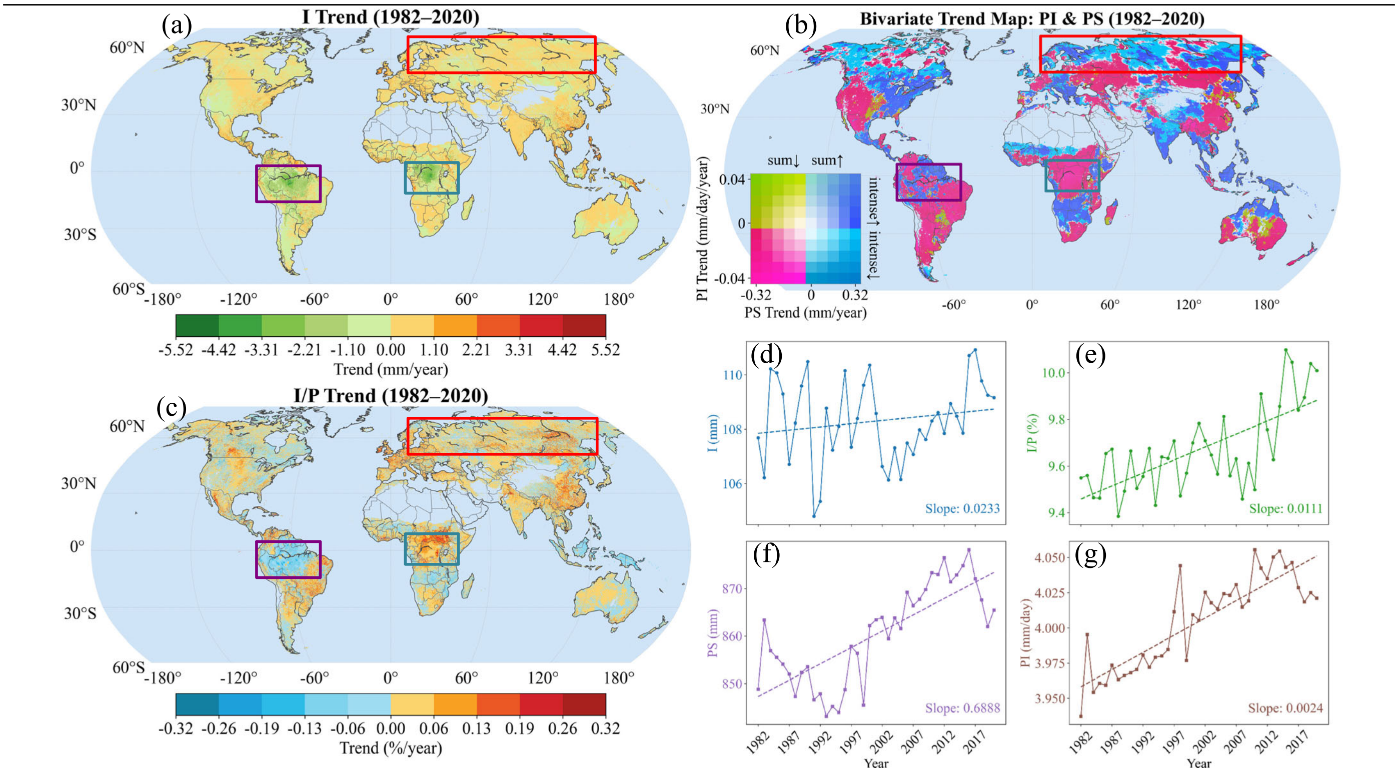

Fig. 10 | Trends in canopy interception and precipitation characteristics during 1982–2020. a, **c** Spatial distributions of trends in interception loss (I) and interception loss ratio (I/P) during 1982–2020. **b** Joint spatial distribution of trends in precipitation intensity (PI) and precipitation sum (PS) for 1982–2020. **d–g** Time series and associated trends of I, I/P, PI, and PS during 1982–2020, respectively. To avoid visual misinterpretation arising from differences in magnitude and units, each variable is presented in a separate panel.

delivery. The value of incorporating three-dimensional canopy structure, therefore, extends beyond interception itself to the broader terrestrial water cycle.

The global analyses reveal another important implication by showing that canopy morphology changes the inferred large-scale geography of interception. The differences relative to existing products in Fig. 9, together with the global spatial patterns in Figs. 7–8, show that our estimates yield substantially higher I/P in tropical regions while maintaining broadly consistent patterns of absolute interception. This divergence is scientifically important because it suggests that current large-scale datasets may underestimate the role of complex, multilayered tropical canopies in regulating precipitation partitioning. In our framework, large CW and CD allow dense tropical forests to sustain high storage capacity and high interception efficiency, which LAI-based schemes cannot capture adequately. By contrast, the high interception efficiency found in temperate and boreal coniferous forests, evident from the combined structural and interception patterns in Figs. 7–8, appears to arise from a different pathway, namely deep and vertically organized canopies that maximize rainfall contact time. Similar hydrological outcomes can therefore emerge from different structural configurations, reinforcing the need to represent canopy morphology explicitly rather than relying on a single vegetation proxy.

The long-term trend analysis further shows that interception responses to global change are nonlinear and regionally differentiated. The spatial and temporal trends in Fig. 10 indicate that in northern Eurasia, increases in both I and I/P are consistent with the combined effects of regional wetting and vegetation greening, suggesting that climate and vegetation are acting in the same direction to strengthen canopy regulation of rainfall. In contrast, tropical declines in I and I/P, despite variable rainfall trends, indicate that changes in canopy structure, likely associated with degradation or climate stress, can offset or even outweigh the hydrological effects of precipitation change. The Congo example in Fig. 10 is especially informative because it shows that total interception and interception efficiency can respond differently when rainfall amount and rainfall intensity evolve in different ways. These results suggest that future changes in terrestrial water balance cannot be inferred reliably from precipitation or LAI trends alone. Changes in 3D canopy architecture must also be considered.

Despite the advances provided by the morphology-driven framework, several limitations point to important directions for future work. First, the compiled observational dataset (Table S6) is not fully consistent in definition, with interception reported on either a crown-projected-area basis or a ground-area basis. Because canopy cover fraction and related structural information are rarely reported, these observations cannot be harmonized unambiguously. The parameterization developed here targets grid-scale, area-averaged interception loss, so the observations are interpreted as approximate constraints at that scale rather than as strictly equivalent measurements. Previous large-scale studies[44,54,55] have shown that such inconsistencies do not preclude meaningful evaluation in a grid-based framework. In addition, because FC-DSR is designed to identify dominant relationships across large samples rather than site-specific variability, this source of uncertainty is more likely to increase random scatter than to introduce systematic bias at the grid scale. Future work should improve dataset consistency by incorporating more complete canopy structural metadata.

Second, interception is fundamentally an event-scale process, whereas our parameterization is developed and evaluated at the monthly scale. This choice was necessary because event-based observations remain too sparse and uneven globally, and because global application would otherwise require process-level information, such as antecedent canopy wetness and within-event evaporation, that is not yet available at large scales. Monthly aggregation is therefore a practical compromise that supports global consistency, but it inevitably smooths sub-event dynamics such as canopy saturation timing and short-lived evaporation pulses. Third, the dynamic land-surface evaluation was limited to CONUS. Although this domain spans strong gradients in climate, vegetation, and hydrological regimes, as reflected in Fig. 6, it does not encompass the full diversity of tropical, boreal, and arid ecosystems. Broader evaluation is still needed.

A further limitation concerns the treatment of wind. Our parameterization (Table 2) captures an overall negative relationship between wind and interception efficiency[56], which is physically consistent with droplet removal and reduced canopy water residence time[2], and this dominant effect is supported by the SHAP dependence patterns in Fig. 2. However, wind can also enhance wet-canopy evaporation under specific hydroclimatic conditions, especially during low-intensity, long-duration rainfall[41]. Because these effects are conditional and nonlinear, they are simplified in the present monthly-scale formulation. This simplification may help explain the weaker performance in the Mississippi basin during some low-flow periods, visible in Fig. 5, where land-cover heterogeneity, seasonal vegetation change, and variable rainfall regimes[57] likely produce more context-dependent wind effects. Future work should therefore explore more flexible formulations in which the role of wind varies with rainfall regime, vegetation condition, and hydroclimatic background rather than remaining fixed as a net negative control.

Overall, this study suggests that interception parameterization in ESMs should move beyond LAI-centered simplifications toward morphology-aware and process-consistent formulations. The consistency of the evidence across attribution analysis, equation discovery, hydrological evaluation, coupled modeling, and global diagnostics indicates that 3D canopy architecture is not a secondary refinement, but a first-order control on rainfall partitioning. By combining feature attribution with SR, FC-DSR provides a practical route for deriving explicit equations that are both interpretable and transferable. More broadly, this framework may be useful for other poorly constrained land-surface processes in which structural heterogeneity and environmental forcing interact nonlinearly.

Methods

Development of a new canopy interception scheme

I/P data assembly. To develop and evaluate our canopy interception scheme, we assembled a global, monthly dataset of I/P from peer-reviewed studies. We searched Google Scholar, Web of Science, and the CNKI database (https://oversea.cnki.net) using the keywords "interception loss", "canopy interception", "rainfall interception", and "rainfall partitioning". Studies were included if (1) rainfall was partitioned into throughfall, I, and stemflow, with I reported directly or computed as a residual when necessary; (2) event-scale measurements were aggregated by calendar month to estimate monthly I and I/P; (3) records longer than 15 days but shorter than a full month were treated as monthly values, whereas shorter records were excluded; and (4) If multiple observations of the same type were available for a site, we averaged them prior to comparison. Applying these criteria yielded 1040 monthly I/P observations from 323 publications covering the period from 1982 to 2020 (Supplementary Table S6). To ensure consistency with the forcing data, we retained only records for which the ratio of observed rainfall to MSWEP grid rainfall fell between 0.7 and 1.3. The spatial distribution of study sites is shown in Supplementary Fig. S9, and this curated dataset provides the empirical basis for testing and validating the proposed canopy interception parameterization.

Ancillary data and preprocessing. TH was taken from the <1 m global product of Tolan et al.[58], derived from satellite imagery and machine learning. Canopy structure variables (CW, CD, and DW) were obtained from the 500 m global dataset of Xiang et al.[59]. LAI followed Cao et al.[60], who fused PKU NDVI with Landsat-based LAI samples. We then calculated the height-to-width ratio as $HW = \mathrm{TH}/CW$. Precipitation was taken from MSWEP v2.8[61], which combines gauge, satellite, and reanalysis data. Meteorological variables were obtained from ERA5[62], including air temperature (Tem, K), horizontal wind components (u, v, m s^{-1}), specific humidity (q, kg kg^{-1}), and surface pressure (sp, Pa). Wind was calculated as $Wind = \sqrt{u^2 + v^2}$, vapor pressure as $e = \frac{q \times (sp/100)}{0.378 \times q + 0.622}$, saturation vapor pressure using the Magnus–Tetens equation, and relative humidity as $RH = e/e_s \times 100\%$. FVC was derived from LAI[63] (−) as $FVC = 1 - \exp(-0.5 \times LAI)$. For FC-DSR, air temperature was converted from Kelvin to degrees Celsius. All source and derived variables were resampled to a monthly 0.1° grid to ensure analytical consistency.

FC-DSR framework. We developed FC-DSR to generate δ parameterizations that are both accurate and physically interpretable (Supplementary Figs. S10 and S11). FC-DSR extends DSR[64](see Supplementary Text S2–4) by combining the nonlinear flexibility of SR with explicit process-based constraints.

The target variable I/P had a strongly skewed distribution, with most values between 10 and 20% and a small number of much higher values. To reduce bias, we applied random over-sampling(ROS)[65] to the training set so that rare, high I/P conditions were adequately represented. The same balanced dataset was used to train both the RF baseline and FC-DSR, which allowed a fair comparison of model performance. For interpretability, RF models were fitted separately by vegetation type. We then applied SHAP analysis[37] to rank predictor importance, and used these rankings as priors in the SR step, increasing the selection probability of influential variables and focusing the search on physically meaningful terms.

During expression generation, we constrained the symbolic search to preserve process meaning and limit complexity. Distinct symbol libraries were assigned to different subtrees, tree height was capped, and key canopy interception operators were required. Candidate δ expressions were inserted into a modified Budyko framework and further calibrated with gradient-based optimization. To maintain physical realism, we added a boundary penalty term to the loss function so that solutions respected known limits while still maximizing predictive skill.

By combining systematic oversampling, SHAP-informed feature priors, constrained symbolic search, and physically bounded optimization, FC-DSR yields δ parameterizations that are reproducible, extensible, and mechanistically interpretable.

a. Random over-sampling. The compiled I/P distribution is strongly imbalanced. Most records fall between 10 and 20%, whereas values above 50% are rare. If used directly for training, this distribution pushes models to fit the majority range, which reduces global error but underrepresents high I/P conditions. To reduce this bias, we applied a model agnostic ROS procedure[65]. The target I/P values were divided into ten equal width bins, and we denoted the largest bin count as $h_{\max}$. For every other bin, existing observations were sampled with replacement until each bin contained $h_{\max}$ samples. The resulting training set represents the full I/P spectrum more evenly and allows high I/P conditions to contribute proportionally to parameter estimation. Oversampling repeats observations and can increase the risk of overfitting, but it is simple, efficient, and performs well in regression. In our case, ROS enabled FC-DSR to capture canopy interception behavior across the entire observed range.

b. SHAP-guided feature prioritization. DSR[64] explores a very large expression space. To steer this search toward physically meaningful predictors, we used SHAP to quantify each variable's contribution to I/P and to bias feature selection in FC-DSR. For a given sample, the SHAP value $\phi(x)$ indicates how much a predictor increases or decreases the predicted I/P relative to the model's average prediction, with positive (negative) values representing an increase (decrease). The RF baseline and FC-DSR shared the same train and validation splits. RF hyperparameters were tuned with five-fold cross-validation over a grid with $n_{estimators} \in \{100, 200, 300, 400\}$ and max_$depth \in \{10, 20, 30, \mathrm{None}\}$. We then applied TreeExplainer[37] to the best RF model, using 200 training samples as the background, and computed SHAP values on the test data. SHAP provides additive, interaction aware attributions and handles correlated predictors[66–68].

Predictors were grouped into four physically motivated sets $X_F^{(j)}\,(j = 1, 2, 3, 4)$ (see Fig. S10a). Within each group, we defined the important subset

$$X_{imp}^{(j)} = \left(x \in X_F^{(j)} \mid \varnothing(x) > \min\left(\varnothing\left(X_F^{(j)} \right) \right) or |X_F^{(j)}| = 1 \right) \quad (2)$$

where $\varnothing(x)$ is the SHAP value of predictor x. Variables in $X_{imp}^{(j)}$ were assigned higher inclusion probabilities during symbolic regression. This design narrows the symbolic search to influential, physically relevant terms while preserving interaction information that conventional importance metrics often miss. The SHAP informed prior therefore links a strong RF baseline to a targeted symbolic search, improving computational efficiency and interpretability without compromising physical relevance.

c. Expression generation via FC-DSR. We guide the symbolic search toward physically meaningful predictors by up-weighting the SHAP identified feature sets $X_{imp}^{(j)}\,(j = 1, 2, 3, 4)$. This design raises the sampling probability of variables that have a demonstrated influence on I/P and strengthens the physical relevance of the learned C_{max} expressions.

FC-DSR extends the standard DSR framework[64], where a recurrent neural network (RNN) samples symbols sequentially to grow expression trees (Supplementary Fig. S10d and Supplementary Text S3). Each RNN cell contains a prior module and a sampling module. The prior combines the original DSR *NestedFunctions* library with a custom *CanopyInterception* module that groups predictors by physical category and adjusts selection probabilities according to node depth and position within the tree. To keep the resulting formulas interpretable and computationally efficient, we enforce three global constraints: we cap function nesting depth, restrict variables by physical group, and weight symbol choices by depth and position to encourage concise, process-consistent structures.

Concrete rules

1. **Function nesting**.
2. Exponential nesting depth is limited to one (through *NestedFunctions*). This cap avoids long stacks of exponentials that are difficult to interpret and numerically unstable.
3. **Masked libraries**. Predictors are partitioned into four physically motivated sets $X_F^{(j)}$ (Supplementary Fig. S11, right). Each subtree carries a mask that restricts variable selection to its assigned set (Supplementary Fig. S10b). This constraint keeps sub-expressions internally consistent in terms of process representation.
4. **Node selection rules**. At the root and first layer (*current step* < 2, or when $Occ == 1$ and $\_dangling == 1$), only $\{+, -, \times, /\}$ are available. From the second layer onward, the library expands to $(+, -, \times, /, n^2, \exp, \log, const, X_{Fj}), j \in \{1, 2, 3, 4\}$, where n^2 denotes a squared term and *const* denotes numeric constants.
5. **Position-dependent weighting**. At step i, we apply a Gaussian weight $w_i = \exp(-\left(i - step_{loc}\right)^2 / 2 \times \sigma^2)$, where $step_{loc} = len\left(S_i\right) - 1$ is the current subtree length and σ is the Gaussian scale. If $i < step_{loc}$, candidate symbol probabilities are multiplied by w_i, which favors compact and coherent subtrees.
6. **Prioritizing important variables**. After applying the Gaussian profile, all variables in $X_{imp}^{(j)}$ receive the maximum inclusion probability. SHAP identified predictors are therefore strongly favored during expression generation.
7. **Depth control**. Tree height is capped at d_{max}. Once this limit is reached, probabilities of nonterminal symbols are set to zero so that only terminal symbols can be sampled.

These constraints work together with the default DSR rules to define the final sampling mechanism. At each training step, the next token in every batch is drawn from the joint multinomial distribution over symbols, which produces candidate δ expressions that are compact, physically consistent, and focused on the most influential predictors.

d. Embedding candidate expressions in the physical model and optimization. Each candidate δ expression is embedded in an improved Budyko formulation to simulate I/P, $I/P = \left(1 - (1 - \delta)e^{-\delta}\right) \times FVC \times 100\%$, with $\delta = C_{max}^{(i)}/PI$. Here, i denotes the current training epoch of the FC-DSR framework, and $C_{max}^{(i)}$ represents the optimal C_{max} parameterization obtained at that epoch. Model error is quantified by the normalized RMSE, $NRMSE = RMSE/\sigma$, with $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\left(\frac{I}{P}\right)_{sim,i} - \left(\frac{I}{P}\right)_{obs,i}\right)^2}$, and σ the standard deviation of the observed I/P. We convert error to a reward using $RMSE = 1/(1 + NRMSE)$, so lower error corresponds to higher reward. To enforce physical plausibility, we impose a hard constraint: if any evaluation point yields $C_{max}^{(i)} > PI$ (storage exceeding precipitation), the RMSE is multiplied by a penalty factor of ten before computing NRMSE and the corresponding reward.

Model parameters are updated using a cross-entropy objective combined with risk seeking policy gradients. At each iteration, the RNN samples a batch of candidate expressions, embeds them in the Budyko formulation, evaluates their performance, and assigns rewards. The best performing "elite" expressions form a reference set. We then compute a cross entropy loss between the RNN's symbol probability distribution and the empirical symbol distribution in this elite set, and minimize this loss with the Adam optimizer[69], which biases generation toward high reward structures. In parallel, a risk seeking policy gradient update (see Supplementary Text S4) further shifts sampling toward promising regions of expression space. Repeating these steps concentrates the generator on compact, physically plausible C_{max} formulas that improve predictive accuracy.

In terms of computational efficiency, SR is relatively expensive during training. Under our configuration, with a batch size of 10^4 and both expression search and constant optimization included, each epoch required about 10 min. This cost mainly arises from evaluating a large number of candidate expressions and optimizing their free parameters. By contrast, inference involves only simple algebraic operations and is therefore computationally negligible.

Budyko-based evaluation

This framework was chosen to match the objective of this study, which emphasizes global interception parameterization rather than event-scale dynamics. Event-based models such as Rutter[70] and Gash[13] explicitly simulate individual rainfall events and depend on high-frequency forcing and event-scale parameterization. In contrast, the Budyko-type formulation offers a physically consistent representation of long-term water balance and is directly suited to monthly-scale analysis. We evaluated four interception parameterizations (Schemes 1–4) within a Budyko-type water balance framework that partitions precipitation while conserving mass. At the monthly timescale, canopy storage is transient and does not accumulate, consistent with our observations. Precipitation P is partitioned into transpiration E_T, surface evaporation E_S, I, and runoff Q:

$$P = E_T + E_S + I + Q \quad (3)$$

Interception is represented with a storage-based formulation,

$$\frac{I}{P} = 1 - (1 - \delta)e^{-\delta}, \delta = \frac{C_{max}}{PI} \quad (4)$$

This compact, process-based structure provides a clean testbed for isolating how alternative C_{max} formulations affect rainfall partitioning while all other process assumptions remain fixed (derivations in Supplementary Text S5).

The four schemes differ in how FVC enters the model and in how C_{max} is defined. FVC either acts at the process level through δ or is applied at the grid scale as a modifier of I/P. $C_{\max}$ follows either the empirical baseline or the FC-DSR morphology law. Full equations, coefficients, and vegetation groupings are given in Methods and Supplementary Tables S2–S3.

For ET validation, we compared modeled ET with eddy-covariance ET at 103 flux-tower sites (FLUXNET2015, OzFlux, La Thuile). Performance at each site was evaluated using KGESS (site metadata in Supplementary Table S4). Because the model follows IGBP land-cover classes, we harmonized vegetation types as follows: evergreen and deciduous needleleaf classes were merged into "Needleleaf," evergreen and deciduous broadleaf into "Broadleaf," closed and open shrublands into "Shrubs," and mixed forest into "Mixed." In Schemes 2 and 4, the baseline parameterization was replaced by the DSR-derived parameterization for needleleaf, broadleaf, mixed forests, and shrubs. For mixed forests, we used the arithmetic mean of the needleleaf and broadleaf values. All other vegetation types retained Zhong's observation-based parameters[44], which are widely used, including in GLEAM[71].

For runoff evaluation, we assessed both annual and cumulative behavior. Basin boundaries for each gauge were taken from GSHA[45] vector files, and area-weighted runoff was computed from the overlap between model grid cells and basin polygons. Only years with complete observations (at least 365 days) were retained. Annual runoff in millimeters, Q_{mm}, was converted to discharge Q_{sim} ($\mathrm{m}^3\ \mathrm{s}^{-1}$) as

$$Q_{sim} = Q_{mm} \times A_{basin} \times 10^3 / seconds \tag{5}$$

where A_{basin} is the basin area (km^2) and *seconds* is the number of seconds per year. Modeled and observed discharge were then matched year by year to evaluate Schemes 3 and 4 and to diagnose long-term water-balance differences.

To test the learned formulas directly, we substituted the FC-DSR δ expressions for needleleaf, broadleaf, mixed forest (approximated as the average of needleleaf and broadleaf), and shrub into the interception component and included FVC explicitly, while holding all other Budyko terms fixed to isolate the effect of δ on partitioning. Forcing and ancillary inputs matched the main analysis: precipitation from MSWEP v2.8; LAI from GIMMS LAI4g; vegetation cover from MODIS MCD12C1 v6.1[72]; soil hydraulic parameters from Rosetta[73] driven by SoilGrids2.0[74]; effective rooting depth from Yang et al.[75]; reference ET from FLUXNET2015 (Table S4), and streamflow from GSHA[45].

CoLM evaluation

We conducted process-level experiments with the CoLM coupled to CaMa-Flood (CoLM2024)[76] to simulate exchanges of water and energy between the land surface and the atmosphere, as well as coupled carbon and nitrogen cycling. Because of its modular structure and extensively tested process representations, CoLM is widely used as a land-surface component in climate and ecohydrological studies[77,78].

Simulations covered the contiguous United States (−125° to −66°E, 22° to 54°N) on a 0.1° daily grid for 1998–2010, with evaluation focused on 2001–2010. To better represent runoff in mountainous regions, we replaced the default SIMTOP runoff scheme with XinAnJiang scheme and kept all other settings at their documented defaults (see CoLM documentation for repository details). We compared two configurations: (i) the standard CoLM setup and (ii) CoLM$_{\mathrm{A}}$, which modifies only the canopy interception parameterization by replacing the default $0.1(LAI + SAI)$ with the FC-DSR-based $C_{\max}$ (Supplementary Text S1). All other model components and forcings were identical, so differences between the two configurations reflect only the new $C_{\max}$ formulation.

Outputs from both runs were evaluated using the OpenBench platform[46], and results were re-plotted to enable direct comparison. Reference data included runoff observations from the GRDC and streamflow records from the GRFR[47].

Evaluation metrics

FC-DSR experiments were run on a Linux server with two AMD EPYC 7C13 64-core CPUs (2.82 GHz, 256 threads in total), 1 TB of RAM, and three NVIDIA RTX A5000 GPUs (24 GB each). For both station-based, SHAP-based, and Budyko-based analyses, we evaluated predictions x_i against observations y_i using the Pearson correlation r, R^2, *Bias*, *RMSE*, *and KGESS* Let $\bar{x}$ and $\bar{y}$ denote the sample means and N the sample size.

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - y)^2} \tag{7}$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{9}$$

$$KGESS = \frac{KGE + 0.41}{1 + 0.41} \tag{10}$$

where

$$KGE = 1 - \sqrt{(R-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2},\ \alpha = \frac{\sigma_x}{\sigma_y},\ \beta = \frac{\bar{x}}{\bar{y}} \tag{11}$$

Here, σ_x and σ_y are the standard deviations of predictions and observations. In the KGE metric, r is the correlation between predictions and observations, α measures relative variability, and β measures the mean bias ratio.

To evaluate CoLM, which generates continuous time series, we adopted $nRMSEScore(g)$, defined as:

$$nRMSEScore(g) = \exp\left(-\frac{CRMSE(g)}{CRMS(g)}\right) \tag{12}$$

where

$$CRMS(g) = \sqrt{\frac{\int_{t_0}^{t_f}\left(v_{ref}(t,g) - \bar{v}_{ref}(g)\right)^2 dt}{t_f - t_0}} \tag{13}$$

$$CRMSE(g) = \sqrt{\frac{\int_{t_0}^{t_f}((v_{sim}(t,g) - \bar{v}_{sim}(g)) - (v_{ref}(t,g) - \bar{v}_{ref}(g)))^2 dt}{t_f - t_0}} \tag{14}$$

Here, g denotes the grid cell, and t denotes time, with t_0 and t_f represent the beginning and end of the evaluation period, respectively. $v_{\mathrm{sim}}(t,g)$ and $v_{\mathrm{ref}}(t,g)$ are the simulated and reference values at time t and grid cell g, and the overbar $(\bar{\cdot})$ denotes the temporal mean.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset is available at https://doi.org/10.5281/zenodo.17732373. The crown structure data are available at https://doi.org/10.5281/zenodo.16940138. The LAI4g data are available at https://doi.org/10.5281/zenodo.

7649107. The MSWEP v2.8 data are available at https://www.gloh2o.org/. The ERA5 data are available at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5. The vegetation cover data are available at https://www.earthdata.nasa.gov/data/catalog/lpcloud-mcd12c1-061. The SoilGrids 2.0 data are available at https://files.isric.org/soilgrids/latest/. The FLUXNET2015 data are available at https://files.isric.org/soilgrids/latest/. The GRDC data are available at https://www.bafg.de/GRDC. The GRFR data are available at https://www.reachhydro.org/home/records/grfr. The GSHA data are available at https://zenodo.org/records/8090704.

Code availability

The code is available at https://github.com/2023ATAI/FC-DSR.

Received: 13 January 2026; Accepted: 20 May 2026;
Published online: 29 May 2026

References

1. Murray, S. J. Trends in 20th century global rainfall interception as simulated by a dynamic global vegetation model: implications for global water resources. *Ecohydrology* **7**, 102–114 (2012).
2. Lian, X., Zhao, W. & Gentine, P. Recent global decline in rainfall interception loss due to altered rainfall regimes. *Nat. Commun.* **13**, 7642 (2022).
3. Shao, R., Shao, W., Gu, C. & Zhang, B. Increased interception induced by vegetation restoration counters ecosystem carbon and water exchange efficiency in China. *Earth's Future* **10** https://doi.org/10.1029/2021ef002464 (2022).
4. Yue, K. et al. Global patterns and drivers of rainfall partitioning by trees and shrubs. *Glob. Change Biol.* **27**, 3350–3357 (2021).
5. Davies-Barnard, T., Valdes, P. J., Jones, C. D. & Singarayer, J. S. Sensitivity of a coupled climate model to canopy interception capacity. *Clim. Dyn.* **42**, 1715–1732 (2014).
6. Astuti, H. P. & Suryatmojo, H. Water in the forest: rain-vegetation interaction to estimate canopy interception in a tropical Borneo rainforest. *IOP Conf. Ser. Earth Environ. Sci.* **361**, 012035 (2019).
7. Hassan, S. M. T., Ghimire, C. P. & Lubczynski, M. W. Remote sensing upscaling of interception loss from isolated oaks: Sardon catchment case study, Spain. *J. Hydrol.* **555**, 489–505 (2017).
8. Schellekens, J., Bruijnzeel, L. A., Scatena, F. N., Bink, N. J. & Holwerda, F. Evaporation from a tropical rain forest, Luquillo Experimental Forest, eastern Puerto Rico. *Water Resour. Res.* **36**, 2183–2196 (2000).
9. van der Ent, R. J., Wang-Erlandsson, L., Keys, P. W. & Savenije, H. H. G. Contrasting roles of interception and transpiration in the hydrological cycle – Part 2: moisture recycling. *Earth Syst. Dyn.* **5**, 471–489 (2014).
10. Wang-Erlandsson, L., van der Ent, R. J., Gordon, L. J. & Savenije, H. H. G. Contrasting roles of interception and transpiration in the hydrological cycle – Part 1: temporal characteristics over land. *Earth Syst. Dyn.* **5**, 441–469 (2014).
11. Brandt, J. The transformation of rainfall energy by a tropical rain forest canopy in relation to soil erosion. *J. Biogeogr.* **15**, 41–48 (1988).
12. Ma, M. et al. Rainfall intensities determine accuracy of canopy interception simulation using the Revised Gash model. *Agric. Forest Meteorol.* **362** https://doi.org/10.1016/j.agrformet.2025.110389 (2025).
13. Gash, J. H. C., Lloyd, C. R. & Lachaud, G. Estimating sparse forest rainfall interception with an analytical model. *J. Hydrol.* **170** https://doi.org/10.1016/0022-1694(95)02697-N (1995).
14. Good, S. P., Moore, G. W. & Miralles, D. G. A mesic maximum in biological water use demarcates biome sensitivity to aridity shifts. *Nat. Ecol. Evol.* **1**, 1883–1888 (2017).
15. Dai, Y., Dickinson, R. E. & Wang, Y.-P. A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. *J. Clim.* **17**, 2281–2299 (2004).
16. Lawrence, D. M. et al. The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *J. Adv. Model. Earth Syst.* **11**, 4245–4287 (2019).
17. Oleson, K. et al. Technical description of version 4.0 of the Community Land Model (CLM). *NCAR Technical Note NCAR/TN-478+STR, 257 pp* (2010).
18. Wiltshire, A. J. et al. JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2. *Geosci. Model Dev.* **13**, 483–505 (2020).
19. Yoshimura, K., Miyazaki, S., Kanae, S. & Oki, T. Iso-MATSIRO, a land surface model that incorporates stable water isotopes. *Glob. Planet. Change* **51**, 90–107 (2006).
20. Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R. & Mao, Y. The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility. *Geosci. Model Dev.* **11**, 3481–3496 (2018).
21. He, C. et al. Modernizing the open-source community Noah with multi-parameterization options (Noah-MP) land surface model (version 5.0) with enhanced modularity, interoperability, and applicability. *Geosci. Model Dev.* **16**, 5131–5151 (2023).
22. Roth, B. E., Slatton, K. C. & Cohen, M. J. On the potential for high-resolution lidar to improve rainfall interception estimates in forest ecosystems. *Front. Ecol. Environ.* **5**, 421–428 (2007).
23. Fischer-Bedtke, C., Metzger, J. C., Demir, G., Wutzler, T. & Hildebrandt, A. Throughfall spatial patterns translate into spatial patterns of soil moisture dynamics – empirical evidence. *Hydrol. Earth Syst. Sci.* **27**, 2899–2918 (2023).
24. Nanko, K., Levia, D. F., Iida, S. I., Shinohara, Y. & Sakai, N. Machine learning reveals the contrasting roles of rainfall and canopy structure metrics on the formation of canopy drip and splash throughfall. *J. Geophys. Res. Biogeosci.* **130** https://doi.org/10.1029/2024jg008340 (2025).
25. Li, Z. & Tian, F. Derivation and validation of a theoretical canopy interception model based on raindrop microphysical processes. *Water Resour. Res.* **61** https://doi.org/10.1029/2024wr038296 (2025).
26. Kochkov, D. et al. Neural general circulation models for weather and climate. *Nature* **632**, 1060–1066 (2024).
27. Eyring, V., Gentine, P., Camps-Valls, G., Lawrence, D. M. & Reichstein, M. AI-empowered next-generation multiscale climate modelling for mitigation and adaptation. *Nat. Geosci.* **17**, 963–971 (2024).
28. Nearing, G. et al. Global prediction of extreme floods in ungauged watersheds. *Nature* **627**, 559–563 (2024).
29. Udrescu, S.-M. & Tegmark, M. AI Feynman: a physics-inspired method for symbolic regression. *Sci. Adv.* **6** https://doi.org/10.1126/sciadv.aay2631 (2020).
30. Makke, N. & Chawla, S. Interpretable scientific discovery with symbolic regression: a review. *Artif. Intell. Rev.* **57** https://doi.org/10.1007/s10462-023-10622-0 (2024).
31. Chen, M. et al. Iterative integration of deep learning in hybrid Earth surface system modelling. *Nat. Rev. Earth Environ.* **4**, 568–581 (2023).
32. Tenachi, W., Ibata, R. & Diakogiannis, F. I. Deep symbolic regression for physics guided by units constraints: toward the automated discovery of physical laws. *Astrophys. J.* **959** https://doi.org/10.3847/1538-4357/ad014c (2023).
33. Zhang, M., Kim, S., Lu, P. Y. & Soljacic, M. Deep learning and symbolic regression for discovering parametric equations. *IEEE Trans. Neural Netw. Learn Syst.* **35**, 16775–16787 (2024).
34. Hafner, D., Gemmrich, J. & Jochum, M. Machine-guided discovery of a real-world rogue wave model. *Proc. Natl. Acad. Sci. USA*. **120**, e2306275120 (2023).
35. Li, Q. et al. Advancing symbolic regression for earth science with a focus on evapotranspiration modeling. *npj Clim. Atmos. Sci.* **7** https://doi.org/10.1038/s41612-024-00861-5 (2024).

36. Yamazaki, D., de Almeida, G. A. M. & Bates, P. D. Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resour. Res.* **49**, 7221–7235 (2013).
37. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
38. Grunicke, S., Queck, R. & Bernhofer, C. Long-term investigation of forest canopy rainfall interception for a spruce stand. *Agric. Forest. Meteorol.* **292-293** https://doi.org/10.1016/j.agrformet.2020.108125 (2020).
39. Herbst, M., Rosier, P. T. W., McNeil, D. D., Harding, R. J. & Gowing, D. J. Seasonal variability of interception evaporation from the canopy of a mixed deciduous forest. *Agric. For. Meteorol.* **148**, 1655–1667 (2008).
40. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008).
41. Murakami, S. A proposal for a new forest canopy interception mechanism: Splash droplet evaporation. *J. Hydrol.* **319**, 72–82 (2006).
42. Spracklen, D. V., Arnold, S. R. & Taylor, C. M. Observations of increased tropical rainfall preceded by air passage over forests. *Nature* **489**, 282–285 (2012).
43. Li, Z., Tian, F., Wang, D. & Peng, Z. A stochastic simulation method for estimating vegetation interception capacity based on mechanical-geometric analysis. *Water Resour. Res.* **61** https://doi.org/10.1029/2025wr040267 (2025).
44. Zhong, F. et al. Revisiting large-scale interception patterns constrained by a synthesis of global experimental data. *Hydrol. Earth Syst. Sci.* **26**, 5647–5667 (2022).
45. Yin, Z. et al. A synthesis of Global Streamflow Characteristics, Hydrometeorology, and Catchment Attributes (GSHA) for large sample river-centric studies. *Earth Syst. Sci. Data* **16**, 1559–1587 (2024).
46. Wei, Z. et al. OpenBench: a land model evaluation system. *Geosci. Model Dev.* **18**, 6517–6540 (2025).
47. Yang, Y. et al. Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bull. Am. Meteorol. Soc.* **102**, E2086–E2105 (2021).
48. Li, Q. et al. Three-dimensional canopy morphology and wind dynamics govern global rainfall interception. *Zenodo* https://doi.org/10.5281/zenodo.17732373 (2025).
49. Miralles, D. G. et al. GLEAM4: global land evaporation and soil moisture dataset at 0.1 resolution from 1980 to near present. *Sci. Data* **12**, 416 (2025).
50. Porada, P., Van Stan, J. T. & Kleidon, A. Significant contribution of non-vascular vegetation to global rainfall interception. *Nat. Geosci.* **11**, 563–567 (2018).
51. Flores, B. M. et al. Critical transitions in the Amazon forest system. *Nature* **626**, 555–564 (2024).
52. Liu, Y. et al. Global greening drives significant soil moisture loss. *Commun. Earth Environ.* **6** https://doi.org/10.1038/s43247-025-02470-3 (2025).
53. Park, A. & Cameron, J. L. The influence of canopy traits on throughfall and stemflow in five tropical trees growing in a Panamanian plantation. *For. Ecol. Manag.* **255**, 1915–1925 (2008).
54. Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M. & Dolman, A. J. Global canopy interception from satellite observations. *J. Geophys. Res. Atmos.* **115** https://doi.org/10.1029/2009jd013530 (2010).
55. Oliveira, S., Cunha, J., Nóbrega, R. L. B., Gash, J. H. & Valente, F. Enhancing global rainfall interception loss estimation through vegetation structure modeling. *J. Hydrol.* **631** https://doi.org/10.1016/j.jhydrol.2024.130672 (2024).
56. Muzylo, A. et al. A review of rainfall interception modelling. *J. Hydrol.* **370**, 191–206 (2009).
57. Mohammed, H. & Hansen, A. T. Spatial heterogeneity of low flow hydrological alterations in response to climate and land use within the Upper Mississippi River basin. *J. Hydrol.* **632** https://doi.org/10.1016/j.jhydrol.2024.130872 (2024).
58. Tolan, J. et al. Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sens. Environ.* **300** https://doi.org/10.1016/j.rse.2023.113888 (2024).
59. Xiang, J. et al. A Global 500 m resolution tree crown structure dataset. *Zenodo* https://doi.org/10.5281/zenodo.16940138 (2025).
60. Cao, S. et al. Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020. *Earth Syst. Sci. Data* **15**, 4877–4899 (2023).
61. Beck, H. E. et al. MSWEP V2 Global 3-Hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Am. Meteorol. Soc.* **100**, 473–500 (2019).
62. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
63. Bruijnzeel, A. I. J. M.vD. L. A. Modelling rainfall interception by vegetation of variable density using an adapted analytical model. Part 1. Model description. *J. Hydrol.* **247**, 230–238 (2001).
64. Petersen, B. K. et al. Deep symbolic regression: recovering mathematical expressions from data via risk-seeking policy gradients. *International Conference on Learning Representations* (2021).
65. Branco, P., Torgo, L. & Ribeiro, R. P. Pre-processing approaches for imbalanced distributions in regression. *Neurocomputing* **343**, 76–99 (2019).
66. Liu, M. et al. Diverging responses of terrestrial ecosystems to water stress after disturbances. *Nat. Clim. Change* **15**, 73–79 (2025).
67. Maynard, D. S. et al. Global relationships in tree functional traits. *Nat. Commun.* **13**, 3185 (2022).
68. Green, J. K., Zhang, Y., Luo, X. & Keenan, T. F. Systematic underestimation of canopy conductance sensitivity to drought by Earth System Models. *AGU Advances* **5** https://doi.org/10.1029/2023av001026 (2024).
69. Kingma, D. P. & Ba, J. in *3rd International Conference for Learning Representations* (ICLR, 2015).
70. Rutter, A. J., Kershaw, K. A., Robins, P. C. & Morton, A. J. A predictive model of rainfall interception in forests, 1. Derivation of the model from observations in a plantation of Corsican pine. *Agric. Meteorol.* **9**, 367–384 (1971).
71. Martens, B. et al. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **10**, 1903–1925 (2017).
72. Friedl, M. & Sulla-Menashe, D. MODIS/Terra+aqua land cover type yearly L3 Global 0.05Deg CMG V061. *NASA Land Processes Distributed Active Archive Center* https://doi.org/10.5067/MODIS/MCD12C1.061 (2022).
73. Zhang, Y. & Schaap, M. G. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *J. Hydrol.* **547**, 39–53 (2017).
74. Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240 (2021).
75. Yang, Y., Donohue, R. J. & McVicar, T. R. Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resour. Res.* **52**, 8260–8276 (2016).
76. Dai, Y. et al. The common land model. *Bull. Am. Meteorol. Soc.* **84**, 1013–1024 (2003).
77. Bai, F. et al. Global Assessment of atmospheric forcing uncertainties in the common land model 2024 simulations. *J. Geophys. Res. Atmos.* **129** https://doi.org/10.1029/2024JD041520 (2024).
78. Yang, L. et al. Calibration of the high-resolution common land model in simulating the soil moisture over the northeastern China using an adaptive parameter learning method. *J. Geophys. Res. Atmos.* **130** https://doi.org/10.1029/2024jd043230 (2025).

Acknowledgements
We acknowledge the high-performance computing support from School of Atmospheric Science of Sun Yat-Sen University.

Author contributions
Q.L. and Z.W. conceived of the presented idea; Q.L. developed the theory with support from Z.W. and performed the computations; X.J. built and processed the data; W.S., J.Z., Z.Z., X.L., Y.Y., J.W., X.S. and Y.L. verified the analytical methods, the scientific insights, and supervised the findings of this study; Q.L. wrote the manuscript with support from Z.W. and Y.D.; Q.L, X.L. and C.Z., programmed the visualization and explanation tools. All authors discussed the results and contributed to the final manuscript.

Funding
Z.W. discloses support for the research of this work from the Guangdong Basic and Applied Basic Research Foundation [grant number 2024A1515010283] and the National Natural Science Foundation of China [grant number 42075158]. Q.L. discloses support for the research of this work from the National Natural Science Foundation of China [grant numbers 42575159, 42275155 and 42105144]. Y.D. discloses support for the research of this work from the National Natural Science Foundation of China [grant number 42088101] and the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China [grant number JYB2025XDXM902]. W.S. discloses support for the research of this work from the National Natural Science Foundation of China [grant number 42205149].

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-026-03694-7.

Correspondence and requests for materials should be addressed to Qingliang Li or Zhongwang Wei.

Peer review information *Communications Earth and Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Rodolfo Nóbrega and ChenRui Diao. A peer review file is available.