Supplement of

Unraveling Local and Upwind Controls on Terrestrial ET-Runoff Partitioning in China: Grid-Scale Attribution and Causal 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)

Contents of this file

Text S1 to S2

Figures S1 to S8

Introduction

This Supporting Information provides technical details and supplementary diagnostics for the grid-scale attribution analysis of blue-green water partitioning across mainland China. It is intended to make the modelling workflow, interpretation diagnostics, and data harmonization steps reproducible, while keeping the main text focused on the principal local-versus-upwind attribution results. Text S1 describes the geographically weighted random forest framework, including the adaptive bisquare kernel, temporal block cross-validation, partial dependence profiles, SHAP computation, entropy-based dominance index, and harmonization of climate, land-surface, and moisture-tracking datasets. Text S2 describes the statistical counterfactual perturbation experiments used as directional sensitivity diagnostics for upwind covariates.

The supplementary figures document exploratory associations, spatial interpretation outputs, seasonal water-balance consistency checks, and additional sensitivity results. Figures S1–S3 summarize local and upwind covariate associations and SHAP-based spatial contributions. Figures S4 and S5 provide seasonal comparisons between the P − ET proxy and upwind attribution diagnostics. Figure S6 shows the spatial distribution and heterogeneity of the upwind permutation-drop diagnostic, Figure S7 reports the counterfactual perturbation responses, and Figure S8 summarizes the relative composition of local, upwind, and seasonal controls.

Text S1. Technical details of model fitting, interpretation, and data harmonization

S1.1 Geographically Weighted Random Forest: kernel function and bandwidth

The Geographically Weighted Random Forest (GW-RF) assigns spatial weights to neighboring observations when fitting a local Random Forest at each target location. For a target grid cell i with great-circle distances d_{ij} to all other observations j, the adaptive bisquare kernel is defined as:

$$w_{ij}(h_i) = \begin{cases} \left[1 - \left(\dfrac{d_{ij}}{h_i}\right)^2\right]^2, & d_{ij} < h_i \\ 0, & d_{ij} \geq h_i \end{cases} \tag{1}$$

where h_i is the adaptive bandwidth, the great-circle distance to the k-th nearest neighbor of grid cell i. Weights are normalized so that $\sum_j w_{ij} = 1$. The local training set for grid i is $S_i = \{j : w_{ij} > 0\}$. An adaptive rather than fixed bandwidth was used to ensure that each local model draws on a consistent number of informative neighbors regardless of spatial data density. The bandwidth parameter k was selected by minimizing the out-of-bag error across a candidate range using the full 120-month sample.

S1.2 Five-fold temporal block cross-validation

To reduce temporal autocorrelation and information leakage, the 120-month time series (2008–2017) was partitioned into five contiguous 24-month blocks. For fold f, all grid-level local GW-RF models were trained on the four remaining blocks and validated on the held-out block at the same grid locations. Block cross-validated R^2 for a given grid was computed as:

$$R_{cv}^2 = 1 - \frac{\sum_f \sum_{t \in fold\ f} (y_t - \hat{y}_t)^2}{\sum_f \sum_{t \in fold\ f} (y_t - \bar{y}_{train,f})^2} \tag{2}$$

where y_t is the observed BG coefficient, $\hat{y}_t$ is the held-out prediction from fold f's local model, and $\bar{y}_{train,f}$ is the training-set mean for fold f. This formulation penalizes predictions relative to the within-fold training mean, giving a conservative estimate of out-of-sample explanatory power.

S1.3 Partial dependence profile computation

For a fitted grid-specific GW-RF model f, the one-dimensional partial dependence profile (PDP) for predictor x_j is the marginal expectation $E[f(x_j, X_{-j})]$, where the expectation is taken over the distribution of all other predictors X_{-j}.

In practice, we stratified the full predictor table by aridity-index class (AI < 0.20; 0.20–0.40; 0.40–0.65; > 0.65) via a latitude–longitude join, then randomly sampled up to 3,000 rows per class as the background distribution over which non-target covariates were marginalized. The seasonal cycle predictors ($\sin_{\text{month}}$, $\cos_{\text{month}}$) were retained in the background sample so that seasonality remained encoded. For each target predictor, we evaluated f at 20 evenly spaced points spanning the observed range of that predictor within each class. Grid cells with fewer than five background rows in a given class were excluded from that class's pooled average. PDP amplitude was defined as max($\hat{y}$) – min($\hat{y}$) across the 20 evaluation points.

S1.4 SHAP value computation

Shapley Additive Explanation (SHAP) values were computed at each grid cell using a locally weighted Random Forest that mirrors the GW-RF fitting design. For each grid, a local RF was re-fitted using the same adaptive bisquare kernel weights (Eq. S1) and the same 12 covariates (five local predictors, five upwind predictors, $\sin_{\text{month}}$, $\cos_{\text{month}}$). TreeExplainer was then applied to this local model to obtain per-sample SHAP values $\phi_{it}^{(k)}$ for each covariate k and month t. Grid-level SHAP maps were produced by averaging over the 120 months:

$$\bar{\phi}_i^{(k)} = \frac{1}{120} \sum_{t-1}^{120} \phi_{it}^{(k)} \tag{3}$$

The mean signed SHAP value $\bar{\phi}_i^{(k)}$ indicates the direction (positive = greener; negative = bluer) and time-average magnitude of predictor k's contribution to the fitted BG coefficient at grid i.

S1.5 Entropy-based dominance index

To summarize the concentration of explanatory weight across predictors, an entropy-based dominance index was computed at each grid cell from the normalized impurity-based feature-importance vector $p = (p_1, \ldots, p_k)$ for the $K = 10$ non-cyclic predictors (five local, five upwind), where $\sum_k p_k = 1$:

$$H_i = -\sum_{k=1}^{K} p_{ik} \log\left(p_{ik}\right) \tag{4}$$

Lower entropy ($H_i \rightarrow 0$) indicates that a single predictor accounts for nearly all explanatory weight (strong single-factor dominance); higher entropy ($H_i \rightarrow \log K$) indicates a more evenly distributed importance profile across predictors.

S1.6 Data harmonization

All predictor and response fields were harmonized to a common 0.5° × 0.5° monthly grid covering January 2008 to December 2017. ERA5 fields (runoff, evaporation, T2M) were obtained at 0.25° resolution and bilinearly aggregated to 0.5°. GLEAM v4.2 fields (actual ET, SMRZ) were provided at 0.25° and bilinearly resampled to 0.5°. CRU TS v4.09 precipitation and PET were provided at 0.5° and used without spatial resampling. GLOBMAP LAI was provided at approximately 0.072° (8 km) and aggregated to 0.5° by area-weighted averaging of valid (non-fill) pixels. GLDAS-2.1 runoff was provided at 0.25° and bilinearly aggregated to 0.5°. UTrack source-region weighting fields were provided at 0.5° and applied directly without resampling. Grid cells were included in the analysis if they fell within China's administrative boundary (including boundary-intersecting cells) and had valid, non-fill values for all predictor and response variables across all 120 months, yielding 4,205 valid grid cells. The aridity index for each cell was computed as:

$$AI_i = \frac{\bar{P}_i}{P\bar{E}T_i} \tag{5}$$

where $\bar{P}_i$ and $P\bar{E}T_i$ are the 2008–2017 mean annual precipitation and potential evapotranspiration at grid *i*, respectively (both from ERA5). Grid cells were classified as hyper-arid (AI<0.20), arid–semi-arid (0.20-0.40), dry sub-humid (0.40-0.65), or humid (>0.65).

Text S2. Counterfactual perturbation experiments

To probe the directional sensitivity of the blue–green coefficient to upwind forcing beyond the permutation necessity tests, we designed counterfactual experiments in which upwind predictor values were replaced by fixed reference values and the GW-RF was retrained.

Two counterfactual conditions were evaluated at each grid cell:

1) Median substitution. All five upwind predictors (LAI_{upwind}, $Precip_{upwind}$, $SMRZ_{upwind}$, PET_{upwind}, $T2M_{upwind}$) were replaced by their grid-wise temporal median across the 120-month record, simulating a hypothetical scenario in which source-region conditions are frozen at their long-run average state.
2) Upper-quartile substitution. All five upwind predictors were replaced by their grid-wise 75th-percentile values, simulating systematically wetter and more vegetated source regions.

Under each condition, local GW-RF models were retrained with the same adaptive bandwidth and kernel weights as the baseline model. Block cross-validated predictions were obtained under the substituted predictor sets. The counterfactual impact was quantified as:

$$\Delta BG_i^{CF} = BG_i^{substituted} - BG_i^{baseline} \tag{6}$$

Positive ΔBG_i^{CF} indicates a shift toward a greener (higher ET share) partitioning state under the substituted upwind conditions; negative values indicate a shift toward bluer (higher runoff share) partitioning. Spatial maps of the mean counterfactual impact under both conditions are provided in Figure S7.

The median-substitution experiment removes inter-annual variability in upwind forcing and isolates whether the long-run source-region state systematically biases the partitioning ratio. The upper-quartile experiment perturbs source regions toward plausibly greener and wetter states and tests whether the directional response is consistent with the inverted-U PDP shapes identified in Section 3.5, that is, whether grid cells already in the intermediate LAI_{upwind} or $SMRZ_{upwind}$ range respond positively to increases, while those already in the high-value range do not.

These experiments are presented as supplementary directional sensitivity evidence. They share the limitation of the main analysis in that the GW-RF is retrained on observational data; the counterfactual scenarios are statistical extrapolations rather than coupled land-atmosphere perturbation runs, and their results should be interpreted accordingly.

Figure List

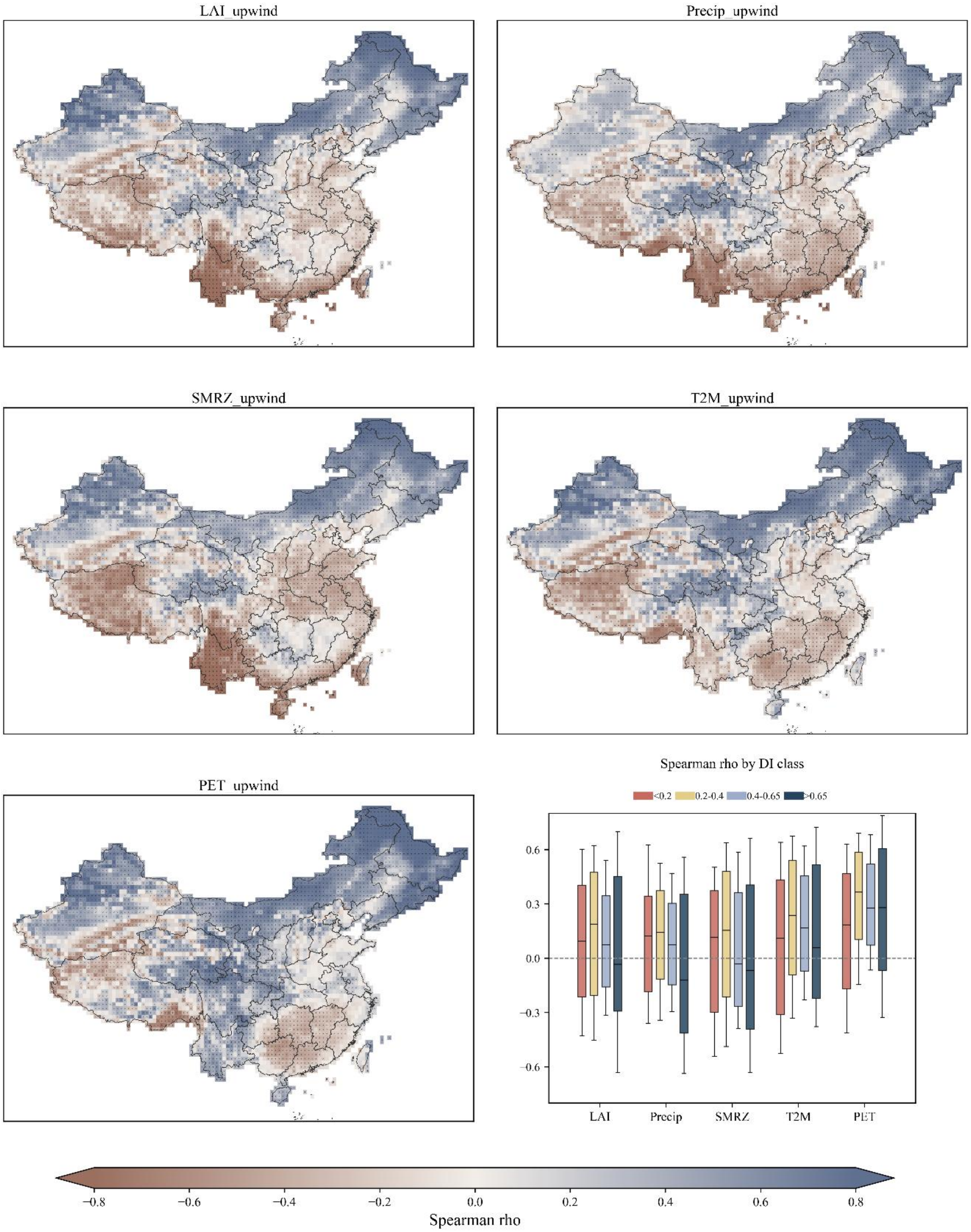

Figure S1. Spatial patterns of Spearman correlation (ρ) between the blue–green water coefficient and local drivers across China. Red denotes positive and blue negative correlations; "+" marks $p<0.01$ and "×" marks $0.01 \le p<0.05$; black lines indicate provincial boundaries; the boxplots summarize each driver's value distribution by aridity classes.

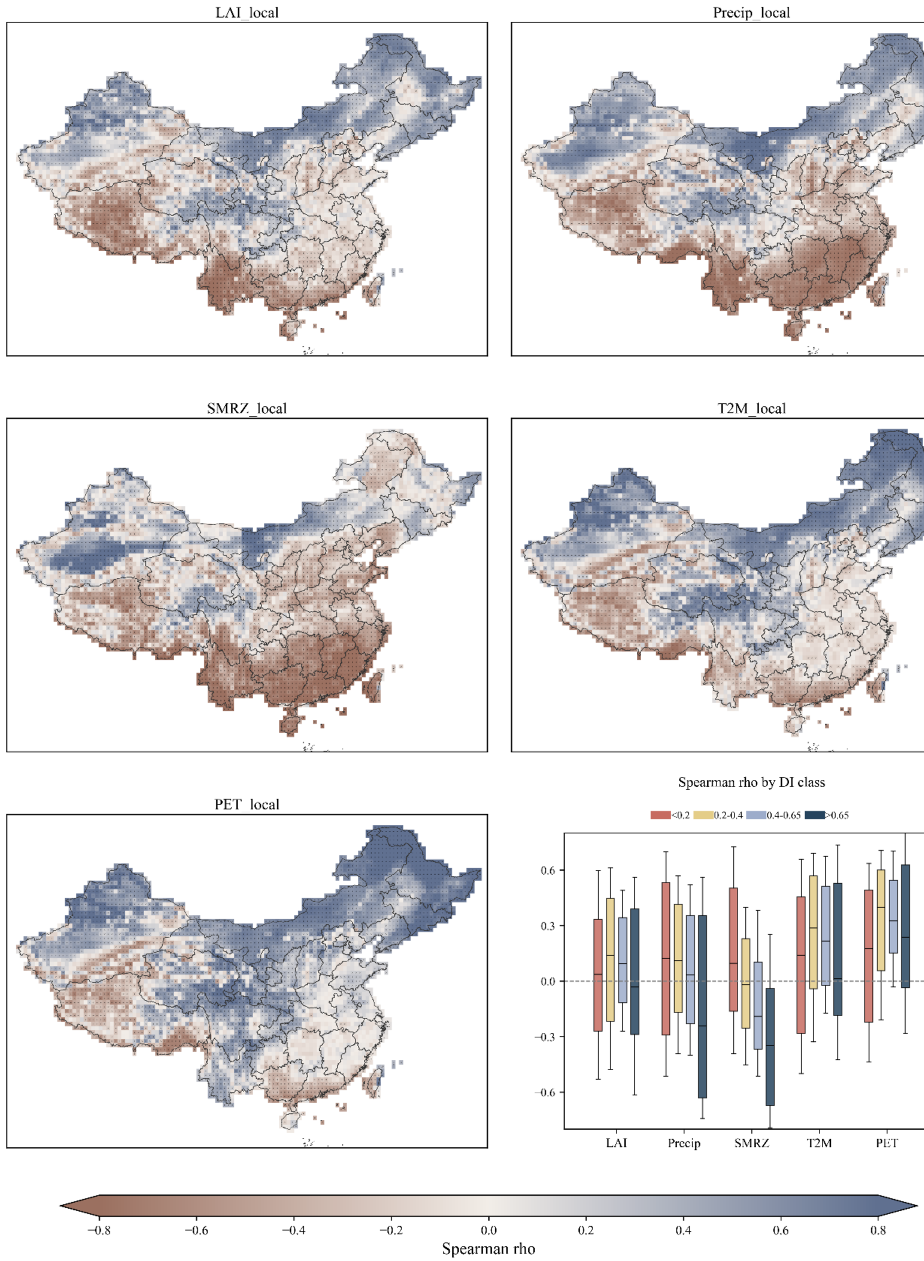

Figure S2. Spatial patterns of Spearman correlation (ρ) between the blue–green water coefficient and upwind drivers across China. Red denotes positive and blue negative correlations; "+" marks $p<0.01$ and "×" marks $0.01 \le p<0.05$; black lines indicate provincial boundaries; the boxplots summarize each driver's value distribution by aridity classes.

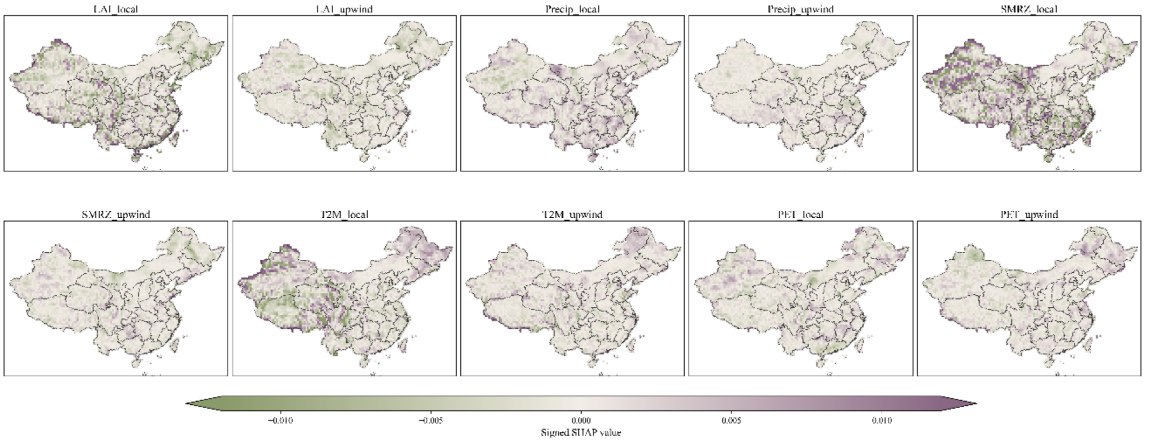

Figure S3. Spatial patterns of signed SHAP values for local and upwind factors of the blue–green water coefficient. The panels show the spatial distribution of grid-cell mean SHAP values for the ten primary covariates (five local: precipitation, potential evapotranspiration, root-zone soil moisture, leaf area index, near-surface temperature; five upwind: matched aggregates from continental moisture-source regions).

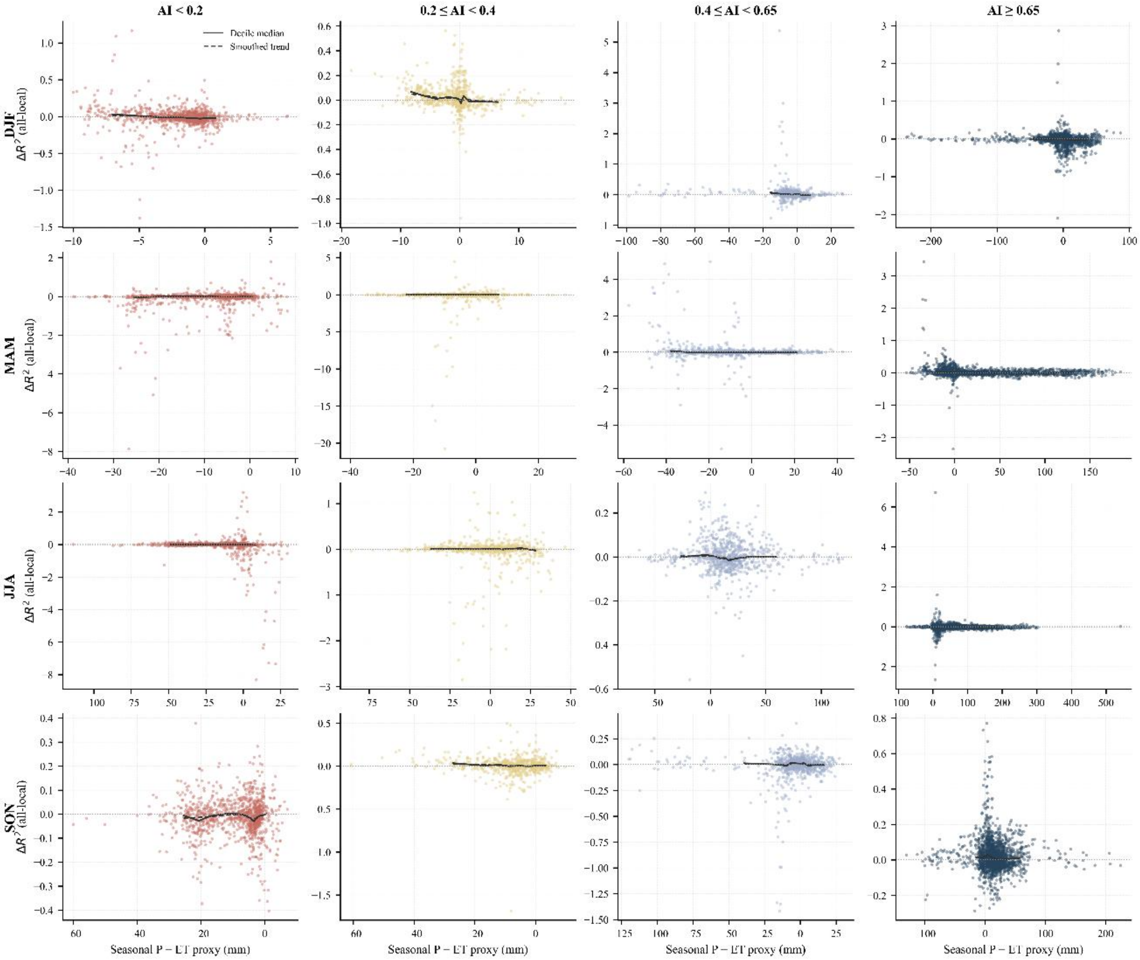

Figure S4. Seasonal relationship between the P − ET proxy and the incremental explanatory gain of upwind predictors. Each panel shows the relationship between the seasonal P − ET proxy and ΔR² (all-local) for one season and aridity class. Rows denote seasons (DJF, MAM, JJA, and SON), and columns denote aridity classes. Points represent grid cells; solid black lines show decile medians, and dashed black lines show smoothed trends. The P − ET proxy is used as an auxiliary water-balance consistency diagnostic rather than as an independent causal validation of the upwind pathway.

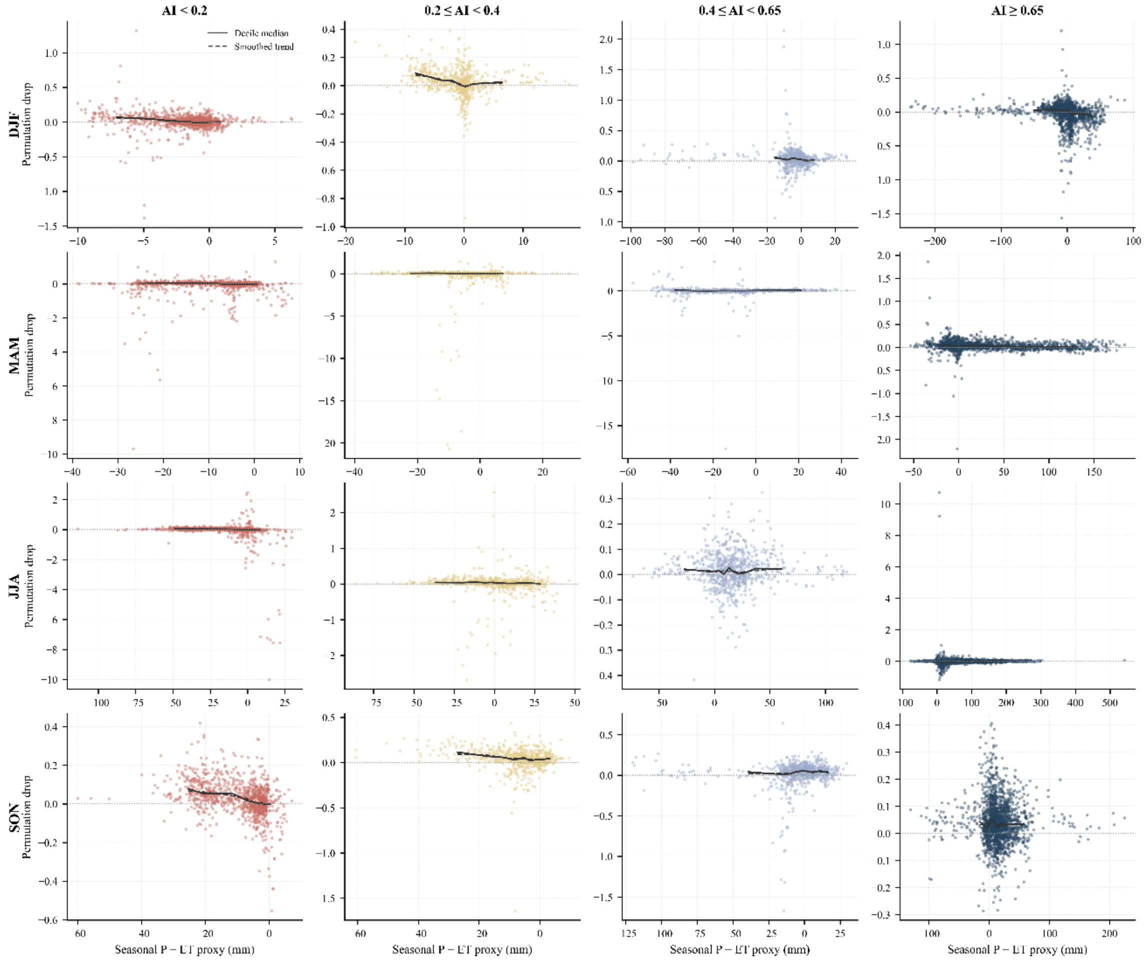

Figure S5. Seasonal relationship between the P − ET proxy and the permutation-based importance of upwind predictors. Each panel shows the relationship between the seasonal P − ET proxy and the permutation drop of upwind predictors for one season and aridity class. Rows denote seasons (DJF, MAM, JJA, and SON), and columns denote aridity classes. Points represent grid cells; solid black lines show decile medians, and dashed black lines show smoothed trends. The P − ET proxy provides seasonal physical context for the upwind attribution diagnostics but should be interpreted as an auxiliary diagnostic consistency check, not as independent causal validation.

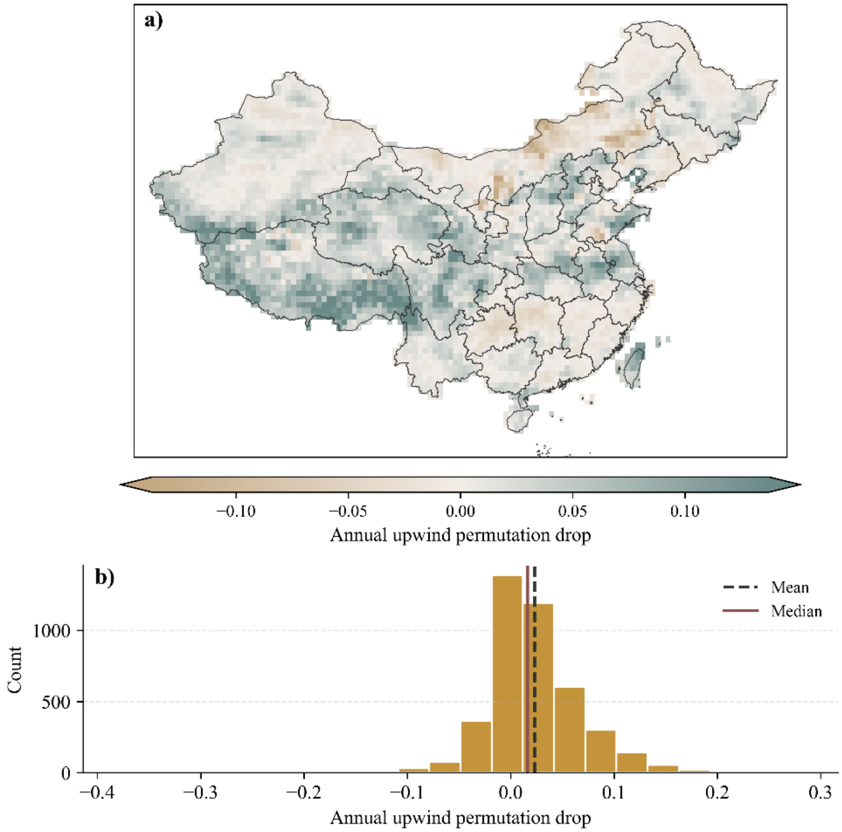

Figure S6. Upwind permutation-drop: spatial map and heterogeneity histogram. (a) Spatial map of the mean loss in block-CV R^2 when all upwind predictors are permuted within folds. (b) Across-grid histogram of permutation-drop values, summarizing spatial heterogeneity. Values are averaged over cross-validation folds; units are absolute R^2 loss.

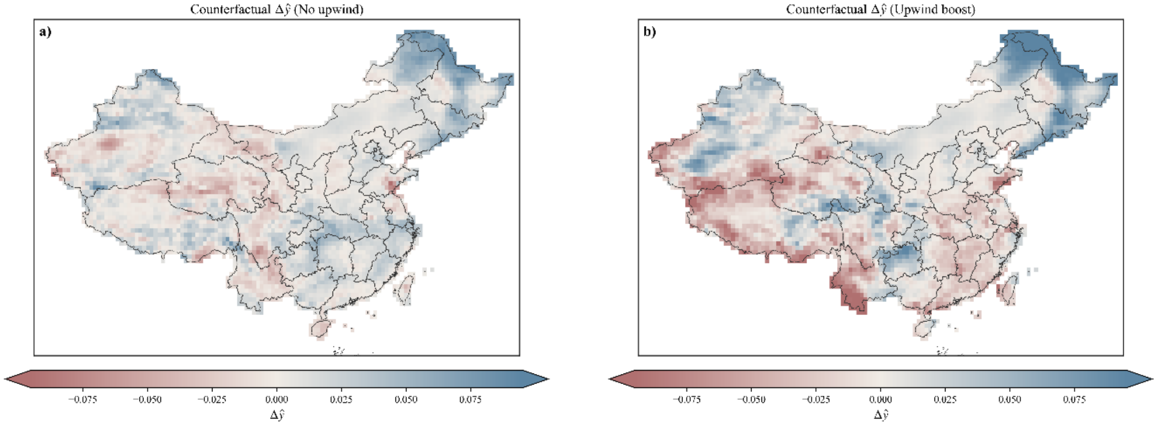

Figure S7. Counterfactual perturbations to upwind conditions: spatial responses of monthly blue–green partition. (a) Δŷ under the "no-upwind" scenario, where all upwind predictors are set to their within-fold median; (b) Δŷ under the "upwind-boost" scenario, where all upwind predictors are set to their within-fold upper quartile; Δŷ denotes the difference between the counterfactual and the baseline model; positive values indicate a higher green-water share and negative values indicate a shift toward blue-water.

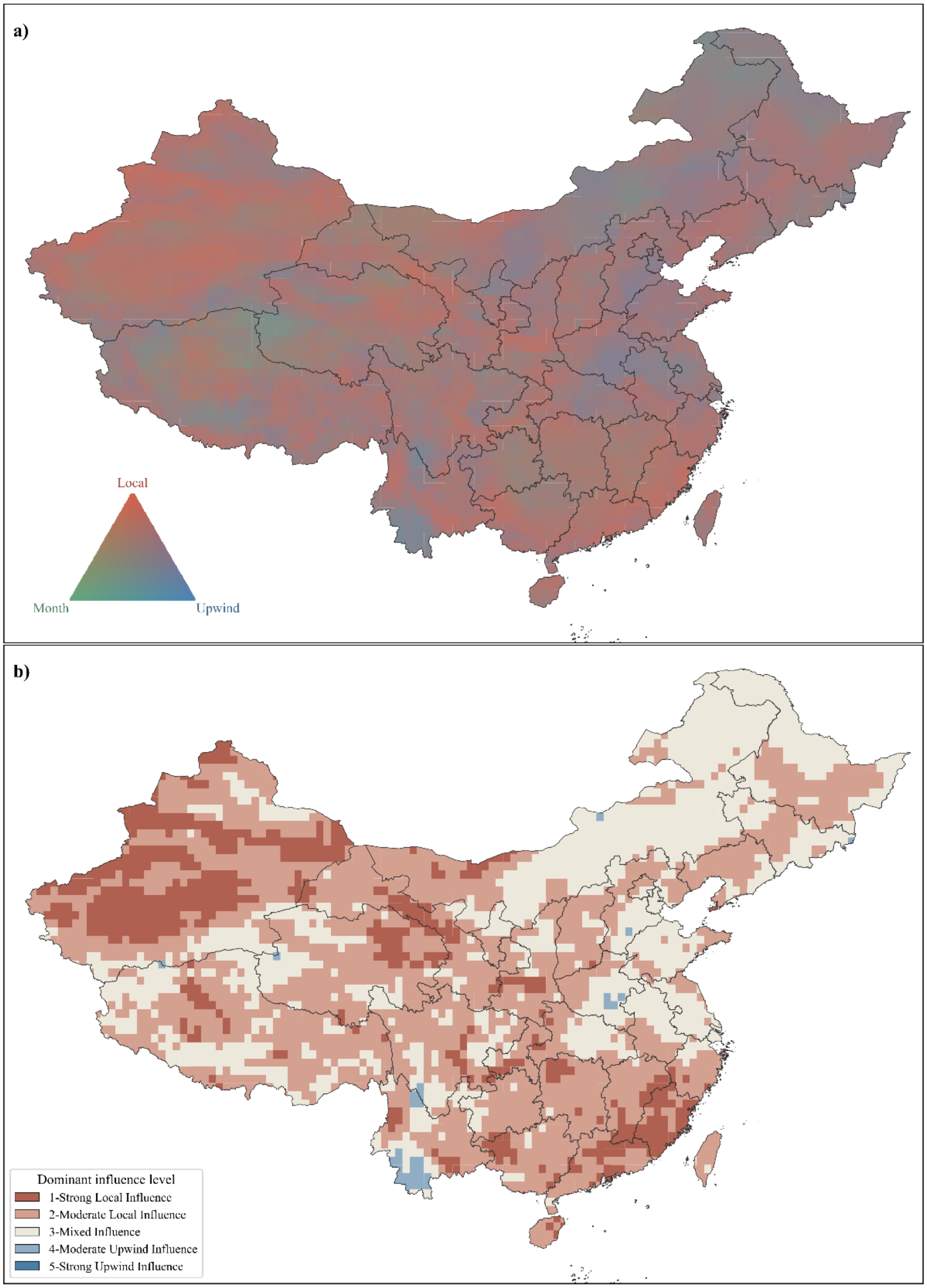

Figure S8. Ternary composition of local, upwind, and seasonal controls on the blue–green water coefficient and SHAP-based dominance classes across China. RGB composite of group-wise importance, where red encodes the normalized contribution of local factors, green encodes upwind land-sourced factors, and blue encodes seasonal harmonics.