



Multidimensional drought-propagation modes across U.S. catchments under methodological uncertainty

Tolga Barış Terzi¹, İsmail Buğra Bölükbaşı², Mehmet Asaf Düzen²

¹Department of engineering, Karadeniz Technical University, Trabzon, 61080, Türkiye

5 ²Department of engineering, Yalova University, Yalova, 77200, Türkiye

Correspondence to: İsmail B. Bölükbaşı (ismail.bolukbasi@yalova.edu.tr)

Abstract. Meteorological drought is transformed as it propagates through catchments, but estimated propagation behaviour depends on how drought events are defined and linked. We analysed meteorological-to-hydrological drought propagation across CAMELS-US using four non-redundant response dimensions: propagation probability, onset lag, severity
10 transmission, and recovery extension. Susceptibility was evaluated across 162 combinations of 27 event definitions, three criterion-weighting methods, and two multicriteria ranking methods. Of 488 event-reliable catchments, 438 had sufficient coverage for the methodological ensemble. Catchment classifications derived from a single baseline event definition differed substantially from those obtained after integrating across event definitions (adjusted Rand index = 0.256). The integrated analysis identified two broad response modes: a rapid transmission–attenuation mode and a delayed persistence mode. These
15 modes were comparatively stable under bootstrap resampling (median adjusted Rand index = 0.937), although they overlapped within a continuous multidimensional response space. Continuous susceptibility was largely insensitive to including duration transmission as a fifth criterion (Spearman $\rho = 0.990$), while approximately half of rank uncertainty arose from interactions among methodological choices. Susceptibility remained moderately predictable under spatially blocked validation, with random-forest R^2 decreasing from 0.424 under random cross-validation to 0.245 under leave-one-HUC2-out
20 validation. Climate provided most of the transferable predictive information. The transmission–persistence association was weak and became uncertain after adjustment for hydroclimate. Overall, integrating plausible methodological choices produced more reproducible drought-propagation structure than reliance on a single event definition.

1 Introduction

Meteorological drought is not transmitted unchanged through catchments. Precipitation and atmospheric water-demand
25 anomalies are filtered through soil moisture, snow accumulation and melt, evapotranspiration, subsurface storage, and runoff-generation processes before appearing as hydrological drought. These processes can delay hydrological drought onset, attenuate or intensify the propagated deficit, modify its duration, and extend hydrological recovery beyond the end of the meteorological drought. Consequently, similar meteorological deficits can produce markedly different streamflow responses among catchments (Van Loon and Van Lanen, 2012; Van Loon et al., 2012). Understanding this transformation is
30 important not only for characterizing drought hazard but also for identifying where hydrological impacts may emerge rapidly or persist after meteorological conditions begin to recover.



Large-sample studies increasingly demonstrate that drought propagation is multidimensional and strongly heterogeneous in space. Xiong et al. (2025), using matched meteorological and hydrological drought events across 1089 catchments, showed that drought severity and intensity are frequently attenuated during propagation and that the magnitude of this transformation varies with climatic and catchment characteristics. Aryal et al. (2024) found substantial spatial variation in both propagation and recovery lags across 407 Australian catchments and demonstrated that estimated lags depend on drought timescale and on the criteria used to define drought onset and cessation. At the global scale, Liu et al. (2026) documented pronounced variability in meteorological-to-runoff and meteorological-to-soil-moisture propagation using multiple datasets and alternative propagation frameworks. Together, these studies show that no single characteristic such as onset lag, propagation probability, or severity transformation can fully represent how a catchment converts meteorological drought into hydrological drought.

Catchment-scale drought behaviour has therefore also been examined using multivariate classification. Konapala and Mishra (2020) identified distinct hydrological drought regimes across 652 minimally disturbed U.S. catchments from drought frequency, duration, and intensity and related those regimes to climatic and catchment controls. Process-based analysis by Apurv and Cai (2020) further demonstrated that contrasting hydroclimatic settings and storage–discharge behaviour can produce markedly different hydrological drought characteristics across the contiguous United States. More recently, multivariate event metrics, dimensionality reduction, and clustering have been used to regionalize drought-propagation behaviour in other propagation pathways, including meteorological-to-agricultural drought (Yan et al., 2026). These studies establish that recurring multivariate drought-response patterns can be identified across heterogeneous landscapes. The existence of a clustering solution, however, does not by itself establish that catchment membership is robust to the methodological decisions used to identify and match the underlying drought events.

This distinction is important because drought-propagation characteristics are conditional on analytical choices. Drought thresholds determine which months are classified as drought, event-merging rules affect event duration and severity, and matching rules determine which meteorological and hydrological events are treated as corresponding events. Aryal et al. (2024), for example, showed that propagation and recovery estimates can vary materially with drought criteria. Similar sensitivity can arise when multiple propagation characteristics are synthesized into an overall susceptibility measure, because criterion weighting and aggregation affect the relative influence of different response dimensions. A classification derived from one defensible combination of these choices may therefore appear precise while representing only one realization of a broader set of plausible methodological assumptions.

Recent studies have therefore moved beyond single fixed formulations. Liu et al. (2026) compared propagation characteristics across multiple datasets and alternative methodological frameworks, while Yan et al. (2026) quantified how a baseline propagation regionalization changed under alternative drought definitions. These approaches make methodological sensitivity explicit, but they primarily compare parallel formulations or perturb a baseline representation. A complementary question is what catchment-level propagation structure emerges when plausible methodological alternatives are integrated into the primary inference itself, and whether its sensitivity to analytical choices is distinct from its stability to catchment



resampling. Here we refer to the former as structural methodological robustness and the latter as sampling stability. A partition may be highly stable under bootstrap resampling while still being strongly conditional on the particular event definition from which it was derived; conflating the two can therefore overstate evidence for sharply defined drought-propagation modes.

70 A related challenge concerns redundancy among propagation metrics. Transmission frequency, timing, magnitude, and recovery describe complementary aspects of drought propagation, but not all commonly used metrics are structurally independent. For a matched meteorological–hydrological event pair, the difference in event duration is algebraically determined by onset lag and recovery extension. Treating all three timing quantities as independent dimensions can therefore give repeated weight to the same event chronology. A multidimensional susceptibility formulation should preserve
 75 physically distinct aspects of propagation while avoiding unnecessary duplication, and its conclusions should be tested against the excluded metric rather than assuming that additional criteria necessarily provide additional information.

Physical interpretation presents a further test of whether an integrated propagation structure is hydrologically meaningful. Climate, snow influence, vegetation, topography, soils, geology, and catchment storage have all been linked to hydrological drought behaviour, but explanatory relationships obtained within a spatially heterogeneous sample may not transfer to
 80 hydrologically distinct regions. Random cross-validation can place geographically and hydroclimatically similar catchments in both training and test sets, potentially overstating generalization. Spatially blocked validation provides a stronger test of whether identified controls contain information that can be transferred beyond the regions used to fit the model. Separating within-sample explanatory power from out-of-region transferability is therefore important when interpreting the physical controls of drought-propagation susceptibility.

85 Here, we analyse meteorological-to-hydrological drought propagation across the contiguous United States using the Catchment Attributes and MEteorology for Large-sample Studies dataset (CAMELS-US; Newman et al., 2015; Addor et al., 2017a). Meteorological drought is represented using a precipitation–evapotranspiration balance with Hargreaves–Samani potential evapotranspiration, and hydrological drought is represented from monthly runoff. Soil-moisture drought is additionally used to diagnose the intermediate meteorological–soil-moisture–hydrological pathway. We describe primary
 90 meteorological-to-hydrological propagation using four non-redundant characteristics: propagation probability, onset lag, severity transmission ratio, and recovery extension. Duration transmission is retained as a diagnostic sensitivity metric rather than as a primary susceptibility dimension because of its exact timing relationship with onset lag and recovery extension.

Rather than assigning catchments a susceptibility or response mode from one event definition, we integrate information across 27 plausible event definitions and evaluate continuous susceptibility across 162 combinations of event definition,
 95 criterion weighting, and multicriteria ranking method. We then compare the response structure obtained from a single baseline definition with the structure that emerges after event-definition integration, quantify basin-specific methodological uncertainty and its interactions, and distinguish this structural robustness from bootstrap sampling stability. Finally, we relate the resulting response modes and continuous susceptibility to catchment attributes and evaluate predictive transfer using leave-one-HUC2-out validation.



100 Accordingly, we address three questions: (1) What multidimensional drought-propagation structure emerges when plausible event definitions are integrated, and how does it differ from the structure obtained under a single event definition? (2) How robust is continuous catchment susceptibility to event-definition, weighting, ranking, and criterion-selection choices, and which methodological choices contribute most to rank uncertainty? (3) Which hydroclimatic and physical catchment characteristics distinguish the resulting response modes and provide transferable information about drought-propagation
 105 susceptibility across major hydrological regions? By separating methodological robustness from sampling stability, the analysis aims to identify drought-propagation patterns that are reproducible without treating broad response modes as intrinsically discrete catchment classes.

2 Data and methods

2.1 Data sources, study period, and catchment screening

110 We analysed meteorological-to-hydrological drought propagation across the contiguous United States using the Catchment Attributes and MEteorology for Large-sample Studies dataset (CAMELS-US; Newman et al., 2015; Addor et al., 2017a). CAMELS-US contains hydroclimatic time series and climatic, topographic, vegetation, soil, geological, and hydrological attributes for 671 catchments with comparatively limited direct human influence.

The principal analysis period was January 1981–December 2014. Data from January–December 1980 were retained as a
 115 warm-up period for calculation of the meteorological drought index but were excluded from event statistics. Basin identifiers were stored as eight-digit gauge identifiers throughout the workflow.

Monthly precipitation and streamflow were taken from the CAMELS-US hydroclimatic records used to construct the monthly basin dataset. Streamflow expressed as mean daily runoff depth was converted to monthly runoff depth as

$$Q_t = q_t d_t \quad (1)$$

120 where q_t is mean runoff depth in mm d^{-1} and d_t is the number of days in month t .

Root-zone soil moisture was represented using basin-aggregated ERA5-Land volumetric soil-water content (Muñoz-Sabater et al., 2021). The first three soil layers, representing depths of 0–7, 7–28, and 28–100 cm, were combined by depth weighting to obtain a 0–100 cm root-zone soil-moisture series.

Potential evapotranspiration (PET) was calculated using the Hargreaves–Samani formulation (Hargreaves and Samani,
 125 1985). Daily minimum and maximum air temperatures were obtained from the CAMELS-US Daymet basin-mean forcing files. For each basin, daily minimum and maximum temperatures were aggregated to monthly mean daily minimum and maximum temperature and supplied to the hargreaves function in the R package SPEI using the corresponding CAMELS forcing latitude. PET was expressed in mm month⁻¹.

Data completeness was assessed independently for this study. During 1981–2014, precipitation, root-zone soil moisture,
 130 monthly runoff, and PET were each required to have at least 98 % temporal coverage. Soil moisture and runoff additionally required at least 30 finite observations for every calendar month to support month-specific empirical standardization.



Catchments were retained after index construction only when the meteorological, soil-moisture, and runoff drought indices each had at least 98 % finite observations.

A further event-reliability criterion was applied within each drought-event definition. Catchments were required to contain at least eight meteorological drought events and at least eight matched meteorological-to-hydrological drought-event pairs, together with finite values for all propagation characteristics required by the corresponding analysis. The primary baseline sample contained 488 catchments. For analyses integrating across alternative event definitions, catchments were retained when sufficient information was available under at least 80% of the 27 event definitions, yielding 438 ensemble-eligible catchments.

2.2 Drought indices

Three monthly drought states were considered: meteorological drought, soil-moisture drought, and hydrological drought represented by runoff.

Meteorological drought was quantified using the 1-month Standardized Precipitation Evapotranspiration Index (SPEI-1; Vicente-Serrano et al., 2010). Monthly climatic water balance was calculated as:

$$B_t = P_t - PET_t \quad (2)$$

where P_t is monthly precipitation and PET_t is Hargreaves PET. SPEI was fitted using the log-logistic distribution and unbiased probability-weighted moments implemented in the R package SPEI (Beguería et al., 2013), using 1981–2014 as the reference period. A 1-month accumulation scale was used to avoid introducing additional persistence through overlapping multi-month accumulation windows.

Root-zone soil moisture was standardized separately for each calendar month using non-parametric normal scores (Farahmand and AghaKouchak, 2015). For a reference sample of size n_m in calendar month m , the plotting position associated with value x was:

$$p(x) = \frac{r(x) - 0.44}{n_m + 0.12} \quad (3)$$

where $r(x)$ is the midpoint rank of x within the corresponding calendar-month reference sample (Gringorten, 1963).

Probabilities were restricted to the open interval (0,1), and the standardized soil-moisture index was:

$$SSMI = \Phi^{-1}[p(x)] \quad (4)$$

where Φ^{-1} denotes the inverse standard-normal cumulative distribution function.

Monthly runoff was standardized analogously to obtain a standardized runoff index (SRI; Shukla and Wood, 2008), while explicitly treating zero and near-zero runoff as a discrete probability mass. Runoff values ≤ 0.01 mm month⁻¹ were classified as zero flow. For each calendar month, let q_0 denote the fraction of reference-period runoff observations at or below this threshold. Zero-flow observations were assigned probability $q_0/2$. Positive-flow observations were ranked within the positive component and assigned probabilities conditional on $1-q_0$, after which the combined probability was transformed using Φ^{-1} .



SPEI, SSML, and SRI were therefore all represented on approximately standard-normal scales, with increasingly negative values denoting increasingly dry conditions.

165 2.3 Drought-event identification

Drought events were identified separately from SPEI, SSML, and SRI using run theory (Yevjevich, 1967). A month was classified as drought when:

$$I_t \leq \tau \quad (5)$$

where I_t denotes the corresponding standardized drought index and τ is the selected drought threshold.

170 Three thresholds were examined:

$$\tau \in \{-0.5, -0.8, -1.0\} \quad (6)$$

Consecutive drought months formed an initial event. Adjacent drought runs separated by no more than g non-drought months were merged, where:

$$g \in \{0, 1, 2\} \quad (7)$$

175 The baseline event definition used $\tau = -0.5$ and $g = 2$. Event duration was measured from the first to the last month of the merged event and therefore included any intervening non-drought months contained within the merged event. At least one actual drought month was required for an event to be retained.

Event severity was calculated from standardized drought deficits only. For event i ,

$$S_i = \sum_{t \in D_i} (-I_t) \quad (8)$$

180 where D_i contains only months within the event for which $I_t \leq \tau$. Non-drought months included within a merged event contributed to event duration but not to severity.

2.4 Event matching and propagation characteristics

Meteorological-to-hydrological ($M \rightarrow H$), meteorological-to-soil-moisture ($M \rightarrow SM$), and soil-moisture-to-hydrological ($SM \rightarrow H$) events were linked using a target-driven matching algorithm.

185 For a source event i and target event j , the target was eligible when:

$$t_{i,start} \leq t_{j,start} \leq t_{i,end} + w \quad (9)$$

where w is a post-source matching window of 0, 1, or 2 months. Targets were processed chronologically and linked to the most recent eligible unused source event. When multiple source events were eligible, the source with the smallest non-negative onset lag was selected. Source reuse was not allowed, so each source drought could contribute to at most one target

190 response. The baseline matching definition used $w = 0$.

For $M \rightarrow H$ propagation, five catchment-level characteristics were calculated.

Propagation probability was:



$$P_{MH} = \frac{N_{MH}}{N_M} \quad (10)$$

where N_M is the number of meteorological drought events and N_{MH} is the number of matched M→H pairs. The primary denominator included all meteorological droughts, including events whose onset occurred while a hydrological drought was already active. This convention provides a consistent event-count denominator across catchments but does not imply that every meteorological event represents an isolated causal forcing episode.

For matched event i , onset lag was:

$$L_i = t_{H,start,i} - t_{M,start,i} \quad (11)$$

and catchment-level onset lag L was the median across matched pairs. Smaller values indicate more rapid propagation.

Severity transmission for matched event i was:

$$R_{S,i} = \frac{S_{H,i}}{S_{M,i}} \quad (12)$$

with catchment-level severity transmission R_S defined as the median matched-event ratio. Values greater than 1 indicate greater standardized hydrological drought severity relative to the matched meteorological drought, whereas values below 1 indicate attenuation.

Recovery extension was:

$$E_i = t_{H,end,i} - t_{M,end,i} \quad (13)$$

with catchment-level E represented by the median across matched pairs. Positive values indicate that hydrological drought persisted beyond meteorological-drought termination.

Duration transmission was also calculated as:

$$R_{D,i} = \frac{D_{H,i}}{D_{M,i}} \quad (14)$$

and summarized by its catchment median R_D . However, duration transmission was not included in the primary susceptibility formulation because matched-event timing obeys the exact identity:

$$D_H - D_M = E - L \quad (15)$$

Thus, duration difference contains timing information that is already determined by onset lag and recovery extension. Duration transmission was therefore retained as a diagnostic sensitivity characteristic rather than treated as a fifth independent primary criterion.

The primary multidimensional propagation representation consequently comprised four characteristics: P_{MH} , L , R_S , E .

Propagation probability, severity transmission, and recovery extension were treated as benefit criteria for susceptibility, whereas onset lag was treated as a cost criterion because shorter lag indicates more rapid propagation.

M→SM and SM→H propagation probabilities and lags were also calculated as pathway diagnostics. Complete M→SM→H chains were identified by linking M→SM and SM→H matches through the same intermediate soil-moisture event. These



quantities, together with the fraction of hydrological droughts without an eligible meteorological precursor, were used for
 225 process interpretation but were not included in the primary susceptibility index.

2.5 Methodological ensemble and consensus propagation susceptibility

Methodological uncertainty was evaluated across 27 drought-event definitions formed by the full factorial combination of
 three drought thresholds, three event-merging gaps, and three matching windows.

Within each event-definition scenario, a catchment was rankable only when it contained at least eight meteorological
 230 drought events, at least eight matched M→H event pairs, and finite values for all four primary propagation criteria.

Before multicriteria aggregation, each criterion was min–max normalized. For a benefit criterion x_j ,

$$x'_{ij} = \frac{x_{ij} - x_{j,min}}{x_{j,max} - x_{j,min}} \quad (16)$$

whereas onset lag was cost-oriented as:

$$x'_{ij} = 1 - \frac{x_{ij} - x_{j,min}}{x_{j,max} - x_{j,min}} \quad (17)$$

235 A criterion with no between-catchment variation within a configuration was assigned a normalized value of 0.5.

Three weighting approaches were used: equal weighting, entropy weighting (Shannon, 1948), and Criteria Importance
 Through Intercriteria Correlation (CRITIC; Diakoulaki et al., 1995). Equal weighting assigned a weight of 0.25 to each
 primary criterion. Entropy weighting assigned larger weights to criteria with greater information divergence, whereas
 CRITIC weighting incorporated both criterion variability and intercriterion correlation.

240 Two multicriteria ranking methods were applied independently: Technique for Order Preference by Similarity to Ideal
 Solution (TOPSIS; Hwang and Yoon, 1981) and VIKOR (Opricovic and Tzeng, 2004). VIKOR used $v=0.5$, and its score
 was reoriented so that larger values consistently denoted greater susceptibility.

The complete susceptibility ensemble therefore comprised 162 methodological configurations.

For each configuration, catchments were ranked from highest to lowest susceptibility. Rank percentile for basin b under
 245 configuration c was:

$$r_{bc} = \frac{R_{bc} - 1}{N_c - 1} \quad (18)$$

where R_{bc} is the basin rank and N_c is the number of catchments rankable in configuration c . Thus, $r_{bc}=0$ denotes the most
 susceptible catchment.

Consensus susceptibility was defined from the median rank percentile across all valid configurations:

$$S_b = 1 - \text{median}_c(r_{bc}) \quad (19)$$

Values approaching 1 therefore denote consistently high propagation susceptibility. A catchment was considered robust-
 eligible when rankings were available for at least 80 % of the 162 successful configurations.



Rank uncertainty for each basin was summarized using the interquartile range of r_{bc} . Rank acceptability was additionally represented by the fraction of valid configurations in which a catchment occurred in the upper susceptibility quartile:

$$A_{25,b} = \frac{1}{C_b} \sum_c 1(r_{bc} \leq 0.25) \quad (20)$$

where C_b is the number of valid configurations for basin b .

As a criterion-selection sensitivity analysis, the complete susceptibility calculation was compared with the corresponding five-criterion formulation that additionally included duration transmission.

2.6 Decomposition of methodological rank uncertainty

For catchments represented in every configuration of the complete factorial design, basin-specific rank variability was decomposed among five methodological factors: drought threshold, event-merging gap, matching window, criterion-weighting method, and multicriteria ranking method. The final balanced design consisted of 162 configurations.

Sum-to-zero contrasts were used to construct model subspaces for every main effect and interaction. For basin b , the contribution of methodological term T was quantified as:

$$V_{b,T} = \frac{SS_{b,T}}{SS_{b,total}} \quad (21)$$

where $SS_{b,T}$ is the squared projection of the centred rank-percentile vector onto the subspace associated with term T , and $SS_{b,total}$ is total rank-percentile variation across the complete factorial design.

Because the design was balanced, the components partitioned total variation apart from numerical closure error. Contributions were summarized across basins using means, medians, and interquartile ranges. Interaction terms were additionally grouped by interaction order into two-, three-, four-, and five-way components. Because summary statistics were calculated independently for each methodological component, medians across components were not expected to sum exactly to 100 %.

2.7 Event-definition-integrated propagation modes

Discrete response structure was analysed separately from the multicriteria susceptibility ranking. Rather than clustering catchments under one baseline drought-event definition, propagation characteristics were first integrated across the 27 event definitions.

For each basin, the median value of each propagation characteristic was calculated across all event definitions for which that basin satisfied the event-reliability criteria. Event-definition coverage was defined as the number of valid definitions divided by 27. Catchments with at least 80 % coverage and finite propagation characteristics were retained, yielding 438 ensemble-eligible catchments.

The preferred clustering used the four non-redundant propagation characteristics P_{MH} , L , R_s , and E . Criteria were benefit-oriented using the same direction convention as the susceptibility analysis and then standardized across catchments. K-means



clustering (Hartigan and Wong, 1979) was evaluated for $k=2$ to $k=6$. Candidate solutions were compared using mean and median silhouette width, minimum silhouette width, the proportion of negative silhouette values, and the between-cluster to total sum-of-squares ratio (Rousseeuw, 1987). The preferred number of clusters was determined primarily from mean silhouette width together with the remaining diagnostics.

The final $k=2$ solution was fitted using 200 random starts. Cluster numbers had no intrinsic hydrological meaning. The cluster with greater propagation probability and faster onset was labelled the rapid transmission–attenuation mode, whereas the contrasting cluster was labelled the delayed persistence mode after inspection of its multidimensional criterion profile.

A five-criterion clustering that additionally included duration transmission was retained as a structural sensitivity analysis. Agreement between four- and five-criterion partitions and between the single baseline-definition partition and the event-definition-integrated partition was quantified using the adjusted Rand index (ARI; Hubert and Arabie, 1985).

For visualization, a four-criterion response space was constructed after benefit orientation and standardization. A transmission axis was obtained from propagation probability and oriented onset rapidity, whereas a persistence–severity axis was obtained from severity transmission and recovery extension. These axes were used only to display the preferred multidimensional ensemble structure and were not used to determine cluster membership.

2.8 Sampling stability and transmission–persistence diagnostics

Sampling stability of the preferred event-definition-integrated $k=2$ solution was evaluated using 500 bootstrap resamples (Efron, 1979) of the 438 ensemble-eligible catchments. For each replicate, catchments were sampled with replacement and a two-cluster k-means solution was fitted to the standardized four-criterion matrix. All original catchments were then assigned to the nearest bootstrap cluster centroid. Binary cluster labels were aligned with the reference solution before comparison.

Bootstrap stability was summarized using the ARI between each bootstrap-derived partition and the reference partition, overall membership agreement, cluster-specific Jaccard similarity, and the proportion of bootstrap replicates in which each catchment retained its reference mode. This bootstrap therefore assessed sampling stability conditional on the event-definition-integrated criterion matrix; it was not treated as a full bootstrap of the raw hydroclimatic time series or of all methodological choices.

A separate diagnostic analysis examined the continuous association between transmission and persistence using the baseline five-criterion response axes retained from the original propagation framework. This diagnostic involved all 488 event-reliable catchments and was kept separate from the four-criterion ensemble clustering.

The global transmission–persistence association was summarized using Spearman's rank correlation and a percentile-bootstrap confidence interval based on 5000 resamples. Hydroclimatic conditioning was evaluated by stratifying catchments into tertiles of aridity and snow fraction, with 2500 bootstrap resamples per stratum and Benjamini–Hochberg adjustment of the corresponding p values.

A climate-adjusted partial Spearman analysis jointly controlled for aridity and snow fraction. Its confidence interval was obtained from 3000 bootstrap resamples, and statistical evidence was assessed using a two-sided permutation test with 5000



permutations. These analyses were treated as supporting diagnostics rather than as criteria for selecting the number of response modes.

2.9 Physical and hydrological interpretation

Physical controls were evaluated using CAMELS-US climatic, topographic, vegetation, soil, and geological attributes.

320 Differences between the two event-definition-integrated response modes were tested using Wilcoxon rank-sum tests (Mann and Whitney, 1947). Effect magnitude was summarized using rank-biserial correlation,

$$r_{rb} = \frac{2U}{n_1 n_2} - 1 \quad (22)$$

where U is the Mann–Whitney statistic and n_1 and n_2 are the numbers of catchments in the two modes. Positive effects were oriented to denote larger values in the rapid transmission–attenuation mode. Multiple-comparison p values were adjusted

325 using the Benjamini–Hochberg false-discovery-rate procedure (Benjamini and Hochberg, 1995).

Hydrological signatures were evaluated separately for process interpretation and were not used to define the propagation modes or primary susceptibility. These included mean runoff, runoff ratio, baseflow index, flow-duration-curve slope, half-flow date, high- and low-flow characteristics, zero-flow frequency, streamflow elasticity, and soil-moisture characteristics.

Continuous four-criterion susceptibility was related to catchment attributes using generalized additive models (GAMs; 330 Wood, 2011) and random forests (Breiman, 2001) implemented with the *ranger* package (Wright and Ziegler, 2017).

Candidate primary predictors were screened for missingness, non-degenerate information content, and pairwise redundancy. Predictors with excessive missingness were excluded, and pairwise correlation pruning used an absolute Spearman-correlation threshold of 0.85. A maximum of 12 predictors was retained for the primary susceptibility models. The final predictor set comprised aridity, snow fraction, precipitation seasonality, mean elevation, mean slope, catchment area, forest 335 fraction, soil depth, soil porosity, soil conductivity, maximum soil water content, and clay fraction.

Continuous predictors were standardized using parameters estimated from the training data. GAMs used penalized cubic-regression splines fitted by restricted maximum likelihood with term selection enabled. The random forest used 1200 trees, square-root predictor subsampling, and a minimum node size of 5. Random-forest permutation importance and GAM partial-response functions were used for interpretation.

340 Predictive skill was quantified using root-mean-square error,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (23)$$

mean absolute error,

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (24)$$

the out-of-sample coefficient of determination,



$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (25)$$

and Spearman correlation between observed and predicted susceptibility.

2.10 Random and spatially blocked validation

To distinguish interpolation within the catchment sample from spatial transferability, identical predictor sets and model specifications were evaluated using both leakage-safe random 10-fold cross-validation and leave-one-HUC2-out cross-validation (Roberts et al., 2017).

For random validation, catchments were assigned to 10 folds. For spatial validation, each two-digit Hydrologic Unit Code region was withheld in turn and the model was trained on all remaining HUC2 regions. All preprocessing was conducted independently within each cross-validation fold. Missing predictor values were imputed using medians calculated from the training catchments only. Predictor means and standard deviations were likewise estimated from the training data and then applied unchanged to the held-out catchments. This procedure prevented information from the validation subset from entering preprocessing or model fitting.

The comparison between random and HUC2-blocked validation used the same 438-basin susceptibility sample and the same 12 retained predictors for both GAM and random-forest models.

2.11 Physical-family attribution

Transferable information was further evaluated for five a priori physical-attribute families: climate, topography, vegetation, soil, and geology. This analysis used a broader screened predictor set than the 12-variable primary model so that each physical domain could be represented explicitly.

Family-level attribution was interpreted primarily using random forests because the high-dimensional full-family GAM showed limited blocked-prediction skill. For each family f , the complete random-forest model was compared with an otherwise identical model in which all predictors belonging to family f had been removed. The loss in transferable skill was

$$\Delta R_f^2 = R_{full}^2 - R_{-f}^2 \quad (26)$$

Positive values indicate deterioration in HUC2-blocked prediction after removal of that family. Negative values indicate that the omitted family did not provide additional transferable predictive information conditional on the remaining predictors; they were not interpreted as evidence that the corresponding physical processes are hydrologically unimportant.

As a complementary descriptive analysis, predictor families were added sequentially in the fixed order climate → topography → vegetation → soil → geology. Because incremental contributions depend on entry order, these models were used only to assess whether increasingly complex physical descriptions improved out-of-region prediction.



3 Results

3.1 Continental variation in consensus propagation susceptibility

Of the 671 CAMELS-US catchments considered initially, 488 satisfied the baseline event-reliability criteria. Of these, 438 retained valid propagation information for at least 80 % of the 27 event definitions and were therefore eligible for the methodological ensemble and subsequent susceptibility modelling.

Consensus drought-propagation susceptibility varied widely across the contiguous United States (Fig. 1a). High- and low-susceptibility catchments occurred across multiple hydroclimatic regions rather than forming a single contiguous spatial gradient. Among the 438 robust-eligible catchments, the median configuration coverage was 1.00, and the median catchment had valid rankings under all 162 methodological configurations.

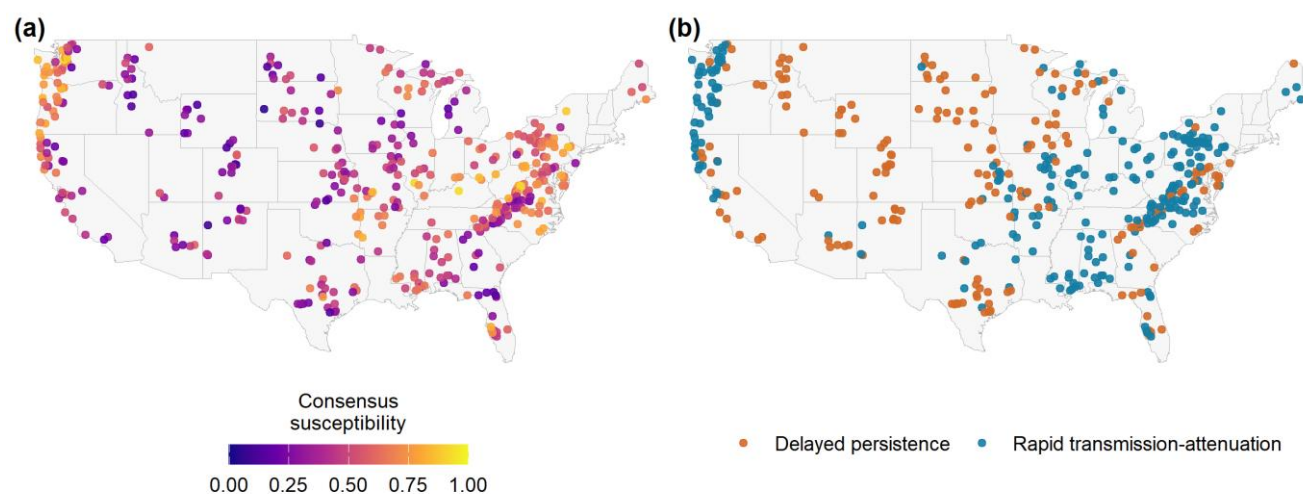
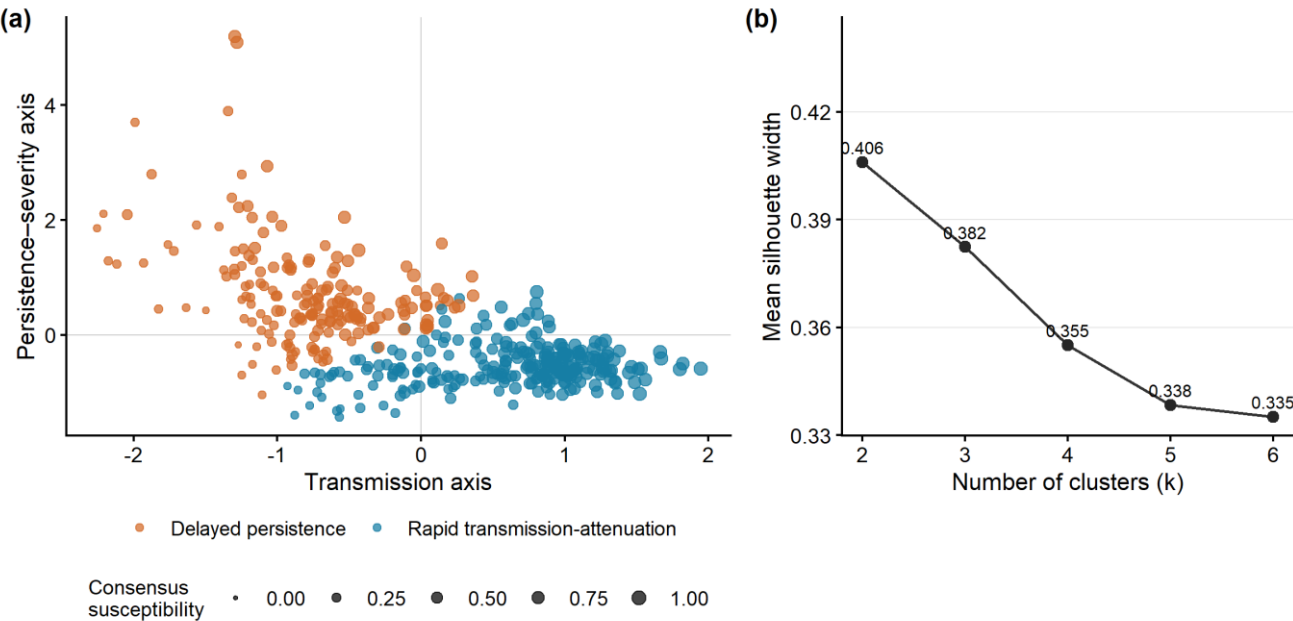


Figure 1: Continental patterns of consensus drought-propagation susceptibility and event-definition-integrated response modes. Catchment outlet locations are shown across the contiguous United States. (a) Consensus propagation susceptibility, defined as one minus the median rank percentile across the 162 event-definition $\times$ weighting $\times$ multicriteria configurations; larger values indicate consistently greater susceptibility. (b) Preferred four-criterion event-definition-integrated response modes for the 438 ensemble-eligible catchments. The rapid transmission–attenuation and delayed persistence labels describe the corresponding multivariate propagation profiles rather than discrete hydrological catchment classes.

The event-definition-integrated response modes were likewise geographically distributed across the study domain (Fig. 1b). The rapid transmission–attenuation mode contained 251 catchments, whereas 187 catchments were assigned to the delayed persistence mode. Neither mode was confined to a single region, indicating that the response structure was not simply a geographic partition of the contiguous United States.

3.2 Event-definition-integrated drought-propagation modes

395 The four primary propagation characteristics separated the 438 ensemble-eligible catchments into two broad response modes (Fig. 2; Table 1).



400 **Figure 2: Multidimensional structure of the event-definition-integrated drought-propagation modes. (a) Four-criterion response space for the 438 ensemble-eligible catchments. The transmission axis combines standardized propagation probability and onset rapidity, whereas the persistence-severity axis combines standardized severity transmission and recovery extension. Point size denotes consensus propagation susceptibility, and colour denotes the preferred (k=2) response mode. (b) Mean silhouette width for candidate four-criterion k-means solutions from (k=2) to (k=6). The maximum mean silhouette width occurred at (k=2).**

405 **Table 1: Propagation characteristics of the two event-definition-integrated response modes. Values are medians [interquartile ranges]. Positive rank-biserial effects indicate larger values in the rapid transmission-attenuation mode. P values are Benjamini-Hochberg adjusted. Duration transmission is retained only as a diagnostic sensitivity characteristic because of its exact timing dependence on onset lag and recovery extension.**

Characteristic	Rapid transmission-attenuation (n = 251)	Delayed persistence (n = 187)	Rank-biserial effect	FDR-adjusted p	Role
Propagation probability	0.479 [0.418–0.522]	0.333 [0.280–0.379]	0.813	<0.001	Primary criterion
Onset lag (months)	0.000 [0.000–0.250]	1.000 [1.000–1.000]	-0.744	<0.001	Primary criterion
Duration transmission ratio	1.000 [1.000–1.000]	1.000 [1.000–1.000]	-0.145	<0.001	Diagnostic sensitivity
Severity transmission	0.938 [0.886–1.004]	1.055 [0.975–	-0.557	<0.001	Primary criterion



Characteristic	Rapid transmission– attenuation (n = 251)	Delayed persistence (n = 187)	Rank- biserial effect	FDR- adjusted p	Role
ratio		1.170]			
Recovery extension (months)	0.000 [0.000–0.000]	1.000 [1.000– 1.000]	-0.910	<0.001	Primary criterion

Candidate k-means solutions from $k = 2$ to $k = 6$ gave the highest mean silhouette width for $k = 2$ (0.406). Mean silhouette width decreased to 0.382 for $k = 3$, 0.355 for $k = 4$, 0.338 for $k = 5$, and 0.335 for $k = 6$ (Fig. 2b; Fig. A2; Table A3). The $k =$
 410 2 solution was therefore retained as the primary event-definition-integrated representation. When duration transmission was added as a fifth criterion, the silhouette optimum shifted to $k = 3$ (mean silhouette = 0.396, compared with 0.381 for $k = 2$; Fig. A2; Table A4), showing that the preferred discrete partition depends on criterion formulation. The five-criterion $k = 2$ profiles nevertheless retained the same broad higher- versus lower-transmission contrast (Table A2) and were used only for a like-for-like comparison with the primary $k = 2$ solution.

415 Catchments in the rapid transmission–attenuation mode had a higher median M→H propagation probability than those in the delayed persistence mode (0.479 versus 0.333; rank-biserial effect = 0.813; Table 1). Median onset lag was 0 months in the rapid mode and 1 month in the delayed mode (rank-biserial effect = −0.744). Recovery extension showed the largest separation, with medians of 0 and 1 month, respectively (rank-biserial effect = −0.910). Severity transmission was also lower in the rapid mode, with a median ratio of 0.938 compared with 1.055 in the delayed mode (rank-biserial effect = −0.557).
 420 Thus, the two modes differed jointly in transmission frequency, timing, severity transformation, and recovery rather than along a single propagation characteristic.

Duration transmission, which was excluded from the primary clustering because of its timing redundancy, had a median of 1.0 in both modes and substantially weaker separation than the primary criteria (rank-biserial effect = −0.145; Table 1). The distributions of individual characteristics nevertheless overlapped between modes (Fig. A1), consistent with a continuous
 425 multidimensional response space rather than sharply separated catchment classes.

The response structure also depended materially on whether event definition was fixed or integrated. Agreement between the cluster assignment produced under the single baseline event definition and the preferred four-criterion ensemble partition was low (ARI = 0.256; Table 2). For the fixed-k criterion-sensitivity comparison, the four-criterion and five-criterion $k = 2$ ensemble partitions agreed strongly (ARI = 0.876; Table A5). Of the 438 ensemble-eligible catchments, 424 (96.8%)
 430 received the same broad mode label in the four- and five-criterion $k = 2$ solutions.

Table 2: Summary of methodological robustness, structural sensitivity, sampling stability, and the diagnostic transmission–persistence analysis. ARI denotes adjusted Rand index. Transmission–persistence results use the separate five-criterion diagnostic axes for the 488 baseline event-reliable catchments.

Analysis	Result	Interpretation
Baseline event-reliable sample	488 catchments	Primary Hargreaves sample



Analysis	Result	Interpretation
Ensemble-eligible sample	438 catchments	≥80% coverage across 27 event definitions
Median valid event definitions	27 of 27	Full median event-definition coverage
Preferred four-criterion clustering	k = 2; mean silhouette = 0.406	Event-definition-integrated solution
Baseline vs ensemble partition	ARI = 0.256	Single-definition vs method-integrated partition
Four- vs five-criterion ensemble partition	ARI = 0.876	Sensitivity to including duration transmission
Four- vs five-criterion susceptibility	Spearman ρ = 0.990; median $ \Delta S $ = 0.0166	Continuous susceptibility sensitivity
Ensemble-mode bootstrap stability	Median ARI = 0.937 (95% CI: 0.776–1.000)	500 bootstrap resamples
Basin membership stability	Median = 1.000; 93.8% ≥ 0.90	Bootstrap same-mode probability
Global transmission–persistence association	ρ = −0.124 (95% CI: −0.214 to −0.027)	Diagnostic five-criterion axes; n = 488
Climate-adjusted association	Partial ρ = −0.080 (95% CI: −0.168 to 0.008); p = 0.088	Adjusted for aridity and snow fraction; permutation p value

435 Sampling stability of the preferred ensemble partition was high. Across 500 bootstrap resamples, the median ARI relative to the reference partition was 0.937, with a 95% interval of 0.776–1.000 (Table 2; Table A6; Fig. A11). Median basin-specific same-mode probability was 1.00, and 93.8% of catchments had membership stability of at least 0.90. These bootstrap results quantify sampling stability conditional on the event-definition-integrated representation and are distinct from the structural sensitivity to the choice of event definition.

440 3.3 Methodological uncertainty in consensus susceptibility

All 162 combinations of drought threshold, event-merging gap, matching window, weighting approach, and ranking method produced valid susceptibility configurations. A total of 438 catchments satisfied the ≥80% configuration-coverage requirement, while 368 had complete observations across the full factorial design used for the exact uncertainty decomposition. Pairwise agreement among the six weighting–ranking combinations under the baseline event definition is shown in Fig. A3.

445 Although consensus susceptibility provided a continuous ordering of catchments, exact rank positions varied appreciably among methodological configurations (Fig. 3b). Among the 438 robust-eligible catchments, the median rank-percentile interquartile range was 0.367. Top-quartile membership probability also changed gradually across the susceptibility ranking rather than separating catchments into deterministic high- and low-susceptibility groups.

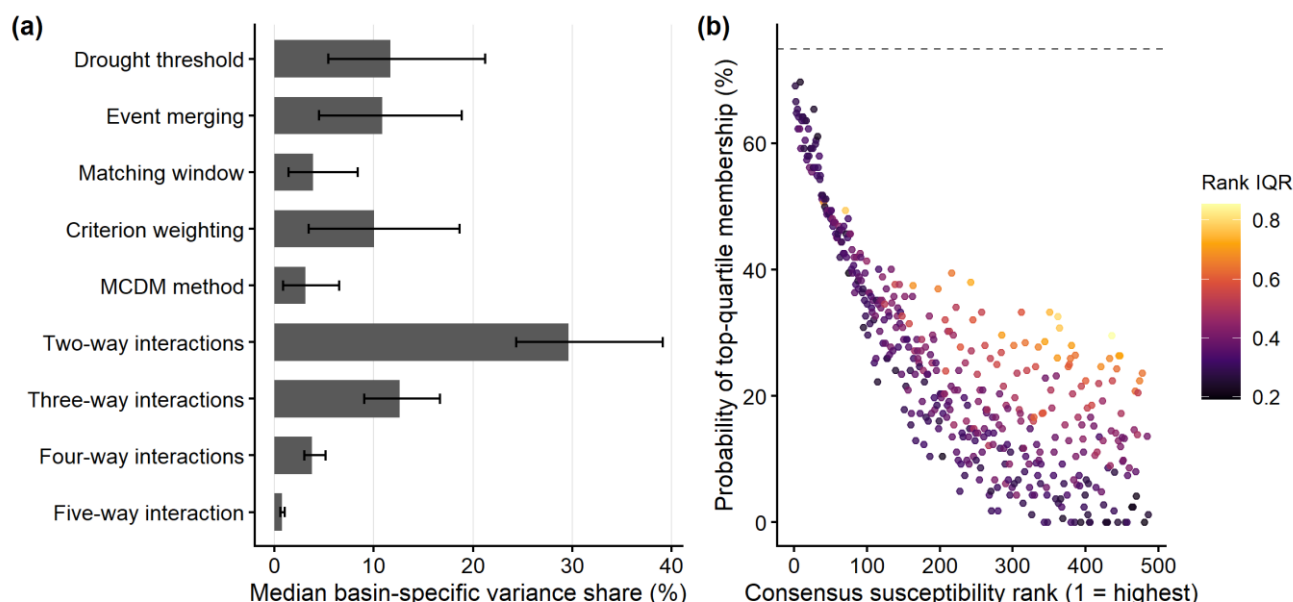


Figure 3: Methodological uncertainty in consensus drought-propagation susceptibility. (a) Basin-specific variance shares from the exact five-factor decomposition of drought threshold, event-merging gap, matching window, criterion-weighting method, multicriteria ranking method, and their interactions for the 368 catchments represented in all 162 configurations. Bars show median shares across catchments and error bars show interquartile ranges. Because components are summarized separately across catchments, the displayed medians do not necessarily sum to 100 %. (b) Probability of occupying the most susceptible quartile across valid methodological configurations as a function of consensus susceptibility rank. Colour denotes the interquartile range of rank percentile across configurations; the dashed line marks a top-quartile membership probability of 75%.

The exact factorial decomposition showed that uncertainty was distributed across both individual methodological choices and their interactions (Fig. 3a; Table A1). Main effects accounted for 49.4% of rank variance on average, whereas interactions accounted for 50.6%. Two-way interactions were the largest grouped component, accounting for a mean 32.0% of basin-specific rank variance, followed by three-way interactions at 13.5 %.

Among individual main effects, drought threshold contributed the largest mean share (15.0%), followed by event-merging gap (13.0%) and weighting method (11.6%). Matching window contributed 5.7%, whereas the independent contribution of the TOPSIS-versus-VIKOR ranking method was 4.0 %. Four- and five-way interactions contributed smaller mean shares of 4.2% and 0.9%, respectively. No single analytical choice therefore accounted for most of the observed rank uncertainty.

Excluding duration transmission from the primary susceptibility calculation had little effect on the resulting continental ordering. Four- and five-criterion susceptibility estimates were correlated at Spearman $\rho = 0.990$ across the 488 common catchments, with Pearson $r = 0.990$. The median absolute difference in susceptibility was 0.0166. Agreement was also high for basin-specific rank uncertainty ($\rho = 0.953$) and top-quartile membership probability ($\rho = 0.974$; Table A7). Thus, removing the redundant duration criterion materially simplified the criterion set without materially altering consensus susceptibility.

3.4 Physical and hydrological contrasts between response modes

The largest physical contrasts between the two event-definition-integrated modes involved hydroclimate and vegetation (Fig. 4a; Table A8). Catchments in the rapid transmission–attenuation mode had higher median maximum green vegetation fraction (0.847 versus 0.662; rank-biserial effect = 0.590) and maximum leaf area index (4.35 versus 2.07; rank-biserial effect = 0.575). Mean precipitation was also higher in the rapid mode (3.47 versus 2.59; rank-biserial effect = 0.545), whereas aridity was lower (0.775 versus 1.052; rank-biserial effect = −0.566).

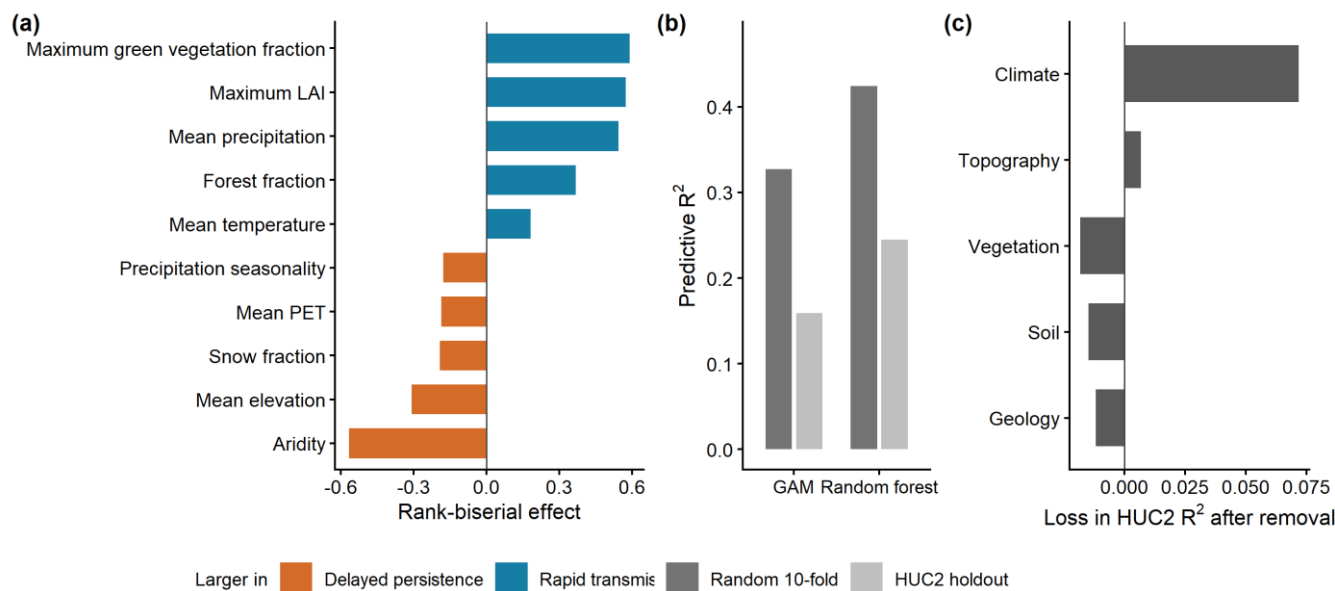


Figure 4: Physical contrasts and spatial transferability of consensus drought-propagation susceptibility. (a) Ten largest physical-attribute contrasts between the rapid transmission–attenuation and delayed persistence modes. Rank-biserial effects are positive when the attribute is larger in the rapid transmission–attenuation mode and negative when it is larger in the delayed persistence mode. (b) Predictive (R^2) of the generalized additive model (GAM) and random forest under leakage-safe random 10-fold cross-validation and leave-one-HUC2-out validation using the same 12 predictors and catchment sample. (c) Change in random-forest HUC2-blocked (R^2) after removing each physical attribute family from the full family model. Positive values indicate loss of transferable predictive information after removal; negative values indicate that the omitted family did not add unique transferable skill conditional on the remaining predictors.

Additional contrasts were consistent with these hydroclimatic differences. Median forest fraction was higher in the rapid mode (0.901 versus 0.589), while median elevation was lower (343 versus 551 m). Snow fraction was also lower in the rapid mode (0.078 versus 0.113). Differences in PET, mean temperature, and precipitation seasonality were smaller than the leading contrasts but followed the broader hydroclimatic separation shown in Fig. A5.

Hydrological signatures independently showed coherent differences between the two modes (Table A9; Fig. A6). Median half-flow date occurred earlier in the rapid mode (164 versus 199 d; rank-biserial effect = −0.593), while high-flow duration was shorter (2.25 versus 5.79 d; rank-biserial effect = −0.462). Rapid-mode catchments also had higher mean runoff (1.33 versus 0.82; rank-biserial effect = 0.438), higher mean root-zone soil moisture (0.354 versus 0.320; rank-biserial effect =



0.405), a steeper flow-duration curve (1.47 versus 1.11; rank-biserial effect = 0.379), and a higher runoff ratio (0.383 versus 0.300; rank-biserial effect = 0.307). Baseflow index was modestly lower in the rapid mode (0.491 versus 0.530; rank-biserial effect = -0.187).

These hydrological signatures were not used to construct either the response modes or susceptibility and therefore provide independent descriptive context for the resulting partition. Complementary M→SM and SM→H pathway diagnostics, including complete-chain and unmatched-hydrological-drought fractions, are shown in Fig. A4; these quantities were not used to define the response modes.

3.5 Prediction and spatial transferability of susceptibility

Predictive performance declined consistently when models were evaluated across hydrological regions rather than using random folds (Fig. 4b; Table 3). Using exactly the same 12 predictors, the GAM achieved $R^2 = 0.327$, Spearman $\rho = 0.567$, RMSE = 0.159, and MAE = 0.129 under leakage-safe random 10-fold cross-validation. Under leave-one-HUC2-out validation, performance decreased to $R^2 = 0.159$, $\rho = 0.432$, RMSE = 0.178, and MAE = 0.144. Univariate attribute–susceptibility associations are summarized in Fig. A7; these descriptive correlations do not account for covariance among predictors.

Table 3: Predictive performance of four-criterion consensus propagation susceptibility under random and spatially blocked validation. Generalized additive model (GAM) and random-forest performance is reported for leakage-safe random 10-fold cross-validation and leave-one-HUC2-out validation using the same 438 catchments and 12 retained predictors. Performance metrics are out-of-sample R^2 , Spearman correlation between observed and predicted susceptibility, root-mean-square error (RMSE), and mean absolute error (MAE).

Model	Validation	n	R^2	Spearman ρ	RMSE	MAE
GAM	Random 10-fold	438	0.327	0.567	0.159	0.129
GAM	Leave-one-HUC2-out	438	0.159	0.432	0.178	0.144
Random forest	Random 10-fold	438	0.424	0.655	0.147	0.119
Random forest	Leave-one-HUC2-out	438	0.245	0.494	0.169	0.138

The random forest showed higher predictive skill under both validation schemes but a similar decline under spatial blocking. Random 10-fold validation gave $R^2 = 0.424$, $\rho = 0.655$, RMSE = 0.147, and MAE = 0.119. Leave-one-HUC2-out validation reduced performance to $R^2 = 0.245$, $\rho = 0.494$, RMSE = 0.169, and MAE = 0.138. Thus, random validation overstated the degree of generalization, although a moderate susceptibility signal remained transferable to withheld HUC2 regions. Random-forest permutation importance for the primary 12-predictor model is shown in Fig. A8.

The broader 21-predictor random-forest family model gave a leave-one-HUC2-out R^2 of 0.255 and Spearman $\rho = 0.512$. Climate produced the largest loss in transferable skill when removed: R^2 decreased to 0.183 and ρ to 0.425, corresponding to



$\Delta R^2 = 0.072$ and $\Delta p = 0.087$ (Fig. 4c; Table 4). Removing topography produced only a small decline ($\Delta R^2 = 0.007$). Removing vegetation, soil, or geology did not decrease blocked-validation performance; their omission slightly increased R^2 .

525 **Table 4: Random-forest physical-family attribution under leave-one-HUC2-out validation. The full family model and models obtained after removing climate, topography, vegetation, soil, or geology are compared using HUC2-blocked R^2 and Spearman correlation. ΔR^2 and Δp are calculated relative to the full model; positive values indicate a loss of transferable predictive skill following removal of the corresponding family.**

Family removed	Predictors retained	R^2	Spearman ρ	ΔR^2	Δp
Climate	14	0.183	0.425	0.072	0.087
Topography	18	0.249	0.503	0.007	0.009
Vegetation	18	0.273	0.524	-0.018	-0.012
Soil	15	0.270	0.521	-0.015	-0.009
Geology	19	0.267	0.520	-0.012	-0.007

530 The incremental random-forest analysis gave the same broad pattern (Fig. A10). Climate variables alone produced $R^2 = 0.274$ and $\rho = 0.525$ under HUC2 validation. Adding topography increased R^2 only slightly to 0.278. Subsequent addition of vegetation, soil, and geology did not improve spatially blocked prediction, with the full 21-predictor model ending at $R^2 = 0.255$. The corresponding high-dimensional GAM family models showed substantially poorer blocked predictive performance and were therefore not used for primary family-level attribution. GAM smooth-term diagnostics and partial-response functions are provided in Table A10 and Fig. A9, respectively.

535 **3.6 Transmission–persistence association**

The diagnostic transmission–persistence analysis showed a weak negative association across the 488 event-reliable catchments (Fig. 5a). The global Spearman correlation was $\rho = -0.124$ with a 95 % bootstrap confidence interval of -0.214 to -0.027 (Table 2).

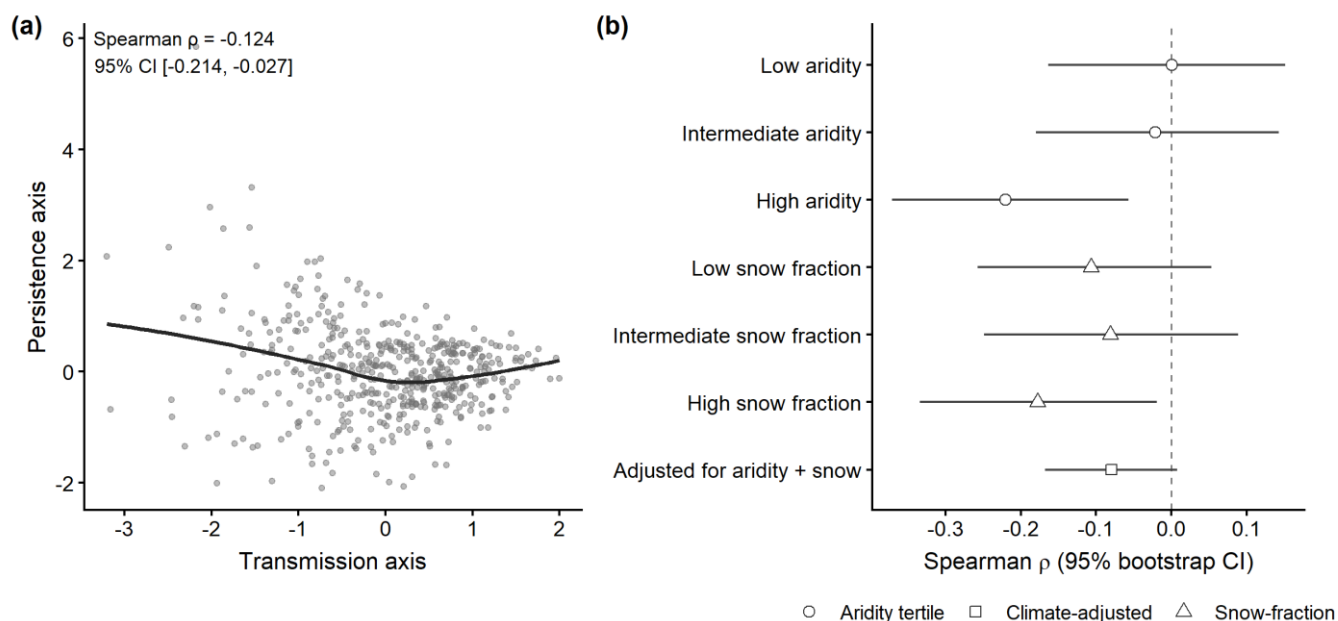


Figure 5: Robustness of the diagnostic transmission–persistence association. (a) Relationship between the baseline diagnostic transmission and persistence axes across the 488 event-reliable catchments. Points represent catchments, the curve is a LOESS smooth, and the annotation gives the global Spearman correlation and percentile-bootstrap 95 % confidence interval. (b) Spearman correlations and bootstrap 95 % confidence intervals within aridity and snow-fraction tertiles, together with the partial Spearman correlation after joint adjustment for aridity and snow fraction. The vertical dashed line denotes $\rho=0$. Climate-stratum p values were adjusted using the Benjamini–Hochberg procedure; the adjusted analysis uses a permutation p value.

After jointly accounting for aridity and snow fraction, the association weakened to a partial Spearman correlation of $\rho = -0.080$, with a 95% bootstrap interval of -0.168 to 0.008 . The corresponding two-sided permutation test gave $p = 0.088$, providing no clear evidence for a residual association after hydroclimatic adjustment.

The association also varied among hydroclimatic strata (Fig. 5b; Fig. A12; Table A11). The strongest negative relationship occurred in the high-aridity tertile ($\rho = -0.220$, 95% CI -0.371 to -0.057 , FDR-adjusted $p = 0.029$). Correlations were near zero in the low- and intermediate-aridity tertiles. The high-snow-fraction stratum also showed a negative correlation ($\rho = -0.178$, 95% CI -0.334 to -0.019), although its FDR-adjusted p value was 0.071. These results show that the global transmission–persistence association was weak and heterogeneous across climatic settings.

4 Discussion

4.1 Method integration changes the inferred drought-propagation structure

The main result of this study is not simply that two groups of catchments can be identified. Rather, it is that the inferred catchment structure depends strongly on whether drought-event definition is fixed or treated as a source of methodological



uncertainty. The partition obtained under the single baseline event definition agreed only weakly with the event-definition-integrated four-criterion partition ($ARI = 0.256$). By contrast, once the propagation characteristics were integrated across plausible event definitions, the resulting two-mode representation was comparatively stable under bootstrap resampling, with a median ARI of 0.937 and more than 93% of catchments retaining their reference mode in at least 90% of bootstrap realizations.

These results separate two forms of robustness that are easily conflated. Bootstrap stability asks whether a classification is reproducible when the catchment sample is perturbed while the analytical definition of drought propagation is held fixed. Methodological robustness instead asks whether the structure persists when defensible choices concerning drought thresholds, event merging, event matching, weighting, and aggregation are changed. A partition can perform well under the first test while remaining conditional on the second. The low agreement between the baseline and ensemble partitions demonstrates this distinction directly. The ensemble modes are therefore better interpreted as response structures that emerge after integrating methodological uncertainty, rather than as validation of a particular single-definition classification.

This distinction is particularly relevant because drought propagation is known to depend on the criteria used to define drought onset, termination, and correspondence between drought types. Aryal et al. (2024) showed that propagation and recovery lags across Australian catchments vary with drought criteria and accumulation timescale. Liu et al. (2026) likewise demonstrated sensitivity to dataset and methodological framework at the global scale, while Yan et al. (2026) reported parameter-dependent stability of a meteorological-to-agricultural drought regionalization under alternative accumulation scales and drought thresholds. Our analysis extends this line of work by integrating alternative event definitions before deriving the primary catchment representation, rather than treating each alternative as a separate regionalization relative to a baseline. The results show that methodological choices influence not only the magnitude of a lag or propagation probability but also which catchments appear hydrologically similar.

The preferred $k = 2$ solution should nevertheless not be interpreted as evidence for two intrinsically discrete types of catchments. Although $k = 2$ produced the largest mean silhouette width among the four-criterion solutions, the silhouette value of 0.406 indicates appreciable overlap, and the individual propagation characteristics showed continuous distributions across the two modes. The modes are therefore parsimonious summaries of recurring combinations within a continuous multidimensional response space. This interpretation is also consistent with the strong fixed- k agreement between the four- and five-criterion $k = 2$ partitions, while the five-criterion sensitivity shifted the silhouette optimum to $k = 3$. The precise number of clusters is therefore less important than the recurrent contrast between relatively rapid transmission with stronger attenuation and delayed transmission with greater hydrological persistence.

4.2 Rapid transmission and delayed persistence represent contrasting combinations of catchment response

The two event-definition-integrated modes differed simultaneously in propagation probability, timing, severity transformation, and recovery. Rapid transmission–attenuation catchments converted meteorological drought into hydrological drought more frequently and with shorter onset lags, but the resulting hydrological droughts were generally less



severe relative to their meteorological precursors and ended sooner. Delayed persistence catchments propagated fewer meteorological drought events and responded later, but their hydrological droughts retained greater relative severity and extended further beyond meteorological recovery.

This pattern reinforces the value of treating drought propagation as multidimensional. A catchment with a high probability of rapid hydrological response is not necessarily one in which the propagated drought is most persistent or severe. Likewise, slow or less frequent propagation does not necessarily imply weak hydrological drought once propagation occurs. Large-sample event analyses have previously shown substantial attenuation of meteorological-drought severity during propagation and strong heterogeneity among catchments. Xiong et al. (2025), for example, found attenuation of drought severity in approximately two-thirds of the 1089 catchments they analysed. Our results show that such transformation is systematically combined with differences in propagation frequency, timing, and recovery rather than occurring as an isolated drought characteristic.

The physical contrasts between modes suggest a strong hydroclimatic organization of this multidimensional behaviour. Rapid transmission–attenuation catchments were generally wetter, less arid, more densely vegetated, lower in elevation, and less influenced by snow. Delayed persistence catchments were comparatively drier, more arid, less vegetated, higher in elevation, and more snow influenced. These contrasts are consistent with previous studies showing that climate, precipitation seasonality, snow processes, and catchment storage strongly influence hydrological drought characteristics and propagation. Apurv and Cai (2020), for example, showed that differences in climate and storage–discharge behaviour produce contrasting drought-propagation characteristics among U.S. watersheds, while Konapala and Mishra (2020) identified strong climatic and catchment controls on hydrological drought characteristics across 652 minimally disturbed U.S. catchments.

The associated hydrological signatures provide additional process context. Rapid-mode catchments had earlier half-flow dates, greater mean runoff and runoff ratios, steeper flow-duration curves, and somewhat lower baseflow indices. Delayed-mode catchments had later flow timing and substantially longer high-flow durations. Taken together, these differences are compatible with contrasting storage and release behaviour: catchments with greater persistence may buffer the initial meteorological deficit sufficiently to delay streamflow drought onset, while slower release from storage can subsequently prolong the hydrological anomaly.

This interpretation should remain process-consistent rather than causal. Static CAMELS attributes and hydrological signatures do not directly resolve groundwater storage volumes, flow-path connectivity, or dynamic storage–discharge relationships. A lower or higher baseflow index, for example, cannot by itself establish how groundwater storage controls drought persistence. Apurv and Cai (2020) similarly showed that the effect of baseflow-related behaviour depends on climate and the storage–discharge relationship rather than on a single hydrological signature. The present results therefore identify combinations of catchment characteristics associated with the observed response modes but do not uniquely attribute those modes to individual subsurface mechanisms.



4.3 Hydroclimate provides the most transferable information on propagation susceptibility

The predictive analysis reinforces the hydroclimatic organization of propagation behaviour while also showing the importance of spatial validation. Both modelling approaches performed better under random cross-validation than when entire HUC2 regions were withheld. For the random forest, R^2 declined from 0.424 under random 10-fold validation to 0.245 under leave-one-HUC2-out validation, while Spearman correlation decreased from 0.655 to 0.494. The GAM showed the same direction of change. Relationships that appear moderately strong when catchments from similar environmental settings are distributed between training and validation folds therefore transfer less effectively to hydrologically distinct regions.

Nevertheless, predictive performance did not collapse under spatial blocking. The remaining skill indicates that part of the continental variation in consensus propagation susceptibility is associated with catchment characteristics that generalize beyond the regions used for model fitting. Among the physical families examined, climate contained the largest amount of such transferable information. Removing climate from the full random-forest model reduced HUC2-blocked R^2 by approximately 0.072 and Spearman correlation by 0.087, whereas removing topography produced only a small decrease and removing vegetation, soil, or geology did not reduce predictive performance conditional on the remaining variables. The incremental models reached most of their transferable skill using climate alone, with only a small additional gain after including topography.

This result is consistent with recent large-scale drought-propagation studies. Liu et al. (2026) found pronounced global spatial heterogeneity in propagation and identified climatic variables as major controls, including temperature and potential evapotranspiration for meteorological-to-runoff drought propagation. Xiong et al. (2025) similarly documented systematic climatic and catchment influences on event-level propagation characteristics. Our blocked-validation analysis adds a different perspective: hydroclimate is not only associated with drought propagation within the observed sample but also contains the largest amount of information that remains useful when predictions are transferred to withheld major hydrological regions.

The weaker incremental contribution of vegetation, soil, and geology should not be interpreted as evidence that these properties are hydrologically unimportant. Several explanations are possible. Their influence may already be partly represented through correlated climatic or topographic gradients; static catchment attributes may inadequately represent the dynamic properties relevant to drought storage and release; effects may operate through interactions that differ among regions; or the available continental-scale attributes may be too coarse to transfer consistently between HUC2 regions. The family-removal experiment quantifies unique transferable predictive information conditional on the other predictors, not process importance in an absolute sense.

The difference between random and spatial validation is therefore itself an important result. Random cross-validation is useful for describing how well observed between-catchment variability can be reconstructed within the sampled environmental space, but it provides a more optimistic estimate of regional generalization. For applications requiring



inference in hydrologically distinct or poorly observed regions, spatially blocked performance provides the more relevant benchmark.

4.4 Consensus susceptibility is robust in broad ordering but not in exact rank

The continuous susceptibility analysis produced a more robust result than any individual methodological configuration, but it should not be interpreted as a uniquely determined ranking of catchments. Approximately half of the basin-specific rank variance was attributable to main effects and half to interactions among methodological choices. Drought threshold was the largest individual main effect, but event-merging rules and criterion weighting also contributed appreciably. No single methodological choice dominated the complete uncertainty budget.

The substantial contribution from interactions is hydrologically plausible. Changing a drought threshold alters event frequency, duration, severity, and the set of events available for matching. The effect of changing a matching window therefore depends partly on which events were produced by the selected threshold and merging rule. Likewise, criterion weighting acts on propagation characteristics whose distributions themselves depend on the preceding event-definition choices. Methodological choices are consequently not independent perturbations of a fixed dataset; they alter different stages of the same event-identification and synthesis chain.

The relatively small independent contribution of the TOPSIS-versus-VIKOR choice indicates that, within the tested formulation, uncertainty arose more strongly from how drought events were constructed and how propagation dimensions were weighted than from which of these two ranking algorithms was applied. This does not imply that multicriteria methodology is generally irrelevant. Rather, within the present normalized four-criterion problem, the distinction between TOPSIS and VIKOR contributed comparatively little additional variance.

An important complementary result is that removing duration transmission from the primary criterion set hardly changed the continuous susceptibility ordering. Four- and five-criterion susceptibility had a Spearman correlation of 0.990, while rank uncertainty and top-quartile membership probabilities were also strongly preserved. This demonstrates that the four-criterion formulation avoids repeated timing information without materially changing the integrated susceptibility signal.

At the same time, strong agreement in consensus susceptibility does not justify emphasizing individual basin ranks. Rank-percentile interquartile ranges remained substantial for many catchments, and top-quartile membership probability changed continuously rather than dividing the sample into deterministic categories. The ensemble is therefore most informative for identifying catchments that consistently occupy relatively high or low portions of the susceptibility distribution and for quantifying uncertainty around those positions. Reporting an exact rank from a single methodological configuration would imply considerably more precision than the analysis supports.

4.5 Transmission and persistence do not form a universal trade-off

The weak negative global association between the diagnostic transmission and persistence axes provides further evidence that the multidimensional propagation structure cannot be reduced to one simple continuum. Although the global correlation



was negative ($\rho = -0.124$) and its bootstrap confidence interval excluded zero, the magnitude was small. After jointly controlling for aridity and snow fraction, the partial association weakened to $\rho = -0.080$, its confidence interval included zero, and the permutation test did not provide clear evidence for a remaining association.

690 The climate-stratified results showed additional heterogeneity. The clearest negative relationship occurred under high aridity, whereas the relationship was weak in the lower-aridity strata. The high-snow-fraction group also showed a negative estimate, but evidence weakened after correction for multiple comparisons. Transmission and persistence therefore appear to covary differently across hydroclimatic settings rather than obeying a general rule in which faster or more frequent transmission necessarily implies weaker persistence.

695 This result also helps clarify the interpretation of the two response modes. Their separation does not require a strong global bivariate trade-off. Clustering uses combinations of several propagation characteristics simultaneously, whereas a Spearman correlation between two diagnostic axes measures only their monotonic association across all catchments. Broad multivariate modes can therefore be sampling-stable even when the corresponding continuous axes remain strongly overlapping and only weakly correlated. The modes should consequently be understood as recurring combinations of propagation characteristics
 700 rather than opposite endpoints of a universal transmission–persistence law.

4.6 Implications, limitations, and future directions

The analysis has several methodological implications for large-sample drought-propagation studies. First, sensitivity of propagation metrics to drought definition should be propagated into subsequent classification or ranking rather than assessed only after a primary definition has been selected. In the present analysis, integrating across event definitions changed
 705 catchment membership substantially relative to the baseline classification. Second, sampling stability and methodological robustness answer different questions and should be reported separately. Third, criteria used in multidimensional assessments should be examined for algebraic as well as empirical dependence. Adding additional metrics does not necessarily add independent information.

There are also limitations. Event-definition integration reduces dependence on one arbitrary specification but does not
 710 encompass all possible drought definitions. The ensemble varied drought threshold, merging gap, and matching window while retaining monthly temporal resolution, a common threshold across drought types within each scenario, a particular target-driven matching algorithm, and a 1-month meteorological accumulation scale. Alternative accumulation periods, season-dependent definitions, variable-specific thresholds, or probabilistic event-matching approaches could yield additional uncertainty.

715 The event matching also imposes a simplified causal chronology. The primary propagation probability includes all meteorological droughts in its denominator, although some meteorological events may begin while a hydrological drought is already active. Such cases are counted within the denominator without implying an isolated causal response and can therefore influence the absolute probability estimate. Event correspondence remains an analytical approximation rather than direct identification of physical causality.



720 The bootstrap stability analysis is likewise conditional on the event-definition-integrated basin characteristics. It quantifies sensitivity to catchment resampling but does not propagate uncertainty from the underlying hydroclimatic time series, drought-index estimation, PET calculation, or event matching through every bootstrap realization. Its high stability should therefore be interpreted as evidence that the derived ensemble partition is not strongly dependent on the particular set of sampled catchments, rather than as a complete uncertainty bound on the classification.

725 Finally, CAMELS-US emphasizes comparatively minimally disturbed catchments. The response modes and susceptibility relationships identified here may differ in regulated or strongly groundwater-managed systems where reservoirs, abstractions, interbasin transfers, irrigation, or other human influences alter the temporal connection between meteorological and streamflow drought. Extension to managed basins and to other large-sample catchment datasets would provide a stronger test of whether the method-integrated response structure is geographically and institutionally transferable.

730 Despite these limitations, the analysis shows that methodological uncertainty can be incorporated directly into catchment-scale drought-propagation assessment rather than treated only as a secondary sensitivity exercise. The most reproducible result is not an exact rank or rigid two-class taxonomy. It is a continuous susceptibility structure and a pair of broad response modes that persist after integration across plausible event definitions, together with evidence that their continental-scale variation is organized primarily by hydroclimate. This distinction provides a more cautious basis for comparing how

735 catchments transmit, transform, and retain meteorological drought.

5 Conclusions

This study evaluated meteorological-to-hydrological drought propagation across CAMELS-US while treating methodological choices as part of the uncertainty rather than fixing a single drought-event definition. Four non-redundant propagation characteristics—propagation probability, onset lag, severity transmission, and recovery extension—were

740 integrated across 27 event definitions and 162 susceptibility configurations. Duration transmission was retained as a diagnostic sensitivity metric because its timing information is algebraically related to onset lag and recovery extension.

The inferred catchment structure depended strongly on whether event definition was fixed or integrated. The partition obtained from a single baseline definition agreed weakly with the event-definition-integrated partition, whereas the integrated four-criterion solution was comparatively stable under bootstrap resampling. The resulting rapid transmission–

745 attenuation and delayed persistence modes should therefore be interpreted as broad, recurring response patterns within a continuous multidimensional space rather than as sharply separated catchment classes. Continuous consensus susceptibility was even more robust: including duration transmission as a fifth criterion changed the catchment ordering only marginally.

Methodological uncertainty remained important. Approximately half of basin-specific rank variance arose from interactions among analytical choices, and no individual methodological factor dominated the complete uncertainty budget. This

750 indicates that exact catchment ranks are less defensible than ensemble-based susceptibility and rank-acceptability measures.



Hydroclimate provided the strongest transferable information on propagation susceptibility. Predictive skill declined under leave-one-HUC2-out validation relative to random cross-validation, but moderate skill remained, with climate contributing most of the out-of-region predictive information. The transmission–persistence relationship was weak and became uncertain after adjustment for aridity and snow fraction, providing no evidence for a universal trade-off.

Overall, the most reproducible description of drought propagation was not a single deterministic classification or ranking, but a method-integrated susceptibility structure accompanied by broad response modes and explicit uncertainty. This framework provides a more defensible basis for comparing how catchments transmit, transform, and retain meteorological drought across heterogeneous hydroclimatic settings.

Appendix A

This appendix provides supporting analyses for the response-mode identification, methodological robustness assessment, and physical-control analysis. These results supplement the main-text figures and tables but are not required for interpretation of the principal conclusions.

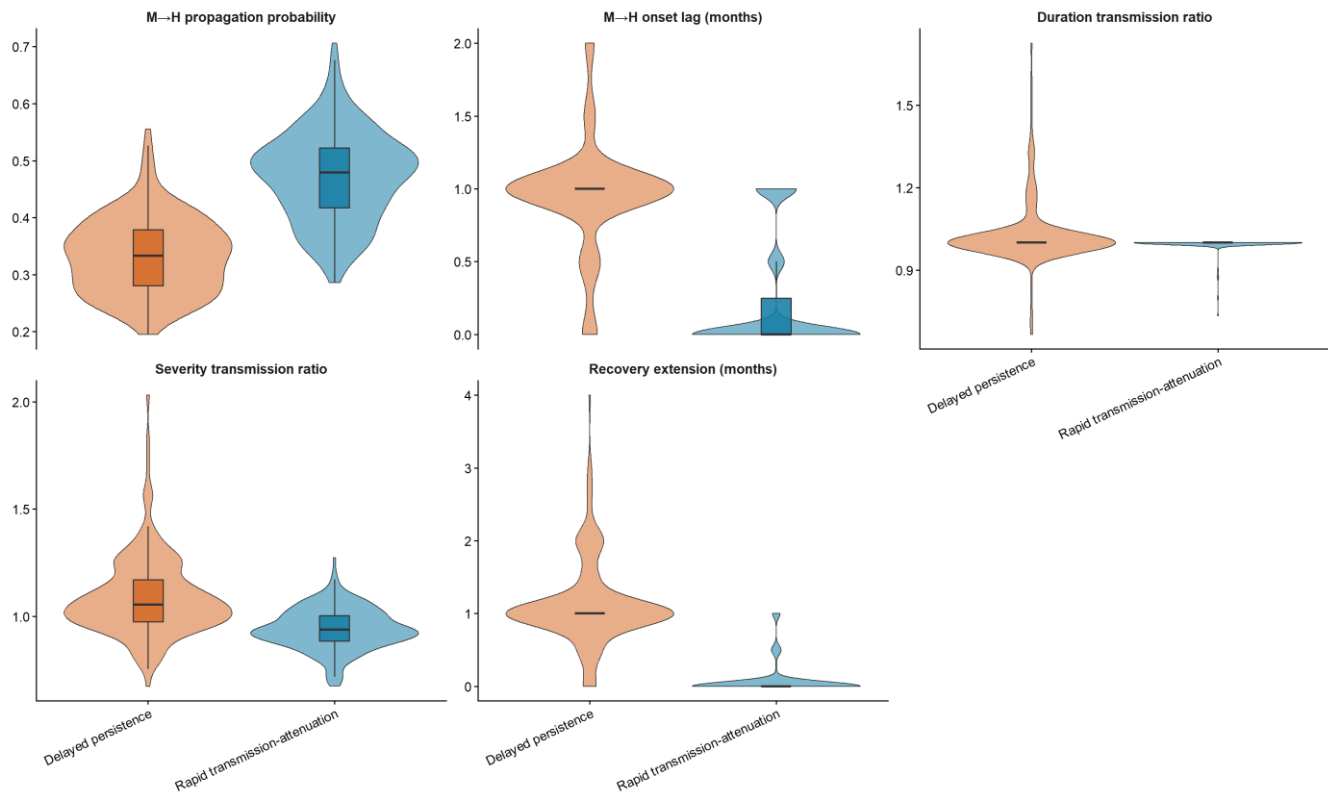


Figure A1: Distributions of drought-propagation characteristics by event-definition-integrated response mode. Violin plots show the distributions of propagation probability, onset lag, duration transmission ratio, severity transmission ratio, and recovery



extension across the rapid transmission–attenuation and delayed persistence modes. Embedded boxplots show medians and interquartile ranges. Duration transmission is shown as a diagnostic sensitivity characteristic and was not included in the primary four-criterion clustering.

770

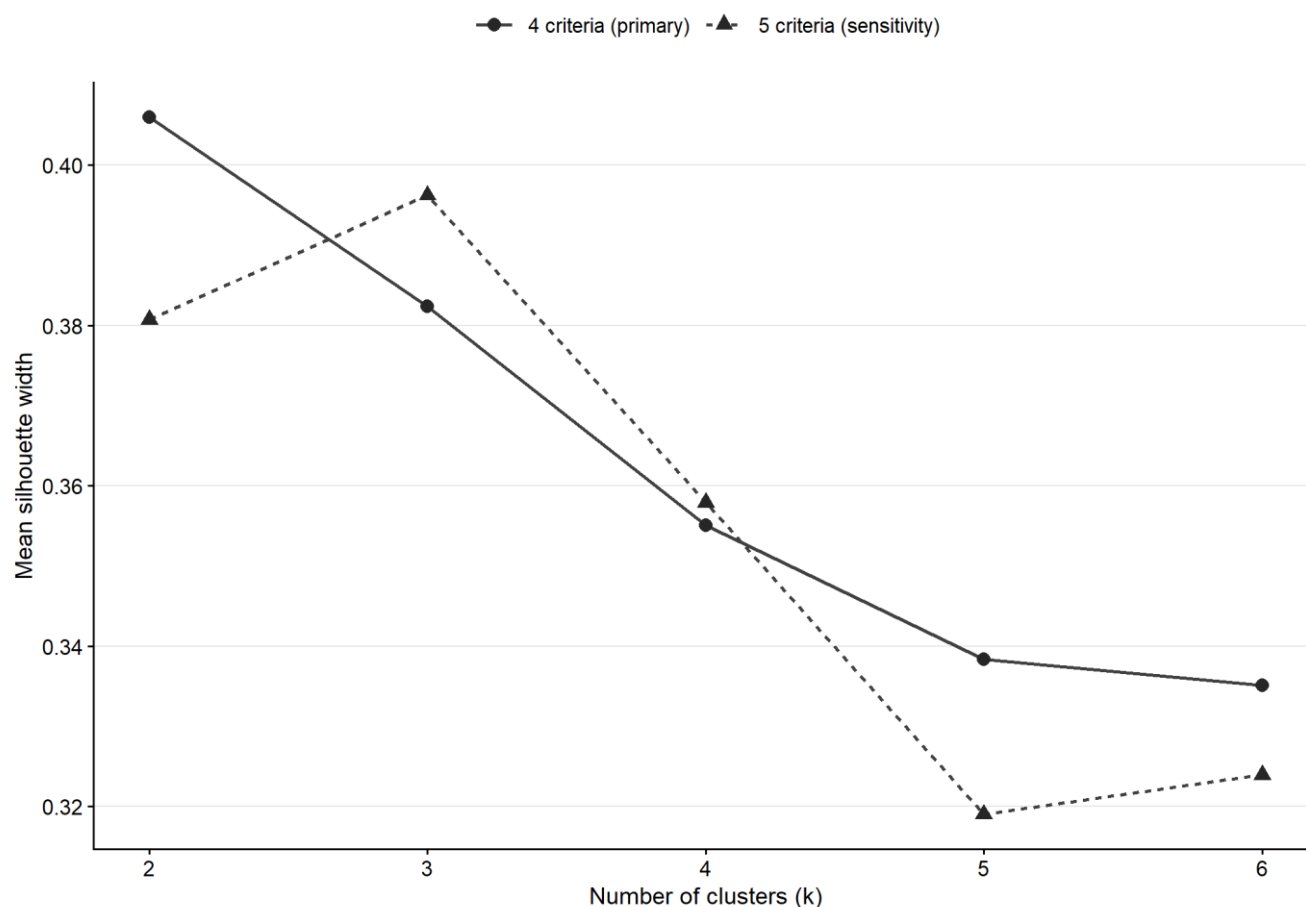


Figure A2: Sensitivity of cluster-number selection to the propagation-criterion formulation. Mean silhouette width is shown for $k=2$ to $k=6$ using the four primary propagation criteria and, as a sensitivity analysis, the corresponding five-criterion formulation including duration transmission. The four-criterion formulation is the primary analysis.

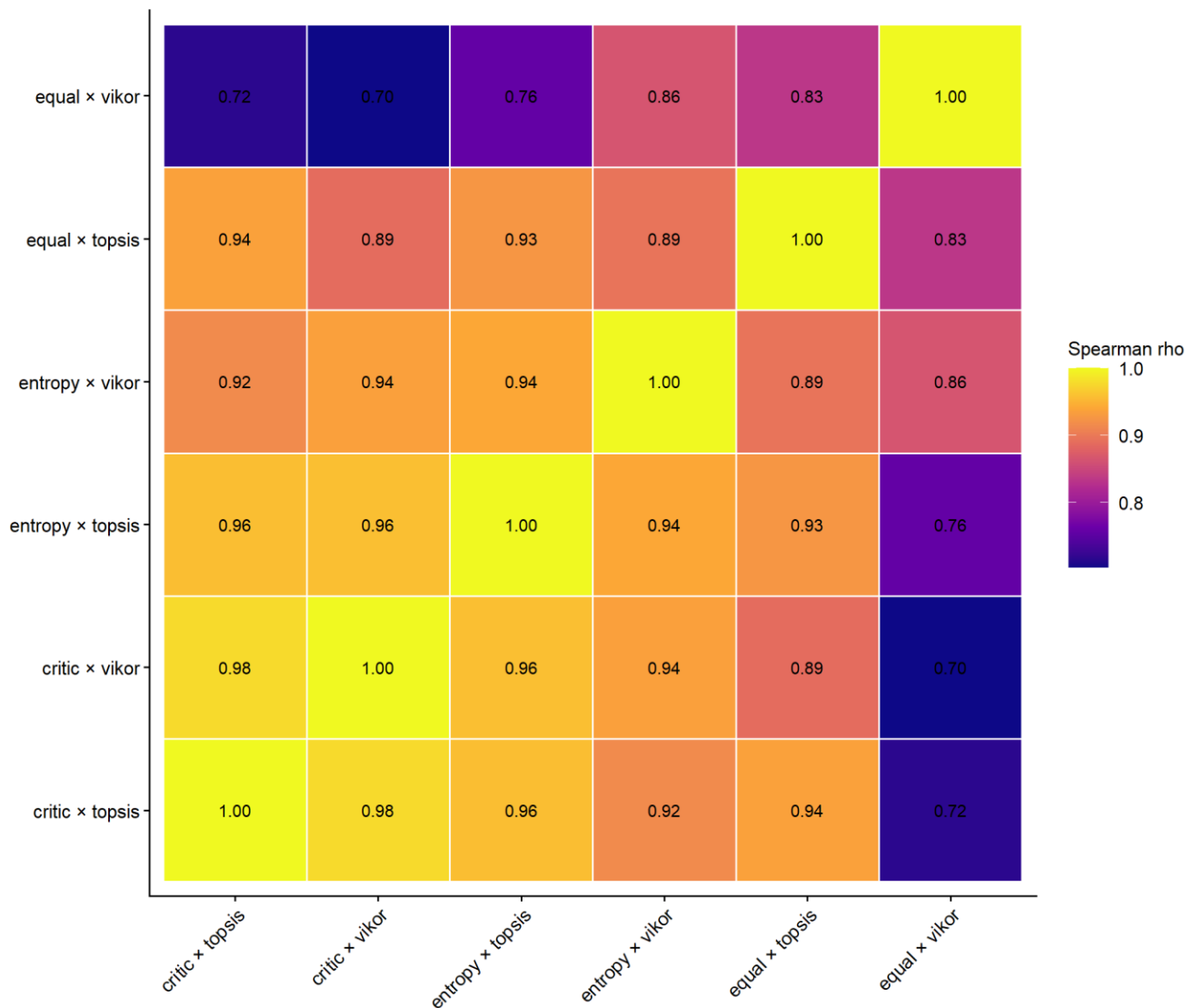


Figure A3: Agreement among baseline four-criterion susceptibility rankings obtained with alternative weighting and multicriteria methods. Cells show pairwise Spearman correlations of catchment rank percentiles among the six combinations of equal, entropy, and CRITIC weighting with TOPSIS and VIKOR. The baseline event definition uses a drought threshold of -0.5, a 2-month event-merging gap, and a 0-month post-source matching window.

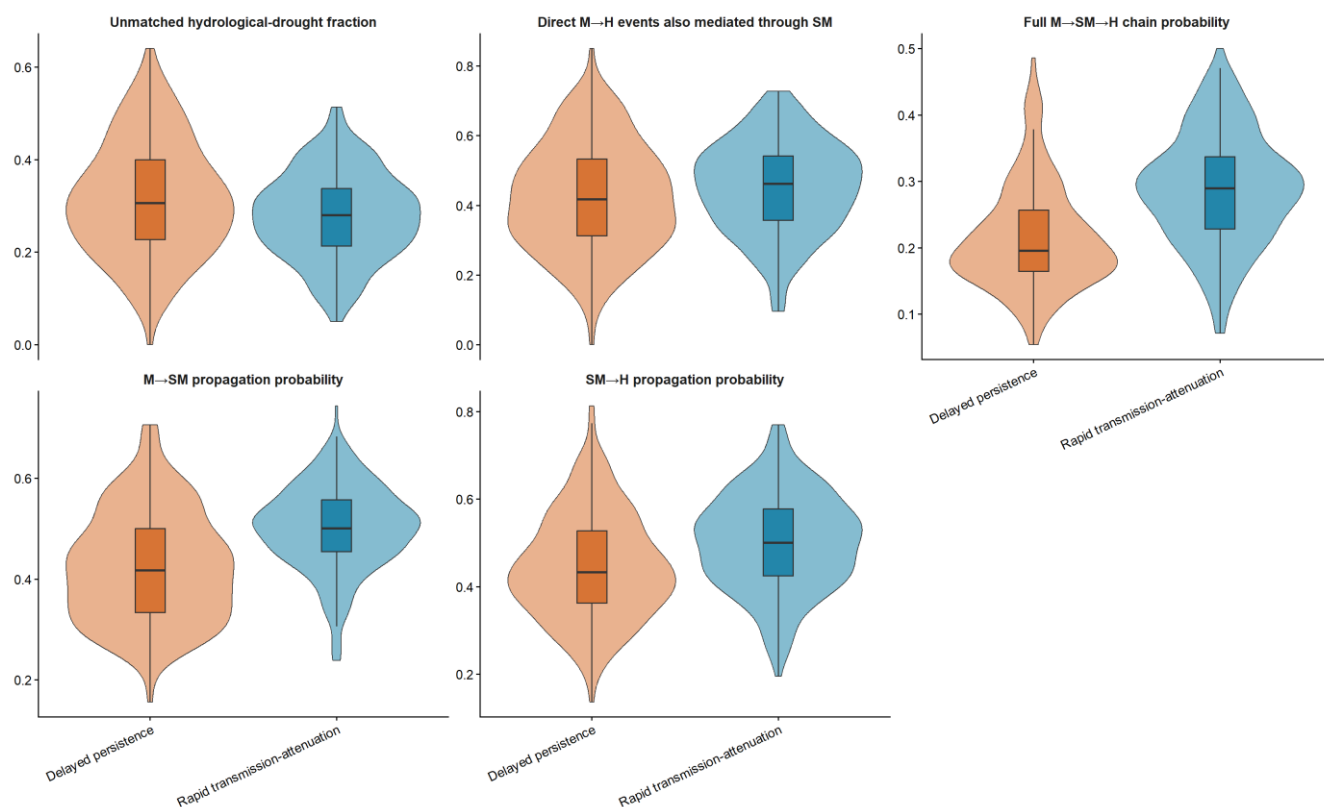
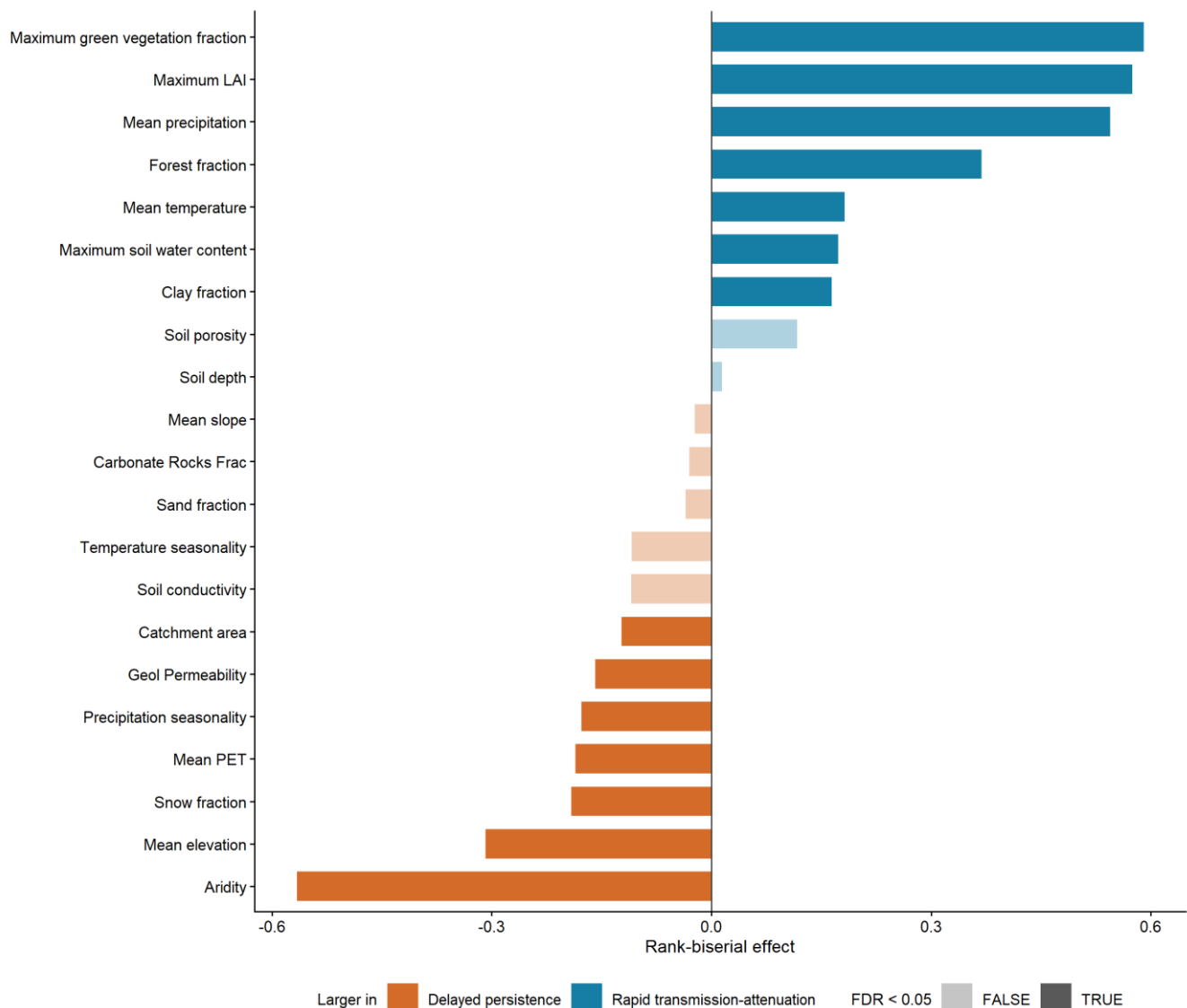
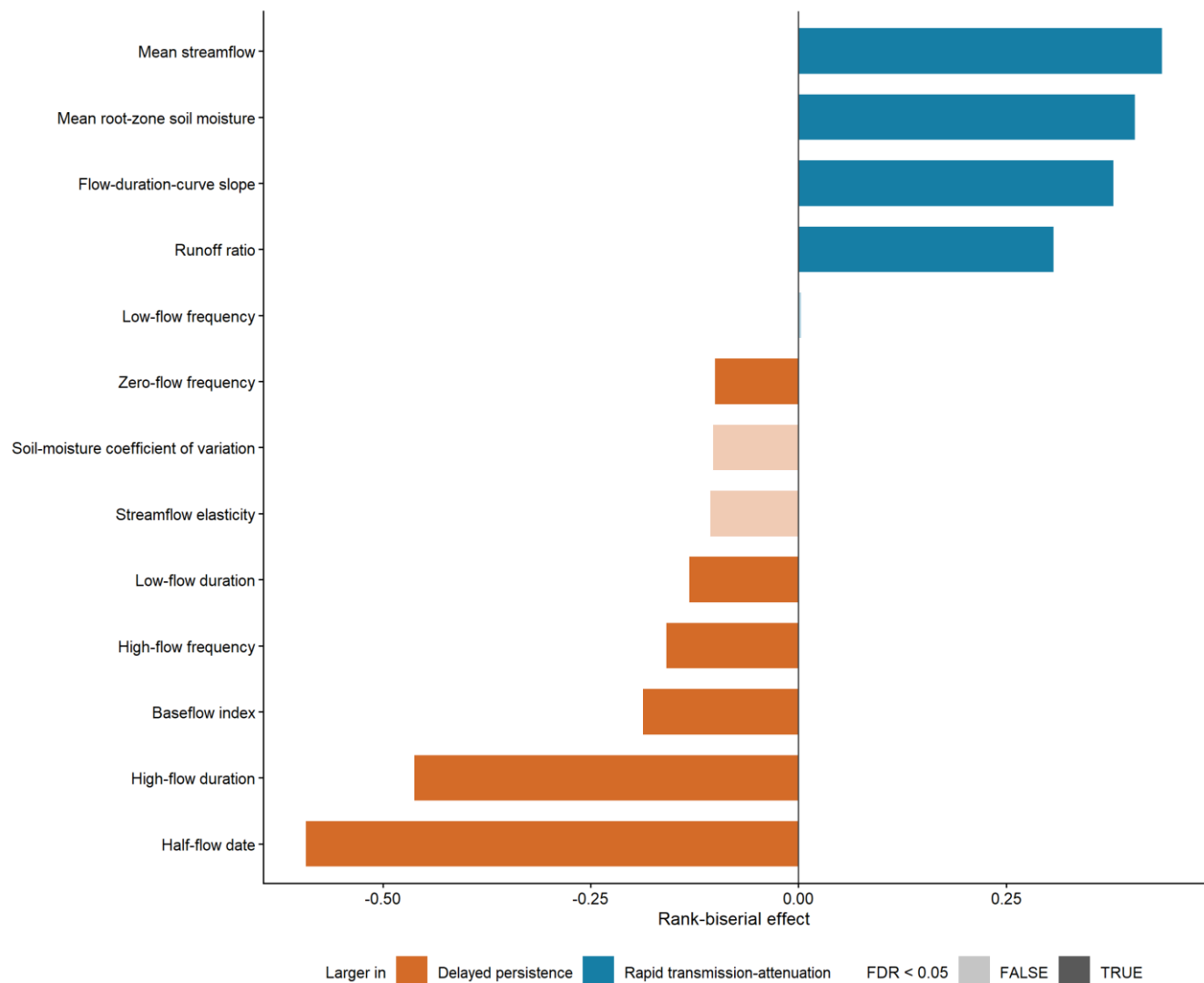


Figure A4: Soil-moisture pathway and unmatched-hydrological-drought diagnostics by response mode. Distributions are shown for the fraction of hydrological droughts without an eligible meteorological precursor, the fraction of directly matched M→H events reproduced through the same M→SM→H chain, the probability of a complete M→SM→H chain, M→SM propagation probability, and SM→H propagation probability. Violin plots show distributions and embedded boxplots show medians and interquartile ranges. These variables were used for process diagnosis and were not included in the primary susceptibility or clustering analyses.

785



790 **Figure A5: Physical-attribute contrasts between the event-definition-integrated response modes. Bars show rank-biserial effect sizes for the full set of evaluated physical attributes. Positive effects indicate larger values in the rapid transmission–attenuation mode and negative effects indicate larger values in the delayed persistence mode. Bar transparency distinguishes contrasts with and without Benjamini–Hochberg-adjusted $p < 0.05$.**



795 **Figure A6: Hydrological-signature contrasts between the event-definition-integrated response modes. Bars show rank-biserial effect sizes for hydrological and soil-moisture signatures evaluated for process interpretation. Positive effects indicate larger values in the rapid transmission-attenuation mode and negative effects indicate larger values in the delayed persistence mode. Bar transparency distinguishes contrasts with and without Benjamini-Hochberg-adjusted $p < 0.05$. These signatures were not used to define the response modes.**

800

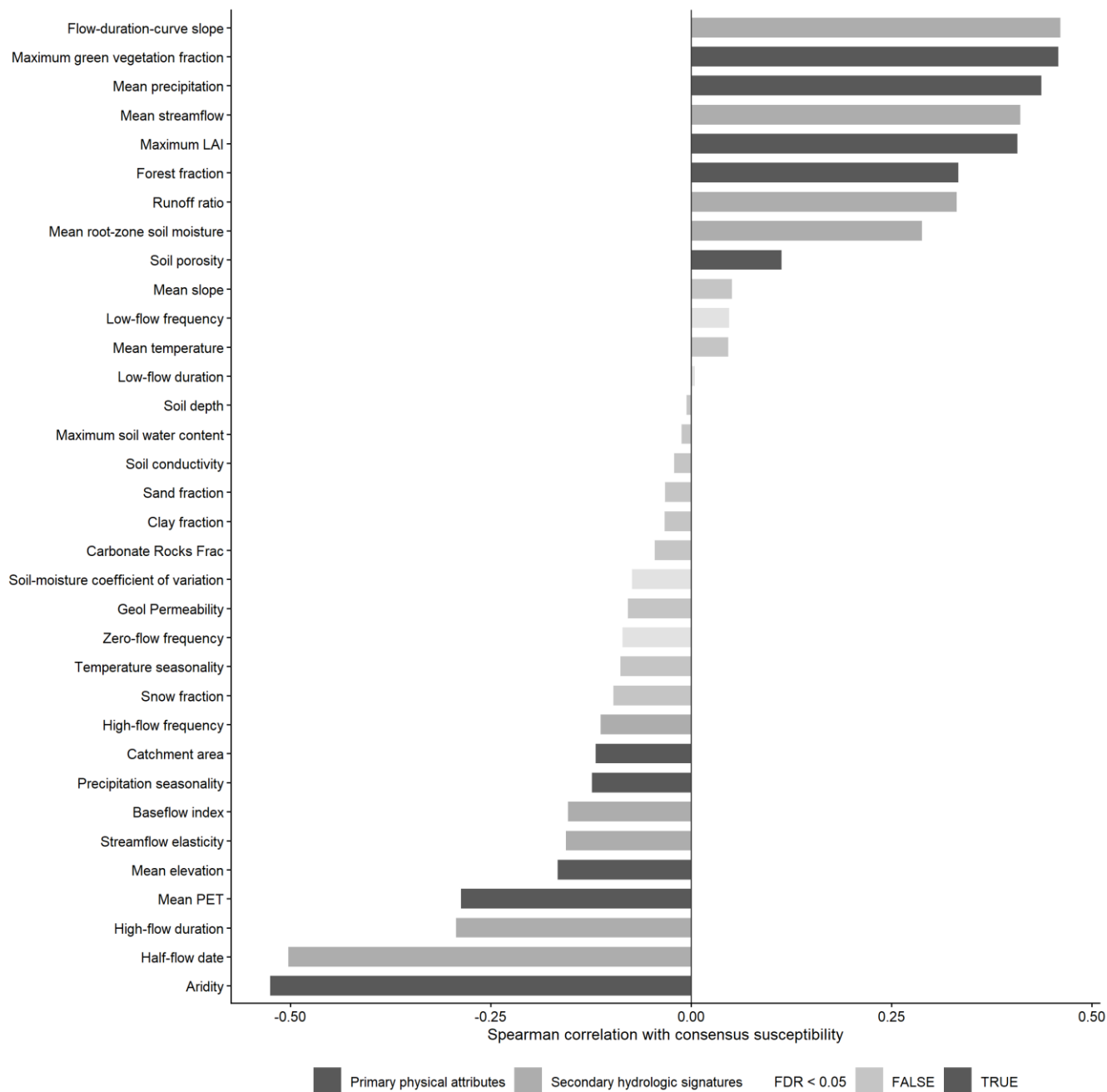


Figure A7: Univariate associations between catchment attributes and four-criterion consensus propagation susceptibility. Bars show Spearman correlations for primary physical attributes and secondary hydrological signatures. Positive values indicate increasing susceptibility with increasing attribute values. Bar transparency distinguishes associations with and without Benjamini–Hochberg-adjusted $p < 0.05$. These univariate associations are descriptive and do not account for covariance among predictors.

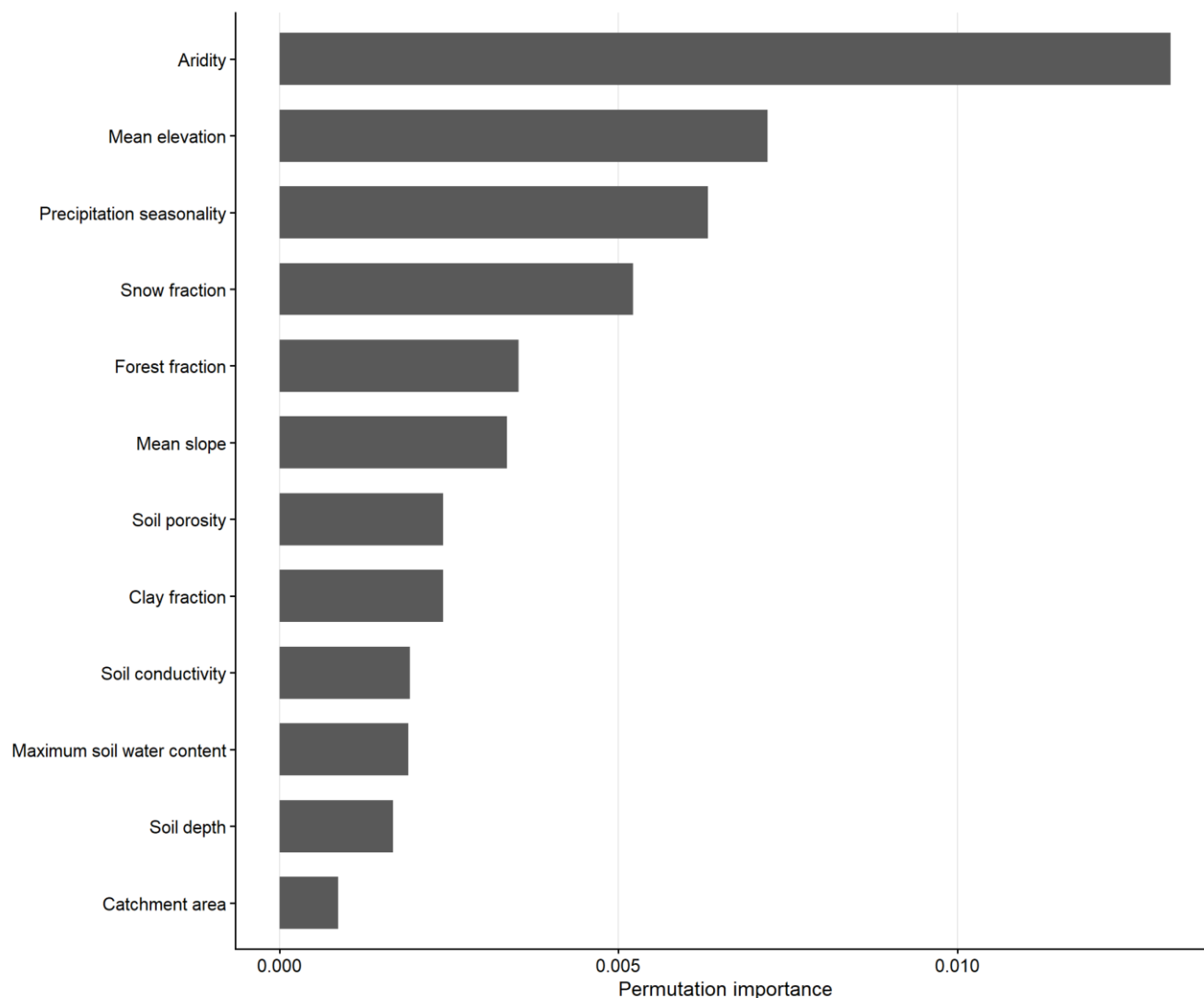


Figure A8: Random-forest permutation importance for prediction of four-criterion consensus propagation susceptibility. Bars show permutation importance for the retained predictors in the primary random-forest susceptibility model. Larger values indicate a greater reduction in model predictive information when the corresponding predictor is permuted. Importance values represent predictive relevance rather than causal effects.

810

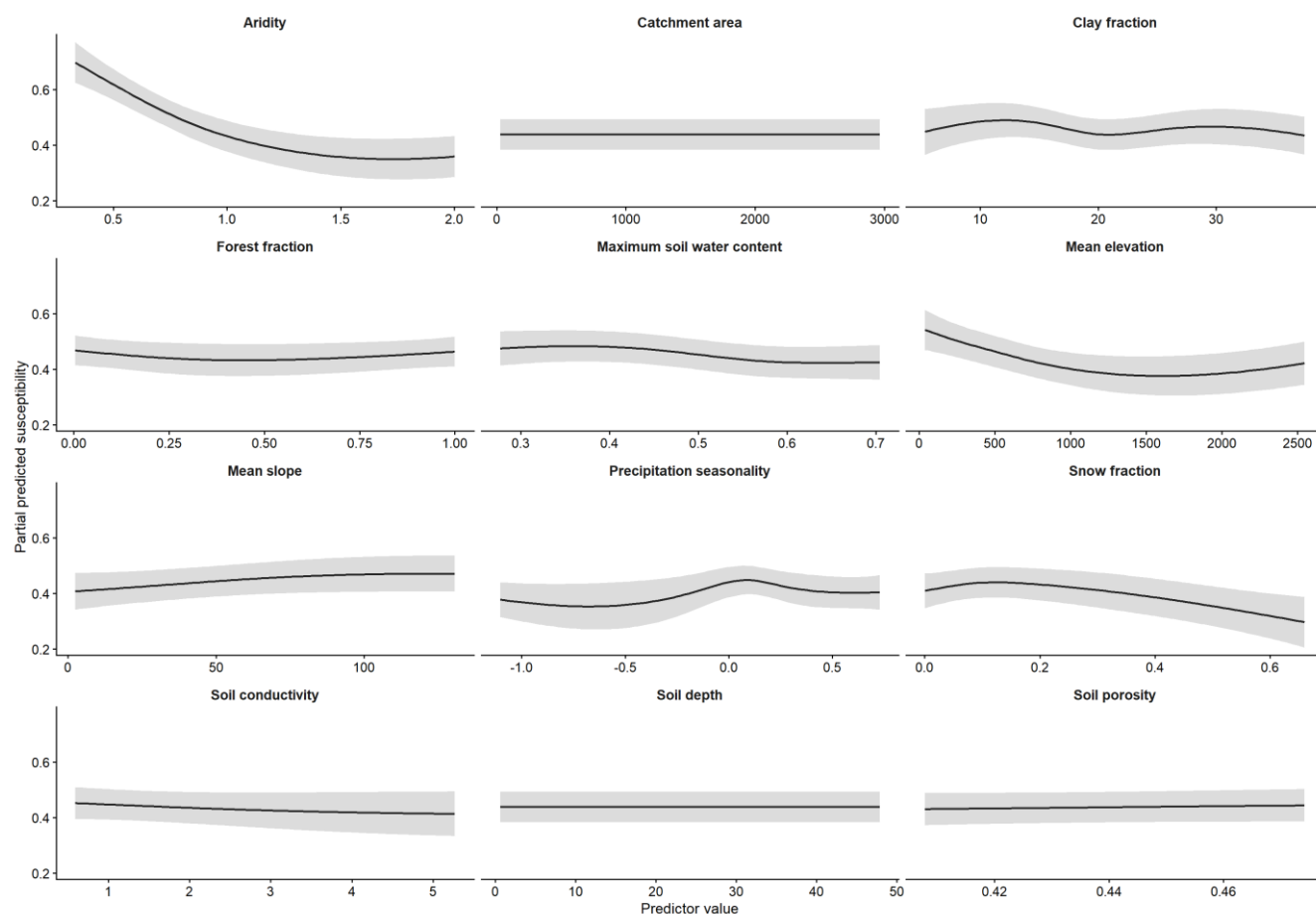
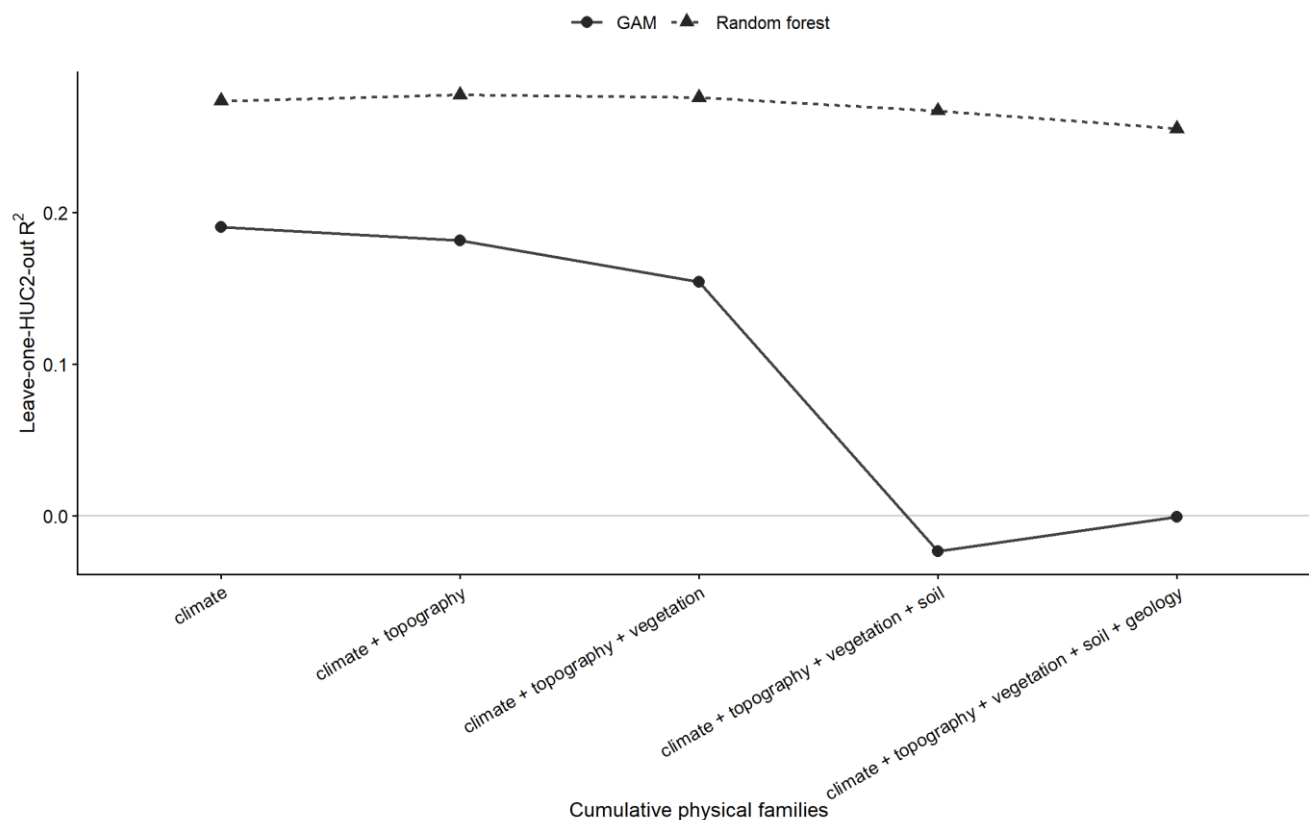
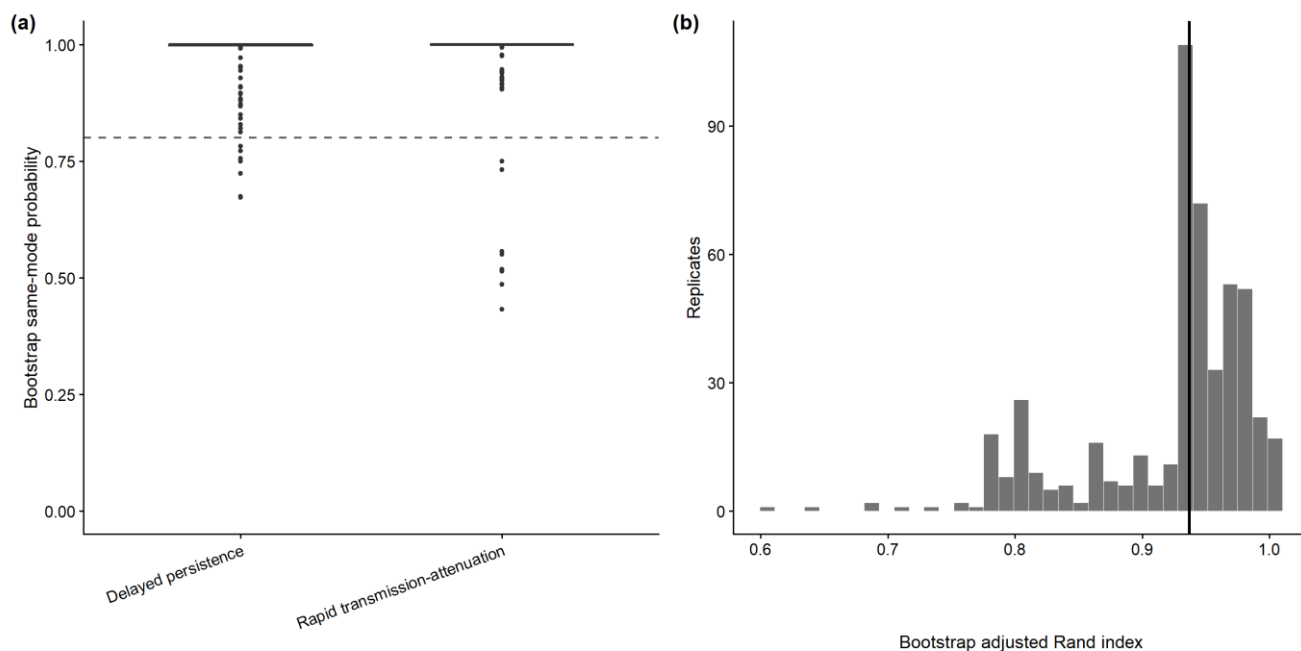


Figure A9: Generalized additive model partial responses for four-criterion consensus propagation susceptibility. Curves show partial fitted susceptibility across the observed range of each retained predictor while other standardized predictors are held at their reference values. Predictor axes are shown in original physical units. Shaded bands represent approximate 95 % confidence intervals calculated as the fitted response ± 1.96 standard errors. Partial responses describe model-based conditional associations and are not interpreted as causal effects.



820 **Figure A10: Incremental physical-family models under leave-one-HUC2-out validation. Predictive (R^2) is shown as physical attribute families are added sequentially in the fixed order climate → topography → vegetation → soil → geology for generalized additive and random-forest models. Because incremental contributions depend on the order of entry, this analysis is descriptive and complements the leave-one-family-out attribution in Fig. 4c and Table 4.**



825 **Figure A11: Bootstrap sampling stability of the preferred four-criterion event-definition-integrated modes. (a) Basin-specific**
probability of retaining the reference response mode across 500 bootstrap resamples, summarized separately by mode. The dashed
horizontal line marks a same-mode probability of 0.80. (b) Distribution of adjusted Rand index values between the reference
partition and the bootstrap-derived partitions. The vertical line denotes the median bootstrap ARI. Bootstrap stability is
 830 **conditional on the event-definition-integrated basin characteristics and does not constitute a full resampling of the underlying**
hydroclimatic time series or methodological ensemble.

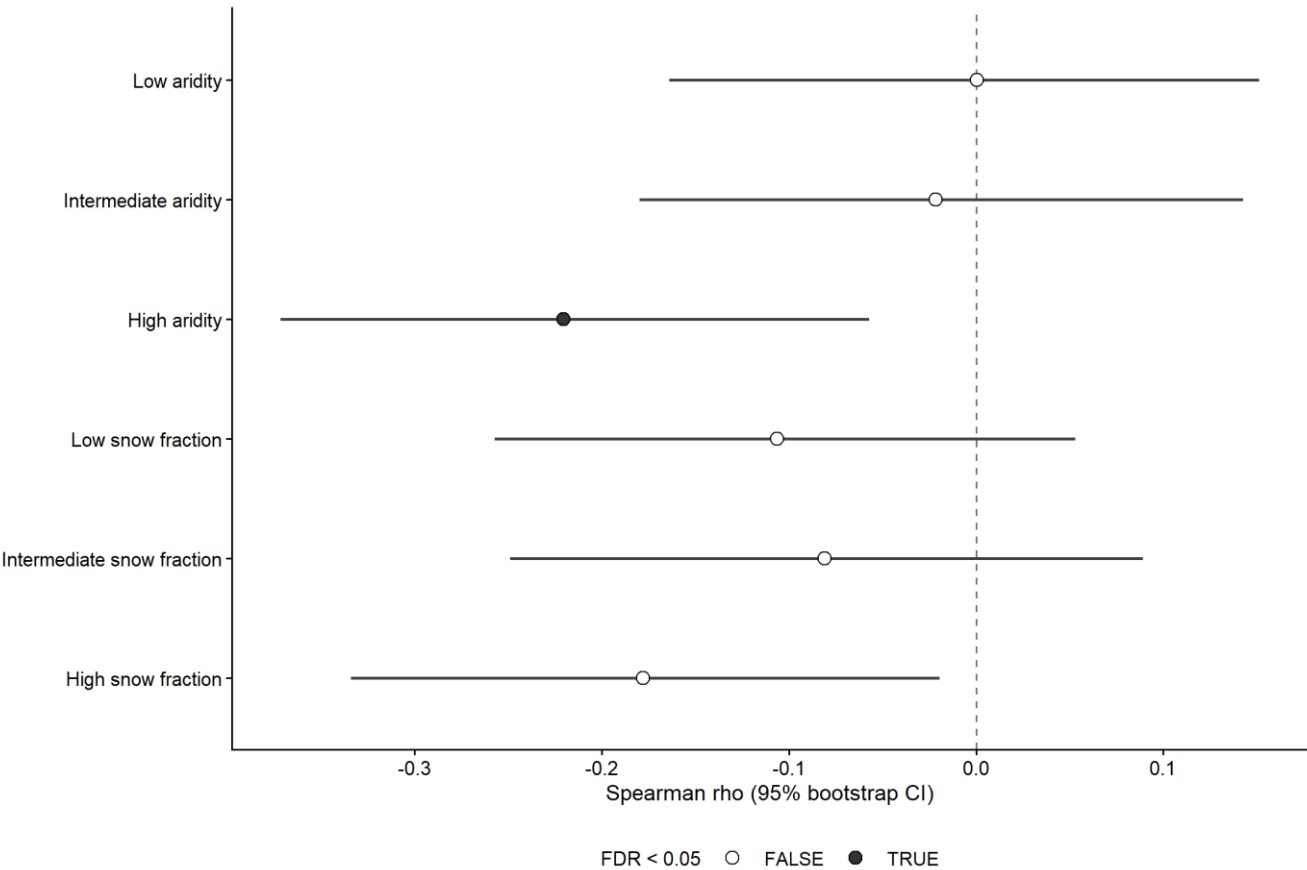


Figure A12: Climate-stratified diagnostic transmission–persistence associations. Points show Spearman correlations within tertiles of aridity and snow fraction, and horizontal bars show percentile-bootstrap 95 % confidence intervals. The vertical dashed line denotes $p=0$. Filled points indicate Benjamini–Hochberg-adjusted $p<0.05$.

835

Table A1: Exact five-factor decomposition of methodological uncertainty in four-criterion consensus susceptibility. Basin-specific rank-percentile variance is partitioned among drought threshold, event-merging gap, matching window, criterion-weighting method, multicriteria ranking method, and all interaction terms for the 368 catchments represented in all 162 configurations.

Effect	Order	n	Mean share (%)	Median share (%)	IQR (%)
Drought threshold	1	368	15.02	11.66	5.41–21.25
Merge gap	1	368	12.97	10.86	4.52–18.87
Weighting method	1	368	11.63	10.02	3.47–18.67
Matching window	1	368	5.70	3.92	1.42–8.38
MCDM method	1	368	4.04	3.12	0.89–6.52
Drought threshold × Merge gap	2	368	12.02	9.93	5.24–16.19



Effect	Order	n	Mean share (%)	Median share (%)	IQR (%)
Merge gap × Weighting method	2	368	4.29	3.89	2.04–6.10
Drought threshold × Matching window	2	368	4.05	2.78	1.23–5.41
Merge gap × Matching window	2	368	3.36	2.21	1.17–4.14
Weighting method × MCDM method	2	368	1.97	1.70	0.60–3.13
Drought threshold × Weighting method	2	368	2.17	1.42	0.65–2.79
Merge gap × MCDM method	2	368	1.16	0.95	0.45–1.61
Matching window × Weighting method	2	368	1.13	0.85	0.45–1.59
Drought threshold × MCDM method	2	368	1.19	0.64	0.19–1.62
Matching window × MCDM method	2	368	0.67	0.49	0.16–1.10
Drought threshold × Merge gap × Matching window	3	368	5.54	4.40	2.75–7.24
Drought threshold × Merge gap × Weighting method	3	368	2.51	2.03	1.44–3.04
Drought threshold × Matching window × Weighting method	3	368	1.03	0.88	0.54–1.37
Merge gap × Matching window × Weighting method	3	368	0.90	0.71	0.38–1.15
Drought threshold × Merge gap × MCDM method	3	368	0.90	0.65	0.35–1.17
Merge gap × Weighting method × MCDM method	3	368	0.65	0.54	0.31–0.88
Matching window × Weighting method × MCDM method	3	368	0.59	0.51	0.26–0.89
Drought threshold × Weighting method × MCDM method	3	368	0.61	0.37	0.16–0.85
Drought threshold × Matching window × MCDM method	3	368	0.44	0.32	0.17–0.59
Merge gap × Matching window × MCDM method	3	368	0.34	0.21	0.11–0.42
Drought threshold × Merge gap × Matching window × Weighting method	4	368	1.85	1.65	1.27–2.24
Drought threshold × Merge gap × Weighting method × MCDM method	4	368	0.77	0.64	0.43–0.97
Merge gap × Matching window × Weighting method × MCDM method	4	368	0.57	0.49	0.36–0.72
Drought threshold × Merge gap × Matching window × MCDM method	4	368	0.57	0.48	0.27–0.75
Drought threshold × Matching window × Weighting method × MCDM method	4	368	0.47	0.41	0.28–0.60



Effect	Order	n	Mean share (%)	Median share (%)	IQR (%)
Drought threshold × Merge gap × Matching window × Weighting method × MCDM method	5	368	0.88	0.80	0.60–1.04

840 **Table A2: Median propagation profiles of the five-criterion clustering sensitivity analysis.**

Characteristic	Higher-transmission cluster (n = 265)	Lower-transmission cluster (n = 173)
Propagation probability	0.476	0.321
Onset lag (months)	0.000	1.000
Duration transmission ratio	1.000	1.000
Severity transmission ratio	0.938	1.060
Recovery extension (months)	0.000	1.000

Table A3: Cluster-number diagnostics for the primary four-criterion event-definition-integrated analysis.

k	Mean silhouette	Median silhouette	Minimum silhouette	Negative silhouettes (%)	Between/total SS
2	0.406	0.433	-0.038	2.1	0.442
3	0.382	0.404	-0.071	2.7	0.587
4	0.355	0.386	-0.066	1.1	0.653
5	0.338	0.353	-0.136	4.6	0.702
6	0.335	0.345	-0.136	5.0	0.738

845 **Table A4: Cluster-number diagnostics for the five-criterion sensitivity analysis.**

k	Mean silhouette	Median silhouette	Minimum silhouette	Negative silhouettes (%)	Between/total SS
2	0.381	0.379	-0.107	4.1	0.365
3	0.396	0.423	-0.065	3.2	0.565
4	0.358	0.373	-0.109	2.3	0.639
5	0.319	0.333	-0.188	7.3	0.679
6	0.324	0.333	-0.188	6.6	0.716

Table A5: Agreement among alternative drought-propagation partitions.



Comparison	Common catchments	Adjusted Rand index
Baseline vs ensemble four-criterion	438	0.256
Baseline vs ensemble five-criterion	438	0.280
Ensemble four- vs five-criterion	438	0.876

Table A6: Bootstrap sampling stability of the preferred four-criterion event-definition-integrated modes.

Metric	Result
Bootstrap replicates	500 / 500 successful
Catchments	438
Median ARI	0.937
Mean ARI	0.922
ARI 95 % interval	0.776–1.000
Median overall agreement	0.984
Overall agreement 95 % interval	0.941–1.000
Median basin same-mode probability	1.000
Basins with stability ≥ 0.80	96.6 %
Basins with stability ≥ 0.90	93.8 %

850

Table A7: Sensitivity of consensus susceptibility to inclusion of duration transmission.

Metric	Result
Common catchments	488
Eligible under both formulations	438
Susceptibility Spearman ρ	0.990
Susceptibility Pearson r	0.990
Median absolute susceptibility difference	0.0166
Maximum absolute susceptibility difference	0.155
Rank-IQR Spearman ρ	0.953
Top-quartile probability Spearman ρ	0.974

Table A8: Physical-attribute contrasts between the rapid transmission–attenuation and delayed persistence modes.

Attribute	Rapid transmission–attenuation median	Delayed persistence median	Rank-biserial effect	FDR-adjusted p
Maximum green vegetation fraction	0.847	0.662	0.590	<0.001



Attribute	Rapid transmission–attenuation median	Delayed persistence median	Rank-biserial effect	FDR-adjusted p
Maximum LAI	4.353	2.072	0.575	<0.001
Mean precipitation	3.468	2.586	0.545	<0.001
Aridity	0.775	1.052	-0.566	<0.001
Forest fraction	0.901	0.589	0.369	<0.001
Mean elevation	342.710	551.060	-0.309	<0.001
Snow fraction	0.078	0.113	-0.191	0.002
Mean PET	2.621	2.735	-0.186	0.002
Mean temperature	10.768	9.651	0.182	0.003
Precipitation seasonality	0.094	0.124	-0.178	0.003
Maximum soil water content	0.604	0.531	0.173	0.004
Clay fraction	20.194	18.657	0.164	0.006
Geologic permeability	-14.100	-13.764	-0.159	0.007
Catchment area	360.990	425.310	-0.123	0.042
Soil porosity	0.446	0.440	0.117	0.051
Soil conductivity	1.254	1.378	-0.109	0.064
Temperature seasonality	8.157	8.018	-0.109	0.064
Sand fraction	33.388	34.845	-0.036	0.580
Carbonate-rock fraction	0.000	0.000	-0.030	0.580
Mean slope	21.467	16.484	-0.023	0.717
Soil depth	1.190	1.411	0.014	0.797

855 **Table A9: Hydrological and soil-moisture signature contrasts between the event-definition-integrated response modes.**

Attribute	Rapid transmission–attenuation median	Delayed persistence median	Rank-biserial effect	FDR-adjusted p
Half-flow date	164.350	199.200	-0.593	<0.001
High-flow duration	2.248	5.786	-0.462	<0.001
Mean streamflow	1.333	0.823	0.438	<0.001
Mean root-zone soil moisture	0.354	0.320	0.405	<0.001
Flow-duration-curve slope	1.474	1.110	0.379	<0.001
Runoff ratio	0.383	0.300	0.307	<0.001
Baseflow index	0.491	0.530	-0.187	0.001



Attribute	Rapid transmission–attenuation median	Delayed persistence median	Rank-biserial effect	FDR-adjusted p
High-flow frequency	13.600	20.650	-0.159	0.007
Zero-flow frequency	0.000	0.000	-0.100	0.021
Low-flow duration	15.181	19.000	-0.131	0.025
Streamflow elasticity	1.689	1.833	-0.106	0.069
Soil-moisture coefficient of variation	0.139	0.154	-0.103	0.072
Low-flow frequency	100.300	98.750	0.003	0.959

Table A10: Generalized additive model smooth terms for four-criterion consensus propagation susceptibility.

Predictor	Effective d.f.	F statistic	p	FDR-adjusted p
Aridity	2.052	17.140	<0.001	<0.001
Snow fraction	2.265	2.605	0.002	0.007
Precipitation seasonality	2.461	2.758	0.003	0.009
Mean elevation	1.789	4.348	<0.001	<0.001
Mean slope	0.917	0.913	0.023	0.040
Catchment area	0.000	0.000	0.427	0.466
Forest fraction	0.933	0.622	0.069	0.103
Soil depth	0.000	0.000	0.645	0.645
Soil porosity	0.322	0.111	0.177	0.213
Soil conductivity	0.501	0.243	0.103	0.137
Maximum soil water content	2.148	2.257	0.006	0.012
Clay fraction	3.318	2.856	0.006	0.012

Table A11: Diagnostic transmission–persistence associations and ensemble-mode bootstrap robustness.

Diagnostic	n	Estimate	95 % interval	p	p-value type
Global	488	-0.124	-0.214–0.027	0.006	Spearman
Aridity: Low aridity	163	0.001	-0.164–0.151	0.995	BH-adjusted
Aridity: Intermediate aridity	163	-0.022	-0.180–0.143	0.942	BH-adjusted
Aridity: High aridity	162	-0.220	-0.371–0.057	0.029	BH-adjusted
Snow fraction: Low snow fraction	163	-0.106	-0.257–0.053	0.353	BH-adjusted
Snow fraction: Intermediate snow fraction	163	-0.081	-0.249–0.089	0.457	BH-adjusted
Snow fraction: High snow fraction	162	-0.178	-0.334–0.019	0.071	BH-adjusted
Adjusted for aridity + snow fraction	488	-0.080	-0.168–0.008	0.088	Permutation



Diagnostic	n	Estimate	95 % interval	p	p-value type
Adjusted Rand index	438	0.937	0.776–1.000	—	—
Median basin same-mode probability	438	1.000	—	—	—

860

Data availability

The original CAMELS-US hydrometeorological time series are publicly available from Newman et al. (2014) at <https://doi.org/10.5065/D6MW2F4D>, and the CAMELS-US catchment attributes are available from Addor et al. (2017) at <https://doi.org/10.5065/D6G73C3Q>. ERA5-Land monthly soil-water data used to derive the 0–100 cm root-zone soil-
 865 moisture series are available from the Copernicus Climate Change Service Climate Data Store at <https://doi.org/10.24381/cds.68d2bb30>. The analysis code, derived basin-level drought-propagation results, methodological-robustness outputs, physical-control and spatial-validation results, and data underlying all manuscript figures and tables are available from the corresponding author upon reasonable request.

Author contributions

870 T.B.T.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Validation, Visualization, Writing – original draft, Writing – review and editing. Í.B.B.: Conceptualization, Methodology, Writing – review and editing. M.A.D.: Conceptualization, Methodology, Writing – review and editing.

Competing interests

The authors declare that they have no conflict of interest.

875 Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.



880 Acknowledgements

The authors acknowledge the developers and maintainers of the CAMELS-US data set for making the catchment-scale hydroclimatic data and attributes used in this study publicly available.

Financial support

This research received no external funding.

885 Review statement

The review statement will be added by Copernicus Publications listing the handling editor as well as all contributing referees according to their status anonymous or identified.

References

- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology
890 for large-sample studies, *Hydrol. Earth Syst. Sci.*, 21, 5293–5313, <https://doi.org/10.5194/hess-21-5293-2017>, 2017a.
- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: Catchment attributes for large-sample studies,
<https://doi.org/10.5065/D6G73C3Q>, 2017b.
- Apurv, T. and Cai, X.: Drought Propagation in Contiguous U.S. Watersheds: A Process-Based Understanding of the Role of
Climate and Watershed Properties, *Water Resour. Res.*, 56, e2020WR027755, <https://doi.org/10.1029/2020WR027755>,
895 2020.
- Aryal, S. K., Zheng, H., Zhang, Y., and Faiz, M. A.: Drought Propagation and Recovery Behaviors Across 407 Australian
Catchments, *Earth's Future*, 12, <https://doi.org/10.1029/2023ef004190>, 2024.
- Beguiria, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI)
revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, *Int. J. Climatol.*, 34, 3001–
900 3023, <https://doi.org/10.1002/joc.3887>, 2013.
- Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple
Testing, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57, 289–300,
<https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>, 1995.
- Breiman, L.: Random Forests, *Machine Learning*, 45, 5–32, <https://doi.org/10.1023/a:1010933404324>, 2001.
- 905 Diakoulaki, D., Mavrotas, G., and Papayannakis, L.: Determining objective weights in multiple criteria problems: The critic
method, *Computers & Operations Research*, 22, 763–770, [https://doi.org/10.1016/0305-0548\(94\)00059-h](https://doi.org/10.1016/0305-0548(94)00059-h), 1995.



- Efron, B.: Bootstrap Methods: Another Look at the Jackknife, *Ann. Statist.*, 7, <https://doi.org/10.1214/aos/1176344552>, 1979.
- Farahmand, A. and AghaKouchak, A.: A generalized framework for deriving nonparametric standardized drought indicators, *Advances in Water Resources*, 76, 140–145, <https://doi.org/10.1016/j.advwatres.2014.11.012>, 2015.
- Gringorten, I. I.: A plotting rule for extreme probability paper, *J. Geophys. Res.*, 68, 813–814, <https://doi.org/10.1029/jz068i003p00813>, 1963.
- Hargreaves, G. H. and Samani, Z. A.: Reference Crop Evapotranspiration from Temperature, *Applied Engineering in Agriculture*, 1, 96–99, <https://doi.org/10.13031/2013.26773>, 1985.
- Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, *Applied Statistics*, 28, 100, <https://doi.org/10.2307/2346830>, 1979.
- Hubert, L. and Arabie, P.: Comparing partitions, *Journal of Classification*, 2, 193–218, <https://doi.org/10.1007/bf01908075>, 1985.
- Hwang, C.-L. and Yoon, K.: Multiple Attribute Decision Making, Springer Berlin Heidelberg, <https://doi.org/10.1007/978-3-642-48318-9>, 1981.
- Konapala, G. and Mishra, A.: Quantifying Climate and Catchment Control on Hydrological Drought in the Continental United States, *Water Resour. Res.*, 56, e2018WR024620, <https://doi.org/10.1029/2018WR024620>, 2020.
- Liu, Y., Hu, T., Yang, J., and Yu, L.: Understanding meteorological, runoff, and agricultural drought propagation and their influencing factors in an ensemble of multiple datasets, *Hydrol. Earth Syst. Sci.*, 30, 2775–2795, <https://doi.org/10.5194/hess-30-2775-2026>, 2026.
- Mann, H. B. and Whitney, D. R.: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other, *Ann. Math. Statist.*, 18, 50–60, <https://doi.org/10.1214/aoms/1177730491>, 1947.
- Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1950 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.68d2bb30>, 2019.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, *Earth Syst. Sci. Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>, 2021.
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrol. Earth Syst. Sci.*, 19, 209–223, <https://doi.org/10.5194/hess-19-209-2015>, 2015.
- Newman, A., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., and Blodgett, D.: A large-sample watershed-scale hydrometeorological dataset for the contiguous USA, UCAR/NCAR, Boulder, CO, <https://doi.org/10.5065/D6MW2F4D>, 2014.



- Oprićovic, S. and Tzeng, G.-H.: Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS, *European Journal of Operational Research*, 156, 445–455, [https://doi.org/10.1016/s0377-2217\(03\)00020-1](https://doi.org/10.1016/s0377-2217(03)00020-1), 2004.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data
945 with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, 40, 913–929, <https://doi.org/10.1111/ecog.02881>, 2017.
- Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, 20, 53–65, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7), 1987.
- Shannon, C. E.: A Mathematical Theory of Communication, *Bell System Technical Journal*, 27, 379–423,
950 <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>, 1948.
- Shukla, S. and Wood, A. W.: Use of a standardized runoff index for characterizing hydrologic drought, *Geophysical Research Letters*, 35, <https://doi.org/10.1029/2007gl032487>, 2008.
- Van Loon, A. F. and Van Lanen, H. A. J.: A process-based typology of hydrological drought, *Hydrol. Earth Syst. Sci.*, 16, 1915–1946, <https://doi.org/10.5194/hess-16-1915-2012>, 2012.
- 955 Van Loon, A. F., Van Huijgevoort, M. H. J., and Van Lanen, H. A. J.: Evaluation of drought propagation in an ensemble mean of large-scale hydrological models, *Hydrol. Earth Syst. Sci.*, 16, 4057–4078, <https://doi.org/10.5194/hess-16-4057-2012>, 2012.
- Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, *Journal of Climate*, 23, 1696–1718,
960 <https://doi.org/10.1175/2009jcli2909.1>, 2010.
- Wood, S. N.: Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73, 3–36, <https://doi.org/10.1111/j.1467-9868.2010.00749.x>, 2011.
- Wright, M. N. and Ziegler, A.: ranger: A Fast Implementation of Random Forests for High Dimensional Data in
965 C++ and R, *J. Stat. Soft.*, 77, <https://doi.org/10.18637/jss.v077.i01>, 2017.
- Xiong, H., Han, J., and Yang, Y.: Propagation From Meteorological to Hydrological Drought: Characteristics and Influencing Factors, *Water Resources Research*, 61, <https://doi.org/10.1029/2024wr037765>, 2025.
- Yan, H., Wang, R., Zhang, C., Qin, Z., Zhang, D., Zheng, Z., Wu, H., and Zhang, K.: Regionalizing Meteorological-to-Agricultural Drought Propagation for Agricultural Risk Management Using Event Metrics and Explainable Machine
970 Learning, *Agriculture*, 16, 1660, <https://doi.org/10.3390/agriculture16151660>, 2026.
- Yevjevich, V. M.: An objective approach to definitions and investigations of continental hydrologic droughts. *Hydrology Paper No. 23*, Colorado State University, Fort Collins, 1967.