https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Self-limiting precipitation recycling during event-scale wet episodes in north-western China's semi-arid transition zone

Ruolin Li[1,2,3], Yang Cui[4], Qi Feng[1,2,3]

[1] Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Lanzhou 730000, China
[2] Qilian Mountains Eco-Environment Research Center in Gansu Province, Lanzhou 730000, China
[3] Key Laboratory of Ecohydrology of Inland River Basin, Northwest Institute of Eco- Environment and Re-sources, Chinese Academy of Sciences, Lanzhou 730000, China
[4] Ningxia Institute of Meteorological Sciences. Yinchuan 75002, China

Correspondence to: Yang Cui (cuiyang@cma.gov.cn) & Qi Feng (qifeng@lzb.ac.cn)

Abstract. Recent decades have seen a marked climatic wetting across north-western China's semi-arid transition zone, yet the extent to which this intensification arises from local land–atmosphere feedback or from external moisture inflow remains uncertain. Using hourly station observations and ERA5 reanalysis for 2020–2024, we develop a process-based framework that links event-scale rainfall variability to the dynamic behaviour of precipitation recycling. Hourly recycling rates are derived through a two-reservoir moisture-tracking scheme, and multi-day wet episodes are identified to isolate transient feedback processes. Machine-learning surrogate models (CatBoost, XGBoost, ExtraTrees, Gradient Boosting) emulate the recycling rate as a function of meteorological conditions under a leave-one-event-out design, enabling counterfactual perturbation experiments in which precipitation intensity and key moisture variables are systematically scaled. A dimensionless percentage elasticity ($\eta\%$) is introduced to quantify the relative response of recycled precipitation to rainfall enhancement. Results from nine regional events reveal a robustly negative $\eta\%$, indicating that additional rainfall often suppresses rather than reinforces local recycling efficiency—a self-limiting wetting feedback. Cluster analysis distinguishes three physical regimes: (1) cool–moist episodes with initially strong but rapidly saturating coupling, (2) advection-dominated events with nearly linear, externally controlled responses, and (3) warm–dry episodes with weak coupling and near-zero elasticity. Collectively, these findings depict the atmosphere above north-western China as a self-stabilizing hydrological system in which increased precipitation does not necessarily strengthen, and may even weaken, local moisture recycling. The proposed event-scale elasticity framework provides a transferable diagnostic for short-term land–atmosphere coupling and for assessing hydrological resilience in arid and semi-arid regions.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

1 Introduction

The arid and semi-arid regions of north-western China have experienced a discernible climatic wetting over recent decades, reflected in rising precipitation totals, a higher frequency of multi-day rainfall events, and a gradual increase in atmospheric humidity (Chi et al., 2023; Gu et al., 2025; Touseef et al., 2020). This regional wetting has been widely recognized as one of the most remarkable hydrological signals in inland Asia, yet its underlying mechanisms remain debated (Qi et al., 2022; Xie et al., 2024). While increased rainfall may suggest an intensified water cycle, it is still unclear whether such intensification originates from stronger local land–atmosphere feedbacks or from enhanced advection of external moisture (Lugato et al., 2011; Zhang et al., 2024; Zhou et al., 2022). The degree to which precipitation recycling contributes to this wetting trend—its magnitude, variability, and short-term response to rainfall intensification—has not been well constrained. Previous studies have estimated precipitation recycling at monthly or climatological scales using reanalysis or moisture-tracking methods, providing valuable large-scale averages but largely overlooking the transient, event-scale interactions through which precipitation and evaporation co-evolve (Bisselink and Dolman, 2009; Dey et al., 2021; Goessling and Reick, 2011; Li et al., 2024; Van Der Ent et al., 2014). In the present study, we focus on a representative semi-arid transition zone (38–39.5° N, 106–107° E) during 2020–2024, a period characterized by notably enhanced precipitation in the region (Ma et al., 2025). The frequent occurrence of multi-day rainfall episodes and their spatial coherence during this interval provide favorable conditions for examining short-term hydrometeorological coupling while minimizing the confounding effects of complex topography. This region therefore offers an appropriate setting to investigate how precipitation recycling responds during individual rainfall episodes under an overall wetter climatic background.

To address this problem, we adopt a process-oriented and data-driven framework that links observed precipitation variability to the dynamic behavior of precipitation recycling. The guiding principle is to analyze the event scale, corresponding to multi-day wet episodes in which surface evaporation, boundary-layer humidity, and convection interact most actively. By isolating these episodes and quantifying the co-variability between recycling rate (RR) and total precipitation (P), it becomes possible to examine whether rainfall events locally reinforce recycling or instead rely predominantly on external moisture inflow. This event-resolved approach provides a natural bridge between the synoptic processes that control storm evolution and the climatological wetting tendencies observed at longer timescales.

Within this framework, several complementary analytical components are employed. First, cluster analysis is applied to classify rainfall episodes according to their environmental background. Such unsupervised grouping has been widely used in land–atmosphere interaction studies to delineate distinct coupling regimes—for example, differentiating soil-moisture–precipitation feedback patterns or identifying circulation-based precipitation types (Day et al., 2025; Tian et al., 2025; Yurova et al., 2020; Zeppetello et al., 2024). Grouping episodes into physically coherent regimes (e.g., moisture-rich versus moisture-poor, advection-dominated versus locally responsive) allows subsequent analyses to be interpreted in a regime-specific context and reduces the influence of inter-event heterogeneity.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Second, machine-learning surrogate models are developed to emulate *RR* as a function of concurrent meteorological variables, with attention to model stability and robustness under limited sample sizes. Tree-based ensemble algorithms such as XGBoost, CatBoost, and ExtraTrees are adopted because they have demonstrated strong performance in capturing nonlinear and multivariate relationships in hydrometeorological systems and are well suited for small-sample settings (Ahn et al., 2023; Papacharalampous et al., 2023). The relatively short analysis period (2020–2024) and the scarcity of continuous rainfall events in arid regions necessitate methods that generalize reliably under data constraints. The models are trained using event-blocked cross-validation, ensuring that predictive skill is evaluated across independent rainfall episodes rather than within a single event, thereby avoiding temporal leakage and improving reproducibility. The trained surrogates are then used to perform counterfactual perturbation experiments (Chen et al., 2024, 2022; Jiang et al., 2023; Zhou et al., 2024), in which precipitation intensity is systematically modified while other predictors remain fixed. This procedure provides a consistent and physically interpretable means to quantify the sensitivity of *RR* to precipitation changes under realistic atmospheric conditions.

To quantify this sensitivity in a consistent and interpretable manner, we design a new indicator termed the percentage elasticity ($\eta\%$), derived from the joint response of *RR* and precipitation. From a physical standpoint, $\eta\%$ expresses the relative change in recycled precipitation associated with a fractional change in total precipitation, analogous to an elasticity measure but tailored to the bounded nature of *RR*. This formulation provides a unit-free and scale-consistent metric that summarizes how efficiently local recycling responds to varying rainfall intensity within an event. The derivation and computation of hourly *RR* and $\eta\%$ are described in Section 2, where mathematical details, physical constraints, and validation procedures are presented to ensure transparency and reproducibility.

In summary, this study investigates how precipitation recycling behaves during event-scale wet episodes in a key semi-arid region of north-western China. By combining physically derived hourly recycling diagnostics, event-based classification, machine-learning surrogate modeling, and counterfactual elasticity analysis, we aim to provide a process-based quantification of the short-term interaction between precipitation and recycling under a wetting climate background. The remainder of the paper is structured as follows. Section 2 describes the data sources and methodological framework, including the derivation of hourly *RR* and the definition of $\eta\%$. Section 3 presents the main results of the surrogate modeling, elasticity analysis, and regime classification. Section 4 discusses the physical implications and limitations of the approach, and Section 5 summarizes the principal findings and broader implications for understanding hydrological feedbacks in arid and semi-arid regions.

2 Data and Methods

2.1 Study area and observation data

The study domain covers 38–39.5° N, 106–107° E in north-central China, situated between the Tengger Desert and the northern Loess Plateau (Fig. 1a). This corridor forms a semi-arid transition zone characterized by limited annual

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

precipitation (< 200 mm yr⁻¹), sharp topographic gradients (≈1100 m a.s.l.), and strong summer concentration of rainfall associated with intermittent monsoon incursions (Zhuo-li, 2015). Short-lived but intense convective rainfall dominate the regional water budget (Qu et al., 2024), providing an ideal setting for investigating short-term moisture recycling and wetting feedback.

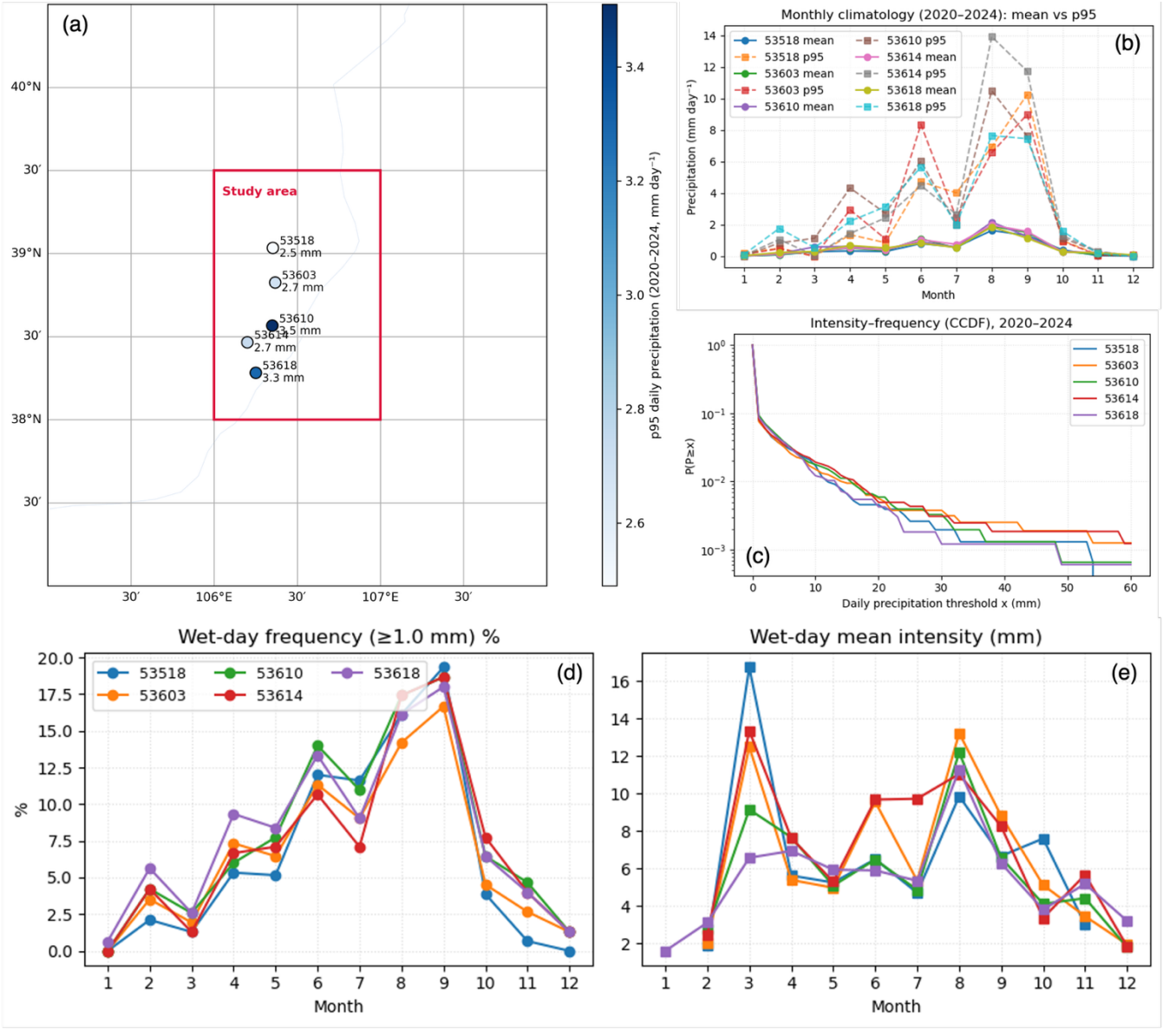


Figure 1. Climatological characteristics of precipitation during 2020–2024 across the semi-arid transition zone of north-central China. (a) Locations of the five CMA meteorological stations within the study domain (red box) and their 95th-percentile (p95) daily precipitation thresholds (mm day⁻¹). (b) Monthly climatology of daily precipitation (mean vs. p95) for each station, showing the strong summer–autumn concentration of rainfall. (c) Complementary cumulative distribution function (CCDF) of daily precipitation, indicating the heavy-tailed nature of rainfall intensity. (d) Monthly frequency of wet days (≥ 1 mm day⁻¹). (e) Monthly mean intensity of wet-day precipitation.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Hourly meteorological observations were collected from 5 national stations (Private communication) operated by the China Meteorological Administration (CMA). Their basic information is listed in Table 1.

Table 1. Geographic information of the five meteorological stations used in this study. The stations are located in Ningxia, north-western China, and provide hourly precipitation and surface meteorological observations for the analysis of wetting events and moisture recycling.

ID	Latitude (°N)	Longitude (°E)	Elevation (m)	Province	Station
53610	38.5667	106.3500	1106.2	Ningxia	Helan
53614	38.4667	106.2000	1110.9	Ningxia	Yinchuan
53618	38.2833	106.2500	1113.7	Ningxia	Yongning
53518	39.0350	106.3533	1126.9	Ningxia	Shizuishan
53603	38.8269	106.3664	1099.9	Ningxia	Shahu

The record extends from 31 December 2019 21:00 UTC to 31 December 2024 20:00 UTC with an hourly time step. Variables include surface pressure, air temperature, vapor pressure, relative humidity, visibility, precipitation, evaporation, wind speed and direction, multi-layer ground temperature, and grass-surface temperature. Rigorous quality control was performed by removing missing or physically inconsistent values. Hourly data were aggregated to daily values when at least 18 valid hours were available to ensure temporal completeness.

Because these five sites provide only point-scale coverage, ERA5 reanalysis was employed to supplement spatial and temporal gaps and to provide vertical atmospheric information not measured at stations. Cross-validation between ERA5 daily precipitation and station observations yielded a correlation coefficient of 0.86 and an RMSE of ≈ 3 mm day^{-1}, confirming that ERA5 reasonably reproduces both timing and magnitude of rainfall events. Hence, ERA5 fields were used as physically consistent background data for subsequent computation of the precipitation recycling rate and for constructing model predictors.

2.2 ERA5 variables and derived predictors

To ensure consistent temporal representation of land–atmosphere coupling, hourly ERA5 reanalysis data (0.25° × 0.25° resolution) were obtained for 2020–2024 from both the surface (C3S, 2018) and pressure-level products (Copernicus Climate Change Service, 2023). Hourly fields were first subset to the target region (38–39.5° N, 106–107° E) and then used to calculate moisture recycling and related diagnostics following our hourly *RR*–calculation framework.

Table 2. ERA5 variables used in this study, retrieved from both surface and pressure-level products for 2020–2024.

Category	Variable	Description / Unit
Hydrological fluxes	*P* (total precipitation), *E* (evaporation), *ro* (runoff)	Surface water fluxes (m h^{-1} or mm h^{-1})
Thermodynamic	*T2M*, sp, *TCWV*	2-m temperature (K), surface pressure (Pa), total column water

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Category	Variable	Description / Unit
state		vapor (kg m^{-2})
Dynamic fields	u10, v10, u, v (at muti-levels)	10-m and integral-level wind components (m s^{-1})
Cloud and radiation	tcc, net_sw, net_lw, sshf_wm2, ssrd_wm2	Total cloud cover (0–1), net short-/long-wave radiation, sensible-heat flux, top-of-atmosphere SW radiation
Soil and precipitation type	swvl1, cp_mph, lsp_mph, cp_frac, lsp_frac	Surface soil moisture, convective/large-scale precipitation and their ratios

From these variables, several diagnostic and harmonic predictors were derived to represent seasonal–diurnal variability and physical interactions (used in the surrogate models):

Table 3. Derived diagnostic and harmonic predictors computed from hourly ERA5 data to capture seasonal–diurnal variability and key physical interactions. These predictors were used as features in the surrogate modelling and elasticity experiments.

Type	Derived variable	Formula / Note
Seasonal & diurnal harmonics	DOY_sin, DOY_cos, HOD_sin, HOD_cos	Sine/cosine of day- and hour-of-year to capture periodicity
Wind harmonics	wind10_dir_sin, wind10_dir_cos, ws10	Wind direction converted to sine/cosine; speed (ws10=)
Precipitation metrics	tp_mph, ln_tp	Hourly precipitation intensity and its natural logarithm for numerical stability
Physical ratios	*E/TCWV*, *P/E*, net_sw, net_lw	Evaporation efficiency, precipitation-to-evaporation ratio, radiative index

Together, these hourly ERA5 variables and their derived predictors provide the meteorological basis for both the calculation of the hourly recycling ratio (*RR*) and the construction of machine-learning surrogate models linking precipitation and moisture recycling.

The target variable for modeling is the hourly precipitation recycling rate (*RR*) computed from ERA5 using the algorithm described in Section 2.3. To facilitate regression within the [0, 1] range, its logit transform

$$y_{\text{logit}} = \ln\left(\frac{RR}{1-RR}\right) \quad (1)$$

was used as the response variable in subsequent machine-learning analyses.

2.3 Hourly calculation of precipitation recycling rate (*RR*)

To quantify the short-term evolution of local precipitation recycling, we computed the hourly precipitation recycling rate (*RR*) over the target domain. The approach follows a two-reservoir moisture-tracking framework that distinguishes

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

atmospheric water vapor originating from local surface evaporation from that advected into the domain by large-scale circulation.

2.3.1 Data and domain configuration

Hourly ERA5 variables—including total precipitation (*P*), evaporation (*E*), specific humidity (*q*), horizontal wind components (*u*, *v*), and surface pressure (*sp*)—were extracted for all pressure levels from 2020 to 2024. The data were cropped to the target region (38–39.5°N, 106–107°E) and converted to SI units. For each pressure layer, the pressure thickness (Δp) was derived from adjacent levels, and the vertical water-vapor mass (kg m^{-2}) was calculated as ($W = q \cdot \Delta p/g$), where ($g = 9.81 ms^{-2}$). The horizontal grid spacing (dx, dy) was estimated using spherical geometry at each grid's latitude, and the temporal resolution was set to 1 h ($\Delta t = 3600 s$).

2.3.2 Two-reservoir moisture tracking

At each time step, the total column water vapor was separated into two prognostic components:

- ($W_{\mathrm{loc}}(x, y, p, t)$): vapor mass of local origin (from regional surface evaporation);
- ($W_{\mathrm{adv}}(x, y, p, t)$): vapor mass of advected origin.

At the initial hour (t_0), all atmospheric water vapor is assumed to be advected (($W_{\mathrm{adv}} = W_{\mathrm{tot}}$)), with ($W_{\mathrm{loc}} = 0$). The system is then updated hourly through surface fluxes, precipitation removal, and horizontal advection.

2.3.3 Source and sink processes

Evaporation input. Surface evaporation (*E*, m hour^{-1}) is converted to mass flux ($E = e \cdot \rho_w$, with $\rho_w = 1000\ kg\ m^{-3}$) and added to the lowest pressure layer of (W_{loc}).

Precipitation removal. At each grid and time step, the total precipitation mass flux $P_{\mathrm{tot}} = tp \cdot \rho_w$ is partitioned between the local and advected reservoirs according to the column-integrated local fraction

$$f_{\mathrm{loc}} = \frac{\sum_p W_{\mathrm{loc}}}{\sum_p (W_{\mathrm{loc}} + W_{\mathrm{adv}}) + \varepsilon} \tag{2}$$

where a small ε = 10^{-6} avoids division by zero. Thus,

$$P_{\mathrm{loc}} = f_{\mathrm{loc}} \cdot P_{\mathrm{tot}}, \quad P_{\mathrm{adv}} = (1 - f_{\mathrm{loc}}) P_{\mathrm{tot}} \tag{3}$$

Each column's precipitation sink is vertically distributed in proportion to its layer-wise mass:

$$\Delta W_{\mathrm{loc}}^{\mathrm{precip}}(p) = P_{\mathrm{loc}} \cdot \frac{W_{\mathrm{loc}}(p)}{\sum_p W_{\mathrm{loc}}}, \quad \Delta W_{\mathrm{adv}}^{\mathrm{precip}}(p) = P_{\mathrm{adv}} \cdot \frac{W_{\mathrm{adv}}(p)}{\sum_p W_{\mathrm{adv}}} \tag{4}$$

Horizontal advection. Horizontal transport for each component is simulated using an **upwind finite-difference scheme** applied separately to *u* and *v*:

$$\Delta W^{\mathrm{trans}} = -\frac{\partial (uW)}{\partial x} - \frac{\partial (vW)}{\partial y} \tag{5}$$

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Open boundary conditions are imposed, allowing inflow and outflow consistent with ERA5 circulation.

2.3.4 Prognostic update and physical constraints

The reservoirs evolve as:

$$W_{\mathrm{loc}}^{t+1} = W_{\mathrm{loc}}^{t} * \Delta W_{\mathrm{loc}}^{\mathrm{evap}} - \Delta W_{\mathrm{loc}}^{\mathrm{precip}} \Delta W_{\mathrm{loc}}^{\mathrm{trans}}, \tag{6}$$

$$W_{\mathrm{adv}}^{t+1} = W_{\mathrm{adv}}^{t} - \Delta W_{\mathrm{adv}}^{\mathrm{precip}} * \Delta_{\mathrm{adv}}^{\mathrm{trans}}. \tag{7}$$

All negative values are clipped to zero to preserve non-negativity. Mass closure is periodically checked to ensure

$$\Delta(W_{\mathrm{loc}} + W_{\mathrm{adv}}) \approx E - P_{\mathrm{tot}} + \nabla \cdot (q\mathbf{V})\Delta t. \tag{8}$$

2.3.5 Hourly and daily recycling indices

For each hour, the **local precipitation amount** is the area integral of (P_{loc}), and the **hourly recycling rate** is defined as:

$$RR(t) = \frac{\sum_{x,y} P_{\mathrm{loc}}(x,y,t)}{\sum_{x,y} P_{\mathrm{tot}}(x,y,t)} \tag{9}$$

To obtain the daily mean *RR*, both (P_{loc}) and (P_{tot}) are first summed over all valid hourly records within the day (≥ 18 hours) before taking their ratio, ensuring consistency with the water-budget definition.

2.3.6 Numerical stability and quality control

To maintain stability, the Courant–Friedrichs–Lewy (CFL) criterion was checked for typical wind speeds and grid spacing. Small negative residuals were truncated ($RR \in [10^{-4}, 1-10^{-4}]$). Diagnostic tests confirmed that domain-integrated moisture mass was conserved within ±0.5 % per day. The resulting hourly *RR* fields were subsequently aggregated for event detection and machine-learning analysis described in later sections.

2.4 Definition and interpretation of the percentage elasticity ($\eta\%$)

The percentage elasticity ($\eta\%$) is introduced in this study as a scale-consistent indicator to describe the sensitivity of precipitation recycling to changes in rainfall intensity. The concept arises from the observation that both the recycling rate (*RR*) and total precipitation (*P*) are dynamic variables that co-evolve during rainfall episodes, yet their relationship is often nonlinear and bounded. Conventional correlation or regression coefficients cannot capture this behavior in a physically interpretable way, particularly when *RR* approaches zero or one.

To overcome these limitations, $\eta\%$ quantifies the relative change in recycled precipitation $R = RR \times P$ in response to a fractional change in total precipitation, defined as:

$$\eta_{R,P} = \frac{\partial \ln(RR \cdot P)}{\partial \ln P} \tag{10}$$

For convenience of interpretation, this elasticity is expressed as a percentage deviation from proportional growth:

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

$$\eta\% = 100 \times (\eta_{R,P} - 1) \quad (11)$$

so that $\eta\% = 0$ represents a proportional (one-to-one) increase of recycled and total precipitation, $\eta\% > 0$ indicates that recycling amplifies rainfall (positive feedback), and $\eta\% < 0$ denotes a self-limiting or externally driven response.

The derivation of $\eta\%$ is grounded in both mathematical consistency and physical interpretability:

1. it is dimensionless, allowing comparison across events and regions;
2. it retains symmetry between amplification and suppression responses;
3. it connects directly to the physical partitioning between local and advected moisture sources.

In practical computation, $\eta\%$ is evaluated from paired simulations or predictions under perturbed precipitation conditions, using the finite-difference form:

$$\eta\% \approx 100 \times \frac{ln(RR' \cdot P') - ln(RR \cdot P)}{lnP' - lnP} - 100 \quad (12)$$

where RR' and P' denote the values after a controlled perturbation (Section 2.8). This formulation ensures numerical stability even for small changes in P and prevents unbounded derivatives near $RR \rightarrow 0$ or 1 by operating in logarithmic space. The robustness and boundary behavior of $\eta\%$ are further assessed by bootstrap resampling and sensitivity tests.

By design, $\eta\%$ serves as a diagnostic indicator rather than a fixed physical constant. It enables concise comparison of recycling responsiveness across meteorological regimes, event types, or model scenarios, providing a quantitative bridge between the instantaneous (hourly) moisture budget and the cumulative hydrological feedback at the event scale.

2.5 Event detection and sample construction

Daily precipitation totals were first derived from the hourly records, and only valid days with ≥ 18 hourly values were retained. To objectively identify region-wide wet episodes, a dual-threshold algorithm was applied:

1. the regional mean of daily precipitation exceeds the 95th percentile (Ricetti et al., 2025) (p95 ≈ 3 mm day^{-1}) calculated from the 2020–2024 baseline; and
2. at least 60 % of the five stations record daily precipitation above their individual p95 thresholds.

Consecutive qualifying days were merged into a single multi-day event if their duration exceeded two days. This procedure ensures both intensity and spatial coherence, effectively distinguishing regional precipitation systems from isolated convective cells. Applying these criteria produced 9 multi-day wet events (E1–E9) between 2020 and 2024. Each event was further verified against ERA5 precipitation to confirm temporal alignment and spatial extent.

For each event window, the corresponding hourly time series of P and recycling rate were extracted and standardized. These event-based samples form the foundation for the cluster classification and surrogate modeling described below.

Throughout this paper, an event-scale wet episode refers to a multi-day (≥2 days) regional rainfall event identified by a dual threshold (regional mean P > p95 and ≥60% stations exceeding their p95)

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

2.6 Machine-learning surrogate modeling

2.6.1 Objective and target variable

To quantify how meteorological factors jointly influence the short-term recycling rate, we trained a suite of machine-learning surrogate models that predict Eq. (1) from concurrent atmospheric conditions.

2.6.2 Predictor matrix

Predictors include all physically relevant variables listed in Section 2.2, supplemented by lagged and smoothed representations to capture short-term memory effects:

1. Lag terms: 1- and 2-hour lags for key variables.
2. Rolling means: 3-hour moving averages of the same variables.
3. Interaction features: e_over_tcw, tp_over_e, uv10m_prod, radiative_idx.
4. Seasonal–diurnal harmonics: sin/cos of DOY and HOD.

2.6.3 Pre-processing and normalization

Only hours with measurable precipitation (`is_rain = 1`) were included. All predictors were standardized using per-event z-score normalization to remove inter-event mean differences while retaining within-event variability. Events with negligible variance in y were excluded to prevent numerical artifacts. After removing NaNs from lag/rolling operations, the final dataset contained several tens of thousands of hourly samples spanning nine events.

2.6.4 Model architecture and training

Four tree-based ensemble regressors were tested—**CatBoost** (Hancock and Khoshgoftaar, 2020), **XGBoost** (Chen and Guestrin, 2016), **ExtraTrees** (Berrouachedi et al., 2019; Martiello Mastelini et al., 2023), and **Gradient Boosting** (Bentéjac et al., 2021)—selected for their strong performance on nonlinear, medium-sized environmental datasets. The models were trained and validated using a **Leave-One-Event-Out (LOEO)** cross-validation scheme, in which each of the nine events served once as the test fold. This structure ensures that temporal autocorrelation within events does not leak into training, preserving independence between calibration and evaluation (Auddy et al., 2024; Vehtari et al., 2017; Zou et al., 2024). Performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root-mean-square error (RMSE) on the *y_logit* scale. Residual-vs-prediction diagnostics confirmed negligible bias, and SHAP (Shapley Additive exPlanations) values (Lundberg and Lee, 2017) were computed to interpret feature contributions consistently across algorithms.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

2.7 Cluster analysis and event-type classification

To identify physically distinct regimes of wetting feedback, we applied a **K-means clustering** (Kodinariya et al., 2013; Likas et al., 2003) procedure to event-averaged diagnostic variables that co-determine recycling elasticity: $\eta\%$, precipitation, evaporation, total column water vapor, 10-m wind speed, and 2-m temperature etc. Prior to clustering, all features were standardized and projected onto the first two principal components (explaining ≈ 75 % of variance) to assess separability.

The optimal cluster number K = 3 was chosen based on silhouette analysis and physical interpretability. The resulting clusters represent:

1. **Cluster 0 – Intermediate/Advection-dominated:** balanced humidity but weak coupling.
2. **Cluster 1 – Cool–Moist:** enhanced local feedback with high total column water vapor, evaporation and moderate 2-m temperature.
3. **Cluster 2 – Warm–Dry weak coupling:** low total column water vapor and strong surface heating.

Composite analyses (± 12 h around event peaks) were performed to examine the joint evolution of precipitation, *RR*, and $\eta\%$. Within each cluster, linear regressions of $\eta\%$ against key covariates (*TCWV, E, WS10,* etc.) quantified their dominant controls. The coherence between composite timing and regression slopes demonstrates that these clusters correspond to genuine physical regimes rather than statistical artifacts. This classification forms the structural basis for the elasticity analyses presented in Section 3.

2.8 Counterfactual experiments and elasticity evaluation

To evaluate how precipitation recycling responds to changes in atmospheric and surface conditions under different environmental regimes, a series of cluster-conditioned counterfactual experiments was conducted using the trained surrogate models (Section 2.6). These experiments follow the ceteris paribus logic at the regime level: within each physically coherent cluster (Section 2.7), a selected set of key meteorological variables is scaled together to represent plausible wetting or drying perturbations, while preserving their internal physical relationships learned from the observations.

For each cluster, a characteristic subset of predictors was identified to represent the dominant moisture and energy controls of precipitation recycling. Specifically, these include *P*, *E*, and *TCWV* as the principal moisture variables, optionally accompanied by near-surface temperature (*T2M*) and surface pressure (*sp*) when thermodynamic adjustment is relevant. Within each cluster, all selected variables were scaled proportionally by constant factors:

$$(X'_j)=s\times(X_j),\quad s\in\{0.80, 0.90, 1.10, 1.20, 1.50, 1.80, 2.00\} \tag{13}$$

where *Xj* denotes the *j-th* variable in the cluster-specific perturbation group. The range of s covers both moderate drying (<1) and wetting (>1) scenarios to test the symmetry of recycling response. Other predictors not directly related to the perturbed group (e.g., wind components, radiation fluxes) were kept fixed to their observed values so that each experiment isolates the effect of changing local moisture and energy supply.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

The surrogate models were then applied to predict the corresponding recycling rates (RR') for every perturbed state in logit space, ensuring bounded and numerically stable outputs. For each event and perturbation level, the changes in RR and recycled precipitation ($R = RR \cdot P$) were used to compute the percentage elasticity ($\eta\%$), following the finite-difference approximation introduced in Section 2.4 (Eq. 13)

Here, P' refers to the perturbed precipitation corresponding to the same scaling factor s. Positive $\eta\%$ values indicate amplification of local recycling relative to total rainfall, while negative $\eta\%$ values denote a reduction of local feedback efficiency.

To account for environmental diversity, $\eta\%$ was evaluated separately for each cluster, and ensemble means together with 95 % confidence intervals were estimated by bootstrap resampling over event members within the cluster. This cluster-conditioned design allows the elasticity of precipitation recycling to be interpreted in a physically meaningful way—for instance, distinguishing between moisture-rich and moisture-limited regimes or between advection-dominated and locally coupled conditions.

The counterfactual framework offers several methodological advantages. First, by perturbing variable groups rather than single predictors, it preserves the internal covariances typical of different atmospheric regimes and yields more realistic composite responses. Second, because the perturbations are implemented through machine-learning surrogates, nonlinear interactions among variables are automatically retained, allowing the derived $\eta\%$ to reflect the full state dependence of recycling efficiency. Third, the framework is data efficient and robust: it reuses observed event states without requiring new model integrations, which is particularly valuable for regions like north-western China where multi-day rainfall events are infrequent and records are short. Together, these features make the counterfactual approach a practical and physically interpretable tool for quantifying short-term precipitation recycling elasticity under diverse environmental conditions.

This section establishes the complete methodological pipeline—from high-resolution data acquisition and hourly computation of precipitation recycling to event identification, surrogate modeling, and cluster-based classification. This framework provides a physically consistent foundation for the subsequent results in Section 3, where the temporal evolution, predictive skill, and regime-specific elasticity of short-term wetting feedback are analyzed in detail.

3 Results

3.1 Event characteristics and selection logic

To capture short-term wetting feedbacks with direct physical relevance, we first established a catalogue of regional wet events during 2020–2024 using hourly station observations. Figure 1 summarizes the climatological background, detection thresholds, and the temporal structure of each selected episode, forming the foundation for the subsequent cluster and elasticity analyses.

Before identifying specific events, it is important to understand the precipitation regime of the study domain, which lies within a semi-arid transition zone between the Tengger Desert and the northern Loess Plateau. Precipitation here is

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

infrequent but highly concentrated in summer. The five national meteorological stations used for event detection (Fig. 1a) exhibit similar seasonal cycles but with distinct intensity levels, reflecting the region's north–south moisture gradient. Their 95th-percentile (p95) daily precipitation thresholds range from 2.5 to 3.4 mm day^{-1}, with higher values toward the southern foothills. The monthly climatology (Fig. 1b) reveals a pronounced wet season from July to September, when both mean and p95 precipitation reach their maxima, indicating that extreme events coincide with the monsoon moisture surge. The complementary cumulative distribution functions (Fig. 1c) further confirm the heavy-tailed distribution of daily precipitation, with exceedance probabilities of 20 mm day^{-1} around 1–2 %. Correspondingly, wet-day frequency and mean intensity (Figs. 1d–e) highlight that a few short-lived storms dominate the annual rainfall budget—conditions under which moisture recycling can vary rapidly and nonlinearly.

Building on this climatological context, regional wet events were identified through a dual-threshold procedure designed to ensure both intensity and spatial coherence. Hourly station data were first aggregated to daily totals, retaining only "valid days" with at least 18 recorded hours. Within the 2020–2024 baseline, a regional p95 threshold was computed from the mean of the five stations. A day was defined as a regional event day when two criteria were simultaneously satisfied: (1) the regional-mean daily precipitation exceeded its p95 (≈ 3 mm day^{-1}), and (2) at least 60 % of stations recorded values above their individual p95 thresholds. Consecutive event days were merged into a single multi-day episode if their duration exceeded two days. This dual-threshold approach ensures that the detected events represent spatially extensive precipitation systems rather than isolated convective cells.

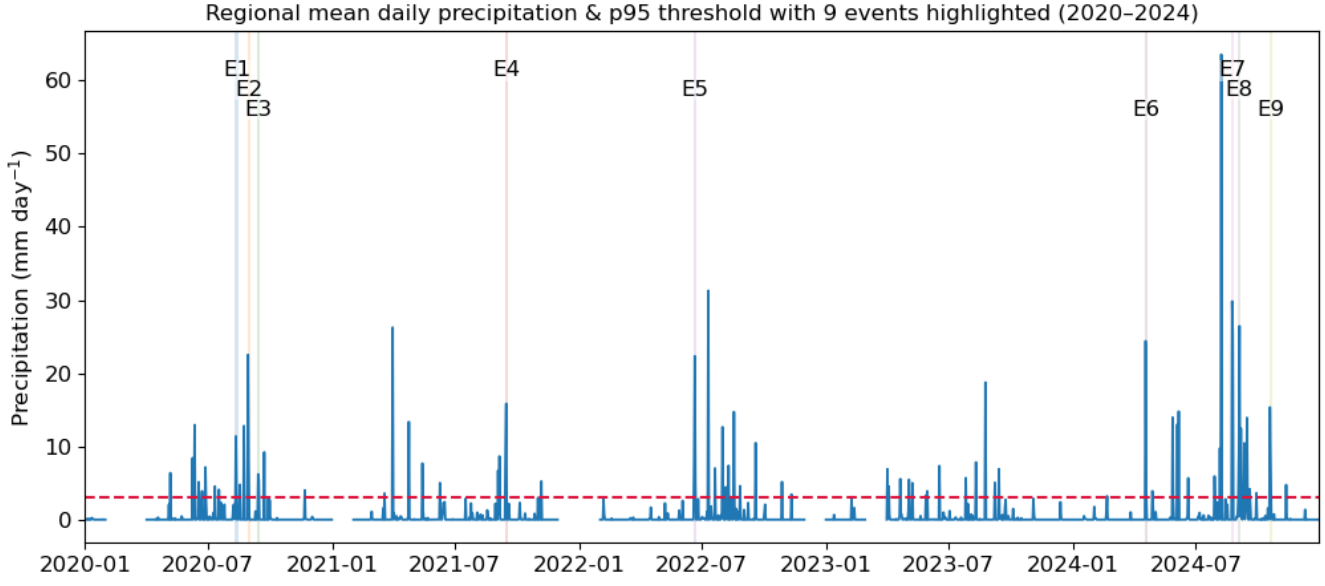


Figure 2. Regional mean daily precipitation (blue bars) and the 95th-percentile (p95) threshold (red dashed line) from 2020 to 2024, with nine multi-day wet events (E1–E9) highlighted by shaded vertical bands. The identified events represent periods when the regional mean precipitation exceeded the p95 threshold and at least 60% of stations recorded concurrent heavy rainfall, ensuring both intensity and spatial coherence. These episodes form the core sample set for subsequent analyses of precipitation recycling and wetting feedback.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Applying these criteria produced nine multi-day events between 2020 and 2024 (E1–E9 in Fig. 2). Their durations range from two to four days (mean = 2.6 days), with peak regional precipitation between 10- and 25-mm day^{-1}. Most events occur in late summer and early autumn, consistent with periods of enhanced total-column water vapor and reduced surface pressure in ERA5. Specifically, E1–E3 cluster in August–September 2020, E4–E5 represent isolated peaks in 2021–2022, and E6–E9 span spring to autumn 2024. The infrequent occurrence of such regionally synchronized anomalies—less than 2% of all days—emphasizes their extremity. Comparison of station and ERA5 daily precipitation yields a correlation of 0.86 and an RMSE of ≈ 3 mm day^{-1}, confirming that ERA5 reliably reproduces both timing and intensity, thereby validating its use in subsequent recycling analyses.

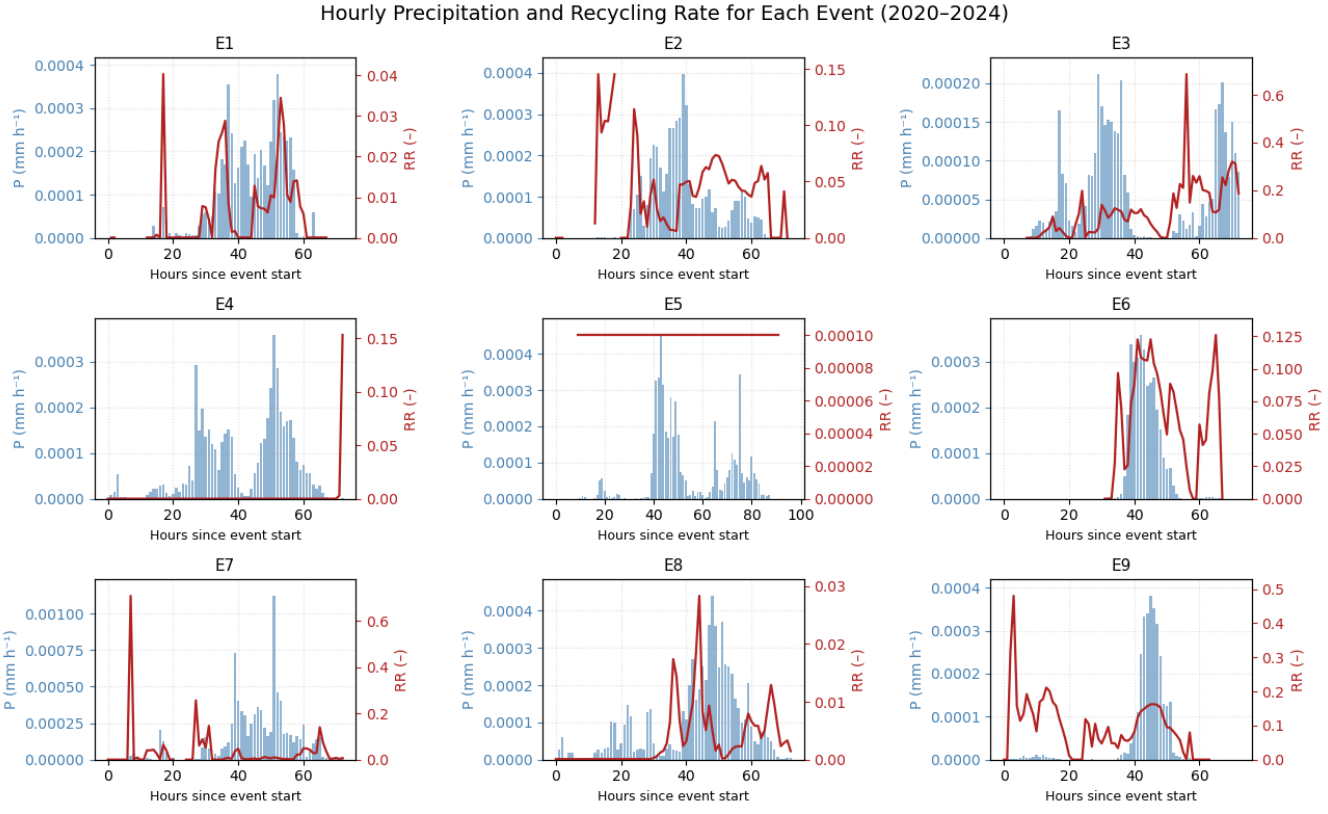


Figure 3. Hourly evolution of precipitation (*P*; blue bars, mm h^{-1}) and precipitation recycling rate (*RR*; red lines, dimensionless) for the nine identified multi-day wet events (E1–E9) during 2020–2024. Each panel shows the temporal evolution of *P* and *RR* since the onset of the event. The results reveal diverse phase relationships between rainfall and recycling across events—some exhibiting near-synchronous peaks, others showing lead–lag structures—indicating varying degrees of local moisture feedback and atmospheric coupling within the region's transient wetting processes.

To further examine how these regional events evolve internally, each was analyzed at the hourly scale for precipitation and the corresponding *RR*. As shown in Fig. 3, *RR* varies coherently with precipitation but exhibits diverse phase relationships among events. In some cases (e.g., E1 and E7), *RR* peaks slightly before or after the main rainfall burst, implying a temporal lag between local evaporation and condensation. In others (e.g., E3 and E4), *RR* increases sharply near the precipitation

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

maximum, indicating rapid recycling triggered by strong convection. Such contrasts highlight that the recycling process evolves dynamically within hours, governed by the interplay between surface evaporation, boundary-layer moistening, and advective inflow. This high-frequency variability provides direct empirical evidence that hourly *RR* diagnostics are essential for resolving short-term wetting feedbacks.

Taken together, these nine region-wide wet episodes offer a physically consistent and well-validated sample set characterized by strong but short-lived rainfall and distinct hourly recycling dynamics. They thus form the empirical foundation for the cluster classification and counterfactual elasticity experiments described in next Sections, where the mechanisms and feedback strength of each event type are examined in detail.

3.2 Event-type classification (cluster analysis)

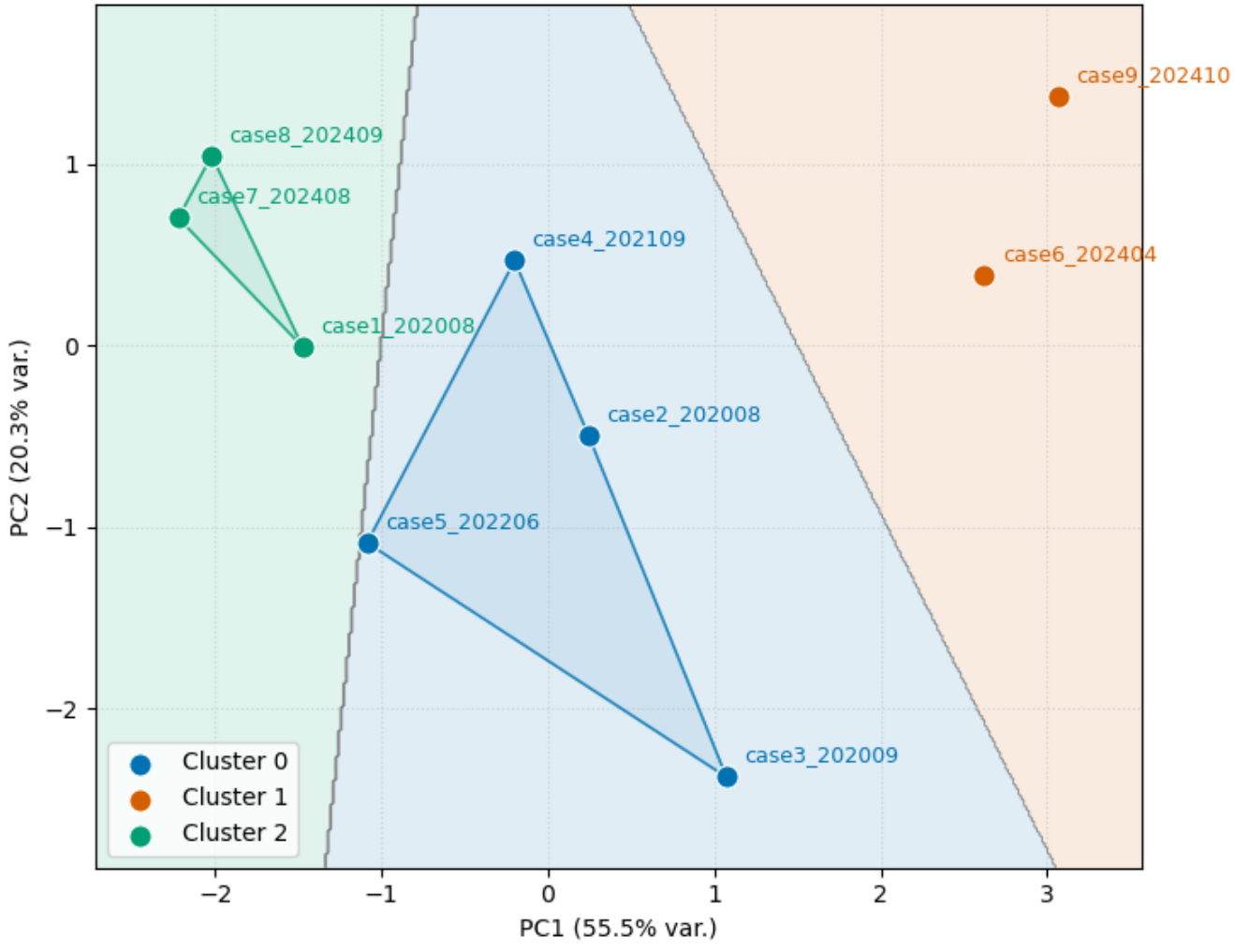


Figure 4. Principal component projection of the nine rainfall events (E1–E9) based on event-mean diagnostic variables that co-determine precipitation recycling elasticity (*η*%, *P*, *E*, *TCWV*, WS_{10}, and T_2M). The first two principal components (PC1 and PC2)

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

explain 55.5% and 20.3% of the total variance, respectively. Three physically distinct clusters are identified: Cluster 0 (blue) — intermediate or advection-dominated events; Cluster 1 (green) — cool–moist events with enhanced local feedback; and Cluster 2 (orange) — warm–dry weak-coupling events. The separation among clusters highlights the contrasting environmental controls governing short-term wetting feedback across regimes.

To partition wet episodes into physically distinct regimes, we applied K-means to event-averaged features that co-determine short-term recycling elasticity: *η%*, *P*, *E*, *TCWV*, *WS10* and *T2M*. The principal-component projection (Fig. 4) indicates that the first two components explain 55.5% and 20.3% of variance, respectively, and already separate three compact groups. Decision boundaries, overlaid as background sectors, show minimal overlap, with the two spring–autumn events (E6, E9) forming a tight warm-side cluster, three late-summer events (E1, E7, E8) occupying the cool–moist sector, and four remaining cases spreading along the intermediate axis. This geometric separation is mirrored in the normalized feature means (Fig. 5): the warm-side group exhibits the highest *T2M* and above-normal winds, the cool–moist group stands out for high *TCWV* but low *T2M*, while the intermediate group shows weak winds and near-neutral humidity.

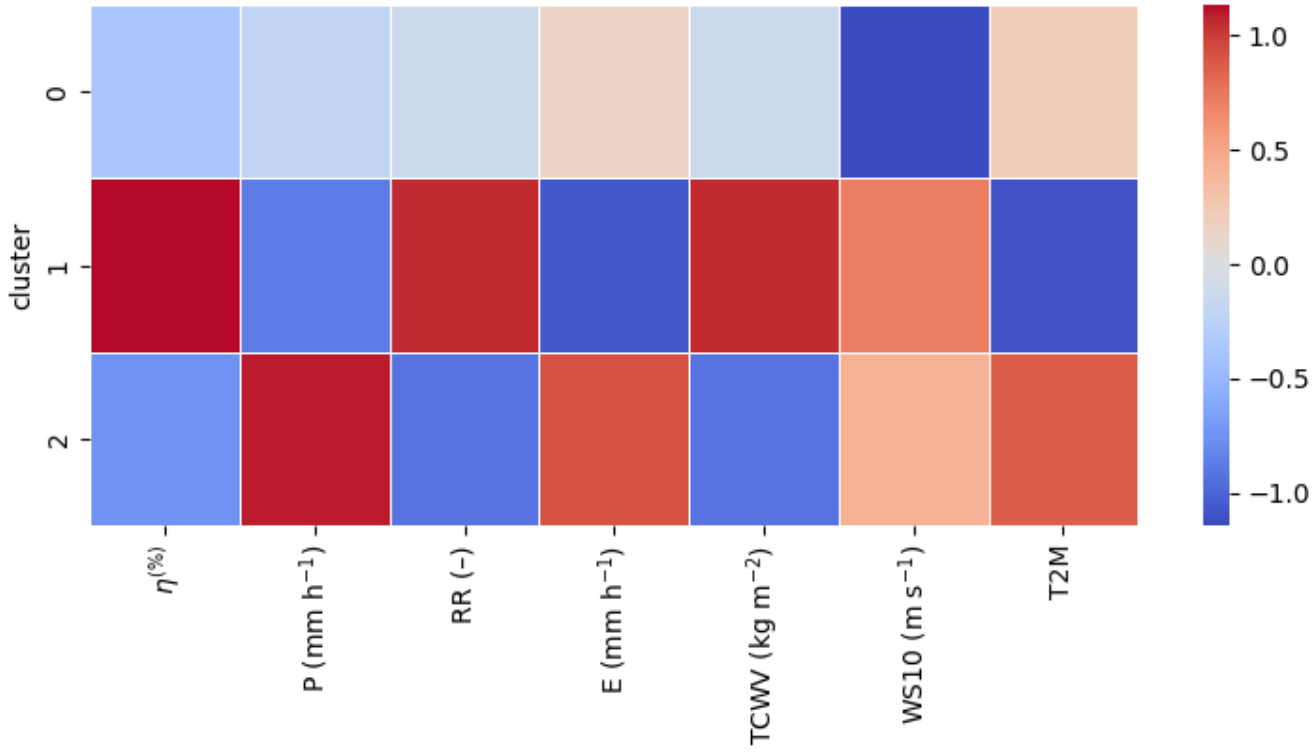


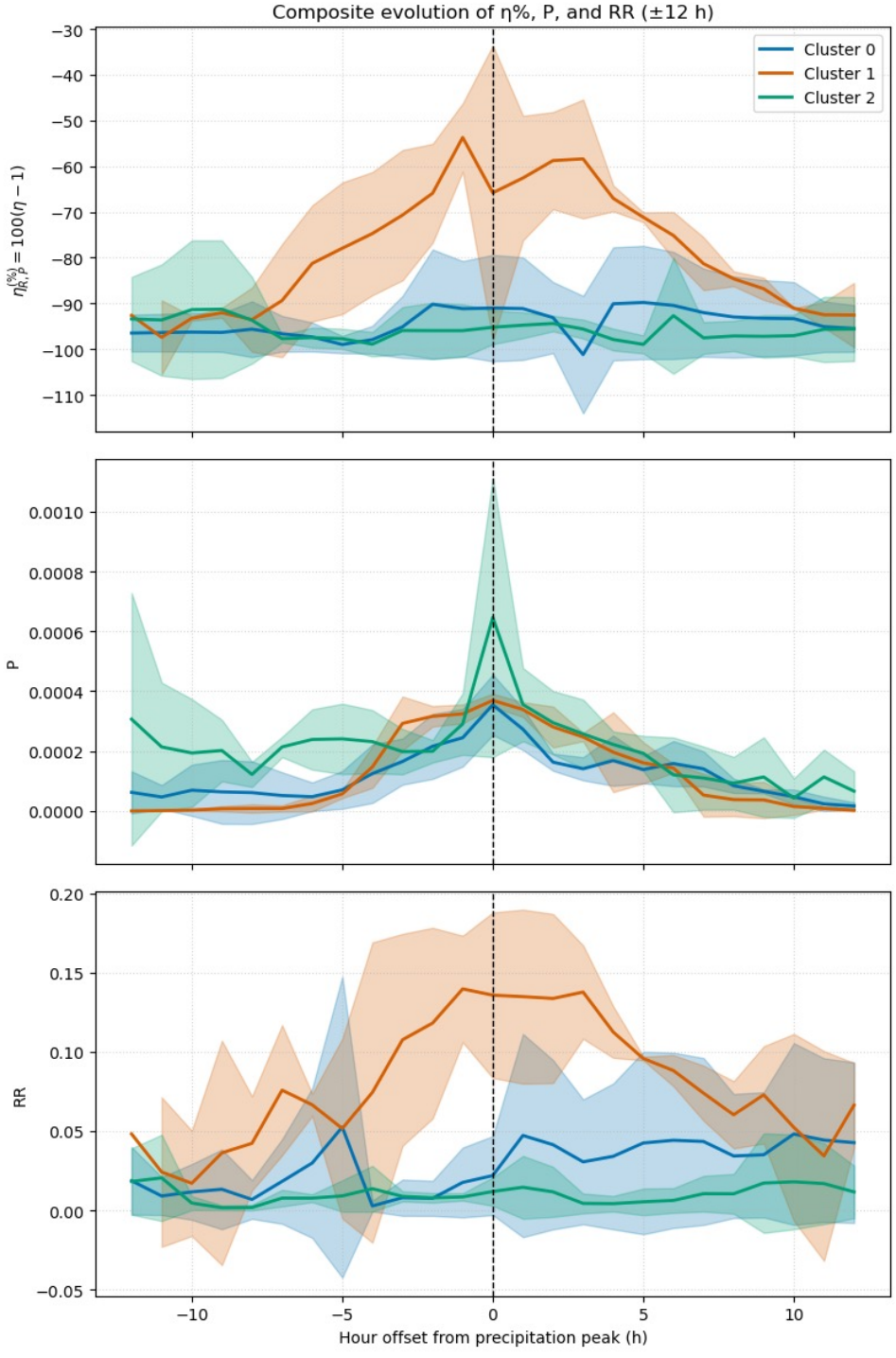
Figure 5. Cluster-mean standardized anomalies of key variables related to precipitation recycling elasticity. Variables include *η%, P, RR, E, TCWV, WS₁₀, and T₂M*. Warm colours denote above-average values and cool colours indicate below-average values relative to the full sample mean. The three clusters exhibit distinct physical regimes: Cluster 0 features weak variability across all factors; Cluster 1 represents humid and high-feedback events with elevated *η%, RR,* and *TCWV*; while *Cluster 2* corresponds to warm, windy, and evaporation-dominated events associated with low recycling efficiency.

Time-resolved composites around the precipitation peak (Fig. 6) clarify how this background states translate into feedback behavior. In the cool–moist group (Cluster 1; orange), *η%* ramps up several hours before the rainfall maximum, reaches its

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

crest near the peak, and then decays, whereas *RR* attains elevated values slightly after the peak. This lead–lag pairing—pre-peak rise in elasticity followed by a delayed *RR* response—indicates that boundary-layer moistening preconditions the column and that the local return fraction materializes as the event matures. The intermediate group (Cluster 0; blue) shows nearly flat $\eta\%$ punctuated by a shallow dip near the peak, while *P* and *RR* evolve almost in lockstep; such phase locking is consistent with externally forced, advection-dominated episodes where local recycling remains secondary. The warm-side group (Cluster 2; green) maintains uniformly low $\eta\%$ with a sharp, short-lived spike in *P* but little response in *RR*, pointing to dry–hot conditions where evaporation–cloud coupling is weak and precipitation arises mainly from dynamical or instability triggers rather than local return.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025



385

Figure 6. Composite evolution of precipitation recycling elasticity (η% = 100 × (η – 1)), *P*, and *RR* within ±12 h of the precipitation peak for each cluster. Shaded envelopes denote ±1 standard deviation among events within each cluster. Cluster 1 (orange) exhibits a pronounced rise in η% and RR preceding and peaking near the maximum rainfall, indicating a transient enhancement

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

of local feedback under humid and low-wind conditions. Cluster 0 (blue) shows moderate coupling with a delayed *RR* response, while Cluster 2 (green) maintains persistently low *η%* and *RR* despite strong precipitation pulses, reflecting weak land–atmosphere interaction in warm, advective environments.

Cluster-wise regressions between *η%* and environmental covariates (Fig. 7) provide further resolution. For the intermediate group, *η%* increases strongly with *TCWV* ($r=0.71$, $p<0.001$) and weakens with wind speed ($r=-0.62$, $p<0.01$), a pattern that fits an import-controlled regime: stronger advection dilutes the local fraction, while humid backgrounds permit modest recycling when import relaxes. In the cool–moist group, *η%* decreases with *TCWV* ($r=-0.69$, $p<0.001$) but rises with P ($r\approx0.38$, $p\approx0.08$), suggesting that once the column is already humid, additional ambient moisture no longer boosts elasticity unless accompanied by precipitation that recycles part of the boundary-layer supply—consistent with the pre-conditioning seen in Fig. 6. In the warm-side group, *η%* grows with *E* ($r=0.50$, $p=0.005$) and with wind speed ($r=0.48$, $p<0.01$), indicating that in otherwise dry conditions only unusually strong surface fluxes and mixing can partially offset the coupling deficit; dependence on *P* is essentially flat. While sample sizes per cluster are limited, the signs and magnitudes are coherent with the composite timing and with the background contrasts in Fig. 5.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

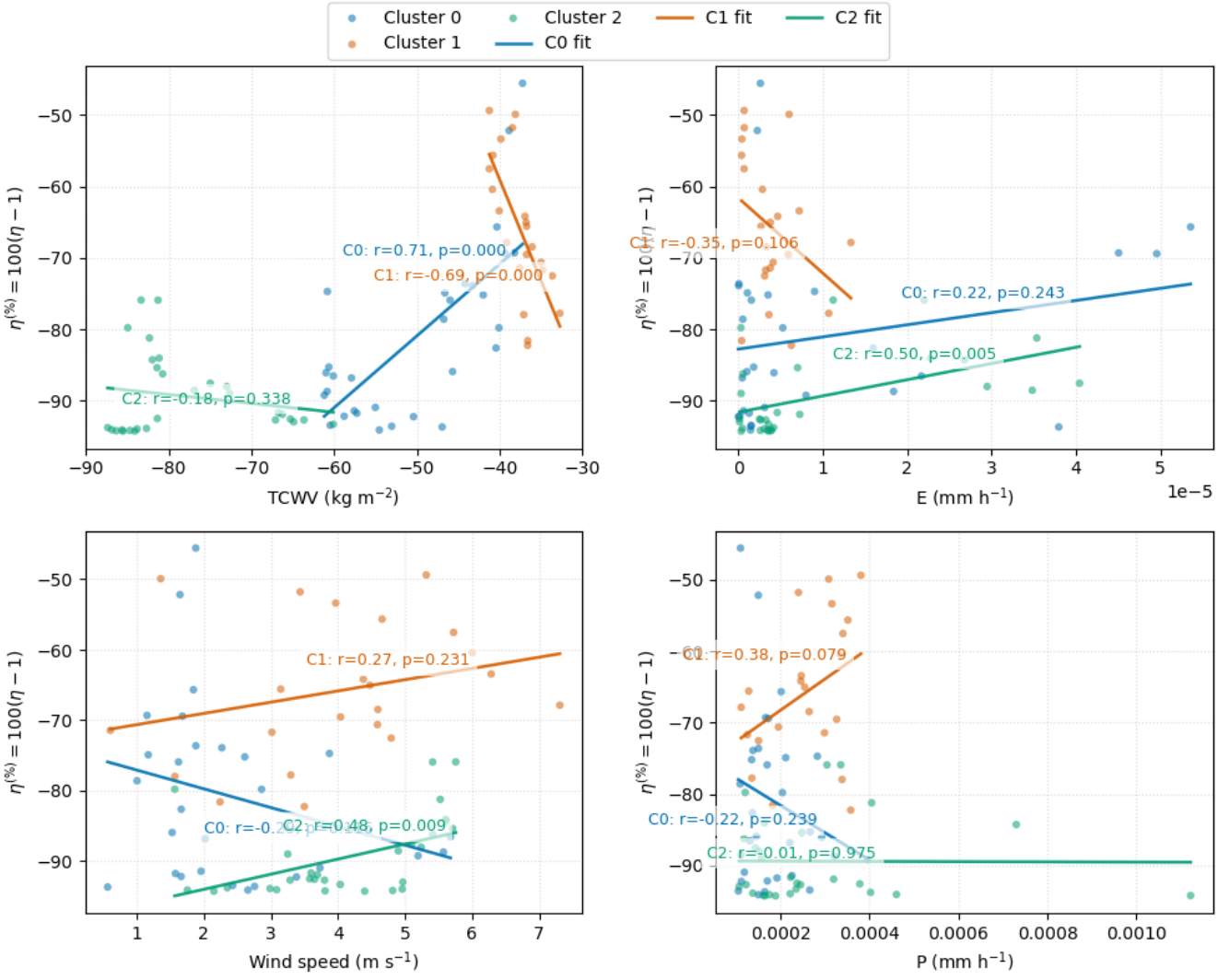


Figure 7. Relationships between precipitation recycling elasticity *η%* and key meteorological factors by cluster. Panels show correlations of *η%* with (a) *TCWV*, (b) *E*, (c) 10 m wind speed, and (d) *P*. Solid lines denote least-squares fits for each cluster, with corresponding Pearson correlation coefficients (r) and p-values annotated. Cluster 0 (blue) exhibits a strong positive dependence of *η%* on *TCWV*, suggesting moisture-limited feedbacks; Cluster 1 (orange) shows a negative *η%–TCWV* relationship characteristic of saturation-limited regimes; and Cluster 2 (green) displays generally weak or inconsistent associations, reflecting a decoupled, advection-dominated hydrological environment.

Distributional evidence underscores that these are not transient differences tied to a single hour. After removing near-zero-recycling states (*η%*<−95%), the filtered histograms (Fig. 8) reveal that the cool–moist group occupies the least negative portion of elasticity space with a discernible right tail toward weakly positive values, whereas the warm-side group remains concentrated near the lowest *η%* and the intermediate group sits between them. Taken together with the composite and regression evidence, the classification can be read as data-driven yet process-consistent: cool–moist episodes favor short-

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

term reinforcement of local recycling; import-dominated episodes are feedback-neutral to negative; dry–hot episodes are most decoupled.

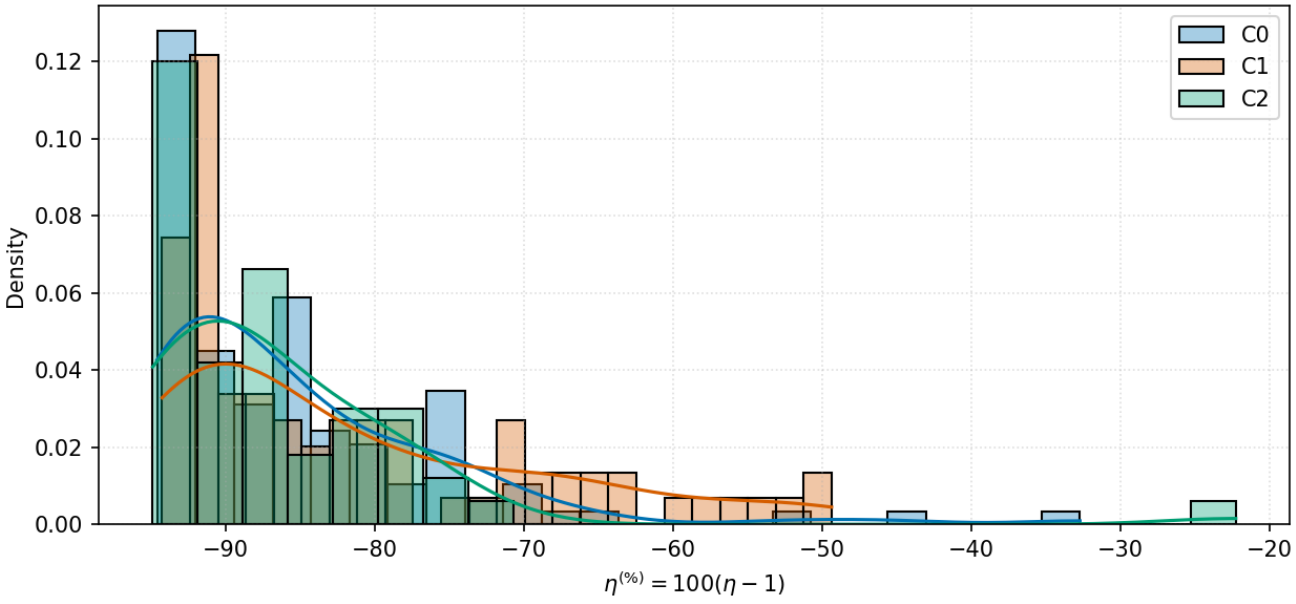


Figure 8. Probability density distributions of precipitation recycling elasticity $\eta\%$ for the three identified clusters (C0–C2). Coloured histograms and kernel density estimates illustrate the frequency and shape of $\eta\%$ variability. All clusters show predominantly negative $\eta\%$ values, confirming the dominance of self-damping wetting responses. Cluster 1 (orange) exhibits a wider and flatter distribution, indicating stronger event-to-event variability and intermittent positive feedback, whereas Clusters 0 (blue) and 2 (green) display more concentrated, negatively skewed distributions, consistent with stable but feedback-limited hydrometeorological regimes.

The conceptual pathways in Fig. 9 synthesize this picture. In advection-dominant events (Cluster 0), external moisture imports sets both P and RR simultaneously, leaving $\eta\%$ nearly invariant. In local wetting-feedback events (Cluster 1), enhanced evaporation and boundary-layer moistening precede cloud development, recycled moisture augments precipitation, and $\eta\%$ temporarily strengthens around the peak. In weak-coupling dry–hot events (Cluster 2), low E and humidity suppress the BL–cloud linkage; precipitation, when it occurs, is largely decoupled from local return, yielding the lowest $\eta\%$. This mechanistic scaffold motivates the stratified analyses that follow: because each cluster represents a distinct balance between local supply and advective control, identical wetting perturbations in Section 3.3 will not project onto elasticity in the same way across regimes.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

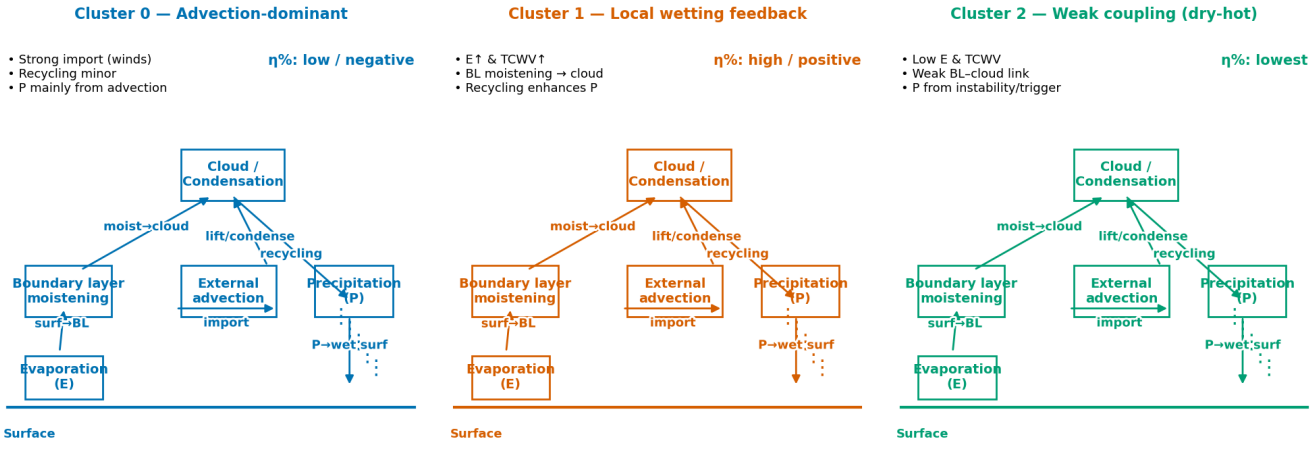


Figure 9. Conceptual schematic of the three precipitation–recycling regimes identified through cluster analysis. Cluster 0 – Advection-dominant: Characterized by strong external moisture import and wind-driven advection. Recycling is minor, and $\eta\%$ remains low or negative as precipitation is primarily sustained by imported moisture rather than local feedback. Cluster 1 – Local wetting feedback: Exhibits enhanced E and $TCWV$, with efficient boundary-layer moistening leading to condensation and precipitation reinforcement. Recycling amplifies rainfall, producing higher and occasionally positive $\eta\%$. Cluster 2 – Weak coupling (dry–hot): Features low E and $TCWV$, with minimal boundary-layer–cloud interaction. Precipitation arises mainly from dynamic instability or external triggers, yielding the lowest $\eta\%$ and negligible local feedback.

3.3 Model performance and predictive skill

We assessed surrogate skill using four tree-based learners trained on the logit-transformed recycling rate, ($y = logit(RR)$): CatBoost, XGBoost, ExtraTrees, and Gradient Boosting. Figure 10 compares observed and predicted (y) for each model. All algorithms reproduce the first-order spread of the data with comparable accuracy (CatBoost: R^2=0.608, MAE = 0.682, RMSE = 1.057; XGBoost: R^2=0.614, MAE = 0.671, RMSE = 1.049; ExtraTrees: R^2=0.631, MAE = 0.652, RMSE = 1.026; Gradient Boosting: R^2=0.598, MAE = 0.708, RMSE = 1.071). ExtraTrees attains the highest coefficient of determination, but the differences among models are small, indicating that the signal linking meteorological predictors to RR is robust across ensemble-type methods. Visually, points align closely with the one-to-one line over the densest range of observations (roughly $y \in [-4,-2]$), where most events occur. Dispersion increases toward the dry tails (very negative (y)), consistent with intrinsically low and noisy recycling fractions in those conditions and the effect of bound-preserving transformations near ($RR \to 0$).

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

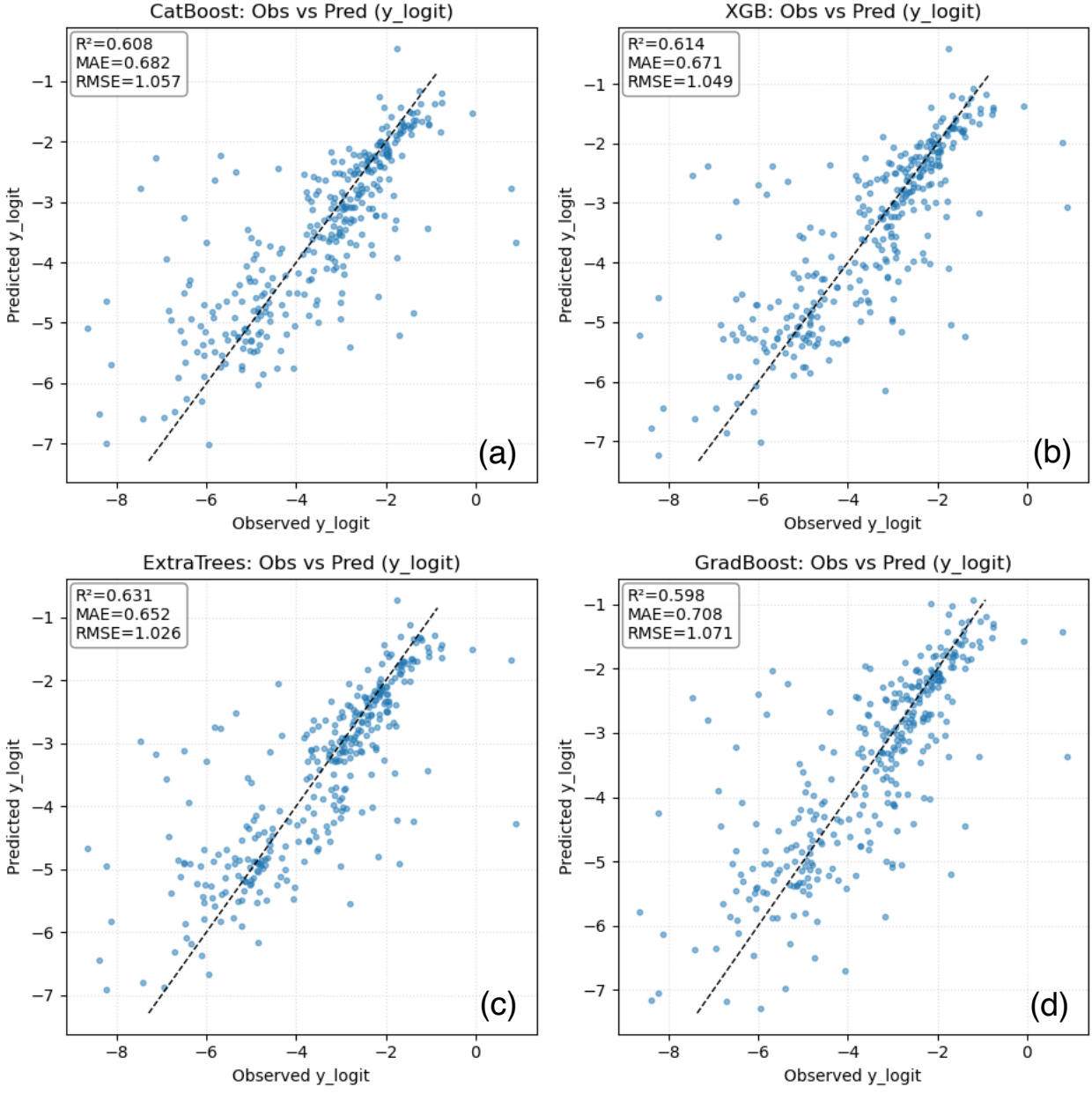


Figure 10. Observed versus predicted logit-transformed recycling rate (y_logit = logit(RR)) from four machine-learning surrogate models under event-blocked cross-validation. (a) CatBoost, (b) XGBoost, (c) ExtraTrees, and (d) Gradient Boosting. Each panel shows the one-to-one comparison between model predictions and observed values, with coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) annotated.

To examine bias and variance structure, Fig. 11 plots residuals (Obs–Pred) against predictions. Median biases are negligible in all cases (CatBoost −0.007; XGBoost −0.019; ExtraTrees +0.001; Gradient Boosting −0.011), and the smoothed residual curves fluctuate around zero without systematic drifts. A modest widening of the confidence envelope appears at the

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

extremes of (y), reflecting heteroscedasticity typical of event-scale hydrometeorological data: when *RR* is very small or, conversely, approaches the upper tail on the logit scale, small absolute errors translate into larger deviations. Importantly, no persistent over- or under-prediction is evident across the central range, supporting the use of these models as unbiased surrogates for counterfactual perturbations.

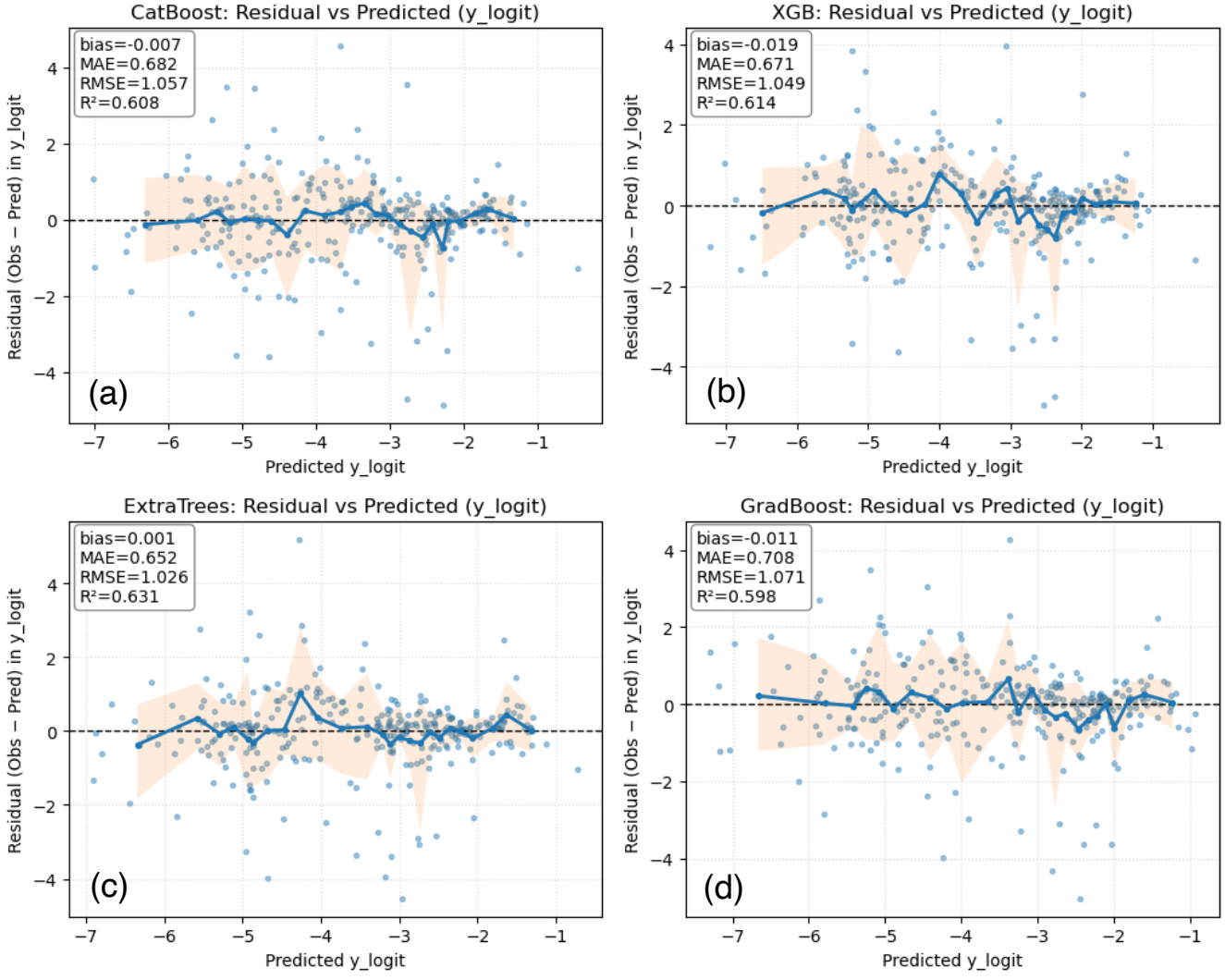


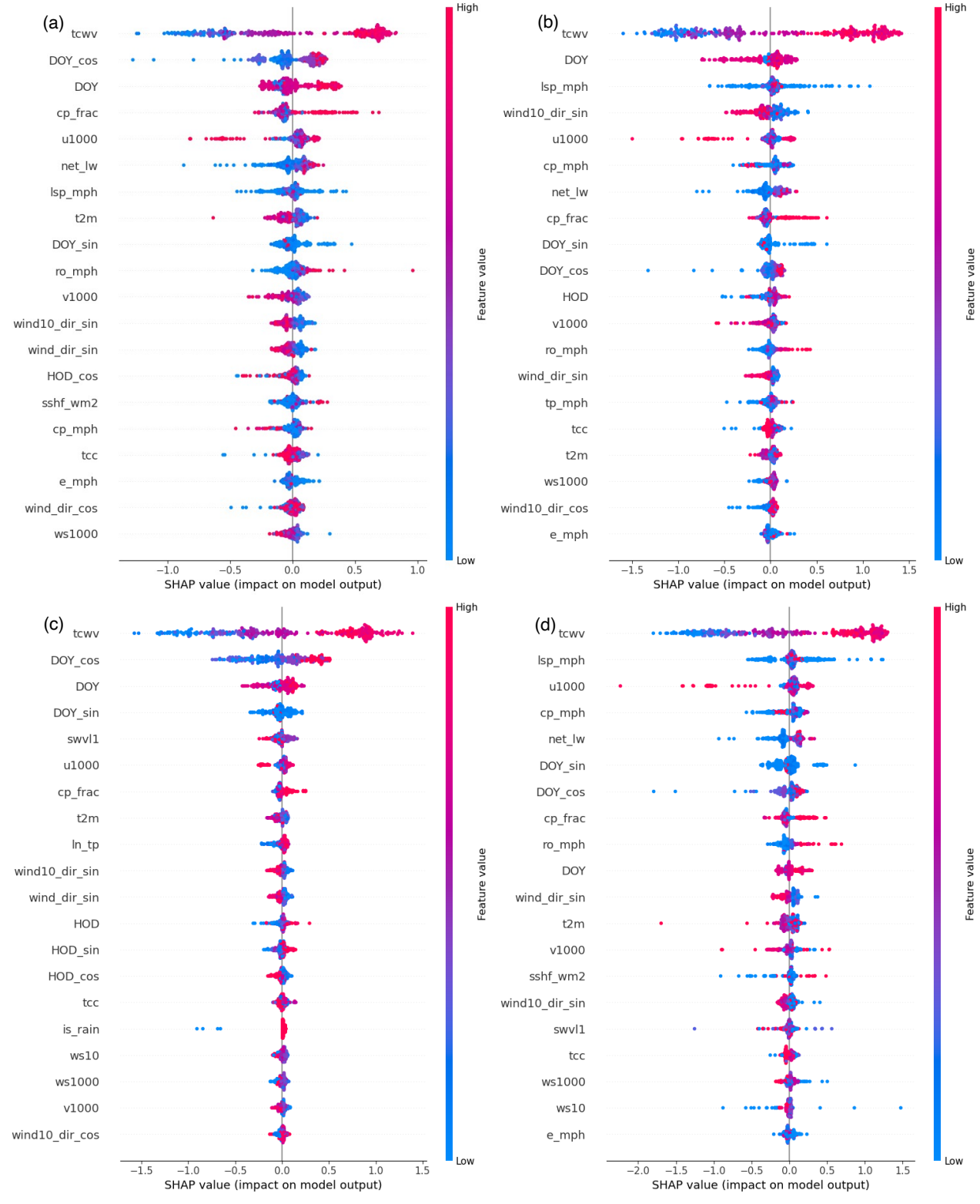
Figure 11. Residual–prediction relationships for logit-transformed recycling rate (*y_logit*) across four machine-learning surrogate models: (a) CatBoost, (b) XGBoost, (c) ExtraTrees, and (d) Gradient Boosting. Residuals are defined as (Observed − Predicted) and plotted against predicted y_logit, with smoothed mean residuals (solid line) and ±1σ spread (shaded area).

Model interpretability is consistent with physical expectations. Figure 12 shows SHAP summary plots for the four learners; despite algorithmic differences, the ranking of influential predictors is stable. Total column water vapor dominates across models, with high *TCWV* values pushing predictions toward larger (y) (higher *RR* on the probability scale). Large-scale and convective precipitation terms (e.g., lsp_mph, cp_mph or cp_frac) also exert strong positive influence, indicating that the

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

occurrence and character of precipitation co-vary with the recycled fraction at event time scales. Lower-tropospheric kinematics (e.g., *u10m*, wind-direction harmonics) contribute meaningfully, consistent with the role of advection in supplying moisture and organizing ascent. Net longwave flux and near-surface temperature (*T2M*) show mixed but non-negligible effects, likely modulating surface–boundary-layer coupling. In contrast, instantaneous evaporation proxies (e_mph) carry comparatively weak importance in the hourly setting, which accords with the short response time of column moisture and the dominance of humidity/advection controls during precipitation episodes. The broad agreement of SHAP patterns across Fig. 4–7 indicates that our conclusions about drivers are not artifacts of a single algorithm.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025



https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Figure 12. SHAP summary plots showing the relative importance and direction of predictor effects on the modelled recycling rate (y_logit) across four machine-learning algorithms: (a) CatBoost, (b) XGBoost, (c) ExtraTrees, and (d) Gradient Boosting. Each point represents an individual hourly sample, coloured by the magnitude of the input feature value. Features are ordered by their mean absolute SHAP value (importance).

Taken together, the accuracy diagnostics (Fig. 10), residual behavior (Fig. 11), and feature attributions (Fig. 12) point to a consistent picture: the surrogate ensemble captures the central variability of RR with moderate skill on the logit scale ($RR^2 \approx 0.60\sim0.63$), little systematic bias, and physically plausible dependence on humidity, precipitation character, and low-level flow. These properties provide sufficient fidelity for the counterfactual wetting experiments that follow, where small differences among algorithms are unlikely to dominate the estimated elasticity $\eta\%$. For transparency, subsequent sections report cluster-conditioned results as well, acknowledging that predictive difficulty increases in regimes with extremely low recycling.

3.4 Elasticity responses and cluster-specific mechanisms under wetting perturbations

Having established that the surrogate models reproduce the observed variability of precipitation recycling with little bias (Section 3.3), we now apply them to counterfactual "wetting experiments" to explore how the recycling elasticity $\eta\%$ responds to progressive increases in precipitation. These experiments provide a dynamic test of short-term wetting feedback: whether additional rainfall tends to amplify, suppress, or leave unchanged the local moisture recycling efficiency under different environmental regimes.

3.4.1 Quantitative responses across clusters

The multi-model mean results (Fig. 13) reveal systematic but non-linear contrasts among clusters. For the **wetter-local type (Cluster 1)**, $\eta\%$ declines sharply with moderate wetting: at a 50 % precipitation increase, the mean elasticity falls by roughly −180%. This steep drop indicates that once the boundary layer becomes nearly saturated, further rainfall yields diminishing returns to local recycling—the classic signature of a self-limiting feedback. At stronger wetting (+100 %–200 %), $\eta\%$ recovers slightly, suggesting that very intense events can briefly reactivate evaporation and recycling through enhanced surface fluxes.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

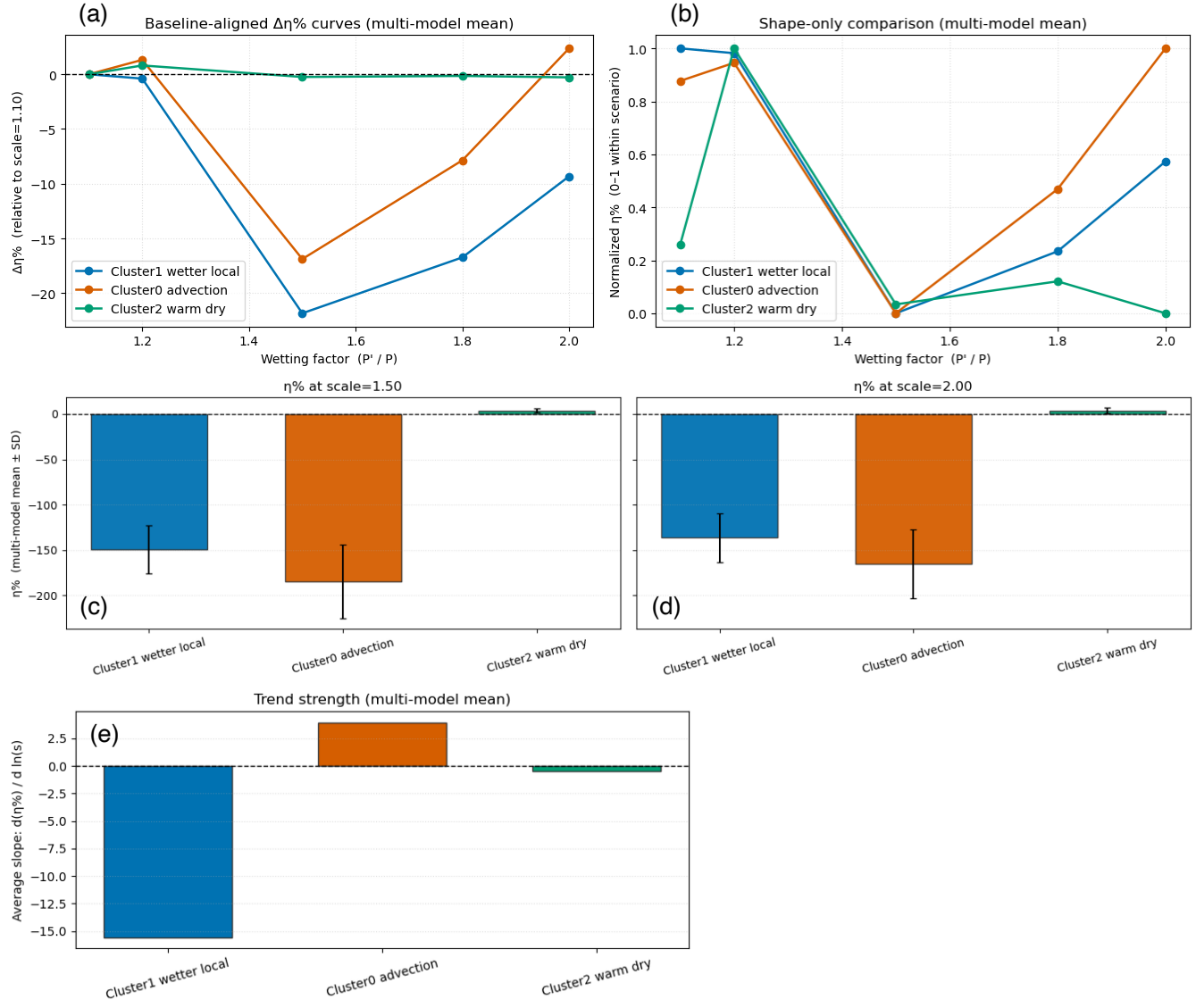


Figure 13. Cluster-dependent responses of *η%* to synthetic wetting perturbations based on multi-model ensemble results. (a) Baseline-aligned *Δη%* curves showing deviations relative to the +10% precipitation scenario (scale = 1.10). (b) Normalized *η%* curves (0–1 scaling within each cluster) highlighting the relative shape and sensitivity of the response. (c–d) Mean *η%* (±1 SD) at moderate (+50%) and strong (+100%) wetting levels (scales = 1.50 and 2.00). (e) Multi-model mean slope of *ln(η%)* with respect to ln(s), representing the trend strength of feedback under progressive wetting. Cluster 1 ("wetter–local") shows a pronounced non-linear recovery pattern consistent with self-reinforcing but transient feedback; Cluster 0 ("advection–dominant") exhibits an inverted U-shaped response indicating self-limiting feedback; and Cluster 2 ("warm–dry") remains nearly invariant, confirming its weak coupling and insensitivity to enhanced precipitation.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

The **advection-dominated type (Cluster 0)** also exhibits negative elasticity but with smaller amplitude (≈ −120 % at +50 % P), consistent with precipitation being primarily controlled by external moisture import; additional rainfall thus proportionally strengthens advection while leaving the local fraction almost unchanged.

In contrast, the **warm–dry weak-coupling type (Cluster 2)** remains essentially flat ($\eta\% \approx 0 \pm 20$ %) across the entire wetting range, implying that precipitation changes in this regime do not translate into appreciable modifications of the local moisture-recycling share.

3.4.2 Curve shape and trend strength

Normalized curves (Fig. 13b) highlight the differing sensitivities: Cluster 1 shows a pronounced concave-down shape, Cluster 0 a mild quasi-linear decrease, and Cluster 2 an almost flat response. Bar plots at representative scales (Figs. 13c–d) confirm that $\eta\%$ stays substantially negative for Clusters 1 and 0 but hovers around zero for Cluster 2. Averaged slopes (Fig. 13e) quantify this gradient: $d\eta\%/dlns \approx -8$ for Cluster 1, −3 for Cluster 0, and ≈ 0 for Cluster 2.

Hence, as the environment becomes wetter, the elasticity of local recycling weakens most rapidly where coupling is already strong (Cluster 1), moderately where advection dominates (Cluster 0), and hardly at all in the dry–hot weak-coupling background (Cluster 2).

3.4.3 Physical interpretation

These contrasting sensitivities point to distinct controlling mechanisms. In **Cluster 1**, abundant ambient moisture and strong boundary-layer coupling initially favor positive feedback, but incremental wetting quickly shifts the balance toward external import, lowering the fractional recycled component. In **Cluster 0**, precipitation scales with advection rather than local recycling, producing a nearly linear yet consistently negative $\eta\%$. In **Cluster 2**, limited total-column water vapor and high temperature decouple evaporation from cloud formation, yielding near-neutral elasticity regardless of precipitation change.

The schematic synthesis in **Fig. 14** encapsulates these processes.

- **Cluster 1 – Wetter-Local:** enhanced E and $TCWV$ drive initial moistening, but progressive wetting dilutes the recycled share ($\eta\% \approx -150 \rightarrow -200$).
- **Cluster 0 – Advection-Dominated:** external import rises with P, keeping the recycling ratio almost constant ($\eta\% \approx -100 \rightarrow -150$).
- **Cluster 2 – Warm-Dry Weak Coupling:** low humidity and strong heating suppress BL–cloud linkage, leading to neutral or slightly positive $\eta\%$ ($\approx 0 \rightarrow +20$).

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

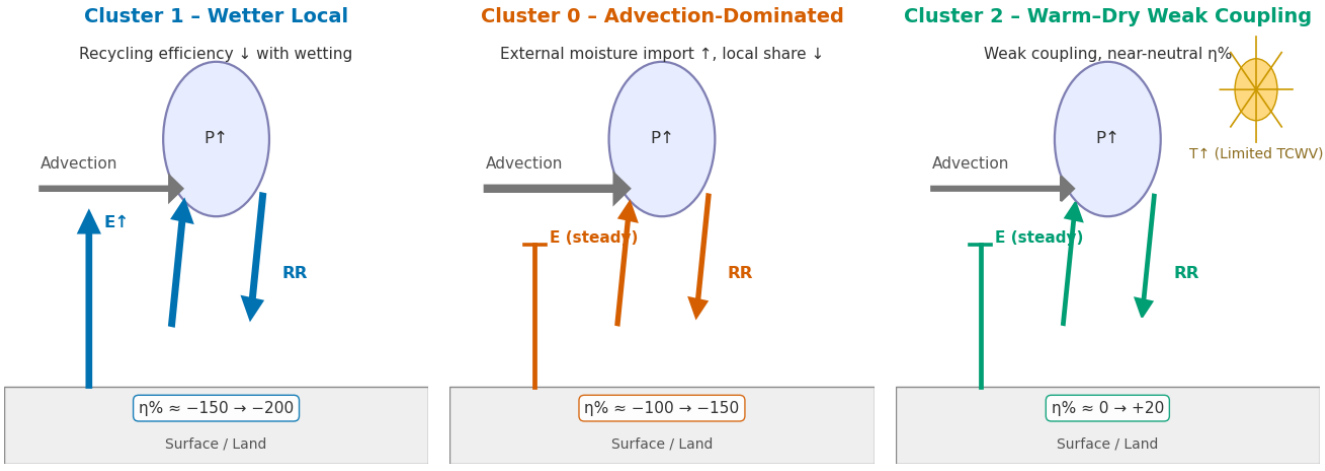


Figure 14. Schematic summary of cluster-specific feedback mechanisms governing the response of precipitation recycling elasticity (*η%*) to enhanced wetting. Cluster 0 – Advection-Dominated: Precipitation growth is mainly sustained by enhanced external moisture import, while local recycling remains secondary (*η%* ≈ −100 → −150). Cluster 1 – Wetter Local: Increasing *E* and *P* amplify near-surface *RR*, but the recycling efficiency declines with further wetting (*η%* ≈ −150 → −200), indicating a transient, self-dampening feedback. Cluster 2 – Warm–Dry Weak Coupling: Characterized by limited total column water vapour (TCWV) and weak land–atmosphere interaction; *η%* stays near-neutral (≈ 0 → +20), implying negligible feedback under dry–hot conditions.

Together, these results show that **short-term wetting does not uniformly amplify local recycling**; instead, elasticity responses depend critically on environmental background. What appears at regional scale as a weakly negative mean *η%* actually conceals three physically distinct behaviors—from strongly self-limiting feedbacks to nearly decoupled regimes. This hierarchy (Cluster 1 < Cluster 0 < Cluster 2 in *η%*) underscores that the potential for short-term "self-reinforcing wetting" exists only under specific thermodynamic conditions, and that continued rainfall in humid events can paradoxically weaken, rather than intensify, local moisture recycling.

These findings set the stage for the broader discussion (Section 4) on how such short-term elasticity patterns integrate to influence seasonal-scale water-cycle efficiency across Northwest China's arid transition zone.

4 Discussion

4.1 Physical nature of short-term wetting feedback

The counterfactual experiments reveal that short-term wetting in the semi-arid/arid region transition zone is not a self-reinforcing process. Instead, additional rainfall frequently reduces the efficiency of local moisture recycling, as indicated by negative elasticity values *η%*. This response—most evident in the wetter-local regime (Cluster 1)—arises from rapid boundary-layer saturation and energy constraints that limit the return of evaporated moisture to precipitation.

At the onset of a wet episode, enhanced evaporation and boundary-layer moistening promote convective development and short-term positive feedback (Su and Dickinson, 2017; Taylor et al., 2010, 2011) Yet once humidity increases further, latent-heat release stabilizes the column, the lapse rate weakens, and convective inhibition rises. Evaporation continues, but its

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

contribution is confined to the lower troposphere rather than feeding precipitation growth. Consequently, $\eta\%$ declines sharply with modest wetting (≈ –180 % at +50 % P) and only recovers slightly at the highest perturbations (Fig. 13). The atmosphere thus acts as a self-limiting system that damps excessive moistening over hours to days.

This short-term stabilization differs fundamentally from the positive land-atmosphere feedbacks inferred at monthly or climatic timescales. On event scales, moisture and energy constraints dominate; on seasonal scales, soil-moisture replenishment and circulation feedbacks may restore a positive tendency. The negative $\eta\%$ documented here therefore represents the intrinsic regulation of the local hydrological loop rather than a contradiction of long-term wetting trends.

4.2 Dynamic and thermodynamic controls across regimes

Comparisons among clusters clarify how external dynamics and internal coupling jointly shape $\eta\%$. In the advection-dominated regime (Cluster 0), precipitation scales almost linearly with imported moisture. When tp increases, both external vapor flux and total rainfall rise together, leaving the recycled share—and hence $\eta\%$—nearly unchanged ($d\eta\%/dlns \approx -3$). This negative but modest elasticity reflects passive dilution: local evaporation contributes a constant proportion while advection determines overall magnitude.

The warm–dry weak-coupling regime (Cluster 2) represents the opposite limit. Low $TCWV$ and high surface temperature produce large vapor-pressure deficits and shallow boundary layers. Even substantial increases in P fail to enhance coupling between E and clouds, yielding nearly flat $\eta\%$ (≈ 0 ± 20 %). Precipitation originates mainly from dynamic lifting or instability triggers rather than recycled moisture.

Together, the three regimes illustrate a continuum from internally limited (Cluster 1) to externally controlled (Clusters 0 and 2) behavior. The sign and magnitude of $\eta\%$ provide a diagnostic of control dominance: strong negative values indicate saturation or advective dilution, whereas near-zero elasticity marks thermodynamic decoupling. This framework links the elasticity metric to physical moisture pathways—clarifying whether rainfall variability stems from local recycling or regional transport.

4.3 From event-scale feedbacks to regional wetting and hydrological implications

Although derived from individual events, these short-term feedbacks scale up to influence regional water-balance efficiency. Event-scale negative elasticity implies that each storm converts a smaller fraction of surface evaporation into rainfall as the atmosphere becomes moister. When aggregated over months, this behavior manifests as rising precipitation totals but declining recycling ratios—a “wetter but externally fed” regime already evident in reanalysis and satellite observations across north-western China.

This interpretation reconciles our findings with earlier climatological studies (Hua et al., 2016; Jiao et al., 2024; Manisha et al., 2023; Zhang et al., 2025), which reported weakly positive or neutral recycling trends at large spatial and temporal scales. Those long-term means represent the integrated outcome of many events whose internal feedbacks alternate between pre-conditioning and saturation. The hourly elasticity resolved here therefore bridges a critical temporal gap between synoptic

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

and seasonal perspectives, explaining why the western Loess Plateau and Alxa corridor can experience more frequent rainfall events yet little change in evaporation or latent-heat flux.

For hydrology and water resources, a declining $\eta\%$ indicates reduced atmospheric water retention within basins. Even if annual precipitation increases, the proportion recycled locally shrinks, making the region more dependent on moisture inflow from surrounding humid zones. Consequently, future wetting will hinge on large-scale circulation rather than local feedbacks. Should monsoonal or westerly transports weaken, precipitation may drop abruptly despite a warming surface (Gao et al., 2023; Guo et al., 2024).

The cluster framework also explains spatial heterogeneity in wetting responses. Cluster 1-type oases or irrigated areas may show transient positive feedbacks followed by suppression; Cluster 0-type mountain corridors exhibit stable but externally forced wetting; Cluster 2-type dry basins remain largely inert. Importantly, stronger rainfall does not equate to a tighter local hydrological loop—the efficiency of internal recycling may even decline as total moisture increases. The associated energy redistribution further moderates temperature extremes: latent-heat consumption during storms cools the surface, stabilizing the boundary layer and reducing the likelihood of consecutive convective events.

4.4 Methodological limitations and conceptual synthesis

While the overall consistency among models and clusters strengthens confidence in our conclusions, several methodological factors should be acknowledged. The event catalog (N = 9) is necessarily limited; expanding it across multiple years will allow more robust statistics. The surrogate models isolate precipitation perturbations while holding other predictors fixed, capturing instantaneous rather than fully coupled responses. Future work should perturb E and $TCWV$ jointly to explore multi-variable elasticities. Cluster assignment is static, based on event-mean conditions; dynamic clustering could reveal within-event regime transitions. Extreme RR values near 0 or 1 introduce additional uncertainty in logit space, despite bounding corrections. Finally, orographic and mesoscale convective effects were not explicitly represented but likely influence local recycling efficiency.

Despite these limitations, the coherence of results across independent algorithms and the physically interpretable ordering of clusters confirm that the detected elasticity patterns are robust. Integrating these findings yields a unified conceptual picture of short-term moisture recycling in the arid–semiarid transition zone (Fig. 14).

Precipitation recycling operates through the interplay of moisture availability, advective import, and thermodynamic coupling.

1. In local-wetting events, positive feedback initiates but quickly saturates ($\eta\% \approx -150 \rightarrow -200$).
2. In advection-dominated events, external moisture inflow scales precipitation linearly ($\eta\% \approx -100 \rightarrow -150$).
3. In warm–dry events, coupling is weak and $\eta\%$ remains near zero.

These regimes represent a continuum from internal stabilization to external control. Negative $\eta\%$ signifies that the atmosphere actively regulates its own moisture budget, preventing excessive amplification of local wetting. In aggregate, the

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

atmosphere above north-western China behaves as a self-stabilizing hydrological system—one in which increased precipitation does not necessarily strengthen, and may even weaken, local moisture recycling.

Recognizing this self-limiting character is crucial for interpreting future climate projections. Regional wetting trends should be assessed not only by the magnitude of precipitation change but also by its elasticity with respect to recycling, which encapsulates the transient balance between land and atmosphere. Accounting for this elasticity will improve the realism of water-resource projections and deepen our understanding of how arid-zone hydrological systems maintain stability amid a changing climate.

5 Conclusion

This study developed an integrated framework to quantify short-term precipitation recycling and wetting feedback in the semi-arid transition zone during 2020–2024. Using hourly station and ERA5 reanalysis data, the two-reservoir moisture-tracking scheme provided physically consistent estimates of the hourly recycling rate. Event-based sampling and machine-learning surrogate models revealed how local evaporation, atmospheric humidity, and circulation jointly regulate the efficiency of moisture recycling under varying wetting conditions.

Three main conclusions emerge:

Short-term wetting acts as a self-limiting process. Counterfactual experiments demonstrate that additional rainfall generally weakens local recycling efficiency rather than reinforcing it. The elasticity $\eta\% < 0$ indicates that once boundary-layer humidity rises, energy constraints suppress further convective development. The atmosphere therefore damps excessive moistening over hourly-to-daily scales.

Distinct regimes govern the variability of feedback strength. Cluster analysis identifies a continuum from internally saturated to externally controlled systems. In cool–moist (Cluster 1) events, recycling begins positive but quickly saturates, producing strong negative elasticity. In advection-dominated (Cluster 0) events, imported moisture scales linearly with rainfall, leaving the recycled fraction nearly constant. In warm–dry (Cluster 2) events, weak coupling between evaporation and convection yields near-zero elasticity. These regimes reflect transitions from local stabilization to external control of precipitation.

Regional wetting depends increasingly on large-scale moisture inflow. Aggregating event-scale behavior shows that although total precipitation may rise, the recycled share declines. The region is thus entering a “wetter-but-externally-fed” state in which atmospheric self-stabilization limits internal feedback. Future precipitation trends will therefore hinge more on the strength of monsoonal and westerly transports than on local evaporation.

Methodologically, the study bridges the gap between process-level modeling and statistical interpretation. The combination of hourly moisture-tracking, event detection, and machine-learning surrogates provides a transferable template for diagnosing short-term land–atmosphere coupling in other data-sparse arid regions. Despite the small number of events and the simplified perturbation design, the consistency across models confirms the robustness of the negative elasticity signal.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Overall, the results portray the atmosphere above north-western China as a self-stabilizing hydrological system—one in which enhanced rainfall does not necessarily intensify, and may even weaken, local moisture recycling. Recognizing this behavior refines our understanding of precipitation efficiency under transient wetting and provides a physically grounded benchmark for evaluating climate-model projections and water-resource sustainability in arid–semi-arid zones.

Competing Interests

The authors declare that they have no competing financial or non-financial interests that could have influenced the results, interpretations, or conclusions of this study.

Code and Data Availability

The ERA5 hourly reanalysis data used in this study are publicly available from the Copernicus Climate Data Store (CDS) operated by the European Centre for Medium-Range Weather Forecasts (ECMWF) at https://cds.climate.copernicus.eu/. The in-situ meteorological station data were provided by the Ningxia Meteorological Bureau and are not publicly available due to institutional data-sharing policies. These data can be obtained from the authors upon reasonable request.

To facilitate reproducibility and comply with FAIR principles, all Python scripts developed for data preprocessing, hourly moisture recycling calculation, event identification, machine-learning surrogate modelling, and counterfactual elasticity (η%) experiments are openly available. The source code is maintained on GitHub at https://github.com/ruolinliTama/2025-Precipitation-Recycling-response-wetting-trend and is permanently archived at Zenodo with the digital object identifier https://doi.org/10.5281/zenodo.17670690.

Interactive Computing Environment

All analyses were performed in an interactive computing environment based on **Python 3.12** under a **Linux (CentOS Steam)** system. The workflow integrated the following key packages and tools:

- Data handling and analysis: xarray, pandas, numpy, scipy
- Visualization: matplotlib, cartopy, seaborn
- Machine learning and statistical modeling: scikit-learn, xgboost, catboost, joblib, shap
- Model interpretation and counterfactual experiments: custom scripts for elasticity ($\eta_(R,P_)$) and moisture recycling diagnostics
- Reproducible computing and documentation: Jupyter Notebook, VS Code, and institutional Linux HPC servers

All codes used for data preprocessing, event identification, surrogate modeling, and elasticity analysis were written in Python and are available from the authors upon request.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Author contributions

RL: writing – original draft, visualization, software, methodology, conceptualization. QF: validation, resources, data curation. YC: investigation, formal analysis.

Financial support

This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDB0720201, Ningxia Hui Autonomous Region key research and development plan of general projects (No 2024BEG03003,2025BEG02028), the National Natural Science Foundation of China under Grant No. 41801015; China Meteorological Administration Innovation Development Special Project under Grant CXFZ2024J043 and the Natural Science Foundation of Ningxia Province under Grant No. 2022AAC05065.

Acknowledgements

This study was supported by the Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences. The authors sincerely thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA5 reanalysis datasets and the Ningxia Meteorological Bureau for supplying the station-based meteorological observations. We also appreciate the valuable discussions and technical assistance from colleagues within our research group

Reference

Ahn, J. M., Kim, J., and Kim, K.: Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting, Toxins, 15, 608, https://doi.org/10.3390/toxins15100608, 2023.

Auddy, A., Zou, H., Rad, K. R., and Maleki, A.: Approximate leave-one-out cross validation for regression with ℓ_1 regularizers, IEEE Transactions on Information Theory, 70, 8040–8071, https://doi.org/10.1109/TIT.2024.3450002, 2024.

Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G.: A comparative analysis of gradient boosting algorithms, Artif Intell Rev, 54, 1937–1967, https://doi.org/10.1007/s10462-020-09896-5, 2021.

Berrouachedi, A., Jaziri, R., and Bernard, G.: Deep Cascade of Extra Trees, in: Trends and Applications in Knowledge Discovery and Data Mining, vol. 11607, edited by: U., L. H. and Lauw, H. W., Springer International Publishing, Cham, 117–129, https://doi.org/10.1007/978-3-030-26142-9_11, 2019.

Bisselink, B. and Dolman, A. J.: Recycling of moisture in Europe: contribution of evaporation to variability in very wet and dry years, Hydrology and Earth System Sciences, 13, 1685–1697, https://doi.org/10.5194/hess-13-1685-2009, 2009.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

C3S: ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/CDS.ADBB2D47, 2018.

Chen, H., Hsu, P., and Hu, S.: Role of Quasi-Biweekly Cloud-Radiative Feedback in Modulating and Simulating Extreme Rainfall Intensity Over Asian Monsoon Regions, Geophysical Research Letters, 51, e2024GL111671, https://doi.org/10.1029/2024GL111671, 2024.

Chen, J.-P., Tsai, T.-C., Tzeng, M.-D., Liao, C.-S., Kuo, H.-C., and Hong, J.-S.: Microphysical Perturbation Experiments and Ensemble Forecasts on Summertime Heavy Rainfall over Northern Taiwan, Weather and Forecasting, 37, 1641–1659, https://doi.org/10.1175/WAF-D-22-0004.1, 2022.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Chi, H., Wu, Y., Zheng, H., Zhang, B., Sun, Z., Yan, J., Ren, Y., and Guo, L.: Spatial patterns of climate change and associated climate hazards in Northwest China, Sci Rep, 13, 10418, https://doi.org/10.1038/s41598-023-37349-w, 2023.

Copernicus Climate Change Service: Complete ERA5 global atmospheric reanalysis, https://doi.org/10.24381/CDS.143582CF, 2023.

Day, J. J., Vitart, F., Stockdale, T., De Rosnay, P., Ardilouze, C., Peano, D., Sanna, A., Fröhlich, K., and Andrews, M.: Soil-moisture-atmosphere coupling hotspots and their representation in seasonal forecasts of boreal summer, Clim Dyn, 63, 275, https://doi.org/10.1007/s00382-025-07753-1, 2025.

Dey, D., Aldama Campino, A., and Döös, K.: A complete view of the atmospheric hydrologic cycle, https://doi.org/10.5194/hess-2021-509, 1 December 2021.

Gao, J., Zhou, B., and Zhai, P.: Changes in Persistent Precipitation in Northwest China and Related Large-Scale Circulation Features, J Meteorol Res, 37, 604–616, https://doi.org/10.1007/s13351-023-3030-9, 2023.

Goessling, H. and Reick, C. H.: What do moisture recycling estimates tell? Lessons from an extreme global land-cover change model experiment, Hydrology and Earth System Sciences, 8, 3217–3235, 2011.

Gu, J., Zhao, C., Xu, M., Ma, Y., Hu, Z., Jin, C., Guo, J., Geng, T., and Cai, W.: Fast Warming Over the Mongolian Plateau a Catalyst for Extreme Rainfall Over North China, Geophysical Research Letters, 52, e2024GL113737, https://doi.org/10.1029/2024GL113737, 2025.

Guo, Y., Yan, X., and Song, S.: Spatiotemporal variability of extreme precipitation in east of northwest China and associated large-scale circulation factors, Environ Sci Pollut Res, 31, 11749–11765, https://doi.org/10.1007/s11356-023-31790-0, 2024.

Hancock, J. T. and Khoshgoftaar, T. M.: CatBoost for big data: an interdisciplinary review, J Big Data, 7, 94, https://doi.org/10.1186/s40537-020-00369-8, 2020.

Hua, L., Zhong, L., and Ke, Z.: Precipitation recycling and soil–precipitation interaction across the arid and semi-arid regions of China, International Journal of Climatology, 36, 3708–3722, 2016.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Jiang, M., Li, Y., Hu, W., Yang, Y., Brasseur, G., and Zhao, X.: Model-based insights into aerosol perturbation on pristine continental convective precipitation, Atmos. Chem. Phys., 23, 4545–4557, https://doi.org/10.5194/acp-23-4545-2023, 2023.

Jiao, Y., Zhu, G., Lu, S., Ye, L., Qiu, D., Meng, G., Wang, Q., Li, R., Chen, L., Wang, Y., Si, D., and Li, W.: The Cooling Effect of Oasis Reservoir-Riparian Forest Systems in Arid Regions, Water Resources Research, 60, e2024WR038301, https://doi.org/10.1029/2024WR038301, 2024.

Kodinariya, T. M., Makwana, P. R., and others: Review on determining number of cluster in k-means clustering, International Journal, 1, 90–95, 2013.

Li, R., Feng, Q., and Cui, Y.: Hydrological dynamics in the China-Mongolia arid region: An integrated analysis of precipitation recycling and water vapor conversion, Heliyon, 10, e32839, https://doi.org/10.1016/j.heliyon.2024.e32839, 2024.

Likas, A., Vlassis, N., and J. Verbeek, J.: The global k-means clustering algorithm, Pattern Recognition, 36, 451–461, https://doi.org/10.1016/S0031-3203(02)00060-2, 2003.

Lugato, E., Alberti, G., Gioli, B., Kaplan, J. O., Peressotti, A., and Miglietta, F.: Ecological research and large scale land-atmosphere feedbacks: lesson from the Bouchet's complementary relationship, https://doi.org/10.5194/bgd-8-6077-2011, 29 June 2011.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, https://doi.org/10.48550/arXiv.1705.07874, 25 November 2017.

Ma, T., Cheng, W., and Yao, W.: Spatiotemporal evolution and multi-scenario simulation of the land-use cover change and habitat quality in arid and semi-arid areas: a case study of the urban agglomeration along the Yellow River in Ningxia, China, Front. Environ. Sci., 13, 1649302, https://doi.org/10.3389/fenvs.2025.1649302, 2025.

Manisha, M., Verma, K., Ramesh, N., Anirudha, T. P., Santrupt, R. M., and Rao, L.: Water, sanitation, and hygiene implications of large-scale recycling of treated municipal wastewater in semi-arid regions, Science of The Total Environment, 904, 166631, https://doi.org/10.1016/j.scitotenv.2023.166631, 2023.

Martiello Mastelini, S., Nakano, F. K., Vens, C., and de Leon Ferreira de Carvalho, A. C. P.: Online extra trees regressor, IEEE Transactions on Neural Networks and Learning Systems, 34, 6755–6767, https://doi.org/10.1109/TNNLS.2022.3212859, 2023.

Papacharalampous, G., Tyralis, H., Doulamis, A., and Doulamis, N.: Comparison of tree-based ensemble algorithms for merging satellite and earth-observed precipitation data at the daily time scale, https://doi.org/10.48550/ARXIV.2301.01214, 2023.

Qi, Z., Cui, C., Jiang, Y., Chen, Y., Ju, J., and Guo, N.: Changes in the spatial and temporal characteristics of China's arid region in the background of ENSO, Sci Rep, 12, 17826, https://doi.org/10.1038/s41598-022-21712-4, 2022.

Qu, Z., Yao, S., and Liu, D.: Response of drought to climate extremes in a semi-arid inland river basin in China, J. Arid Land, 16, 1505–1521, https://doi.org/10.1007/s40333-024-0064-5, 2024.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Ricetti, L., Hurtado, S. I., Zaninelli, P. G., and Agosta, E. A.: Determining the percentile threshold of daily extreme precipitation, methods evaluation, Stoch Environ Res Risk Assess, 39, 2887–2902, https://doi.org/10.1007/s00477-025-02998-y, 2025.

Su, H. and Dickinson, R. E.: On the Spatial Gradient of Soil Moisture–Precipitation Feedback Strength in the April 2011 Drought in the Southern Great Plains, Journal of Climate, 30, 829–848, https://doi.org/10.1175/JCLI-D-13-00185.1, 2017.

Taylor, C. M., Harris, P. P., and Parker, D. J.: Impact of soil moisture on the development of a Sahelian mesoscale convective system: a case-study from the AMMA Special Observing Period, Quart J Royal Meteoro Soc, 136, 456–470, https://doi.org/10.1002/qj.465, 2010.

Taylor, C. M., Gounou, A., Guichard, F., Harris, P. P., Ellis, R. J., Couvreux, F., and De Kauwe, M.: Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns, Nature Geosci, 4, 430–433, https://doi.org/10.1038/ngeo1173, 2011.

Tian, Q., Lu, J., and Chen, X.: An innovative method for measuring the hysteresis effects of soil moisture on meteorological variables at various time scales and climate conditions, Geo-spatial Information Science, 28, 671–684, https://doi.org/10.1080/10095020.2023.2280574, 2025.

Touseef, M., Chen, L., Yang, K., and Chen, Y.: Long-Term Rainfall Trends and Future Projections over Xijiang River Basin, China, Advances in Meteorology, 2020, 1–18, https://doi.org/10.1155/2020/6852148, 2020.

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

Vehtari, A., Gelman, A., and Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, Stat Comput, 27, 1413–1432, https://doi.org/10.1007/s11222-016-9696-4, 2017.

Xie, X., Liu, X., Shi, Z., Li, X., Xie, X., Sun, H., He, J., Che, H., Zhang, X., An, Z., Wang, D., and Liu, Y.: Sharp decline of dust events induces regional wetting over arid and semi-arid Northwest China in the NCAR Community atmosphere model, Environ. Res. Lett., 19, 014061, https://doi.org/10.1088/1748-9326/ad16a5, 2024.

Yurova, A., Kozlov, D., and Zhu, Y.: Drought occurrence in key regions of soil moisture-atmosphere interaction in temperate Eurasia, , https://doi.org/10.5194/egusphere-egu2020-8877, 2020.

Zeppetello, L. R. V., Zhang, L. N., Battisti, D. S., and Laguë, M. M.: How Much Does Land–Atmosphere Coupling Influence Summertime Temperature Variability in the Western United States?, Journal of Climate, 37, 3457–3478, https://doi.org/10.1175/JCLI-D-23-0716.1, 2024.

Zhang, Y., Wagner, N., Goergen, K., and Kollet, S.: Summer evapotranspiration-cloud feedbacks in land-atmosphere interactions over Europe, Climate Dynamics, 62, 10767–10783, https://doi.org/10.1007/s00382-024-07475-w, 2024.

Zhang, Z., Chen, X., Jiang, K., Pan, Z., Zhang, J., Zhang, Y., and Wang, J.: Dynamics and interactions of soil moisture and temperature in semi-arid regions of northern China, Journal of Arid Environments, 227, 105327, https://doi.org/10.1016/j.jaridenv.2025.105327, 2025.

https://doi.org/10.5194/egusphere-2025-5314
Preprint. Discussion started: 24 November 2025

Zhou, S., Williams, A. P., Lintner, B. R., Findell, K. L., Keenan, T. F., Zhang, Y., and Gentine, P.: Diminishing seasonality of subtropical water availability in a warmer world dominated by soil moisture–atmosphere feedbacks, Nat Commun, 13, 5756, https://doi.org/10.1038/s41467-022-33473-9, 2022.

Zhou, X., Letson, F., Crippa, P., and Pryor, S. C.: Urban Effect on Precipitation and Deep Convective Systems Over Dallas-Fort Worth, JGR Atmospheres, 129, e2023JD039972, https://doi.org/10.1029/2023JD039972, 2024.

Zhuo-li, C.: Characteristics and developing potential of cloud water resource in Ningxia with the CERES data, Arid Land Geography, 2015.

Zou, H., Auddy, A., Rad, K. R., and Maleki, A.: Theoretical analysis of leave-one-out cross validation for non-differentiable penalties under high-dimensional settings, in: International conference on artificial intelligence and statistics, Citation Key: Zou2024TheoreticalAO, 2024.