https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

Unraveling Local and Upwind Controls on Terrestrial ET-Runoff Partitioning in China: Grid-Scale Analysis and Explanatory Diagnostics

Zhiwei Luo[1], Ling Ji[1], Yulei Xie[2]

[1]School of Economics and Management, Beijing University of Technology, Beijing, 100124, China.
[2]Institute of Environmental and Ecological Engineering, Guangdong University of Technology, Guangzhou, Guangdong, 510006, China.

Correspondence to: Ling Ji (hdjiling@bjut.edu.cn); Yulei Xie (xieyulei@gdut.edu.cn)

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

Abstract. How water is divided between evapotranspiration and runoff controls regional water availability, but this partition is usually attributed mainly to local climate and land-surface conditions. Because part of precipitation is supplied by evaporation from upwind land areas, downwind partitioning may also reflect the ecohydrological state of terrestrial moisture-source regions. Here we analyze monthly evapotranspiration-runoff partitioning across mainland China from 2008 to 2017 using a geographically weighted random forest that separates local hydroclimatic and land-surface covariates from upwind land-surface covariates aggregated along atmospheric moisture trajectories. The evapotranspiration share varies non-monotonically with aridity, remaining near 0.80-0.82 from hyper-arid to dry sub-humid regions before declining to 0.62 in humid regions. Local controls dominate the explainable variability nationally, with mean cross-validated R2 increasing only from 0.468 to 0.471 when upwind covariates are added. However, the terrestrial upwind signal is geographically and seasonally concentrated, being most evident in dry sub-humid transition zones and in autumn. Upwind vegetation density, soil wetness, and precipitation show inverted-U relationships with the downwind green-water share, whereas upwind evaporative demand shows a monotone-negative relationship. These patterns persist across alternative evapotranspiration and runoff product combinations, but their interpretation remains limited by gridded product uncertainty, the terrestrial-only definition of upwind covariates, and the observational attribution design. The results suggest that assessments of blue-green water partitioning may benefit from considering upwind land-surface conditions alongside local hydroclimate, particularly in dry sub-humid transition zones.

1 Introduction

Freshwater resources are increasingly stressed by the combined effects of global climate change and intensifying human pressures (Rodell et al., 2018; Zhang et al., 2025). Rising temperatures (Ravinandrasana and Franzke, 2025), shifting precipitation patterns (Konapala et al., 2020), and rapid land use transformation (Chagas et al., 2022) are reorganizing terrestrial water cycles, while growing demands for food, energy, and ecosystem services further strain water availability. The distinction between blue water (surface runoff, groundwater recharge, and snowmelt that sustain human water use) and green water (precipitation stored in the unsaturated zone and returned to the atmosphere through evapotranspiration, ET) is widely used in ecohydrology and water sustainability science (Althoff and Destouni, 2023; Pan et al., 2023; Song et al., 2025). The partitioning of precipitation into these two fluxes governs both the resilience of ecosystems to hydroclimatic extremes and the reliability of water supply across sectors (Asselin et al., 2024; Stephens et al., 2023). Consequently, understanding how land-atmosphere interactions allocate water between ET and runoff at decision-relevant spatial and temporal scales remains a core scientific and management challenge (Te Wierik et al., 2020).

Despite this importance, methods for attributing spatial and temporal variability in blue–green water partitioning remain limited (Yu et al., 2025). Most prior studies have examined ET or runoff in isolation, typically at watershed scale and over multi-year periods (Li et al., 2025). While such work has clarified the roles of climate and land-cover change in modulating

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

individual fluxes, it treats these fluxes independently rather than as a coupled, bounded partition (Wang-Erlandsson et al., 2022). Few studies have represented blue and green water as a joint monthly grid-scale ratio that can be used to compare local and nonlocal drivers against the same response metric.

One under-examined dimension of blue-green partitioning is the role of upwind terrestrial moisture sources (Gao et al., 2025). Precipitation at a given location is not solely local in origin. A substantial fraction can originate as evaporation from distant land surfaces, including vegetation, soils, and inland water bodies, before being transported through the atmosphere to a downwind location (Tuinenburg et al., 2020b). The land-surface state of those source regions including vegetation cover, root-zone soil moisture, and surface energy balance, can affect moisture export along continental trajectories and the water available for downwind partitioning between ET and runoff (Posada-Marín et al., 2024; Wang-Erlandsson et al., 2018). Moderate vegetation density in source regions can enhance plant transpiration, increasing moisture export and potentially raising the green-water share at downwind locations. At higher vegetation density, canopies may intercept more rainfall locally and shorten moisture-recycling distances, which could reduce net export to distant sinks. Whether this produces an intermediate optimum or a monotone response at regional scales remains unresolved (Van Der Ent et al., 2014; Wunderling et al., 2022). More broadly, the extent to which upwind land-surface conditions are detectable in downwind blue-green partitioning, and under what hydroclimatic and seasonal conditions this influence is most expressed, is not known. The sensitivity of partitioning to advective moisture supply is expected to vary with local aridity and the water-energy balance, and the seasonal composition of moisture-source regions shifts substantially as large-scale atmospheric circulation evolves through the year, modulating how much of the incoming precipitation carries a detectable terrestrial land-surface signal (Fu et al., 2024). A systematic explanatory framework is therefore needed to characterize both the geographic and temporal structure of this detectability across hydroclimatic gradients.

Theoretical frameworks for water partitioning have not fully incorporated this continental perspective. The Budyko hypothesis links aridity to the long-term evaporative ratio at basin scale (Hunt et al., 2024; Peng et al., 2025; Rice and Emanuel, 2019). Its direct application to monthly grid-scale partitioning is limited because it does not account for transient soil-water storage, seasonal hysteresis, or atmospheric moisture transport between locations (Greve et al., 2015). Conventional moisture tracking approaches quantify transport pathways and volumes but largely overlook the ecohydrological state of source regions, including the productivity and moisture-export efficiency of those surfaces (Fahrländer et al., 2024; Weng et al., 2019). Without linking source-region land-surface properties to moisture export, analyses may attribute partitioning dynamics mainly to transport volume and overlook nonlinear contributions from source-region ecohydrology.

Existing methods have different limitations. Process-based hydrological models embed physical principles but require extensive parameterization and struggle to isolate the contributions of local versus remote drivers at fine spatial resolution (Sharma et al., 2023; Tan et al., 2024; Veettil and Mishra, 2016). In contrast, machine learning methods such as Random Forest (RF) effectively capture nonlinear relationships among variables but assume spatial stationarity, thereby masking geographic heterogeneity in driver–response relationships (Li et al., 2020; Liu et al., 2024; Mao et al., 2025). Geographically Weighted

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

Regression (GWR) accommodates spatial non-stationarity but is restricted to linear forms (Tong et al., 2024; Wan et al., 2024). The Geographically Weighted Random Forest (GW-RF) combines the nonlinear flexibility of Random Forest with spatially local weighting (Fayaz and Galasso, 2024), making it well suited for explanatory modeling of spatially heterogeneous ecohydrological processes (Georganos et al., 2021).

To address these gaps, we developed a grid-scale explanatory framework that explicitly separates local and continental upwind controls on monthly blue-green water partitioning across China (2008–2017). Local factors represent the hydrometeorological and land-surface state at each grid cell. Upwind factors represent the weighted land-surface conditions of terrestrial continental moisture-source regions identified from atmospheric moisture trajectories. We used incremental explanatory gain, permutation tests, counterfactual perturbations, and a water-balance proxy to assess whether upwind information improved model explanation beyond local factors and whether the inferred relationships were consistent with terrestrial water-balance expectations. Within this framework, we address three central questions: (1) What is the relative contribution of local versus upwind land-surface controls to monthly blue-green water partitioning variability across China's diverse hydroclimatic regimes? (2) How do ecohydrological conditions in upwind source regions modulate downwind water partitioning, and under what aridity and seasonal conditions is this influence most detectable? (3) Are the inferred explanatory patterns consistent with terrestrial water-balance expectations, including the role of atmospheric moisture transport from terrestrial source regions?

This study provides a spatially explicit attribution of blue-green water partitioning that links moisture-recycling information with ecohydrological modelling. The resulting framework may inform future assessments of ecological compensation, land-use planning, and transboundary water governance in water-stressed regions.

2 Materials and Methods

2.1 Study area

This study focuses on mainland China, one of the world's most hydrologically heterogeneous regions. Extending from the humid tropics in the south to the arid continental interiors in the northwest, China spans a full gradient of moisture regimes, including humid, semi-humid, semi-arid, and arid zones, and includes complex topography and land-cover transitions. Precipitation and evapotranspiration generally decrease from southeast to northwest, whereas aridity increases along the same transect. In addition, the country's complex terrain, from the high-elevation Tibetan Plateau and vast inland basins to extensive plains and coastal lowlands, creates pronounced contrasts in water availability, vegetation cover, and land-atmosphere interactions. Terrestrial moisture supply to China is anisotropic, dominated by a Eurasian mid-latitude corridor and subtropical Asian pathways, which is precisely the geography where state-sensitive upwind effects and corridor-mediated modulations are expected to be most pronounced. This spatial variability makes China a useful setting for examining blue-green water partitioning across diverse hydroclimatic regimes.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

2.2 Response variable and covariate construction

The response variable is the monthly blue-green water (BG) coefficient, defined at each grid cell as ET / (ET + Runoff). Under the main analysis line, ET is taken from GLEAM v4.2 actual evaporation (Miralles et al., 2025) and Runoff from ERA5 monthly fields (Hersbach et al., 2020) . This ratio is dimensionless and bounded between 0 and 1, allowing comparisons across hydroclimatic regimes. It represents the allocation of water between evapotranspiration and runoff. Two additional response lines were used for robustness testing: ERA5 ET + ERA5 Runoff (internal consistency baseline) and GLEAM ET + GLDAS Runoff (Beaudoing et al., 2020; maximum product independence).

Local covariates represent conditions at the target grid cell that directly regulate partitioning there: monthly precipitation (Precip_{local}), root-zone soil moisture (SMRZ_{local}), potential evapotranspiration (PET_{local}), leaf area index (LAI_{local}), and near-surface air temperature ($\mathrm{T2M}_{local}$). Two harmonic terms ($\sin_{month}$, $\cos_{month}$) encode the intra-annual seasonal cycle.

Upwind covariates represent the ecohydrological state of terrestrial moisture-source regions contributing precipitation to the target grid. We used the UTrack Lagrangian moisture-tracking dataset (Tuinenburg et al., 2020a), which provides gridded monthly fractional source–sink linkages at 0.5° resolution for 2008–2017. A land-only mask was applied: each upwind covariate (LAI_{upwind}, Precip_{upwind}, SMRZ_{upwind}, PET_{upwind}, $\mathrm{T2M}_{upwind}$) is the moisture-contribution-weighted mean of the corresponding land-surface variable across terrestrial source grid cells only. Oceanic source contributions are excluded by construction. The upwind covariates therefore capture the marginal contribution of terrestrial source-region land-surface factors, rather than total moisture convergence. This scope is retained throughout the analysis and interpretation.

2.3 Geographically Weighted Random Forest

Standard Random Forest assumes spatial stationarity. The Geographically Weighted Random Forest (Georganos et al., 2021) relaxes this assumption by fitting a locally weighted Random Forest at each target location, using spatially decaying weights assigned to neighboring observations via an adaptive bisquare kernel. The adaptive bandwidth ensures that each local model draws on a consistent number of neighbors regardless of data density, reducing sensitivity to spatial clustering. We adopted two fitting protocols. Full-sample fitting used all 120 months (2008–2017) to produce maps of R^2 and feature importance. Five-fold temporal block cross-validation divided the time series into five contiguous 24-month blocks, trained on four blocks, and validated on the held-out block at each location to reduce temporal autocorrelation and information leakage (Koldasbayeva and Zaytsev, 2025). Block cross-validated R^2 is the primary explanatory-power metric reported throughout. Full technical specifications (kernel function, bandwidth selection procedure, and cross-validation formula) are provided in Text S1.

For each target grid in China, we constructed an upwind-weighted variable set by multiplying the ecohydrological attributes of all source grids by their fractional moisture contributions and aggregating the results. This yielded upwind counterparts to the local factors, representing the integrated influence of continental inputs on local partitioning. These variables allowed the relative roles of local and upwind drivers to be compared for the blue-green water coefficient (Figure 1).

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

The correlation plots of each local and upwind factor with the blue-green water coefficient are documented in the Supporting Information (Figures S1 and S2).

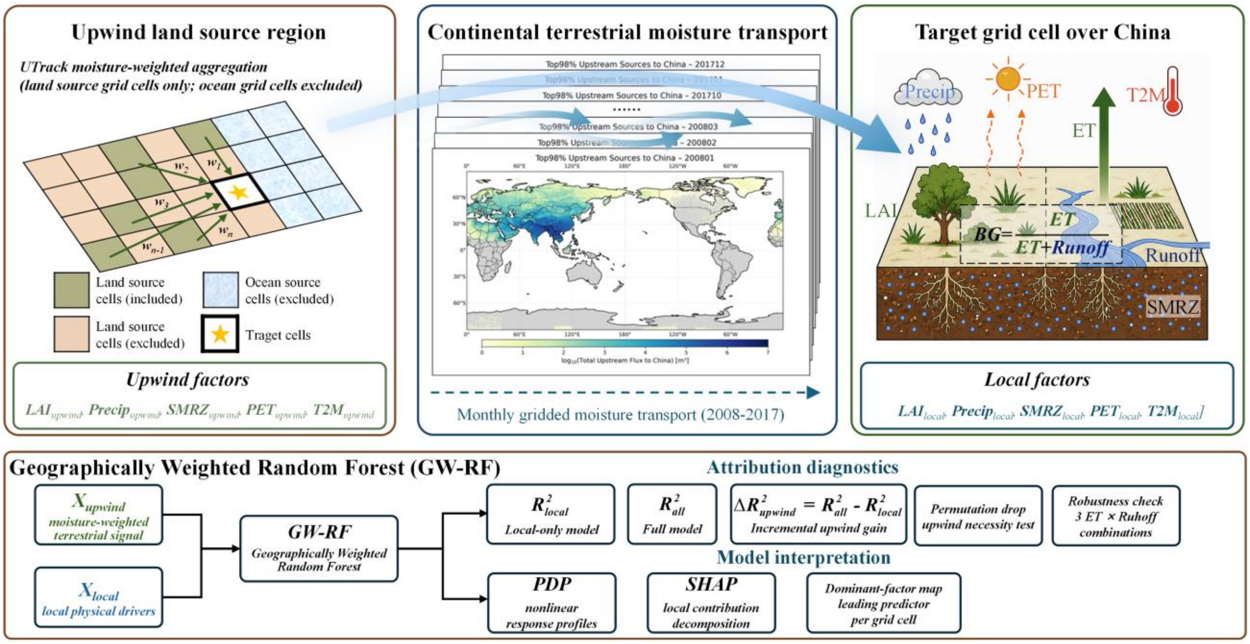

Figure 1. Conceptual framework linking local controls, continental upwind controls, and blue-green water partitioning.

2.4 Explanatory diagnostics

To move beyond descriptive fitting toward explanatory, we applied a four-step diagnostic chain to evaluate local and upwind contributions (Gupta et al., 2008). First, three nested models were fitted at each grid cell under block cross-validation: local-only (five local covariates plus harmonic terms), upwind-only (five upwind covariates plus harmonic terms), and all-factors. Incremental upwind explanatory gain was calculated as $\Delta R^2 = R^2_{all} - R^2_{local}$. Positive ΔR^2 indicates that upwind information improves explanation beyond the local-only model.

Then, permutation-based necessity tests randomly shuffled upwind covariates within year-month blocks and refitted the GW-RF. The mean reduction in cross-validated R^2 across ten permutation draws was defined as the permutation drop. Larger drops indicate greater model dependence on the upwind covariate set. The product-combination robustness analysis replicated the main diagnostics under the two alternative response definitions to test whether the explanatory pattern depends on the specific ET or runoff product chosen. The permutation drop was further disaggregated at a seasonal level: the four standard meteorological seasons (DJF: December–February; MAM: March–May; JJA: June–August; SON: September–November). The stratification was computed at each grid cell and then averaged within aridity classes.

We also compared the explanatory diagnostics with an auxiliary annual P − ET proxy constructed from CRU TS v4.09 precipitation minus ERA5 evaporation (Harris et al., 2020). This benchmark is used as a diagnostic consistency check rather

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

than as independent causal validation. Counterfactual perturbation experiments replaced upwind covariates with grid-wise median or upper-quartile values and then retrained the GW-RF. These analyses provided supplementary evidence on directional sensitivity and are described in Text S2, with spatial responses shown in Figure S7.

2.5 Model interpretation and dominant-factor derivation

Partial dependence profiles (PDPs) characterize the conditional relationship between the modeled BG coefficient and a single covariate, marginalizing over the joint distribution of all other covariates within each aridity class. Shapley Additive Explanations (SHAP) decompose each grid-level fitted value into signed additive contributions from individual covariates, mapping the direction and magnitude of each factor's influence spatially (Chen et al., 2023). PDPs were used to describe nonlinear response shapes, and SHAP was used to map where explanatory weight was concentrated. Dominant-factor maps were derived by identifying, at each grid cell, the non-cyclic covariate with the highest impurity-based feature-importance value; grid cells where the margin between the top two covariates fell below a predefined uncertainty threshold were classified as Uncertain. An entropy-based dominance index was computed from the normalized non-cyclic feature-importance vector to summarize whether explanatory weight was concentrated in one covariate or distributed across several covariates. Lower entropy indicates stronger single-factor dominance, whereas higher entropy indicates a more distributed importance profile. Full definitions are provided in Text S1.

2.6 Data sources

All variables were harmonized to a common 0.5° monthly grid for 2008–2017, covering 4,205 valid grid cells within mainland China. GLEAM v4.2 provided actual ET and SMRZ (Miralles et al., 2025); ERA5 provided runoff, evaporation, and T2M; CRU TS v4.09 provided the precipitation and PET covariates used in the models and the P−ET diagnostic proxy (Harris et al., 2020); GLOBMAP LAI provided vegetation greenness, which was bilinearly aggregated to 0.5° (Liu et al., 2012); GLDAS-2.1 provided the alternative runoff data for the sensitivity branch (Beaudoing et al., 2020). The UTrack atmospheric moisture-tracking dataset supplied source-region weighting fields (Tuinenburg et al., 2020a). The monthly source-region weights were derived from the original UTrack source-sink fractions. A statistically reconciled UTrack-derived product improves annual consistency with ERA5 evaporation and precipitation totals (De Petrillo et al., 2025). However, because our seasonal attribution requires monthly terrestrial source-region fractions, we did not substitute this annual reconciled product in the main workflow. We therefore treat the absence of a transport-product sensitivity test using reconciled monthly weights as a limitation. Aridity index (AI), defined as the ratio of annual precipitation to potential evapotranspiration, was computed from ERA5 2008–2017 mean annual fields and used to classify grid cells into four regimes following the conventional dryland classification: hyper-arid (< 0.20), arid–semi-arid (0.20–0.40), dry sub-humid (0.40–0.65), and humid (> 0.65) (United Nations Environment Programme, 1997). Data harmonization procedures are described in Text S1.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

3 Results

3.1 Spatial patterns of blue-green water coefficient

The BG coefficient showed a spatially coherent but non-monotonic pattern across mainland China (Figure 2). The nationally pooled mean is 0.726 (median 0.753), with 43.4% of grid cells exceeding 0.8. High-BG values are concentrated in semi-arid northeastern China, the Loess Plateau, and the arid northwest, where much of the limited precipitation returns to the atmosphere. Low-BG values (<0.3) are scattered across orographic zones of southwestern China and glacier-fed Tibetan Plateau margins, where runoff generation is enhanced by relief and meltwater.

Class-averaged BG does not decline monotonically with aridity index. Mean values are 0.801 (hyper-arid), 0.802 (arid-semi-arid), and 0.821 (dry sub-humid), near-constant across the three driest classes, before falling to 0.621 in the humid class. Within every class, grid-cell BG spans nearly the full [0, 1] interval (standard deviation 0.154–0.202), indicating that aridity alone cannot explain within-class variability. This motivates the statistical analysis that follows.

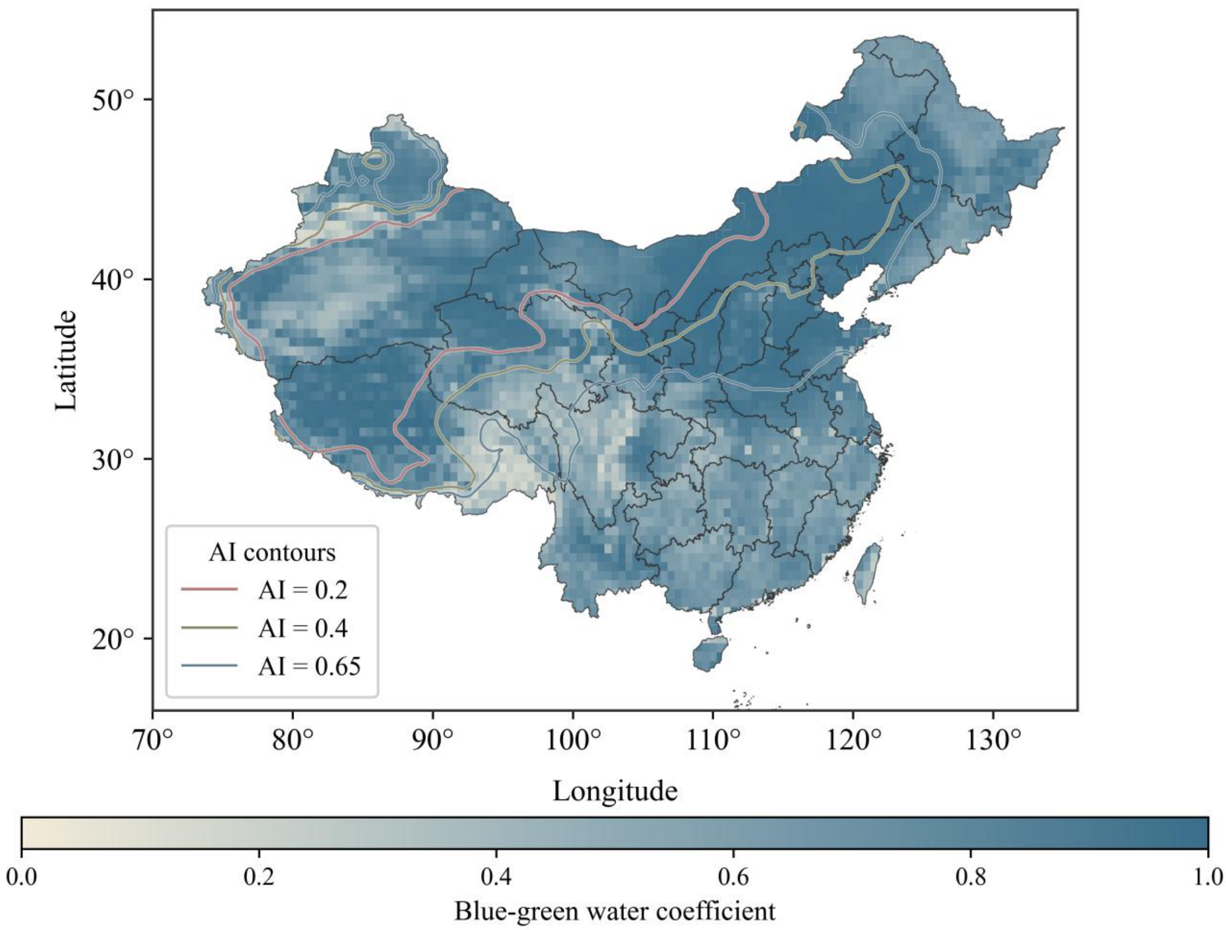

Figure 2. Spatial distribution of the blue-green water coefficient across China. The background colors denote grid-scale values

© Author(s) 2026. CC BY 4.0 License.

of the coefficient, and red contours represent aridity index thresholds at 0.20, 0.40, and 0.65.

Figure 3 shows the ten-year mean distribution of terrestrial evaporation that is transported to and precipitates over China, as derived from the UTrack dataset (2008–2017). The dominant terrestrial source regions are located in low- to mid-latitude Eurasia (10–50°N, 70–140°E), where both the flux intensity and the number of contributing land grids reach their maxima. This belt encompasses China itself and its surrounding countries, including the southern Siberian region of Russia, Mongolia, the Central Asian republics, the Indian subcontinent, and Southeast Asian countries. The spatial continuity of this source zone shows that China's precipitation is not solely dependent on domestic land evaporation and receives substantial upwind continental contributions from neighboring Eurasian regions. It is consistent with recent evidence that China's dryland precipitation is strongly shaped by moisture recycling through surrounding continental pathways (Wei et al., 2024). By contrast, contributions from Africa, North America, and South America are negligible, and the Southern Hemisphere landmasses provide almost no flux to China.

The longitudinal and latitudinal transects (Figure 3b-c) confirm these patterns: total fluxes peak sharply between 60°E and 140°E longitude and at 20–40°N latitude. Outside these Eurasian and monsoon-influenced belts, both the mean and integrated flux values drop to near zero. These results indicate that China's terrestrial moisture supply is anisotropically structured, dominated by the Eurasian mid-latitude corridor and supplemented by subtropical Asian sources. This spatial structure provides the basis for defining upwind factors in the attribution analysis.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

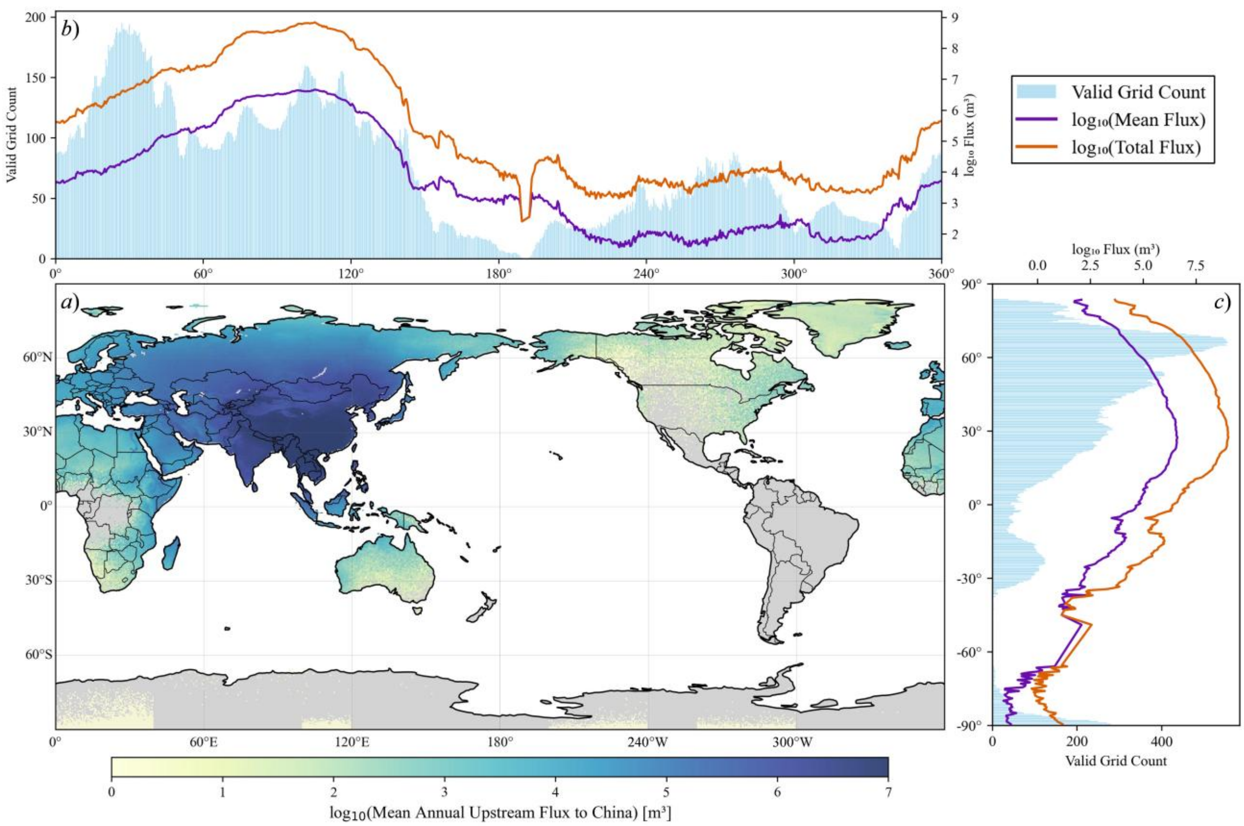

Figure 3. Ten-year mean spatial distribution of upwind terrestrial moisture sources to China (2008-2017). (a) Mean annual upwind fluxes aggregated to a global 0.5° grid; (b) Longitudinal transect of contributing grid count and flux; (c) Latitudinal transect of contributing grid count and flux.

3.2 Local dominance and the geographic concentration of upwind contribution

The GW-RF results showed a nationally local-dominant explanatory structure (Figure 4). The mean block cross-validated R^2 of the local-only model is 0.468; adding upwind covariates raises this to 0.471, a mean incremental gain of $\Delta R^2 = 0.003$ (median = -0.001). Local hydrometeorological and land-surface controls therefore account for the dominant share of explainable monthly partitioning variance. Both upwind explanatory diagnostics revealed a geographic concentration that the national mean obscures, with positive values clustering in particular aridity classes. The mean annual permutation drop is 0.023 nationally (69.9% of grid cells positive). The spatial distribution and across-grid heterogeneity of the upwind permutation-drop diagnostic are provided in Figure S6. Stratifying by aridity class shows a marked concentration in the dry sub-humid zone: the mean permutation drop reaches 0.028 and 79.2% of grid cells show positive values, the highest of any class. The incremental R^2 diagnostic provides a sharper filter: mean ΔR^2 is positive only in the dry sub-humid (0.010) and humid (0.006) classes, and weakly negative in the hyper-arid (-0.004) and arid–semi-arid (-0.003) classes, indicating that upwind information

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

does not improve cross-validated R^2 on average in the two driest regions. Grid cells with $\Delta R^2 > 0.01$ number 1,394 (33.2%) and concentrate in the humid (42.5%) and dry sub-humid (24.2%) classes. These diagnostics indicate that local controls are the primary determinants of monthly BG partitioning. Upwind land-surface conditions provide a secondary contribution, with the clearest incremental explanatory gain in the dry sub-humid transition belt. Grid cells in this zone combine moderate water limitation with sufficient vegetation cover, which may make upwind information more detectable. Because the upwind covariates were based on land-source UTrack weights, this secondary contribution represents a terrestrial source-region signal rather than a complete land-and-ocean moisture-source attribution.

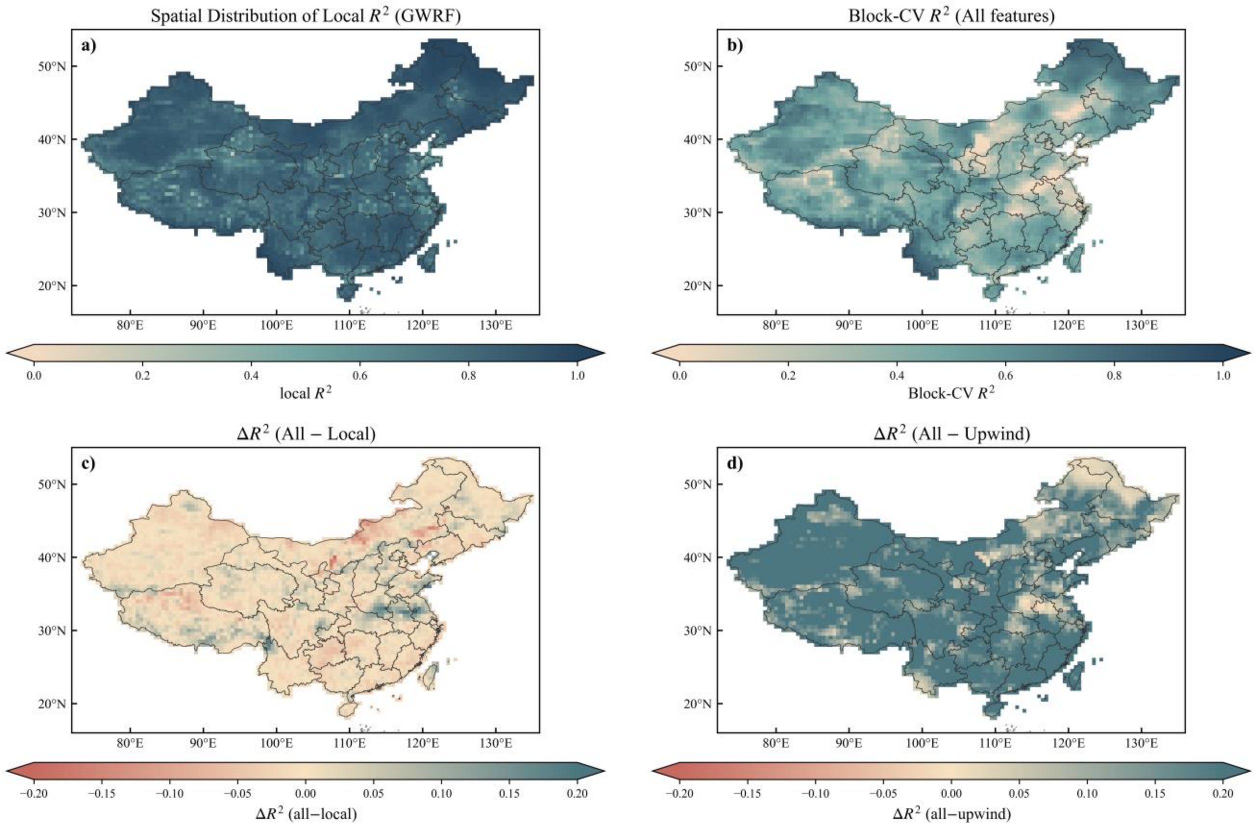

Figure 4. Geographically weighted random forest performance and the local-versus-upwind hierarchy. (a) Local R^2 values under full-sample fitting; (b) Local R^2 values under five-fold temporal block cross-validation; (c) Incremental explanatory power of upwind factors, expressed as ΔR^2(All-Local); (d) Incremental explanatory power of local factors, expressed as ΔR^2 (All-Upwind).

3.3 Robustness across ET and runoff product combinations

Reanalysis-derived variables still appear in parts of the analysis, most notably ERA5 Runoff in the response and ERA5 T2M among the covariates. This creates a risk that part of the fitted structure reflects product-specific consistency rather than

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

a product-robust ecohydrological pattern. To evaluate this, we compared explanatory diagnostics across three response definitions of progressively increasing product independence: ERA5 ET + ERA5 Runoff, GLEAM ET + ERA5 Runoff, and GLEAM ET + GLDAS Runoff (Figure 5). Full-model R^2 declines with product independence (R^2 = 0.492, 0.471, and 0.428), as expected when independent products with their own biases replace a single consistent reanalysis. National mean ΔR^2 appears similar across products (0.009, 0.003, 0.008), but these values mask opposing within-class trends (Table 1).

Table 1. Annual explanatory diagnostics by aridity class and product combination.

AI_{class}	Product	R^2_{all}	ΔR^2_{mean}	ΔR^2>0(%)	$perm_{mean}$	perm>0(%)
<0.20	ERA5+ERA5	0.431	+0.014	65.5%	+0.036	85.9%
	GLEAM+ERA5	0.449	-0.005	41.5%	+0.023	73.6%
	GLEAM+GLDAS	0.452	+0.002	49.7%	+0.024	78.4%
0.20-0.40	ERA5+ERA5	0.450	+0.010	60.1%	+0.027	74.2%
	GLEAM+ERA5	0.412	+0.003	55.6%	+0.023	66.9%
	GLEAM+GLDAS	0.343	+0.002	65.8%	+0.009	77.5%
0.40-0.65	ERA5+ERA5	0.467	+0.005	56.6%	+0.022	75.5%
	GLEAM+ERA5	0.436	+0.009	56.4%	+0.032	80.4%
	GLEAM+GLDAS	0.349	+0.014	67.3%	+0.028	78.3%
>0.65	ERA5+ERA5	0.551	+0.007	48.2%	+0.018	67.9%
	GLEAM+ERA5	0.520	+0.005	48.2%	+0.020	64.3%
	GLEAM+GLDAS	0.478	+0.011	59.3%	+0.013	64.3%

In the hyper-arid class, ERA5+ERA5 yields an apparent upwind gain of ΔR^2 = 0.014, but replacing ERA5 ET with GLEAM inverts this to $\Delta R^2 = -0.006$, suggesting that the hyper-arid signal is more likely to reflect product circularity than product-robust remote influence. In the arid–semi-arid class, the upwind signal attenuates progressively: ΔR^2 declines from 0.010 (ERA5+ERA5) to 0.004 (GLEAM+ERA5) and 0.002 (GLEAM+GLDAS), and the permutation drop follows a similar trajectory (0.027, 0.023, 0.009). Unlike the hyper-arid class, the arid–semi-arid signal does not reverse sign, but the progressive attenuation indicates that a substantial share of the apparent upwind gain under ERA5+ERA5 reflects shared reanalysis variance rather than a product-robust signal. Upwind explanatory signals in these two driest classes are not stable across product combinations and should not be interpreted as product-robust evidence of remote influence.

The dry sub-humid class moves in the reverse direction. ΔR^2 increases with product independence (0.006, 0.009, 0.014), and the permutation drop moves in a similar direction (0.023, 0.032, 0.028). This pattern is more consistent with a product-robust dry sub-humid signal than with an ERA5-only artifact. The humid class is similarly consistent (ΔR^2 = 0.007, 0.005, 0.011).

In the dry sub-humid and humid classes, upwind signals persisted across all three product combinations and strengthen with product independence. The two drier classes show signals that attenuate or reverse sign when ERA5 dependence is

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

reduced: the hyper-arid class shows a sign reversal in ΔR^2, while the arid–semi-arid class shows progressive attenuation without reversal. GLEAM ET + ERA5 Runoff is retained as the primary analysis line because it reduces direct ET circularity while maintaining complete spatial and seasonal diagnostics.

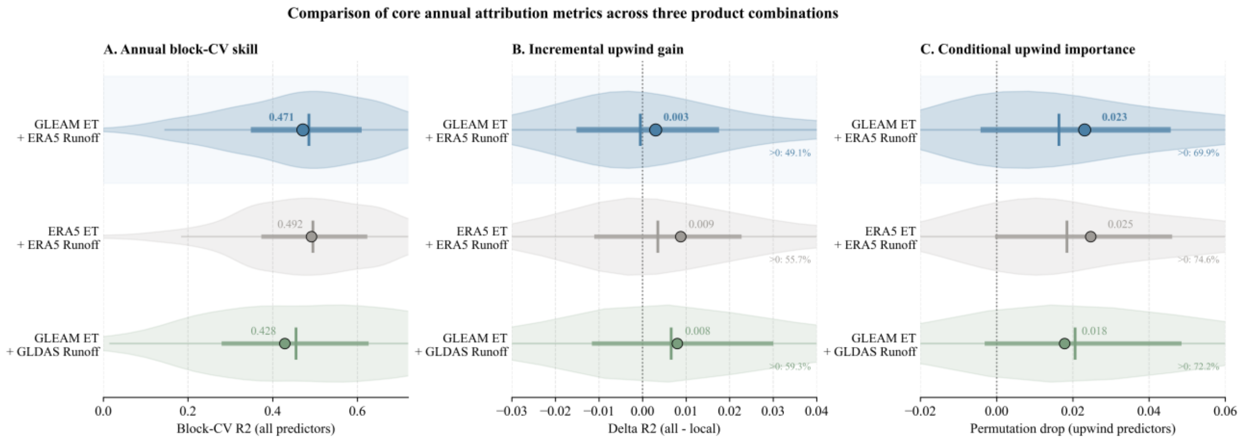

Figure 5. Annual robustness of explanatory metrics across three ET and runoff product combinations. The three panels show grid-level distributions of annual diagnostics across three response definitions: ERA5 ET + ERA5 runoff (single-reanalysis baseline), GLEAM ET + ERA5 runoff (partial product independence), and GLEAM ET + GLDAS runoff (maximum independence of the two flux terms). Panels show (a) annual block-CV R2, (b) incremental upwind gain, and (c) conditional upwind importance.

3.4 Dominant-factor geography and the local-upwind asymmetry

The dominant-factor classification provides a single-rank summary of the GW-RF explanatory structure (Figure 6). Nationally, local covariates achieve dominance at 45.2% of grid cells, upwind covariates at only 1.6%, and 51.7% fall into Uncertain. Among classified cells, $SMRZ_{local}$ is most frequently dominant (18.3% of all grids), followed by $Precip_{local}$ (14.2%), $T2M_{local}$ (6.3%), LAI_{local} (4.5%), and PET_{local} (1.8%). Mean summed importance is 0.526 for local covariates versus 0.369 for upwind, with the same ordering reproduced by SHAP magnitudes (mean absolute local SHAP = 0.116; upwind SHAP = 0.061, Figures S3). A complementary ternary composition of local, upwind, and seasonal control groups, together with SHAP-based dominance classes, is provided in Figure S8.

Dominant-factor composition varied with aridity. In the hyper-arid class, $SMRZ_{local}$ achieves dominance at 45.0% of grid cells, more than four times any other covariate, reflecting the primacy of antecedent root-zone storage when precipitation is so scarce that each event's fate depends critically on prior moisture state. In the dry sub-humid class, this ordering reverses: $Precip_{local}$ dominates 29.4% of grid cells while $SMRZ_{local}$ drops to 7.5%, indicating that direct precipitation supply is the proximal control in this transition zone. The semi-arid class shows the most distributed structure (Uncertain fraction: 54.9%), consistent with competing controls, and the humid class returns to a precipitation-dominated structure ($Precip_{local}$ at 17.1%).

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

Upwind covariates achieve first-rank dominance at only 69 grid cells (1.6%), but their composition is asymmetric. PET_{upwind} leads (42 cells, 60.9% of upwind-dominant grids), with 27 of those concentrated in the semi-arid class. This pattern is consistent with source-region energy availability, rather than vegetation density, being the dominant control on moisture export to downwind semi-arid grids. $SMRZ_{upwind}$ accounts for 20 upwind-dominant cells (29.0%), while LAI_{upwind} dominates at only two cells nationally.

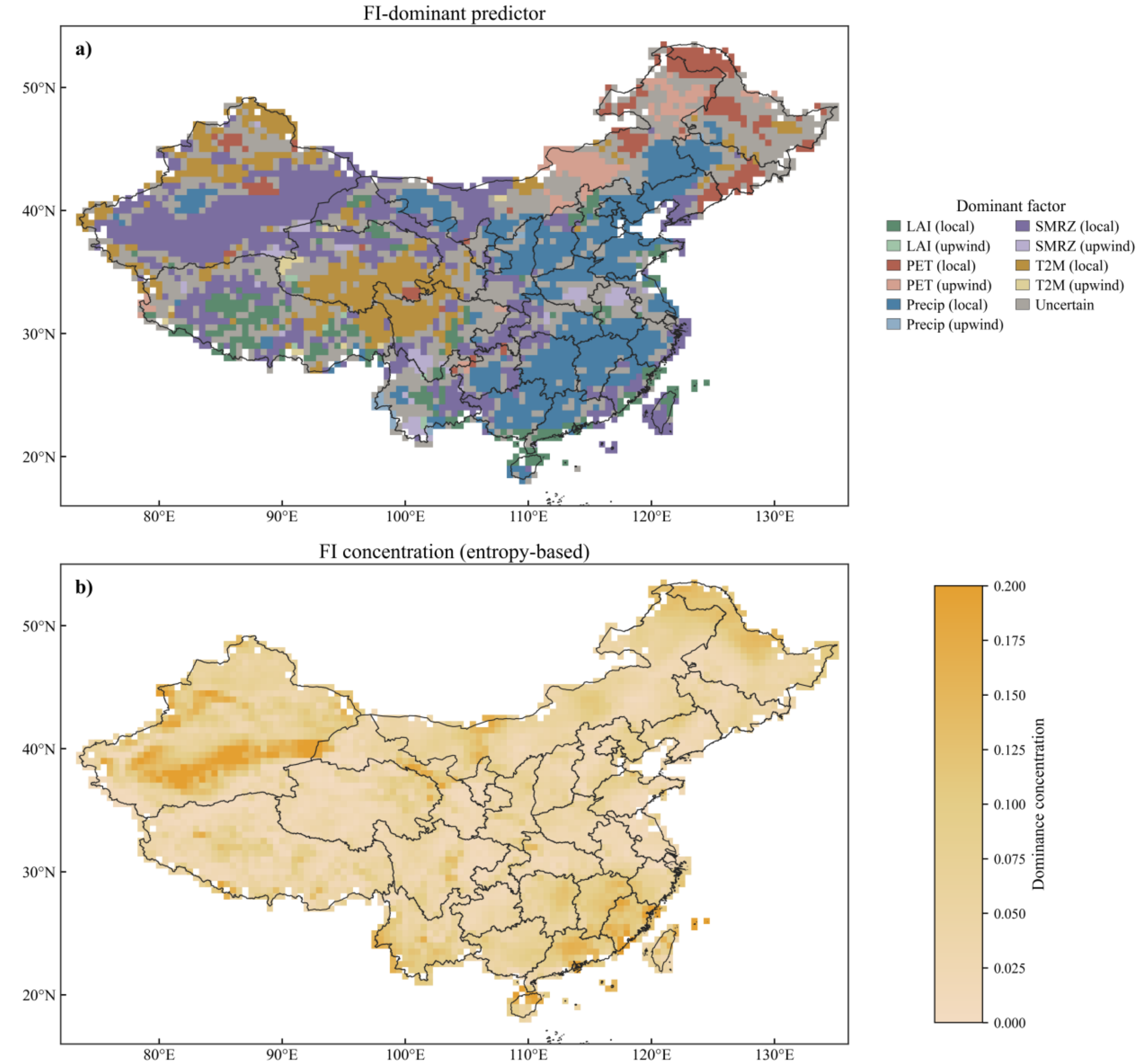

Figure 6. Dominant explanatory factors of the blue-green water coefficient derived from GW-RF. (a) Spatial distribution of dominant factors; (b) Classified dominance strength based on entropy.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

3.5 Nonlinear partial-dependence structure of local and upwind covariates

PDPs describe the fitted relationship between each covariate and the BG coefficient within each aridity class (Figure 7). PET_{upwind} shows the largest PDP amplitude among upwind covariates, with strongly monotone-negative dependence across three non-hyper-arid classes: spans of 0.570 (AI = 0.20–0.40), 0.540 (AI = 0.40–0.65), and 0.129 (AI > 0.65). Higher source-region atmospheric evaporative demand was associated with lower downwind BG values, consistent with reduced advective export to downwind grids. In the hyper-arid class, PET_{upwind} adopts a weak inverted-U shape (span = 0.113).

LAI_{upwind}, $SMRZ_{upwind}$, and $Precip_{upwind}$ each independently exhibit inverted-U responses to downstream BG across all four aridity classes. LAI_{upwind} shows pronounced amplitudes, especially in the semi-arid and dry sub-humid classes (spans 0.344 and 0.472; peak at LAI = 1.8-2.4). $SMRZ_{upwind}$ peaks at 0.19–0.30 (spans 0.14–0.37) and $Precip_{upwind}$ peaks at an intermediate band (spans 0.17–0.55). The parallel inverted-U structure across three water-cycle variables is consistent with maximum moisture export at intermediate source-region wetness, canopy density, and precipitation. Declines at both ends are consistent with moisture limitation at the dry extreme and local interception at the wet extreme. This pattern is compatible with a source-region interpretation, but the mechanism is not directly demonstrated by the PDPs.

Local covariates show aridity-dependent sign reversals. LAI_{local} switches from monotone-negative in the two driest classes (spans 0.078–0.338) to monotone-positive in the dry sub-humid class (span = 0.126), reflecting a water-to-energy limitation transition: denser canopies deplete scarce soil water in arid regimes but increase transpiration without inducing stress once AI ≥ 0.4. $SMRZ_{local}$ shows a low-value inverted-U peak (maxima at 0.09–0.15; semi-arid span = 0.364, the largest of any covariate-class combination) in the two driest classes, transitioning to monotone-negative in wetter classes (0.212–0.234). $Precip_{local}$ is monotone-positive in semi-arid, inverted-U in dry sub-humid, and monotone-negative in humid settings (span = 0.306), the last reflecting preferential runoff partitioning at high precipitation. These aridity-dependent reversals indicate spatially heterogeneous BG sensitivity across the aridity gradient. More broadly, the PDPs show that pronounced nonlinear responses do not necessarily translate into first-rank dominance, a contrast interpreted further in Section 4.2.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

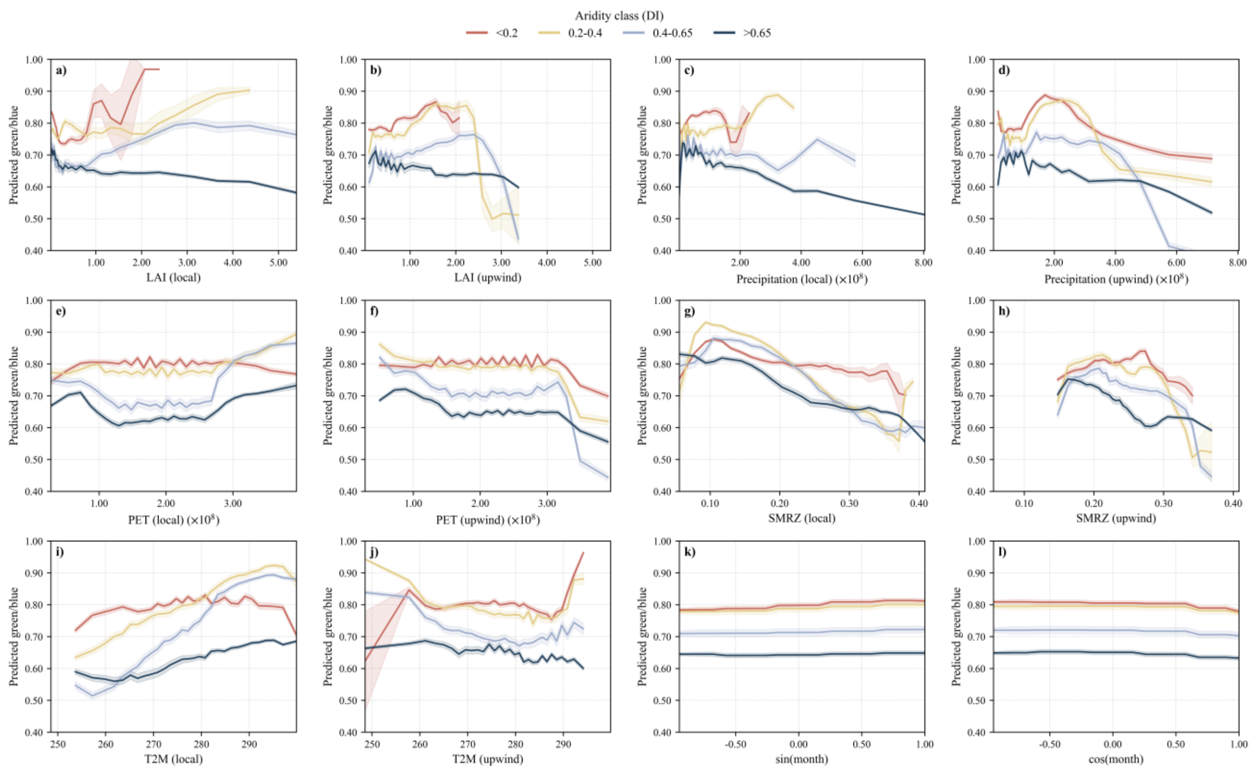

Figure 7. Aridity-stratified partial dependence of the blue–green water coefficient on local and upwind covariates. Solid curves show the AI-class–specific ensemble means of grid-level PDPs; shaded bands show mean ± 1.96 standard errors across grid cells within each AI class (normal-approximation 95% confidence intervals). Partial dependence was computed conditional on AI-specific backgrounds, with all other covariates marginalized over each AI distribution. The PDP evaluation grid was restricted to the 5th–95th percentiles of the corresponding background to avoid extrapolation.

3.6 Physical-benchmark and seasonal consistency of the upwind explanatory diagnostics

We compared ΔR^2 and the permutation drop against an annual P − ET water-balance proxy (Figure 8) and disaggregated both diagnostics by meteorological season (Figures S4 and S5). The annual proxy was constructed from CRU TS precipitation minus ERA5 evaporation and is independent of the GW-RF covariate set. Because it is a reduced water-balance diagnostic rather than a direct measure of terrestrial source-region transport, it is used to assess directional consistency, not causal attribution or magnitude correspondence.

Across all 4,205 grid cells, the annual P − ET proxy correlated weakly negatively with the permutation drop (Spearman $\rho = -0.183$) and incremental R^2 ($\rho = -0.084$). Stratifying by aridity class revealed that these national averages pooled opposing patterns. Within-class correlations with the permutation drop were strongly negative in hyper-arid regions ($\rho = -0.524$), less

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

negative in semi-arid regions ($\rho = -0.378$), positive in dry sub-humid regions ($\rho = 0.235$), and weakly negative in humid regions ($\rho = -0.119$).

The within classes patterns provided additional context. In the two driest classes, the negative association suggests that annual water-balance surplus or deficit alone does not translate into stronger upwind-attributable partitioning, consistent with dominant local storage and water-limitation constraints. In the dry sub-humid class, the positive association indicates that grid cells with larger P − ET proxy values also tend to register stronger upwind explanatory signal. This pattern is consistent with greater detectability in areas where moderate water limitation and vegetation cover coexist. The humid class is consistent with a saturated partitioning regime, where additional water-balance surplus produces little marginal partitioning change. This class-dependent sign reversal is consistent with the dry sub-humid concentration of the upwind signal, but it should be interpreted as an auxiliary water-balance consistency check rather than as independent validation.

Seasonal disaggregation shows a clear temporal structure. Seasonal mean permutation drops span: SON (0.033) > DJF (0.002) > JJA (−0.013) > MAM (−0.049). The SON maximum is consistent with monsoon retreat, stronger westerly continental moisture transport from Eurasia, and partial depletion of antecedent soil moisture. Under these conditions, the terrestrial-only upwind covariates may better represent the relevant moisture supply, and the receiving surface may be more sensitive to that supply. Negative signals in JJA and MAM reflect a shift toward oceanic moisture sources and strong local controls (monsoon precipitation, snowmelt) that mask residual terrestrial upwind contributions.

Together with the product-sensitivity analysis, the annual P − ET benchmark and seasonal diagnostics provide convergent but auxiliary support for the dry sub-humid interpretation. The same regional pattern appears in the P − ET association, the SON maximum in permutation drops, and the response-product comparison, making it less likely to reflect a single diagnostic artifact. These diagnostics support a cautious interpretation of dry sub-humid detectability, but they do not constitute independent causal validation.

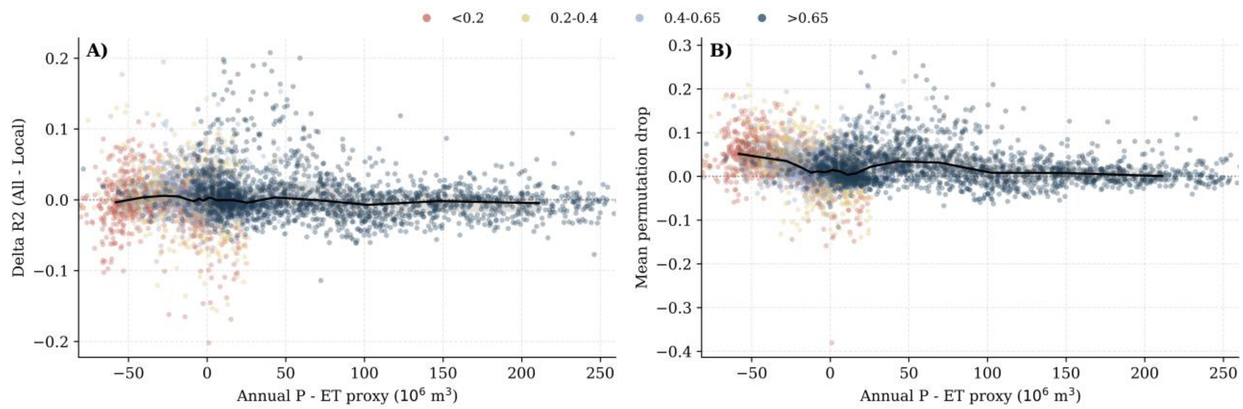

Figure 8. Annual diagnostic benchmark of the upwind explanatory signal against an auxiliary P − ET water-balance proxy. (a) Relationship between the annual P − ET proxy (constructed from CRU TS precipitation minus ERA5 evaporation) and the incremental gain in block cross-validated R^2 from adding upwind covariates to the local-only model. (b) Relationship between

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

the same proxy and the annual permutation drop of upwind covariates.

4 Discussion

4.1 Conditional and attenuated upwind detectability in the dry sub-humid transition zone

Across the diagnostics presented in Sections 3.2–3.6, the most consistent regional pattern is that the upwind signal is most detectable in the dry sub-humid class, especially relative to the hyper-arid and arid–semi-arid classes. This finding is narrow in scope. Even in that class, the incremental gain from adding upwind covariates remains modest in absolute terms. The results therefore do not indicate strong remote control of monthly blue-green partitioning. Instead, upwind information provides a small recurring improvement in explanatory power within a specific hydroclimatic regime, while local controls account for most explainable variance. The same aridity class also shows the highest fraction of positive permutation drops and the clearest strengthening of the upwind signal as the response definition becomes more product-independent. This pattern is less consistent with a purely reanalysis-internal artifact than the patterns in the hyper-arid and arid–semi-arid classes, where ΔR^2 weakens or reverses when product dependence is reduced. At the same time, the product comparison should be interpreted as a robustness stress test rather than as formal validation.

The dry sub-humid concentration may reflect a change in the local partitioning regime across the aridity gradient. In the hyper-arid class, the dominant-factor maps identify $SMRZ_{local}$ as the leading covariate at 45.0% of grid cells, whereas Preciplocal is dominant in less than 4%; this pattern is more consistent with antecedent root-zone storage than with incoming precipitation as the main control on the monthly ratio. In the dry sub-humid class, this ordering is reversed: $Precip_{local}$ dominates 29.4% of grid cells and $SMRZ_{local}$ drops to 7.5%. The local PDP structure reinforces this observation: LAI_{local} switches from monotone-negative in the two driest classes to monotone-positive in the dry sub-humid class, and $SMRZ_{local}$ transitions from a low-value inverted-U peak to a monotone-negative dependence. The within-class association between the annual P − ET proxy and permutation drop reverses sign in the same aridity window, from $\rho = -0.524$ in the hyper-arid class to $\rho = 0.235$ in the dry sub-humid class. These patterns are consistent with a setting in which externally supplied moisture is more strongly buffered by antecedent root-zone storage in hyper-arid areas, but more directly reflected in partitioning in dry sub-humid areas. We interpret the connection between this local-regime shift and enhanced upwind detectability as a synthesis of the combined diagnostics, not as a uniquely identified mechanism. These observational diagnostics locate a transition in detectability more clearly than they isolate the physical process that produces it.

Although the humid class also shows positive mean ΔR^2 across product combinations, we place less interpretive weight on that result. Its positive-gain fraction is lower, its signal is more spatially diffuse, and local precipitation control remains strong. The dry sub-humid class is therefore best described as the clearest region of detectable upwind contribution, not the only region where such contribution appears.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

This interpretation also helps distinguish our findings from the broader literature on semi-arid land–atmosphere coupling. The classical semi-arid hotspot concerns the sensitivity of precipitation or coupling strength, whereas the response analyzed here is the monthly partitioning of water between evapotranspiration and runoff. The three driest aridity classes share a near-constant BG value of 0.80–0.82, indicating that ET/runoff competition changes little across those classes. The dry sub-humid class marks the point at which this pattern begins to change, which is consistent with higher water-balance sensitivity along this part of the aridity gradient. Recent evidence also suggests that dryland land–atmosphere coupling and ecological-restoration effects can shift across aridity thresholds (Zhang et al., 2024). The aridity band in which BG partitioning is most responsive to upwind information therefore need not coincide exactly with the band in which soil-moisture–precipitation coupling is strongest (Koster et al., 2004). In that sense, the dry sub-humid concentration identified here is best viewed as a response-specific observational pattern rather than as a replacement for the established coupling framework.

4.2 A tentative interpretation of the upwind nonlinearities

The conditional and attenuated nature of the upwind signal is also reflected in its nonlinear form: source-region variables are associated with downwind partitioning only within limited state ranges, and no single upwind variable emerges as a widespread first-order control. Figure 7 showed that three upwind covariates (LAI_{upwind}, $SMRZ_{upwind}$, and $Precip_{upwind}$) exhibit inverted-U PDP shapes across aridity classes, whereas PET_{upwind} is predominantly monotone negative. These patterns are physically interpretable, but they do not directly demonstrate a mechanism. PDPs summarize model behavior conditional on the fitted covariate structure; they can be influenced by correlation among covariates, uneven sampling across covariate space, and the specific response definition used in the analysis. The nonlinear shapes are therefore best treated as hypothesis-generating evidence, not as confirmation of a source-region mechanism.

The recurring inverted-U shapes across the three upwind water-state covariates suggest that moisture export may be greatest at intermediate source-region wetness and vegetation density. Very dry source regions may contribute little downwind moisture because evapotranspiration is moisture-limited. At the wet or densely vegetated end of the gradient, a larger share of evaporated water may be retained or recycled locally, consistent with land-cover and land-management perturbation experiments showing that vegetation changes can reorganize atmospheric water-cycle pathways (De Hertog et al., 2024). The monotone-negative PET_{upwind} pattern is likewise consistent with stronger source-region atmospheric demand reducing the amount of moisture available for downwind advection. This tentative source-region-retention interpretation is also consistent with evidence that reduced moisture transport can propagate hydroclimatic impacts downwind, emphasizing that upwind land conditions affect downstream water availability only when moisture remains available for advective export (Herrera-Estrada et al., 2019).

One reason for keeping this interpretation provisional is that the dominant-factor map and the PDP profiles summarize different aspects of the fitted models. The dominant-factor classification identifies the covariate with the largest local importance at each grid cell, whereas PDP amplitude describes the range of fitted BG responses as one covariate is varied over

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

its class-specific distribution. A covariate can therefore show a pronounced nonlinear PDP within an aridity class without becoming the first-ranked covariate across many grid cells, particularly when local covariates explain most variance and upwind covariates are correlated. LAI_{upwind} illustrates this distinction: it shows a pronounced dry sub-humid PDP span (0.472), but achieves first-rank dominance at only two grid cells nationally. The two diagnostics should therefore be read as complementary evidence for a distributed, secondary upwind signal rather than as competing rankings of the same effect.

The numerical peak ranges recovered from the PDPs should therefore be treated with restraint. For example, the intermediate LAI_{upwind} range identified in the present dataset is useful as a descriptive summary of where the fitted model shows the largest response, but it should not be interpreted as a universal threshold or direct management target. These ranges can be used as candidate values for future coupled land-atmosphere experiments or focused observational tests that perturb or evaluate source-region conditions more explicitly. The PET_{upwind} result provides a testable expectation for future studies: a negative monotone relationship should recur in other regions if source-region evaporative demand consistently reduces moisture available for downwind transport. An inverted-U PET_{upwind} response in a replication study would weaken this interpretation.

The upwind nonlinearities should therefore be read as hypothesis-generating evidence rather than as direct mechanism. In the fitted models, intermediate source-region wetness and vegetation states, together with lower source-region evaporative demand, are associated with modestly higher downwind BG values, with the clearest expression in the dry sub-humid class. These associations are consistent with a conditional moisture-export pathway, but their persistence under explicit perturbations of source-region land states remains untested. The present analysis therefore complements moisture-tracking and land-atmosphere coupling studies by identifying specific source-region hypotheses for future experiments.

4.3 Seasonal detectability, limitations, and cautious implications

The same conditionality appears temporally. Upwind information is most detectable in autumn, whereas spring and summer diagnostics are weak or negative because the terrestrial-source signal is diluted by oceanic moisture inputs and strong local controls. We interpret this as evidence about seasonal detectability, not as proof that upwind influence is absent outside SON. Two factors plausibly limit detectability during the warm and transition seasons. The moisture-source mix shifts substantially toward oceanic origins during JJA and, less strongly, during MAM, so the terrestrial-only upwind covariates structurally represent only a fraction of the actual moisture supply during those months. In addition, strong monsoon precipitation in JJA and snowmelt or early-growing-season storage processes in MAM directly dominate the monthly partitioning ratio, masking any residual upwind contribution. The seasonal results therefore indicate when the present diagnostics are most informative, not when continental moisture transport is active in an absolute sense.

Several limitations affect the interpretation of these results. First, the effect sizes are modest, particularly for incremental explanatory power, so the main finding concerns secondary modulation rather than strong remote control. Second, the analysis remains observational and partly product-dependent. The revised main response definition reduces, but does not eliminate,

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

reliance on reanalysis-based fields, and the UTrack source-region weights were held fixed in this study. Third, the upwind covariates aggregate land-surface conditions of terrestrial source regions only; oceanic moisture pathways are not represented in the covariate set, and the diagnosed upwind contribution should accordingly be read as the marginal contribution of the terrestrial source-region pathway rather than as the effect of total land-and-ocean moisture-source variability. Fourth, the annual and seasonal P − ET proxy diagnostics provide physical context but do not constitute independent causal validation (Ombadi et al., 2020). Fifth, monthly temporal resolution and the China-only spatial domain limit how precisely one can infer process timescales or generalize the recovered PDP peak ranges. Finally, anthropogenic water management and other local interventions are not explicitly separated from the gridded flux products, which may blur interpretation in heavily managed basins.

Within those limits, the results still suggest that blue-green water partitioning assessments may benefit from considering upwind terrestrial source-region land-surface conditions in addition to local hydroclimate, particularly in dry sub-humid transition zones. That implication is diagnostic rather than prescriptive. We do not infer that upwind land management should be optimized to a specific LAI target on the basis of this analysis alone. Instead, the recovered nonlinear ranges are better viewed as candidate hypotheses for regional model experiments and targeted observational evaluation. Management applications would require corroboration from coupled land-surface simulations that perturb source-region conditions and track downstream responses.

A near-term value of the analysis is to identify where upwind information changes interpretation and where local controls remain sufficient. The main contribution is to refine the empirical map of where upwind terrestrial source-region land-surface conditions appear relevant to monthly partitioning and where local controls remain dominant.

5 Conclusions

This study presents a grid-scale explanatory of monthly blue-green water partitioning across mainland China (2008–2017) that explicitly separates local hydroclimatic controls from the ecohydrological state of upwind terrestrial moisture-source regions. We used a Geographically Weighted Random Forest with five-fold temporal block cross-validation and a four-step diagnostic chain comprising incremental explanatory gain, permutation-based necessity, multi-product robustness, and an auxiliary P − ET water-balance consistency check. The results show a hierarchical structure in which local hydrometeorological and land-surface conditions account for most explainable monthly partitioning variance, with mean block-CV R^2 increasing from 0.468 in the local-only model to 0.471 when upwind covariates are added.

The upwind contribution, while modest in national-average terms, is geographically and seasonally concentrated. This pattern is most evident in the dry sub-humid class, where mean ΔR^2 is 0.010 in the main analysis and rises to 0.014 under the maximum product-independence test. The direction of this change is opposite to the expectation under an ERA5 circularity artifact, and therefore argues against that explanation. The upwind signal is largest in autumn, when westerly-driven

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

continental moisture transport from Eurasia is more prominent and local monsoon controls are weaker. Partial dependence profiles show nonlinear fitted relationships. Upwind vegetation density, soil wetness, and precipitation each show inverted-U relationships with the downwind BG coefficient, consistent with intermediate source-region states yielding the highest moisture export. Upwind evaporative demand shows a monotone-negative relationship.

The results suggest that blue-green water partitioning assessments in dry sub-humid transition zones may benefit from incorporating upwind terrestrial land-surface conditions alongside local hydroclimate. The multi-product consistency of the dry sub-humid signal, together with the auxiliary P − ET water-balance benchmark, supports further testing of this zone in coupled land-atmosphere experiments that perturb source-region conditions. In the two driest aridity classes and during the summer monsoon season, local controls dominate and upwind information adds little practical value for assessment.

Data Availability

The gridded climate and land-surface datasets used in this study are publicly available from their original providers: CRU TS v4.09 from https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.09/; UTrack from https://doi.pangaea.de/10.1594/PANGAEA.912710; ERA5/ERA5-land from https://cds.climate.copernicus.eu/; GLEAM from https://www.gleam.eu/; GLDAS-2.1 from https://disc.gsfc.nasa.gov/datasets?keywords=GLDAS; and GLOBMAP LAI from the data source described by Liu et al. (2012). The analysis code and scripts are available at: https://figshare.com/s/871b32d2265ead14bba9.

Code availability

The analysis code and scripts used to produce the results reported in this study are available on figshare at https://figshare.com/s/871b32d2265ead14bba9.

Author contributions

ZL: conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing original draft. LJ: conceptualization, funding acquisition, methodology, project administration, supervision, review and editing. YX: supervision, review and editing.

Competing interests

The contact author has declared that none of the authors has any competing interests.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

Financial support

This work was supported by the National Natural Science Foundation of China (No. 52379001 and 52222901), and the Beijing Nova Program (No. 20220484034).

References

Althoff, D. and Destouni, G.: Global patterns in water flux partitioning: Irrigated and rainfed agriculture drives asymmetrical flux to vegetation over runoff, One Earth, 6, 1246–1257, https://doi.org/10.1016/j.oneear.2023.08.002, 2023.

Asselin, O., Leduc, M., Paquin, D., De Noblet-Ducoudré, N., Rechid, D., and Ludwig, R.: Blue in green: forestation turns blue water green, mitigating heat at the expense of water availability, Environ. Res. Lett., 19, 114003, https://doi.org/10.1088/1748-9326/ad796c, 2024.

Beaudoing, H., Rodell, M., and NASA/GSFC/HSL: GLDAS Noah Land Surface Model L4 monthly 0.25 x 0.25 degree, Version 2.1, https://doi.org/10.5067/SXAVCZFAQLNO, 2020.

Chagas, V. B. P., Chaffe, P. L. B., and Blöschl, G.: Climate and land management accelerate the Brazilian water cycle, Nat Commun, 13, 5136, https://doi.org/10.1038/s41467-022-32580-x, 2022.

Chen, H., Covert, I. C., Lundberg, S. M., and Lee, S.-I.: Algorithms to estimate Shapley value feature attributions, Nature Machine Intelligence, 5, 590–601, https://doi.org/10.1038/s42256-023-00657-x, 2023.

De Hertog, S. J., Lopez-Fabara, C. E., Van Der Ent, R., Keune, J., Miralles, D. G., Portmann, R., Schemm, S., Havermann, F., Guo, S., Luo, F., Manola, I., Lejeune, Q., Pongratz, J., Schleussner, C.-F., Seneviratne, S. I., and Thiery, W.: Effects of idealized land cover and land management changes on the atmospheric water cycle, Earth Syst. Dynam., 15, 265–291, https://doi.org/10.5194/esd-15-265-2024, 2024.

De Petrillo, E., Monaco, L., Tuninetti, M., Staal, A., and Laio, F.: Cell-scale atmospheric moisture flows dataset reconciled with ERA5 reanalysis, Sci Data, 12, 629, https://doi.org/10.1038/s41597-025-04964-3, 2025.

Fahrländer, S. F., Wang-Erlandsson, L., Pranindita, A., and Jaramillo, F.: Hydroclimatic Vulnerability of Wetlands to Upwind Land Use Changes, Earth's Future, 12, e2023EF003837, https://doi.org/10.1029/2023EF003837, 2024.

Fayaz, J. and Galasso, C.: Interpretability and spatial efficacy of a deep-learning-based on-site early warning framework using explainable artificial intelligence and geographically weighted random forests, Geoscience Frontiers, 15, 101839, https://doi.org/10.1016/j.gsf.2024.101839, 2024.

Fu, Z., Ciais, P., Wigneron, J.-P., Gentine, P., Feldman, A. F., Makowski, D., Viovy, N., Kemanian, A. R., Goll, D. S., Stoy, P. C., Prentice, I. C., Yakir, D., Liu, L., Ma, H., Li, X., Huang, Y., Yu, K., Zhu, P., Li, X., Zhu, Z., Lian, J., and Smith, W. K.: Global critical soil moisture thresholds of plant water stress, Nat Commun, 15, 4826, https://doi.org/10.1038/s41467-024-49244-7, 2024.

Gao, C., Fu, J., Han, Z., Ji, W., Zhao, L., Wei, X., and Li, Z.: Local and downwind precipitation has been boosted by evapotranspiration change-induced moisture recycling in the Chinese Loess Plateau, Agricultural and Forest Meteorology, 371, 110623, https://doi.org/10.1016/j.agrformet.2025.110623, 2025.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., and Kalogirou, S.: Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling, Geocarto International, 36, 121–136, https://doi.org/10.1080/10106049.2019.1595177, 2021.

Greve, P., Gudmundsson, L., Orlowsky, B., and Seneviratne, S. I.: Introducing a probabilistic Budyko framework, Geophysical Research Letters, 42, 2261–2269, https://doi.org/10.1002/2015GL063449, 2015.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, https://doi.org/10.1002/hyp.6989, 2008.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Sci Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Herrera-Estrada, J. E., Martinez, J. A., Dominguez, F., Findell, K. L., Wood, E. F., and Sheffield, J.: Reduced Moisture Transport Linked to Drought Propagation Across North America, Geophysical Research Letters, 46, 5243–5253, https://doi.org/10.1029/2019GL082475, 2019.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quart J Royal Meteoro Soc, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hunt, A. G., Sahimi, M., Ghanbarian, B., and Poveda, G.: Predicting Ecosystem Net Primary Productivity by Percolation Theory and Optimality Principle, Water Resources Research, 60, e2023WR036340, https://doi.org/10.1029/2023WR036340, 2024.

Koldasbayeva, D. and Zaytsev, A.: Foundation for unbiased cross-validation of spatio-temporal models for Species Distribution Modeling, Ecological Informatics, 92, 103521, https://doi.org/10.1016/j.ecoinf.2025.103521, 2025.

Konapala, G., Mishra, A. K., Wada, Y., and Mann, M. E.: Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation, Nat Commun, 11, 3044, https://doi.org/10.1038/s41467-020-16757-w, 2020.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y., and Yamada, T.: Regions of Strong Coupling Between Soil Moisture and Precipitation, Science, 305, 1138–1140, https://doi.org/10.1126/science.1100217, 2004.

Li, M., Zhang, Y., Wallace, J., and Campbell, E.: Estimating annual runoff in response to forest change: A statistical method based on random forest, Journal of Hydrology, 589, 125168, https://doi.org/10.1016/j.jhydrol.2020.125168, 2020.

Li, Z., Sahotra, H., Ahmad, S., Wang, W., Yang, Z., Wu, P., Khan, E., and Zhuo, L.: A Distributed Machine Learning Model for Blue and Green Water Resources With Transferable Applications in Similar Climatic Zones, Water Resources Research, 61, e2024WR039169, https://doi.org/10.1029/2024WR039169, 2025.

Liu, J., Tian, J., Wu, J., Feng, X., Li, Z., Wang, Y., and Ya, Q.: Driving factors and trend prediction for annual runoff in the upper and middle reaches of the yellow river from 1990 to 2020, Environ. Res. Commun., 6, 085014, https://doi.org/10.1088/2515-7620/ad6bf6, 2024.

Liu, Y., Liu, R., and Chen, J. M.: Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data, Journal of Geophysical Research: Biogeosciences, 117, https://doi.org/10.1029/2012JG002084, 2012.

Mao, X., Zheng, J., Lu, B., Wang, R., Han, W., and Harris, P.: Spatiotemporal drought forecasting in Xinjiang's irrigated agriculture: Model comparison and multi-source data integration, Journal of Hydrology, 660, 133483, https://doi.org/10.1016/j.jhydrol.2025.133483, 2025.

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, Sci Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Ombadi, M., Nguyen, P., Sorooshian, S., and Hsu, K.: Evaluation of Methods for Causal Discovery in Hydrometeorological Systems, Water Resources Research, 56, e2020WR027251, https://doi.org/10.1029/2020WR027251, 2020.

Pan, Z., Yang, S., Lou, H., Li, C., Zhang, J., Zhang, Y., Luo, Y., and Li, X.: Perspectives of human–water co-evolution of blue–green water resources in subtropical areas, Hydrological Processes, 37, e14818, https://doi.org/10.1002/hyp.14818, 2023.

Peng, K., Zhang, Y., Tang, Q., Zhang, Y., Li, Z., Wang, G., and Cao, C.: Decomposing the Effects of Changes in Catchment Characteristics on Runoff Into Chain Transmission Effects of Climate Change and Human Activities Using an Improved Budyko Framework, Earth's Future, 13, https://doi.org/10.1029/2025ef006041, 2025.

Posada-Marín, J., Salazar, J., Rulli, M. C., Wang-Erlandsson, L., and Jaramillo, F.: Upwind moisture supply increases risk to water security, Nat Water, 2, 875–888, https://doi.org/10.1038/s44221-024-00291-w, 2024.

Ravinandrasana, V. P. and Franzke, C. L. E.: The first emergence of unprecedented global water scarcity in the Anthropocene, Nat Commun, 16, 8281, https://doi.org/10.1038/s41467-025-63784-6, 2025.

Rice, J. S. and Emanuel, R. E.: Ecohydrology of Interannual Changes in Watershed Storage, Water Resources Research, 55, 8238–8251, https://doi.org/10.1029/2019WR025164, 2019.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., and Lo, M.-H.: Emerging trends in global freshwater availability, Nature, 557, 651–659, https://doi.org/10.1038/s41586-018-0123-1, 2018.

Sharma, A., Patel, P. L., and Sharma, P. J.: Blue and green water accounting for climate change adaptation in a water scarce river basin, Journal of Cleaner Production, 426, 139206, https://doi.org/10.1016/j.jclepro.2023.139206, 2023.

Song, Z., Zhang, X., Xu, J., Lin, Y., She, D., Wang, Y., Chen, S., and Hu, C.: Vegetation greening promotes the conversion of blue water to green water by enhancing transpiration, Journal of Hydrology, 658, 133181, https://doi.org/10.1016/j.jhydrol.2025.133181, 2025.

Stephens, C. M., Band, L. E., Johnson, F. M., Marshall, L. A., Medlyn, B. E., De Kauwe, M. G., and Ukkola, A. M.: Changes in Blue/Green Water Partitioning Under Severe Drought, Water Resources Research, 59, e2022WR033449, https://doi.org/10.1029/2022WR033449, 2023.

© Author(s) 2026. CC BY 4.0 License.

Tan, X., Liu, B., Tan, X., Huang, Z., and Fu, J.: Projected climate change impacts on the availability of blue and green water in a watershed of intensive human water usage, Journal of Hydrology: Regional Studies, 53, 101827, https://doi.org/10.1016/j.ejrh.2024.101827, 2024.

Te Wierik, S. A., Gupta, J., Cammeraat, E. L. H., and Artzy-Randrup, Y. A.: The need for green and atmospheric water governance, WIREs Water, 7, e1406, https://doi.org/10.1002/wat2.1406, 2020.

Tong, C., Ye, Y., Zhao, T., Bao, H., and Wang, H.: Soil moisture disaggregation via coupling geographically weighted regression and radiative transfer model, Journal of Hydrology, 634, 131053, https://doi.org/10.1016/j.jhydrol.2024.131053, 2024.

Tuinenburg, O. A., Theeuwen, J. J. E., and Staal, A.: Global evaporation to precipitation flows obtained with Lagrangian atmospheric moisture tracking, https://doi.org/10.1594/PANGAEA.912710, 2020a.

Tuinenburg, O. A., Theeuwen, J. J. E., and Staal, A.: High-resolution global atmospheric moisture connections from evaporation to precipitation, Earth System Science Data, 12, 3177–3188, https://doi.org/10.5194/essd-12-3177-2020, 2020b.

United Nations Environment Programme: World Atlas of Desertification: Second Edition, 1997.

Van Der Ent, R. J., Wang-Erlandsson, L., Keys, P. W., and Savenije, H. H. G.: Contrasting roles of interception and transpiration in the hydrological cycle – Part 2: Moisture recycling, Earth Syst. Dynam., 5, 471–489, https://doi.org/10.5194/esd-5-471-2014, 2014.

Veettil, A. V. and Mishra, A. K.: Water security assessment using blue and green water footprint concepts, Journal of Hydrology, 542, 589–602, https://doi.org/10.1016/j.jhydrol.2016.09.032, 2016.

Wan, W., Döll, P., and Müller Schmied, H.: Global-Scale Groundwater Recharge Modeling Is Improved by Tuning Against Ground-Based Estimates for Karst and Non-Karst Areas, Water Resources Research, 60, e2023WR036182, https://doi.org/10.1029/2023WR036182, 2024.

Wang-Erlandsson, L., Fetzer, I., Keys, P. W., van der Ent, R. J., Savenije, H. H. G., and Gordon, L. J.: Remote land use impacts on river flows through atmospheric teleconnections, Hydrology and Earth System Sciences, 22, 4311–4328, https://doi.org/10.5194/hess-22-4311-2018, 2018.

Wang-Erlandsson, L., Tobian, A., van der Ent, R. J., Fetzer, I., te Wierik, S., Porkka, M., Staal, A., Jaramillo, F., Dahlmann, H., Singh, C., Greve, P., Gerten, D., Keys, P. W., Gleeson, T., Cornell, S. E., Steffen, W., Bai, X., and Rockström, J.: A planetary boundary for green water, Nature Reviews Earth & Environment, 3, 380–392, https://doi.org/10.1038/s43017-022-00287-8, 2022.

Wei, F., Wang, S., Fu, B., Li, Y., Huang, Y., Zhang, W., and Fensholt, R.: Quantifying the precipitation supply of China's drylands through moisture recycling, Agricultural and Forest Meteorology, 352, 110034, https://doi.org/10.1016/j.agrformet.2024.110034, 2024.

Weng, W., Costa, L., Lüdeke, M. K. B., and Zemp, D. C.: Aerial river management by smart cross-border reforestation, Land Use Policy, 84, 105–113, https://doi.org/10.1016/j.landusepol.2019.03.010, 2019.

Wunderling, N., Wolf, F., Tuinenburg, O. A., and Staal, A.: Network motifs shape distinct functioning of Earth's moisture recycling hubs, Nat Commun, 13, 6574, https://doi.org/10.1038/s41467-022-34229-1, 2022.

https://doi.org/10.5194/egusphere-2026-3497
Preprint. Discussion started: 3 September 2026
© Author(s) 2026. CC BY 4.0 License.

Yu, D., Zhu, Q., Zhang, J., Wang, L., Wu, X., Jiang, S., Fang, X., Jin, J., Wang, Y., and Ren, L.: Patterns of blue and green water in the Yellow River Basin from 1998 to 2020: Influence of climate change and human activity, Journal of Hydrology: Regional Studies, 59, 102367, https://doi.org/10.1016/j.ejrh.2025.102367, 2025.

Zhang, P., Zhang, H., Wang, S., Woodward, G., O'Gorman, E. J., Jackson, M. C., Hansson, L.-A., Hilt, S., Frenken, T., Wang, H., Zhou, L., Wang, T., Zhang, M., and Xu, J.: Multiple Stressors Simplify Freshwater Food Webs, Global Change Biology, 31, e70114, https://doi.org/10.1111/gcb.70114, 2025.

Zhang, Y., Feng, X., Zhou, C., Sun, C., Leng, X., and Fu, B.: Aridity threshold of ecological restoration mitigated atmospheric drought via land–atmosphere coupling in drylands, Commun Earth Environ, 5, 381, https://doi.org/10.1038/s43247-024-01555-9, 2024.