https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

# Response-specific subsurface thresholds and storage-based cross-catchment scaling of runoff thresholds

Zhen Cui[a,b], Fuqiang Tian[a,b*]

[a] Department of Hydraulic Engineering, State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China

[b] Key Laboratory of Hydrosphere Sciences of the Ministry of Water Resources, Tsinghua University, Beijing 100084, China

*Corresponding author: Fuqiang Tian (tianfq@tsinghua.edu.cn)

## Key points

- Estimated p50 transition coordinates increased with the extent of monitored subsurface responses.
- Deep-soil response timing was non-uniform; widespread activation was associated with, but did not ensure, delayed runoff.
- Storage-range normalization reduced cross-catchment threshold dispersion by 66.9% across four catchments.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.


## Abstract

Runoff generation thresholds depend on catchment wetness and subsurface conditions, yet it remains unclear whether subsurface responses over different spatial extents share common thresholds, how their temporal development relates to runoff generation, and whether thresholds are comparable across catchments. Here, we analyzed 272 rainfall events in the Xitaizi Experimental Watershed (XEW), China, using observations from five deep-soil profiles, six groundwater wells, and outlet streamflow. We quantified response-specific 50% probability (p50) transition coordinates along a combined antecedent soil-water storage and event-rainfall gradient ($ASI_0$+P), examined response timing and joint responses against rainfall–wetness-conditioned references, and evaluated storage-range normalization using published thresholds from four catchments. Among 208 events with complete soil-profile observations, the p50 coordinate increased by 48.2 mm from D1 (at least one of five profiles responding) to D5 (all five profiles responding; 142.7–190.9 mm). Among 34 events with complete six-well observations, it increased by 30.2 mm from G4 (at least four wells responding) to G6 (all six wells responding; 149.0–179.2 mm). Response expansion was temporally non-uniform: in 39 D5 events, the median fourth-to-fifth onset interval was 2.83 h, compared with 0.50 h for the cross-event median of within-event medians of the preceding three intervals. W13 was the last responding well in all 14 G6 events. Higher X80 was associated with a shorter D5 first-to-last response span (Spearman $\rho = -0.613$). Joint extensive soil-water and groundwater responses did not exceed the rainfall–antecedent-storage-conditioned expectation (11 observed versus 10.60 expected; $p = 0.520$). In contrast, all 18 verified delayed-peak events in the matched 139-event cohort occurred when at least four soil profiles responded, exceeding the conditioned expectation of 13.29 events ($p = 0.000052$). Although D4–D5 events had greater median stormflow depths, this association weakened under stricter rainfall–storage matching (61.9% of matched pairs favored $D \geq 4$; $p = 0.127$). The delayed-peak p50 coordinate was 21.3 mm higher than that for D5, indicating that widespread deep-soil activation was associated with, but did not ensure, delayed-peak occurrence. Across four catchments, normalization by the monitored 5th–95th percentile storage range reduced relative between-catchment median absolute deviation (MAD) by 66.9%, yielding dimensionless coordinates of 1.034–1.361. These findings distinguish response-specific subsurface transitions from the conditions associated with delayed runoff and demonstrate the potential of monitored-storage normalization for cross-catchment threshold comparison and hydrological model evaluation.


**Keywords:** Runoff generation thresholds; Subsurface response dynamics; Groundwater response; Deep soil moisture; Storage normalization

## 1. Introduction

Storm runoff often exhibits nonlinear responses to event rainfall and antecedent wetness as catchment storage states, contributing areas, and active flow pathways change (Tromp-van Meerveld and McDonnell, 2006a, 2006b; Detty and McGuire, 2010; Fu et al., 2013). However, the physical meaning of a hydrological

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

threshold depends on the response variable and the spatial support of the observations used to identify it. Local deep-soil wetting, activation across multiple hillslope profiles, groundwater-level rises, and delayed streamflow peaks represent distinct responses and need not occur under the same hydrological conditions. A threshold inferred from an outlet rainfall–runoff relationship therefore cannot be assumed to mark a common transition throughout the subsurface–streamflow system. Identifying response-specific transition conditions is essential for relating empirical thresholds to the processes underlying runoff generation.

Field studies have shown that local subsurface activation does not necessarily coincide with substantial runoff delivery. At Panola, localized transient saturation developed before pronounced subsurface outflow, which increased sharply only after saturated areas became connected to the trench (Tromp-van Meerveld and McDonnell, 2006a, 2006b). More generally, rainfall thresholds for hillslope outflow can emerge from the connectivity of small-scale flow pathways and spatial variations in storage capacity (Lehmann et al., 2007), while the fill-and-spill concept provides a broader framework for understanding such threshold-controlled runoff generation across spatial scales (McDonnell et al., 2021). Detty and McGuire (2010) linked nonlinear quickflow responses to groundwater dynamics and changes in effective runoff-contributing areas, while Fu et al. (2013) identified distinct rainfall–antecedent-moisture thresholds for hillslope interflow and catchment runoff in a fissured granitic catchment. Other studies have highlighted the roles of antecedent wetness and catchment characteristics in shaping runoff thresholds (Scaife et al., 2020; Nanda and Safeeq, 2023; Cui and Tian, 2026b). Together, these findings indicate that runoff thresholds are state-dependent and reflect spatially heterogeneous subsurface conditions. Nevertheless, quantitative evidence remains limited on how rainfall–wetness conditions associated with subsurface responses vary with the extent of monitored response, and how these transitions relate to event-scale timing and outlet stormflow characteristics.

Distributed soil-water and groundwater observations offer a way to resolve these relationships, but monitored response extent, response timing, and hydraulic connectivity must be distinguished. The number of responding profiles or wells indicates the spatial extent of monitored activity, but does not by itself establish connected flow pathways or contributions to streamflow (Bracken et al., 2013; Blume and van Meerveld, 2015; Jencso et al., 2009). In particular, a continuous shallow perched saturated layer does not necessarily imply continuous downslope connectivity to a stream (Klaus and Jackson, 2018). Soil-water and groundwater responses may also occur in different orders and over different timescales: groundwater at some hillslope locations has been observed to respond before substantial wetting was detected at depth, while groundwater–stream response lags vary with antecedent conditions (Salve et al., 2012; Beiter et al., 2020). Moreover, apparent co-occurrence may arise from shared rainfall forcing and antecedent wetness rather than additional hydraulic coupling. Comparing event-scale response timing and evaluating joint responses against rainfall–wetness-conditioned common-forcing expectations can therefore help distinguish observed co-variation from stronger mechanistic interpretations.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

The Xitaizi Experimental Watershed (XEW), a forested headwater catchment underlain by weathered and fractured granite, provides long-term observations of rainfall, soil moisture at five hillslope profiles, groundwater levels in six wells, and outlet streamflow. Previous studies at XEW identified the occurrence conditions and water-source characteristics of bimodal runoff and examined soil-water and groundwater dynamics associated with delayed stormflow (Cui et al., 2024; Cui and Tian, 2025). Studies of double-peak hydrographs in other forested headwaters have also distinguished rapid rainfall-linked runoff from delayed groundwater-mediated responses, with their relative contributions depending on antecedent wetness (Gao et al., 2025). These studies provide a process basis for examining how rainfall–antecedent-wetness conditions associated with deep-soil and groundwater responses vary across monitored response extents, how these responses evolve during individual events, and how they relate to delayed streamflow peaks. In particular, extensive deep-soil activation and delayed streamflow-peak occurrence are treated as distinct responses, rather than assuming that one necessarily implies the other.

At the cross-catchment scale, a related challenge is comparing event stormflow thresholds. Threshold coordinates expressed in millimetres may differ among catchments because monitored soil-water storage ranges, observation depths, and spatial monitoring support differ, even when the stormflow response is defined consistently. Previous inter-catchment studies have documented differences in threshold responses across catchments with contrasting environmental and hydrological characteristics (Ali et al., 2015; Ross et al., 2021). Cui and Tian (2026a) reported catchment-specific $ASI_0$+P threshold coordinates derived from event stormflow-depth relationships for XEW, P301, W18, and W27, providing a published basis for cross-catchment comparison. Previous studies have also shown that soil-water integration depth and monitoring location influence storage–runoff relationships and threshold interpretation (Cui and Tian, 2026a; Nanda and Safeeq, 2024). Expressing each threshold relative to the corresponding monitored soil-water storage interval between dry and wet reference states may account for part of this between-catchment variation. If this normalization reduces the spread of threshold coordinates, it may provide a useful dimensionless basis for comparative analysis and for evaluating storage-related thresholds in hydrological models. However, the normalized coordinate should be interpreted as an observation-based comparative index, not as total catchment storage capacity, a universal threshold, or a directly transferable model parameter. Its sensitivity to storage-reference bounds, soil-water monitoring depth, and spatial location therefore needs to be assessed.

Accordingly, this study has two objectives. First, using 272 rainfall events in XEW, we quantify and compare the $ASI_0$+P transition coordinates associated with deep-soil responses across five monitored profiles (D1–D5), groundwater responses across six wells (G4–G6), and delayed streamflow peaks. We further examine event-scale response timing, associations with stormflow, and whether joint soil-water–groundwater responses and delayed peaks exceed expectations under rainfall–wetness-conditioned common-forcing references. We hypothesize that broader monitored subsurface responses are associated with higher transition coordinates, but that their temporal development is not necessarily uniform. Second,

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

using published $ASI_0$+P thresholds from XEW, P301, W18, and W27, we evaluate whether normalization by the corresponding monitored S5–S95 soil-water storage range reduces between-catchment threshold dispersion. We test the hypothesis that this normalization improves threshold comparability while recognizing that the resulting dimensionless coordinate is an empirical index rather than a universal measure of catchment storage capacity.

## 2. Study sites and data

### 2.1. Xitaizi process-observation network

The Xitaizi Experimental Watershed (XEW) lies in the Yanshan Mountains of northern China, approximately 70 km northeast of Beijing (40°32′ N, 116°37′ E). The drainage area adopted for event stormflow-depth conversion was 4.22 km², and elevations range approximately from 676 to 1,201 m. Forested hillslopes with relatively shallow soils overlie variably weathered and fractured granite (Fig. 1).

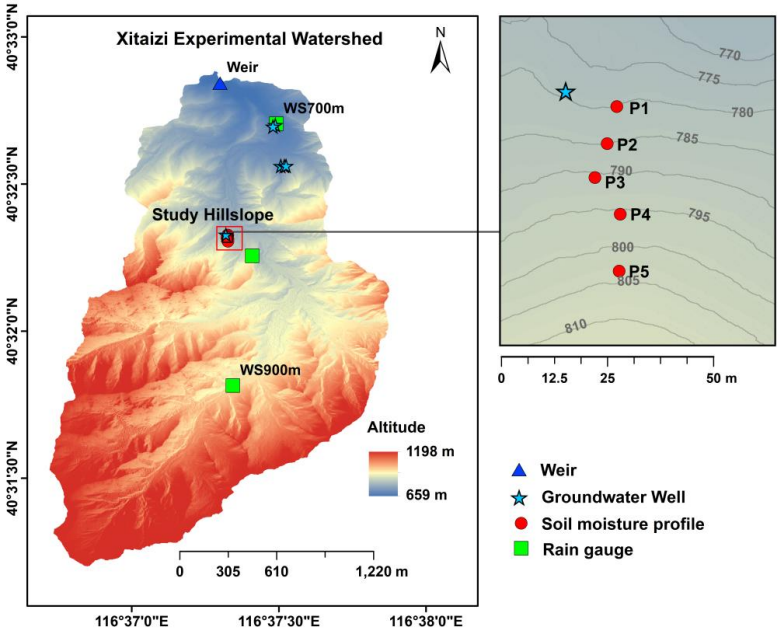


**Figure 1.** Location of the Xitaizi Experimental Watershed (XEW) and spatial distribution of hydrological monitoring sites. The inset shows the five soil-water profiles (P1–P5) on the study hillslope, while six groundwater wells are distributed across three hillslopes. The locations of the outlet weir and rain gauges are also shown.

Rainfall and soil-water content were monitored at 10-min intervals. The process analysis used five soil-water profiles (P1–P5) on Hillslope 1, with sensors at 10-cm intervals between 10 and 80 cm. Six groundwater wells represented three hillslopes: W13 on Hillslope 1; W21–W23 on Hillslope 2; and W31–W32 on Hillslope 3. Hourly groundwater observations recorded water-table depth below ground surface, so a rise corresponds to a decrease in recorded depth. Well characteristics and calibrated response criteria

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

are detailed in Table S1. Event stormflow volumes and delayed-peak labels were taken from the existing discharge-processing archive.

The reviewed rainfall-event catalogue covered 2014 to 21 August 2026, rather than 13 complete calendar years. Other XEW profiles contributed to previously published $ASI_0$+P threshold coordinates associated with event stormflow-depth relationships. Raw-data screening retained endpoint-specific missingness and censoring.

### 2.2. Published cross-catchment thresholds and storage records

The cross-catchment comparison used XEW and three experimental catchments from the earlier depth-specific runoff-threshold study (Cui and Tian, 2026a): P301 (99 ha; Kings River Experimental Watersheds, California), W18 (12.3 ha; Coweeta Hydrologic Laboratory, North Carolina), and W27 (39.8 ha; Coweeta). Previously selected soil-water integration depths were retained: 20 cm for P301, 30 cm for W27, and 60 cm for W18 and XEW. Published site-specific thresholds and daily soil-water storage records were used for the normalization analysis; no new runoff-threshold models were fitted for these catchments.

For the primary normalization, threshold averaging and the monitored-storage reference used the same five XEW Hillslope-1 positions (PO1-1 to PO1-5). The broader published eight-position XEW mean was retained only for a spatial-support sensitivity comparison. The other catchments retained the selected 10 P301, four W27, and six W18 threshold locations (Table S2).

**Table 1.** Study catchments, observation records, and selected soil-moisture integration depths. MID, soil-water integration depth; SWC, volumetric soil-water content; Q, event stormflow depth.

| Catchment | Drainage area | Record represented | Monitoring / analysis resolution | Selected MID | Role in this study |
|---|---|---|---|---|---|
| XEW | 4.22 km² | 2014–Aug 2026 | Rain / SWC: 10 min; groundwater: hourly; Q: event-integrated* | 80 cm for process analysis; 60 cm for normalization | Within-catchment responses and four-catchment comparison |
| P301 | 99 ha | Published 2009–2015 | Daily comparison; hourly soil-water source | 20 cm | Published-threshold comparison |
| W27 | 39.8 ha | Published 1999–2019 | Daily comparison | 30 cm | Published-threshold comparison |
| W18 | 12.3 ha | Published 1999–2019 | Daily comparison | 60 cm | Published-threshold comparison |

## 3. Methods

### 3.1. Event catalogue, observation support, and streamflow characteristics

The manually reviewed rainfall-event catalogue was the sole source of inferential XEW events. Of 299 non-empty candidate records, 298 passed basic time and rainfall reconstruction checks. Events with reconstructed rainfall of at least 5 mm were eligible for inclusion. Adjacent events with rainfall below 10 mm were excluded only when their occurrence prevented the predefined stable antecedent or post-event observation periods. Events with rainfall of at least 10 mm were retained rather than automatically excluded

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

because of their proximity to other events. When subsequent rainfall interrupted an event-specific response window, the affected endpoint was censored at the subsequent event rather than treating the incomplete follow-up as a nonresponse. This screening procedure yielded 272 uniquely identified events, which formed the common event catalogue for all within-XEW analyses.

Each response analysis retained its own eligible observation set. Deep-soil response extent required complete observations from all five profiles. Groundwater outcomes were considered observable only when all six designated wells had a corrected status of 0 (no newly detected response) or 1 (new response) under the fixed 72 h follow-up definition, with the window truncated at the next rainfall step meeting the 0.2 mm per time-step criterion. Pre-event-active, right-censored, indeterminate, missing, or otherwise invalid well outcomes were not recoded as nonresponses. Analyses against X80 additionally required valid antecedent storage and event rainfall. Same-cohort contrasts and timing analyses used only events satisfying the relevant shared-support or full-response criteria.

Event stormflow depth (Q, mm) was obtained from the existing discharge-processing archive, in which baseflow was separated using the HYSEP fixed-interval method with manual hydrograph adjustments. Event stormflow depths were calculated by dividing integrated event stormflow volumes by the 4.22 km² contributing area. The present analysis did not repeat the upstream baseflow separation or discharge integration. Q represents the magnitude of the event-associated streamflow response rather than a binary indicator of first runoff generation. The available joint Q/L sample contains positive stormflow depths and therefore does not determine the fraction of all rainfall events with no runoff.

Delayed-peak occurrence (L) was defined from researcher-reviewed hydrograph classifications. The original post-rainfall detection window extended from 7 to 168 h, and peaks confounded by subsequent rainfall were excluded. All 18 positive L events in the 139-event D–Q–L cohort were individually checked against their hydrographs. An L = 0 label indicates no annotated delayed peak under the operational classification, rather than an absence of any delayed discharge contribution.

### 3.2. Deep-soil and groundwater response definitions

Deep-soil and groundwater responses followed the reviewed event-processing criteria (Table 2). For each soil-water sensor, the baseline was the median volumetric water content during the preceding 6 h. The detection threshold was the maximum of 0.005 $m^3\ m^{-3}$, three times the sensor-specific dry-period noise, and the 99th percentile of positive changes in dry reference periods. A response required the threshold to be exceeded for three consecutive 10-min observations. A deep-soil profile was classified as responding when at least two observable sensors among the 60-, 70-, and 80-cm depths met the response criterion within the event-specific window. The profile onset was defined as the time when the second eligible sensor, ranked by response onset time, met the criterion. Thus, when all three sensors were observable and responded, the profile onset corresponded to the second sensor response; when only two sensors were observable, it corresponded to the later of their two response times. The soil-response window extended to at most 48 h after rainfall cessation and was censored by the next rainfall event.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

The number of responding profiles, $K_D$, ranged from zero to five; $D_k$ denotes the occurrence of at least k responding profiles. These nested criteria describe monitored response extent, not the activation of fixed numbered positions. Conditional two-profile timing (S2h) was recorded when the second responding profile began to respond within 2 h of the first among events with at least two responding profiles and complete five-profile observations. Three- and four-hour windows were examined as sensitivity definitions.

For groundwater, the pre-event baseline was the median water-table depth during the preceding 12 h. A new response required a sustained water-table rise beyond a well-specific calibrated detection threshold for at least two consecutive hourly observations. Calibration of the thresholds in the upstream workflow accounted for minimum detectable rises and dry-period background fluctuations. The adopted thresholds for W13, W21, W22, W23, W31, and W32 were 0.050, 0.020, 0.010, 0.017, 0.060, and 0.021 m, respectively (Table S1). The primary follow-up window extended to 72 h after rainfall cessation and was truncated at the next rainfall step meeting the 0.2-mm-per-time-step criterion. Only corrected well statuses of 0 and 1 were treated as observable; other statuses, including pre-event activation and right censoring, remained unavailable rather than being treated as nonresponses. $K_G$ denotes the number of wells with newly detected responses (zero to six), and $G_k$ indicates $K_G \geq k$ on complete six-well support. The main analyses considered G4–G6; wells already active at event onset were not recoded as new responses.

Response counts refer to instrumented locations: five soil profiles on one hillslope and six wells on three hillslopes. They are not area-weighted fractions of the catchment or direct measures of hydraulic connectivity.

**Table 2.** Operational response criteria and analysis support. Q and L are distinct event-level indicators; 18 events with delayed runoff were confirmed in the 139-event shared D–Q–L sample. Not meeting a response criterion is distinguishable from missing observation support.

| Indicator | Operational criterion | Eligible events, n | Reported estimand |
|---|---|---|---|
| D1–D5 | At least 1–5 of five complete deep-soil profiles meet the two-node rise criterion. | 208 | Occurrence centers; response extent |
| G4–G6 | At least 4–6 of six fully observable wells show new rises. | 34 with X80 | Occurrence centers; response extent |
| S2h | Second soil profile responds within 2 h of first, conditional on ≥2 responding profiles. | 77 | Conditional response synchrony |
| Q | Previously processed event stormflow depth (mm). | 161 with X80 and L; 139 with D | Stormflow-magnitude comparisons |
| L | Researcher-reviewed delayed runoff in the specified post-rainfall window. | 161 with X80; 139 with D | Occurrence centers; D–L co-occurrence |
| First–last span | Time from earliest to latest qualifying onset in full-response events. | 39 D5; 14 G6 | Within-event timing |

### 3.3. Antecedent-storage and rainfall coordinates

Antecedent soil-water storage ($ASI_0$) was derived from the original 10-min volumetric soil-water observations by integrating soil-water storage over the monitored soil column to a specified depth d. For each event, profile-level storage estimates from available monitoring profiles were used to obtain the event-

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

level antecedent-storage coordinate. Events with no valid profile-level storage estimates were treated as unavailable for analyses requiring $ASI_0$. The combined antecedent-storage and event-rainfall coordinate was defined as:

$$X_e^{(d)} = ASI_{0,e}^{(d)} + P_e \quad (1)$$

Here, e denotes the rainfall-event index, and d denotes the soil-water integration depth measured downward from the soil surface. $ASI_{0,e}^{(d)}$ represents the antecedent soil-water storage for event *e* integrated to depth *d*, and $P_e$ is event rainfall, both in millimetres. The within-XEW response analyses used d = 80 cm (X80). The cross-catchment comparison used the previously published XEW threshold at 60 cm (Section 3.7), without substituting one depth-specific threshold for the other (Cui and Tian, 2026a).

### 3.4. Response-specific transition coordinates and matched contrasts

The ordinal count $K_D$ was related to X80 using a proportional-odds cumulative-logit model (McCullagh, 1980). For the k-th nested deep-soil response criterion,

$$\text{logit}\left[\Pr\left(K_{D,e^*} \geq k \cdot | \, X_{80,e}\right)\right] = \alpha_k + \beta \cdot X_{80,e} \quad (2)$$

where ordered cutpoints $\alpha_k$ and a common positive slope β generate noncrossing occurrence curves. The p-th occurrence coordinate is $T_{Dk,p}$ = [logit(p) - $\alpha_k$]/ β. We report the p50 coordinates (p = 0.5) and fitted p10–p90 transition intervals. A transition interval describes the change in fitted occurrence probability, not uncertainty around its center. The proportional-odds model constrains both the ordering and the shared transition width; independent binary-logistic D models were therefore used to examine sensitivity to the common-slope assumption.

Separate binary-logistic models estimated G4–G6, conditional S2h, and delayed-peak (L) occurrence centers. Whole calendar years were resampled with replacement, preserving the events within each year. Main p50 estimates and contrasts used 2,000 year-block bootstrap draws; independent-D-model sensitivity used 600 draws, while supplementary G1–G3 and alternative synchrony-window analyses used 300 draws. Percentile intervals used valid draws, with the numbers of successful fits reported in the tables.

Primary coordinate contrasts comprised D5 − D1 and G6 − G4, together with the final-versus-preceding D adjacent-gap contrast. The L − D5 contrast was estimated from separate binary-occurrence fits on the same 139 events and the same paired year-block resamples. The continuous stormflow-depth variable Q was not assigned an additional fitted threshold in the present study.

### 3.5. Response-onset sequence and fixed-extent timing

The first, second, and subsequent responding profiles or wells were ranked by their detected onset times within each event, independently of their fixed geographical identities. Timing was characterized by cumulative elapsed times from the earliest response, intervals between consecutive onsets, and the first-to-last response span. The intervals between consecutive soil-profile and groundwater responses were calculated separately as:

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$\Delta t_{D,k} = t_{D,(k)} - t_{D,(k-1)}, \quad k = 2, \ldots, m_D \tag{3}$$

$$\Delta t_{G,k} = t_{G,(k)} - t_{G,(k-1)}, \quad k = 2, \ldots, m_G \tag{4}$$

Here $t_{D,(k)}$ and $t_{G,(k)}$ denote the onset times of the k-th responding soil profile and groundwater well, respectively, ranked within an individual event. The cumulative elapsed time for the *k*-th response was calculated relative to the first responding location as $t_{D,(k)} - t_{D,(1)}$ for soil profiles and $t_{G,(k)} - t_{G,(1)}$ for groundwater wells. The first-to-last response span corresponds to the cumulative elapsed time of the final responding location. To compare the same number of responding locations, the soil analysis used 39 D5 events and the groundwater analysis used 14 G6 events. For the D5 cohort, the final adjacent interval ($\Delta t_{D,5}$ between the fourth and fifth profile onsets) was compared with the within-event median of the preceding three intervals ($\Delta t_{D,2}$, $\Delta t_{D,3}$, and$\Delta t_{D,4}$). These event-specific medians were then summarized across the 39 events. This comparison distinguishes the final response interval from the typical duration of preceding intervals within each event.

The 14-event G6 timing cohort comprised events in 2014, 2015, 2021, 2022, and 2025; in each event, all six wells had corrected new-response status and a valid onset time. Across-event medians, interquartile ranges, and year-block bootstrap intervals were calculated for timing summaries. Spearman rank correlations assessed associations between the first-to-last response span and continuous X80.

The identity of the last responding location was tabulated in each full-response cohort. The effect of individual monitoring positions on the completion span was evaluated by excluding each position in turn from the same events. Soil–groundwater signed onset differences were calculated as groundwater onset minus the earliest deep-soil profile onset; negative values indicate that a groundwater well responded first. This latter comparison included nine G6 events with an identified deep-soil onset, eight of which also met D5.

### 3.6. Joint responses, shared-forcing references, and stormflow association

Extensive monitored responses were operationally defined as D ≥ 4 and G ≥ 4. Their co-occurrence was examined in 26 events with complete support for both domains. Null-A fixed the total numbers of D ≥ 4 and G ≥ 4 events while randomizing their event pairing over the full cohort. Null-B fixed the same binary margins within strata of event rainfall (P) and antecedent storage ($ASI_{0,80}$). Because the D–G cohort was small, two quantile-based classes were used for each forcing variable (2 × 2 strata). For the 139-event D–L cohort, the primary conditional reference used 3 × 3 forcing strata. Conditioning additionally on year and alternative forcing definitions provided sensitivity checks.

Under each reference, the null overlap within a stratum followed a hypergeometric distribution with the two binary margins fixed. Stratum-level distributions were combined to obtain the exact total overlap distribution. One-sided p values were calculated as the reference probabilities of observing at least the actual number of co-occurrences. Strata in which a response margin was zero or complete contributed little

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

or no randomization variation. These comparisons assessed observed association relative to the specified shared-forcing reference; they did not estimate flow-path contributions.

Event stormflow depth (Q) was compared across D0, D1–D3, and D4–D5 events, with particular emphasis on extensive ($D \geq 4$) versus less-extensive ($D < 4$) responses. Cross-group event pairs were formed within rainfall–antecedent-storage strata. The comparison statistic was the fraction of pairs in which Q was higher for the extensive-response event, with ties assigned a value of 0.5. Because individual events could contribute to multiple pairs, the pairs were not treated as independent sampling units. Statistical significance was assessed using 5,000 event-level conditional permutations. Sensitivity analyses imposed maximum between-event differences in rainfall and antecedent storage of 20/30, 15/20, or 10/20 mm, or additionally required events to occur in the same year (Table S6).

### 3.7. Cross-catchment monitored-storage normalization and sensitivities

Previously published $ASI_0$+P threshold associated with event stormflow depth coordinates at the selected depth were used for each of the four catchments without refitting (Cui and Tian, 2026a). The representative threshold T was the arithmetic mean of selected site-specific published threshold estimates. For each catchment, daily monitored soil-water storage was averaged across eligible positions, and the 5th and 95th percentiles ($S_5$ and $S_{95}$) were calculated from this daily mean-storage series. The monitored reference range and normalized coordinate were

$$C_{ref} = S_{95} - S_5 \quad (5)$$

$$F = \frac{T - S_5}{C_{ref}} \quad (6)$$

The dimensionless F expresses the published rainfall-plus-antecedent-storage coordinate relative to the monitored dry-to-wet range; $C_{ref}$ is not a measurement of total subsurface storage capacity. For the primary XEW calculation, T used the same five Hillslope-1 locations represented by the daily storage reference, whereas the original eight-location mean was evaluated only for spatial-support sensitivity.

Primary daily-storage coverage required at least eight of ten P301 positions, all four W27 positions, five of six W18 positions, and two of five XEW positions. Sensitivity tests substituted 10th–90th reference percentiles, changed daily-coverage requirements, and examined selected position-support effects. In W18, the threshold-mean sensitivity to omitting the high-threshold PO2_bot profile was calculated while retaining the original six-position storage reference; the full six-position mean remained the primary estimate.

Cross-catchment dispersion was summarized using relative median absolute deviation, defined as the MAD from the four-catchment median divided by that median. Catchments, rather than individual profiles or years, formed the units of comparison. Changes in dispersion were reported descriptively for the four catchments; no confidence intervals were assigned to the dispersion reduction. No retrospective leave-one-catchment threshold reconstruction was used as a validation analysis.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

# 4. Results

## 4.1. Event support and response occurrence

The XEW catalogue contained 272 rainfall events, including 255 with valid X80 (Fig. 2a; Table 2). Complete observations from all five deep-soil profiles were available for 208 events. All six groundwater wells were observable in 36 events, of which 34 also had valid X80. Q and L were jointly available for 177 events, declining to 161 when X80 was required. The shared cohorts contained 139 events with D, Q, and L; 26 with D and G; and 24 with D, G, Q, and L. Conditional S2h could be evaluated for 77 events (Table S3).

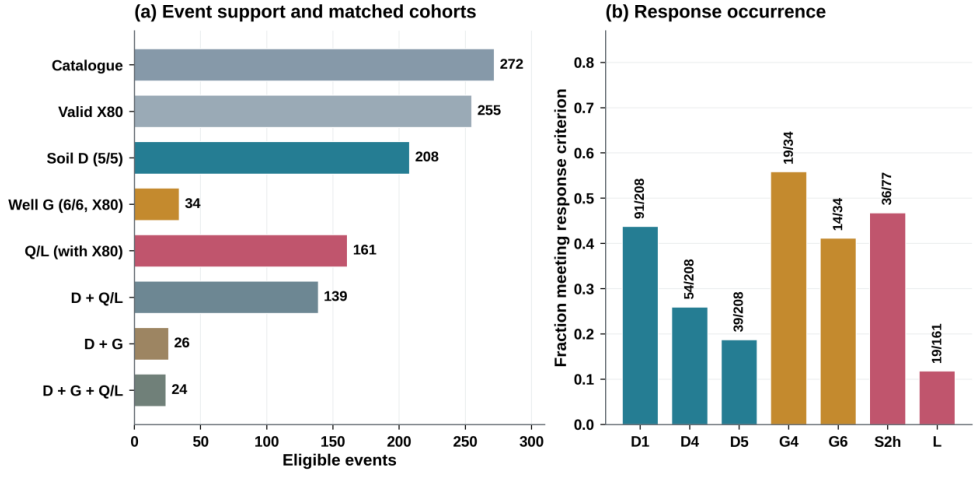


**Figure 2.** Eligible event supports and response occurrence. (a) Counts for the full event catalogue and selected criterion-specific and matched supports. (b) Fractions meeting selected nested soil (D1, D4, D5), new-groundwater (G4, G6), conditional two-profile timing (S2h), and delayed-peak (L) criteria. Each fraction uses the denominator displayed above its bar; supports differ across endpoints.

Of the 208 soil-complete events, 117 showed no detected deep-profile response; 91, 77, 61, 54, and 39 events met the nested D1–D5 criteria, respectively (Fig. 2b). Of the 34 well-complete events with valid X80, 11 showed no new groundwater response, 19 met G4, and 14 met G6. The fixed G6 timing cohort therefore consisted of these 14 fully responding events, not all events with six well rows in a provisional onset export; the 14 events spanned five years (Table S3). S2h occurred in 36 of the 77 eligible events. Delayed runoff events were annotated in 20 of 177 Q/L events, 19 of the 161 with X80, and 18 of the 139 with D support. All 18 L-positive cases in the last cohort were confirmed by hydrograph inspection.

The observation sets also differed in seasonal and hydrometeorological support. The 34 complete six-well events with X80 had median rainfall and $ASI_{0,80}$ of 31.95 and 133.2 mm, compared with 17.1 and 104.2 mm among the 208 soil-complete events. Forty-six soil-complete events in 2018–2019 had no eligible Q/L data. The 2026 record ended in August. Occurrence fractions in Fig. 2b therefore use their own eligible-event denominators.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

### 4.2. Response-specific transition coordinates and contrasts

In the 208 soil-complete events, ordered-model p50 coordinates increased from 142.7 mm for D1 to 152.2, 165.2, 172.4, and 190.9 mm for D2–D5, respectively (Fig. 3a; Table S4). The D5 − D1 difference was 48.2 mm (95% year-block interval, 38.8–55.2 mm; Table 3). The four successive adjacent differences were 9.5, 12.9, 7.2, and 18.5 mm; the final gap exceeded the preceding gap by 11.3 mm (2.2–18.2 mm; Fig. 3d). Independent binary-logistic fits yielded a D5 − D1 difference of 43.1 mm (35.3–49.3 mm) and a final-versus-preceding gap contrast of 10.7 mm (5.4–17.5 mm; Fig. S1). The ordered model imposed the same fitted p10–p90 width of 84.2 mm for every D criterion (Table S4).

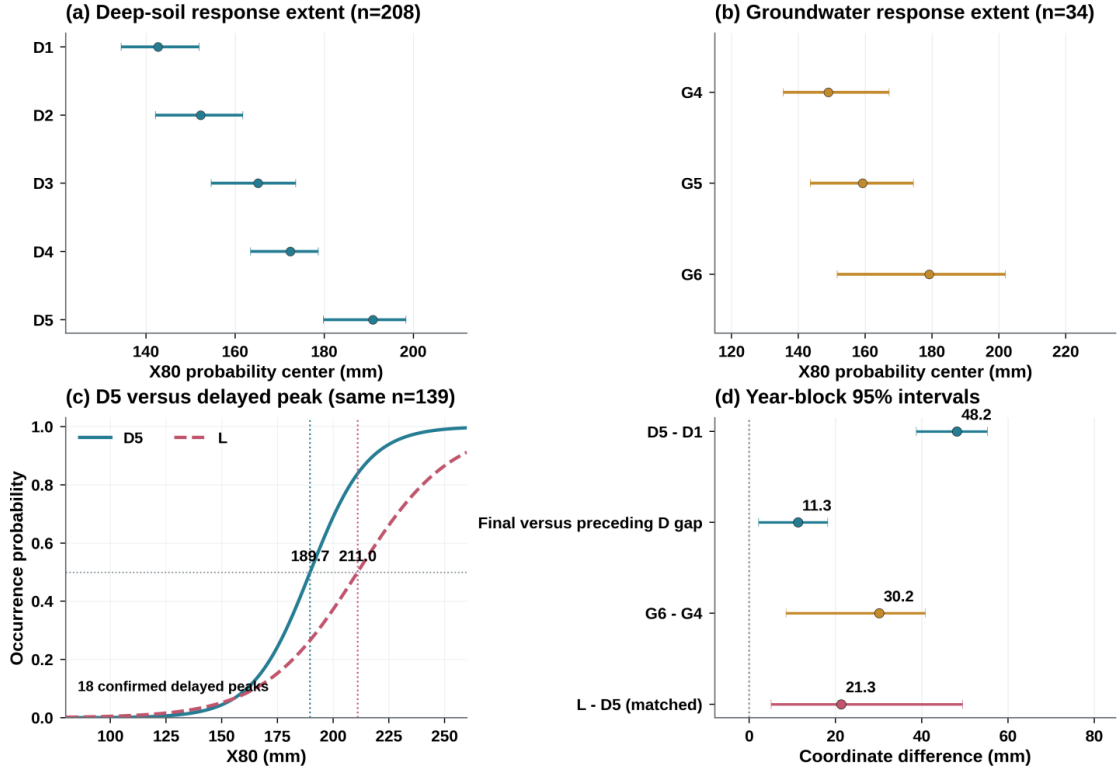


**Figure 3.** Occurrence coordinates for monitored response extent and delayed runoff. (a) Ordered-model p50 values for D1–D5 among 208 five-profile-complete events. (b) Binary-logistic p50 values for G4–G6 among 34 six-well-complete events with X80; error bars in (a,b) are 95% year-block intervals. (c) D5 and researcher-confirmed L occurrence curves refitted on the same 139 events. (d) Within-cohort coordinate contrasts and 95% year-block intervals. The differences are X80-coordinate separations, not response-time lags.

Among the 34 well-complete events with X80, p50 coordinates were 149.0, 159.3, and 179.2 mm for G4, G5, and G6, respectively (Fig. 3b; Table S4); 14 of the 34 events met the G6 criterion. The G6 − G4 difference was 30.2 mm (95% year-block interval, 8.5–40.9 mm), with valid estimates in 1,741 of 2,000 paired resamples (Fig. 3d; Table 3). Its leave-one-year-out estimates ranged from 16.3 to 35.5 mm. In the

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

separate 77-event conditional sample, S2h had a p50 of 185.0 mm (165.3–203.4 mm); using 3- and 4-h definitions gave centers of 154.1 and 147.9 mm, respectively (Table S4).

On the same 139 D–L events, separate binary models yielded p50 coordinates of 189.7 mm for D5 and 211.0 mm for delayed-peak occurrence (Fig. 3c). The paired difference L − D5 was 21.3 mm (95% year-block interval, 5.1–49.5 mm; Fig. 3d; Table 3), with valid estimates in all 2,000 paired resamples. In the broader 161-event L sample, the L center was 211.9 mm (188.7–242.5 mm; Table S4). All coordinate differences refer to positions on the X80 axis rather than elapsed response times.

**Table 3.** Selected contrasts among response-specific X80 occurrence coordinates. D contrasts were evaluated in the 208-event ordered-response cohort; G6 − G4 used the same 34 events; L − D5 used separate binary-occurrence models on exactly the same 139 events. Year-block percentile intervals were obtained from paired model refits.

| Contrast | Matched n | Difference (mm) | 95% year-block interval (mm) | Valid / attempted draws |
|---|---|---|---|---|
| D5 − D1 | 208 | 48.2 | 38.8–55.2 | 2,000 / 2,000 |
| (D5 − D4) − (D4 − D3) | 208 | 11.3 | 2.2–18.2 | 2,000 / 2,000 |
| G6 − G4 | 34 | 30.2 | 8.5–40.9 | 1,741 / 2,000 |
| L − D5 (matched binary models) | 139 | 21.3 | 5.1–49.5 | 2,000 / 2,000 |

### 4.3. Response-onset sequence and timing

Across eligible events, the counts of responding soil profiles and newly responding groundwater wells increased with X80 (Spearman $\rho = 0.712$, $n = 208$, and $\rho = 0.746$, $n = 34$, respectively). Within the fixed set of 39 D5 events, median cumulative elapsed times from the first soil-profile onset to the second through fifth onsets were 1.50, 2.17, 2.83, and 6.83 h, respectively (Fig. 4a; Table S5). The corresponding median intervals between consecutive onsets were 1.50, 0.67, 0.50, and 2.83 h (Table S7). The median first-to-last response span was 6.83 h (event interquartile range, 4.42–11.33 h; 95% year-block interval for the median, 4.50–10.33 h). Across events, the median of the event-specific medians of the first three adjacent intervals was 0.50 h, compared with a median final interval of 2.83 h. The final interval was strictly the longest in 29 of 39 events (74.4%); including two additional events in which it tied for the longest, this occurred in 31 events (79.5%).

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

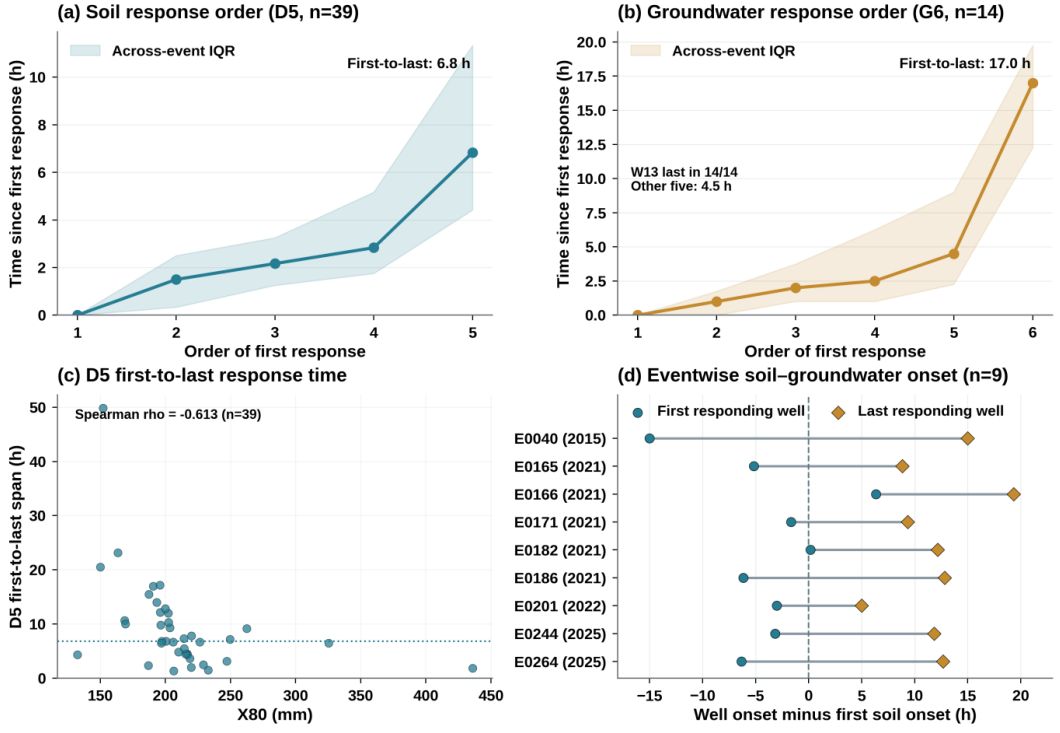


**Figure 4.** Eventwise onset order and timing within fixed-response cohorts. (a) Cumulative times for five deep-soil profiles in 39 D5 events. (b) Cumulative times for six wells in the strict 14-event G6 cohort, in which all six wells had corrected new-response status and complete onset times. Shading shows across-event interquartile ranges. (c) The D5 first-to-last span versus X80. (d) Earliest and latest groundwater onset relative to the first deep-soil onset in nine G6 events. The x-axis sign in (d) denotes observed onset ordering, not water-transit time.

The D5 first-to-last span was negatively associated with X80 (Spearman $\rho = -0.613$, $n = 39$; 95% year-block interval, −0.871 to −0.236; Fig. 4c). Leave-one-year-out correlations ranged from −0.695 to −0.538. Within the fixed 14-event G6 cohort, median cumulative times from the first responding well to the second through sixth wells were 1.0, 2.0, 2.5, 4.5, and 17.0 h, respectively (Fig. 4b; Table S5). The median first-to-last response span was 17.0 h (event IQR, 12.25–19.75 h; 95% year-block interval for the median, 12.0–30.0 h). The final fifth-to-sixth-well interval had a median of 10.5 h (event IQR, 7.25–13.50 h; 95% year-block interval for the median, 8.0–25.0 h), compared with median adjacent intervals of 1.0, 1.0, 0.0, and 1.0 h for the preceding steps. This final interval was strictly the longest in 12 of 14 events (85.7%; Table S7). The G6 span–X80 correlation was −0.342 (95% year-block interval, −0.861 to 0.335).

The last responding location was P2 in 25 of 39 D5 events and W13 in all 14 G6 events. After omission of P2, the median span among the remaining four soil profiles was 3.33 h, compared with 6.83 h for all five. Omitting W13 reduced the median span among the remaining five wells to 4.50 h, compared with 17.0 h for all six (Fig. S2). These omission calculations used the same events within each domain.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

Nine G6 events also had a detected first deep-soil response. The earliest well responded before the earliest deep-soil profile in seven of the nine events, whereas the latest well responded after the first deep-soil onset in all nine (Fig. 4d). Median signed groundwater-minus-soil onset differences were −3.17 h for the earliest well and +12.17 h for the latest well; eight of these nine events also satisfied D5.

### 4.4. Joint responses, shared forcing, and stormflow characteristics

Among the 26 events with complete soil-water and groundwater observations, 11 showed joint extensive responses (D ≥ 4 and G ≥ 4). This exceeded the Null-A expectation of 6.46 events (one-sided p = 0.000462). However, under the rainfall–antecedent-storage-conditioned Null-B reference, the expected count was 10.60 events, and the observed count did not differ significantly from this expectation (p = 0.520; Fig. 5b). Notably, all eight events in the high-rainfall/high-$ASI_{0,80}$ stratum exhibited joint extensive responses, limiting the variation available for distinguishing additional associations within this stratum.

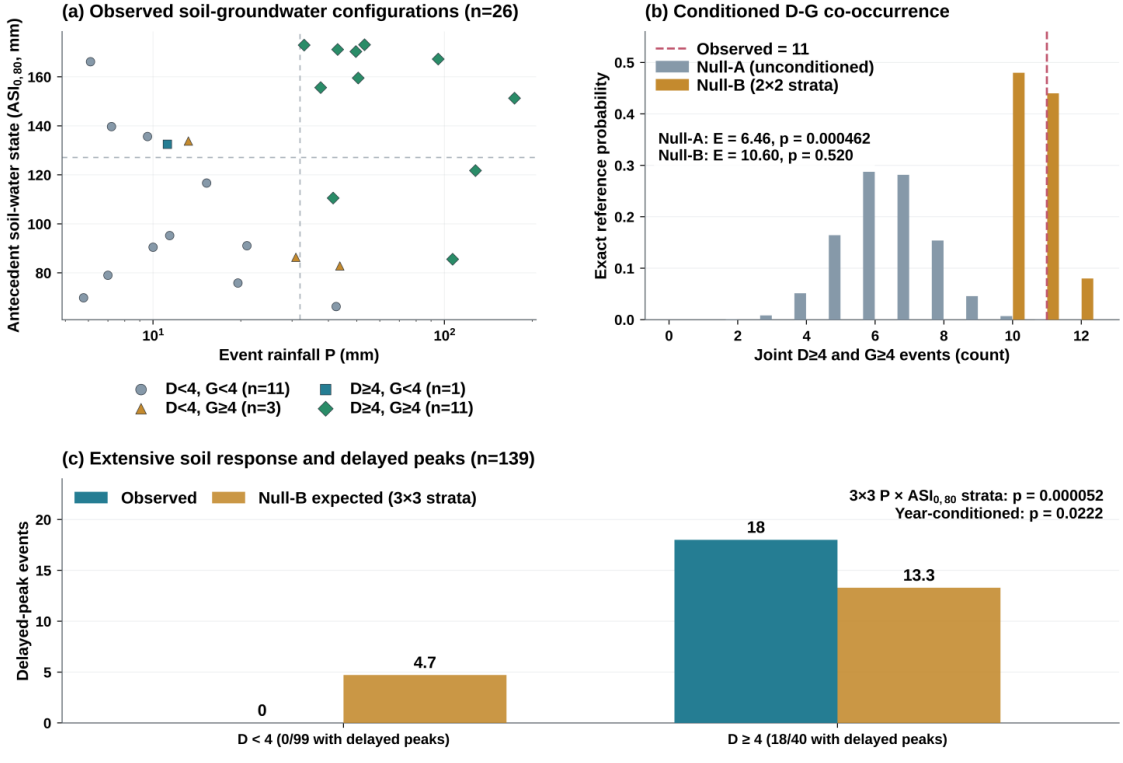


**Figure 5.** Joint soil–groundwater responses and delayed-peak occurrence under shared forcing. (a) Four observed D/G configurations across event rainfall P and antecedent $ASI_{0,80}$ for 26 complete D–G events. Dashed lines indicate the median boundaries of the 2 × 2 conditional strata (P = 31.95 mm; $ASI_{0,80}$ = 127.13 mm). (b) Exact unconditioned Null-A and 2 × 2 conditioned Null-B co-occurrence distributions, compared with 11 observed events (Null-A expected 6.46, p = 0.000462; Null-B expected 10.60, p = 0.520). (c) Delayed-peak counts among 139 D–Q–L events relative to a separate 3 × 3 rainfall–storage conditional reference (18 observed; 13.29 expected; p = 0.000052; year-conditioned p = 0.0222). Panels (a,b) and (c) use different event sets and different conditional strata.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

In the separate 139-event D–Q–L cohort, median stormflow depths were 0.040 mm for D0 (n = 80), 0.051 mm for D1–D3 (n = 19), and 0.544 mm for D4–D5 (n = 40; Fig. S3). Median rainfall across these groups was 12.3, 15.3, and 40.5 mm, and median antecedent $ASI_{0,80}$ was 88.1, 116.7, and 157.5 mm, respectively. Within coarse 3 × 3 forcing strata, Q was greater for the extensive-D member in 76.0% of 179 cross-group pairs (76 distinct events). With pairwise restrictions of $|\Delta P| \leq 10$ mm and $|\Delta ASI_{0,80}| \leq 20$ mm, this fraction was 61.9% among 63 overlapping pairs formed from 44 distinct events (event-level permutation p = 0.127). Same-year, same-stratum comparisons gave 68.6% among 35 pairs from 36 events (p = 0.089; Fig. S3; Table S6).

All 18 researcher-confirmed delayed runoff events in the 139-event cohort occurred among the 40 events satisfying $D \geq 4$; no delayed runoff was annotated among the 99 events with $D < 4$ (Fig. 5c). The other 22 $D \geq 4$ events had no annotated delayed runoff. Under the 3 × 3 rainfall-by-$ASI_{0,80}$ conditional reference, the expected $D \geq 4$ and L co-occurrence count was 13.29, compared with 18 observed (exact one-sided p = 0.000052). Additional year conditioning gave an expected count of 15.53 and p = 0.0222 (Fig. S4). Twenty-four events supported simultaneous D, G, Q, and L observations; in this subset, one event met $D \geq 4$ but $G < 4$ (Table S3).

### 4.5. Cross-catchment monitored-storage normalization

Previously published site-specific thresholds gave representative T values of 57.25 mm for P301 (20 cm), 99.08 mm for W27 (30 cm), 131.18 mm for XEW (60 cm), and 203.38 mm for W18 (60 cm; Fig. 6a; Table 4). The corresponding means included ten, four, five, and six threshold locations. For XEW, the five-profile mean used the same Hillslope-1 positions as the monitored-storage reference; the broader eight-position mean of 135.76 mm was retained only for sensitivity analysis.

The monitored $S_{95} - S_5$ storage-reference ranges were 36.84 mm in P301, 49.17 mm in W27, 92.61 mm in XEW, and 86.07 mm in W18 (Table 4). The corresponding normalized coordinates F were 1.034, 1.283, 1.059, and 1.361, respectively (Fig. 6b). The raw thresholds ranged from 57.25 to 203.38 mm; the four normalized coordinates ranged from 1.034 to 1.361. Each representative T was larger than the corresponding empirical $S_{95}$ reference.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

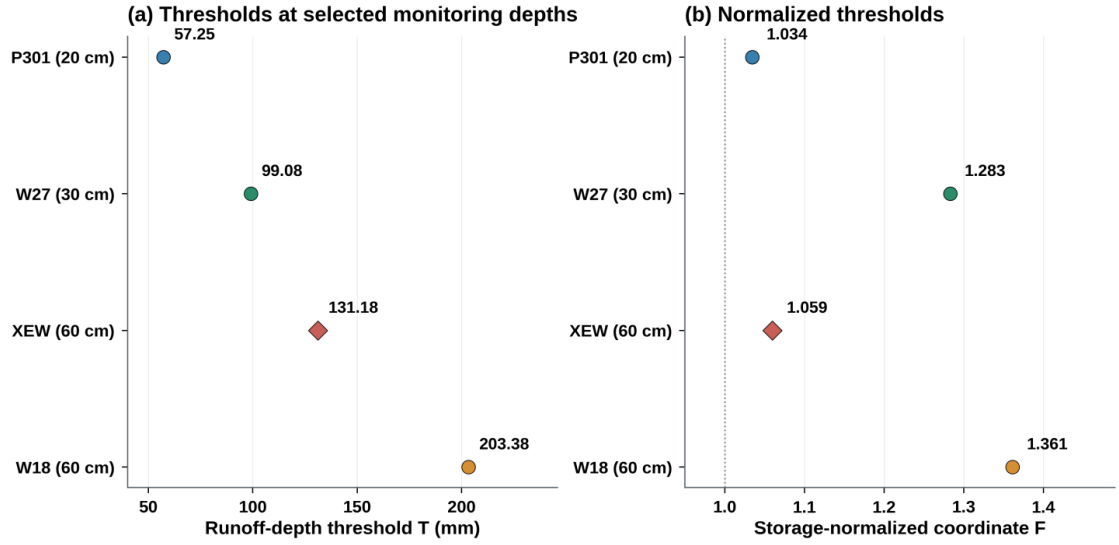


**Figure 6.** Published ASI$_0$+P threshold coordinates and their monitored-storage normalization across four catchments. (a) Representative arithmetic means of previously published site-specific thresholds at the selected integration depth. (b) Normalized F coordinates based on empirical $S_5$ and $S_{95}$ storage references. The XEW primary threshold uses the five spatially matched Hillslope-1 profiles (T = 131.18 mm). The reference percentiles do not measure total catchment storage capacity.

**Table 4.** Published ASI$_0$+P threshold coordinates associated with event stormflow depth and monitored-storage normalization at selected integration depths. $S_5$ and $S_{95}$ are percentiles of the daily monitored-storage series. XEW uses the matched-five primary threshold; the original eight-site mean is not the primary comparison. Values are rounded for display.

| Catchment | MID (cm) | T (mm) | $S_5$ (mm) | $S_{95}$ (mm) | C_ref (mm) | F | Sites (T/storage) |
|---|---|---|---|---|---|---|---|
| P301 | 20 | 57.25 | 19.14 | 55.98 | 36.84 | 1.034 | 10 / 10 |
| W27 | 30 | 99.08 | 35.97 | 85.15 | 49.17 | 1.283 | 4 / 4 |
| XEW | 60 | 131.18 | 33.06 | 125.67 | 92.61 | 1.059 | 5 / 5 |
| W18 | 60 | 203.38 | 86.22 | 172.29 | 86.07 | 1.361 | 6 / 6 |

### 4.6. Storage-reference and monitoring-support sensitivity

For the four catchments using the primary five-site-matched XEW threshold, relative MAD decreased from 0.3211 for T to 0.1062 for F, a descriptive reduction of 66.9% (Fig. 7a). The normalized values still differed among catchments, with F ranging from 1.034 in P301 to 1.361 in W18.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

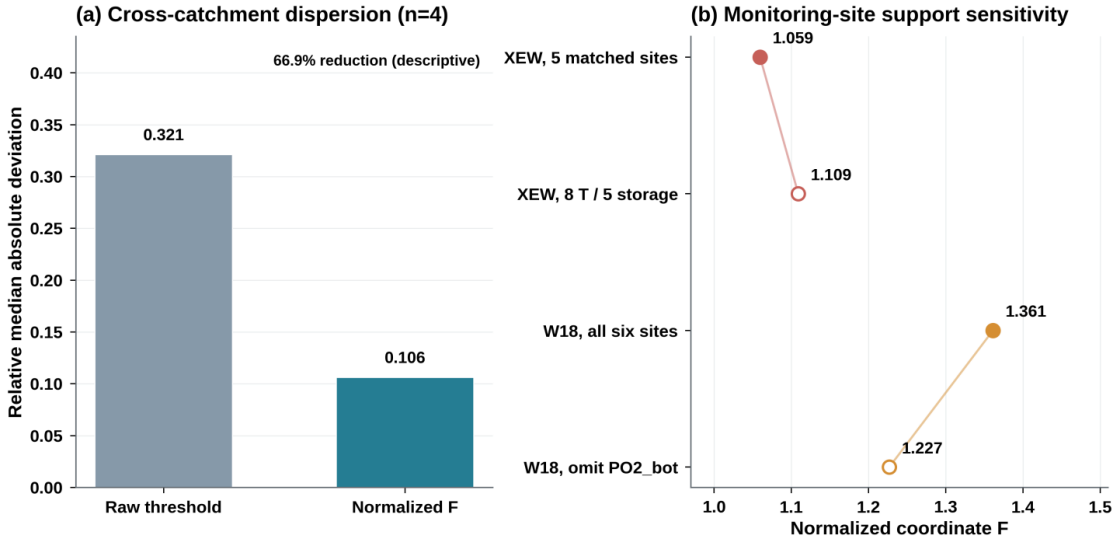


**Figure 7.** Descriptive cross-catchment dispersion and selected monitoring-support diagnostics. (a) Relative MAD of the four raw T values and the four normalized F values (66.9% descriptive reduction). (b) XEW five-position matched versus eight-threshold/five-storage comparison, and W18 full-six versus threshold-only omission of PO2_bot. The W18 calculation retains the original storage reference and is not a fully rematched five-site analysis.

With the published thresholds held fixed, replacing 5th–95th percentile storage references by 10th–90th percentile references changed F from 1.034 to 1.111 in P301, 1.283 to 1.548 in W27, 1.059 to 1.192 in five-site-matched XEW, and 1.361 to 1.564 in W18 (Fig. S5). Daily-monitoring coverage also changed selected values. In P301, the primary ≥8-of-10 criterion yielded 1,826 qualifying days (F = 1.034); allowing seven positions yielded 2,063 days (F = 1.049). In XEW, requiring all five profiles each day gave F = 1.051, compared with 1.059 under the primary daily-coverage rule.

Using the legacy eight-position XEW threshold mean (135.76 mm) with the five-position storage reference gave F = 1.109, compared with 1.059 for the five-position matched threshold (131.18 mm; Fig. 7b). In W18, omission of PO2_bot (site-level T = 261.2 mm) from the threshold mean changed T from 203.38 to 191.82 mm and F from 1.361 to 1.227 while the six-position storage reference was held fixed. The six-position W18 value remained the primary estimate. These sensitivity calculations used fixed published thresholds and did not estimate joint threshold-and-reference uncertainty.

## 5. Discussion

### 5.1. Non-uniform expansion of deep-soil responses

Previous hillslope studies have linked threshold runoff responses to the expansion and connection of locally active or saturated areas (Tromp-van Meerveld and McDonnell, 2006a, b). The present results extend this understanding by revealing temporally non-uniform expansion of monitored deep-soil responses. Among the 39 events in which all five profiles responded, the median fourth-to-fifth onset

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

interval was 2.83 h, compared with 0.50 h for the within-event median of the preceding intervals; the final interval was strictly the longest in 29 events (74.4%). P2 was the last responding profile in 25 events and P4 in another 10. Excluding P2 reduced the median first-to-last response span from 6.83 to 3.33 h. These results indicate that a few recurrently late-responding profiles disproportionately prolonged the completion of monitored deep-soil activation, providing an event-scale temporal complement to threshold-based descriptions of expanding subsurface response areas.

The delayed responses were not consistently ordered by ground-surface elevation. P2 and P4, the last responding profiles in 64.1% and 25.6% of D5 events, respectively, were located at elevations of 783.135 and 793.847 m, whereas the highest profile, P5 (801.496 m), was last in only one event. This suggests that elevation alone cannot explain the observed spatial heterogeneity. Soil depth, hillslope position, bedrock topography, and preferential flow pathways may contribute to spatial differences in subsurface responses (Bachmair and Weiler, 2012; Freer et al., 2002; Hopp and McDonnell, 2009; Penna et al., 2015), although their individual effects at P2 and P4 remain unresolved. The negative association between the D5 response span and X80 ($\rho = -0.613$) further suggests more synchronized profile activation under stronger combined rainfall and antecedent wetness, consistent with previous findings on the influence of hydrological state on subsurface response synchrony (Beiter et al., 2020; Pirastru et al., 2022). However, X80 does not isolate the effects of rainfall and antecedent storage, and the observed onset intervals represent differences in response timing rather than water-particle travel times.

The relationship between extensive deep-soil activation and outlet runoff provides further insight into the distinction between subsurface response and runoff delivery. Median event stormflow depth increased from 0.040 mm for D0 events and 0.051 mm for D1–D3 events to 0.544 mm for D4–D5 events. However, this contrast weakened after controlling for rainfall and antecedent storage: although 76.0% of matched pairs favored greater stormflow depth in the $D \geq 4$ group under the broader 3 × 3 matching scheme, the proportion decreased to 61.9% under stricter matching (rainfall difference ≤10 mm and antecedent-storage difference ≤20 mm; $p = 0.127$). The corresponding proportion under same-year matching was 68.6% ($p = 0.089$), providing limited evidence that extensive deep-soil activation independently increases stormflow magnitude. In the matched 139-event comparison, the D5 and delayed-peak probability centers were 189.7 and 211.0 mm, respectively, with a paired difference of 21.3 mm (95% year-block interval, 5.1–49.5 mm). All 18 manually verified delayed-peak events occurred among events with $D \geq 4$, indicating an association between widespread deep-soil activation and delayed-peak occurrence. However, extensive activation alone was not sufficient to ensure delayed runoff, which may additionally involve groundwater recharge and lateral hydraulic connectivity, consistent with previous XEW findings (Cui and Tian, 2025). The difference between probability-transition centers does not establish a temporal or causal sequence between deep-soil activation and delayed-peak formation.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

### 5.2. Spatially heterogeneous groundwater responses: persistent delays and hydrogeological controls

Groundwater response timing in XEW varied substantially among wells and hillslopes. W21 had a relatively shallow groundwater depth (0.2–2.2 m below ground surface), whereas W23 and W32 exhibited deeper ranges (6.4–12.2 and 6.1–9.9 m, respectively), consistent with previous observations (Cui et al., 2024; Cui and Tian, 2025). However, groundwater depth did not provide a consistent ranking of event-scale response times: the three HS2 wells generally responded rapidly, whereas W31 and W32 at HS3 showed longer peak-response lags. Similar spatial heterogeneity has been associated with topographic position, soil depth, and subsurface geological properties (Penna et al., 2015; Rinderer et al., 2016; Bachmair and Weiler, 2012), while antecedent groundwater conditions and hydraulic connectivity further influence response timing (Beiter et al., 2020; Pirastru et al., 2022).

A persistent late response was identified at W13, which was the last responding well in all 14 events with complete six-well activation (G6). Excluding W13 reduced the median first-to-last response span from 17.0 to 4.5 h; the median final adjacent interval was 10.5 h and was strictly the longest in 12 of 14 events (85.7%). Thus, one recurrently delayed location disproportionately prolonged the completion of the monitored groundwater response network. Unlike the D5 soil-response span, the G6 groundwater-response span showed no clear association with X80 ($\rho = -0.342$; 95% year-block interval, −0.861 to 0.335), suggesting that the temporal coordination of groundwater responses was not consistently related to the combined rainfall–antecedent-storage condition. W13 groundwater levels ranged from approximately 2.8 m below ground surface to below the 10 m borehole bottom (Cui et al., 2024). Its ground elevation (777.497 m) was close to that of the lowest HS1 soil-water profile, PO1-1 (778.149 m), while its screened interval in moderately weathered granite and overlying gravel and blockstone deposits provides a plausible hydrogeological context. Nevertheless, long-term groundwater depths do not establish event-specific antecedent conditions, and the respective effects of local storage, hydraulic conductivity, and subsurface structure remain unresolved.

The comparison of soil-water and groundwater onsets further showed that the earliest groundwater response preceded the first soil-profile response in seven of nine G6 events with identifiable deep-soil responses, whereas the latest groundwater response occurred after it in all nine. These contrasting boundaries indicate asynchronous activation rather than a fixed response sequence, consistent with previous observations of spatially variable groundwater and streamflow responses (Rinderer et al., 2016; Blumstock et al., 2016). Such response timing does not establish water-transfer pathways or travel times, because rapid hydraulic responses can coexist with much longer water-particle transit and storage times (McGuire and McDonnell, 2010). Although joint extensive soil-water and groundwater responses exceeded the unconditioned reference (11 observed versus 6.46 expected; $p = 0.000462$), they did not exceed the rainfall–antecedent-storage-conditioned expectation (10.60 expected; $p = 0.520$). Thus, the observed co-occurrence provided no clear evidence of additional statistical association beyond shared rainfall and antecedent wetness. However, the small sample ($n = 26$) and limited variation in the conditioned reference,

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

particularly in the high-rainfall/high-$ASI_{0,80}$ stratum where all eight events showed joint responses, constrain the ability to detect weaker associations. These results neither establish statistical independence nor exclude hydraulic connectivity between the monitored response domains.

### 5.3. Interpretation and limits of cross-catchment storage normalization

Absolute rainfall–antecedent-storage thresholds are difficult to compare across catchments with different soil-water integration depths, storage definitions, and monitoring support. Scaife et al. (2020) reported gross-precipitation-plus-antecedent-storage thresholds of 124–285 mm across three headwater catchments, while Nanda and Safeeq (2023) identified catchment wetness thresholds of 212–250 mm across three Mediterranean headwater catchments. These values demonstrate between-catchment variability but are not directly comparable with the present thresholds because their storage definitions, integration depths, and runoff-response metrics differ. Soil-moisture integration depth affects threshold coordinates and their explanatory relevance to stormflow generation (Cui and Tian, 2026a), while storage–discharge relationships also depend on monitoring location, measurement depth, and antecedent conditions (Nanda and Safeeq, 2024). More broadly, rainfall, catchment wetness, soil properties, geology, and hydrological connectivity jointly influence runoff thresholds (Cui and Tian, 2026b).

Normalizing threshold coordinates by the monitored S5–S95 storage range reduced the relative MAD across four catchments from 0.3211 to 0.1062, a descriptive reduction of 66.9%, although normalized values still ranged from 1.034 to 1.361. This suggests that differences in monitored storage variability account for part of the dispersion in absolute threshold magnitudes. However, ($C_{ref} = S_{95} - S_5$) represents observed storage variability at the selected integration depth and spatial support, not total soil, weathered-bedrock, or catchment storage capacity. Because T combines antecedent storage and event rainfall whereas $S_5$ and $S_{95}$ represent antecedent storage, $F = (T - S_5)/(S_{95} - S_5)$ can exceed one without implying saturation. Residual differences may reflect soil depth, texture, coarse fragments, bedrock weathering, and preferential flow pathways (Nanda and Safeeq, 2023, 2024; Cui and Tian, 2026b). Moreover, the threshold and storage references derive from related hydrological records, so mathematical scaling and statistical dependence may contribute to the reduced dispersion. The four-catchment comparison therefore supports an empirical comparability hypothesis, not a universal critical storage threshold or transferable physical parameter.

### 5.4. Implications for hydrological modelling, limitations and future research

The response-specific thresholds identified here offer potential observational constraints for evaluating threshold-controlled subsurface processes in hydrological models. Previous studies have demonstrated the value of threshold responses for model calibration and evaluation (Tromp-van Meerveld and McDonnell, 2006a; Penna et al., 2011; Beiter et al., 2020). The results provide empirical relationships between rainfall–antecedent storage conditions and the monitored extent and timing of deep-soil (D1–D5) and groundwater (G4–G6) responses. These could complement existing models: the Xinanjiang model represents spatially distributed tension-water storage capacity through WM and B (Zhao, 1992), whereas

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

the semi-distributed four-source model of Tudaji et al. (2025) represents pervious and impervious areas and vertically differentiated capillary-water (W), subsurface (S), and groundwater (G) layers. The latter provides a more direct framework for examining multiple runoff pathways, including impervious-area runoff, surface runoff, subsurface runoff, and groundwater runoff. Its simplified spatial representation of groundwater storage and implicit treatment of groundwater response timing could be further evaluated using the observed G4–G6 thresholds and persistent delay at W13. Such applications require explicit correspondence between model states and observational locations and depths; monitored response counts are not contributing-area fractions, and F is not a physical storage-capacity parameter.

An important hypothesis is that incorporating response-specific thresholds and event-scale timing alongside conventional outlet-streamflow constraints may reduce parameter equifinality. Different parameter combinations can reproduce similar hydrographs while representing different internal storage states and runoff-generation processes (Beven and Freer, 2001). Subsurface response thresholds and timing may therefore help narrow the parameter sets that simultaneously reproduce outlet discharge and independently observed internal processes. Groundwater-level observations have been shown to improve parameter identification, although the benefits of additional soil-moisture observations depend on their measurement characteristics and compatibility with model structure (Kreye et al., 2019). Multi-objective calibration may also improve internal-variable simulations without necessarily improving streamflow predictions (Mei et al., 2023). Whether these additional constraints reduce equifinality and improve predictive performance remains to be tested.

The main limitations are the concentration of five soil-water profiles on one hillslope, the uneven distribution of six groundwater wells across three hillslopes, and the limited G6 sample of 14 events. Differences in monitoring locations, well depths and screened intervals, and response-detection criteria further constrain spatial connectivity interpretations. Monitoring-network density and configuration also influence the representation of groundwater-level distributions (Ohmer et al., 2019; Allgeier et al., 2020). The cross-catchment normalization is likewise based on only four catchments with different integration depths and spatial supports. Future research should combine expanded, high-frequency soil-water and groundwater monitoring with independent characterization of soil and bedrock properties. Larger cross-catchment datasets with comparable integration depths and storage references are also needed. For modelling, response-specific probability-transition coordinates and timing metrics should be incorporated alongside streamflow, soil-moisture, and groundwater observations using appropriate observation operators. Comparing streamflow-only and multi-objective calibration, followed by validation against independent events, periods, and catchments, would test whether these constraints improve process representation, parameter identifiability, and predictive performance.

## 6. Conclusions

# 6. Conclusions

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

This study investigated response-specific hydrological thresholds and event-scale subsurface response dynamics in the Xitaizi Experimental Watershed (XEW), and evaluated the potential of monitored-storage normalization for cross-catchment threshold comparison. The main conclusions are:

(1) Estimated p50 transition coordinates increased with the spatial extent of monitored subsurface responses, although the increases were not uniform. The coordinates increased from 142.7 to 190.9 mm for deep-soil responses (D1–D5) and from 149.0 to 179.2 mm for groundwater responses (G4–G6). The delayed-peak transition occurred at a higher coordinate than complete five-profile activation, indicating that widespread subsurface activation and delayed runoff represent distinct hydrological transitions.

(2) Subsurface response expansion was temporally non-uniform, with a few recurrently delayed monitoring locations disproportionately prolonging network activation. Widespread deep-soil activation was consistently associated with delayed runoff occurrence, although it was not sufficient to ensure it. In contrast, joint extensive soil-water and groundwater responses did not exceed the rainfall–antecedent-storage-conditioned expectation, providing no clear evidence of additional statistical association beyond shared forcing. The limited sample and conditional reference, however, preclude conclusions about statistical independence or the absence of hydraulic connectivity.

(3) Normalizing published $ASI_0$+P thresholds by the monitored S5–S95 storage range reduced between-catchment relative dispersion by 66.9% across four catchments, suggesting improved comparability of thresholds relative to observed storage variability. This normalization provides an empirical comparative coordinate rather than a measure of relative saturation or a universal critical storage state. Together with the response-specific thresholds and timing characteristics, it offers potential observational constraints for hydrological model evaluation and may help reduce parameter equifinality. Whether these constraints improve process representation and predictive performance remains to be tested through multi-objective calibration and independent cross-catchment validation.

## Author contributions

Zhen Cui: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing - original draft. Fuqiang Tian: Supervision, funding acquisition, project administration, writing - review and editing.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. U2442201 and 52409031).

## Data availability

The expanded XEW observations and reviewed event catalogue are available from the corresponding author upon reasonable request. KREW and Coweeta observations originate from the U.S. Forest Service and Coweeta research archives identified in Cui and Tian (2026a).

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

## Competing interests

Fuqiang Tian is a member of the editorial board of Hydrology and Earth System Sciences.

## Acknowledgements

We thank the U.S. Forest Service and Coweeta research teams for maintaining and sharing the comparative catchment observations.

## References

Ali, G., Tetzlaff, D., McDonnell, J. J., Soulsby, C., Carey, S., Laudon, H., McGuire, K., Buttle, J., Seibert, J., and Shanley, J.: Comparison of threshold hydrologic response across northern catchments, Hydrol. Process., 29, 3575–3591, https://doi.org/10.1002/hyp.10527, 2015.

Allgeier, J., González-Nicolás, A., Erdal, D., Nowak, W., and Cirpka, O. A.: A stochastic framework to optimize monitoring strategies for delineating groundwater divides, Front. Earth Sci., 8, 554845, https://doi.org/10.3389/feart.2020.554845, 2020.

Bachmair, S. and Weiler, M.: Hillslope characteristics as controls of subsurface flow variability, Hydrol. Earth Syst. Sci., 16, 3699–3715, https://doi.org/10.5194/hess-16-3699-2012, 2012.

Beiter, D., Weiler, M., and Blume, T.: Characterising hillslope–stream connectivity with a joint event analysis of stream and groundwater levels, Hydrol. Earth Syst. Sci., 24, 5713–5744, https://doi.org/10.5194/hess-24-5713-2020, 2020.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, https://doi.org/10.1016/S0022-1694(01)00421-8, 2001.

Blume, T. and van Meerveld, H. J.: From hillslope to stream: methods to investigate subsurface connectivity, Wiley Interdiscip. Rev. Water, 2, 177–198, https://doi.org/10.1002/wat2.1071, 2015.

Blumstock, M., Tetzlaff, D., Dick, J. J., Nuetzmann, G., and Soulsby, C.: Spatial organization of groundwater dynamics and streamflow response from different hydropedological units in a montane catchment, Hydrol. Process., 30, 3735–3753, https://doi.org/10.1002/hyp.10848, 2016.

Bracken, L. J., Wainwright, J., Ali, G. A., Tetzlaff, D., Smith, M. W., Reaney, S. M., and Roy, A. G.: Concepts of hydrological connectivity: research approaches, pathways and future agendas, Earth-Sci. Rev., 119, 17–34, https://doi.org/10.1016/j.earscirev.2013.02.001, 2013.

Cui, Z., Tian, F., Zhao, Z., Xu, Z., Duan, Y., Wen, J., and Khan, M. Y. A.: Bimodal hydrographs in a semi-humid forested watershed: characteristics and occurrence conditions, Hydrol. Earth Syst. Sci., 28, 3613–3632, https://doi.org/10.5194/hess-28-3613-2024, 2024.

Cui, Z. and Tian, F.: Delayed stormflow generation in a semi-humid forested watershed controlled by soil water storage and groundwater dynamics, Hydrol. Earth Syst. Sci., 29, 2275–2291, https://doi.org/10.5194/hess-29-2275-2025, 2025.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.

719 Cui, Z. and Tian, F.: Soil moisture thresholds for runoff generation: depth sensitivity and mechanistic
720 controls, J. Hydrol., 677, 135827, https://doi.org/10.1016/j.jhydrol.2026.135827, 2026a.
721 Cui, Z. and Tian, F.: Runoff thresholds, runoff generation mechanisms, and catchment characteristics: a
722 global synthesis, Hydrol. Earth Syst. Sci., 30, 5281–5296, https://doi.org/10.5194/hess-30-5281-2026,
723 2026b.
724 Detty, J. M. and McGuire, K. J.: Threshold changes in storm runoff generation at a till-mantled headwater
725 catchment, Water Resour. Res., 46, W07525, https://doi.org/10.1029/2009WR008102, 2010.
726 Freer, J., McDonnell, J. J., Beven, K. J., Peters, N. E., Burns, D. A., Hooper, R. P., Aulenbach, B., and
727 Kendall, C.: The role of bedrock topography on subsurface storm flow, Water Resour. Res., 38, 1269,
728 https://doi.org/10.1029/2001WR000872, 2002.
729 Fu, C., Chen, J., Jiang, H., and Dong, L.: Threshold behavior in a fissured granitic catchment in southern
730 China: 1. Analysis of field monitoring results, Water Resour. Res., 49, 2519–2535,
731 https://doi.org/10.1002/wrcr.20191, 2013.
732 Gao, H., Pfister, L., and Kirchner, J. W.: Quantifying controls on rapid and delayed runoff response in
733 double-peak hydrographs using ensemble rainfall-runoff analysis (ERRA), Hydrol. Earth Syst. Sci.,
734 29, 6529–6547, https://doi.org/10.5194/hess-29-6529-2025, 2025.
735 Hopp, L. and McDonnell, J. J.: Connectivity at the hillslope scale: identifying interactions between storm
736 size, bedrock permeability, slope angle and soil depth, J. Hydrol., 376, 378–391,
737 https://doi.org/10.1016/j.jhydrol.2009.07.047, 2009.
738 Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.:
739 Hydrologic connectivity between landscapes and streams: transferring reach- and plot-scale
740 understanding to the catchment scale, Water Resour. Res., 45, W04428,
741 https://doi.org/10.1029/2008WR007225, 2009.
742 Klaus, J. and Jackson, C. R.: Interflow is not binary: a continuous shallow perched layer does not imply
743 continuous connectivity, Water Resour. Res., 54, 5921–5932,
744 https://doi.org/10.1029/2018WR022920, 2018.
745 Kreye, P., Gelleszun, M., and Meon, G.: Parameter identification in hydrological models using
746 groundwater-level measurements and satellite-based soil moisture, Hydrol. Sci. J., 64, 633–652,
747 https://doi.org/10.1080/02626667.2019.1599120, 2019.
748 Lehmann, P., Hinz, C., McGrath, G., Tromp-van Meerveld, H. J., and McDonnell, J. J.: Rainfall threshold
749 for hillslope outflow: an emergent property of flow pathway connectivity, Hydrol. Earth Syst. Sci.,
750 11, 1047–1063, https://doi.org/10.5194/hess-11-1047-2007, 2007.
751 McCullagh, P.: Regression models for ordinal data, J. R. Stat. Soc. B, 42, 109–127,
752 https://doi.org/10.1111/j.2517-6161.1980.tb01109.x, 1980.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.



753 McDonnell, J. J., Spence, C., Karran, D. J., van Meerveld, H. J., and Harman, C. J.: Fill-and-spill: a process
754 description of runoff generation at the scale of the beholder, Water Resour. Res., 57, e2020WR027514,
755 https://doi.org/10.1029/2020WR027514, 2021.
756 McGuire, K. J. and McDonnell, J. J.: Hydrological connectivity of hillslopes and streams: characteristic
757 time scales and nonlinearities, Water Resour. Res., 46, W10543,
758 https://doi.org/10.1029/2010WR009341, 2010.
759 Mei, Y., Mai, J., Do, H. X., Gronewold, A., Reeves, H. W., Eberts, S. M., Niswonger, R. G., Regan, R. S.,
760 and Hunt, R. J.: Can hydrological models benefit from using global soil moisture, evapotranspiration,
761 and runoff products as calibration targets?, Water Resour. Res., 59, e2022WR032064,
762 https://doi.org/10.1029/2022WR032064, 2023.
763 Nanda, A. and Safeeq, M.: Threshold controlling runoff generation mechanisms in Mediterranean
764 headwater catchments, J. Hydrol., 620, 129532, https://doi.org/10.1016/j.jhydrol.2023.129532, 2023.
765 Nanda, A. and Safeeq, M.: Nonlinear storage–discharge dynamics of forested headwater catchment: a
766 hysteresis index approach, Hydrol. Process., 38, e15201, https://doi.org/10.1002/hyp.15201, 2024.
767 Ohmer, M., Liesch, T., and Goldscheider, N.: On the optimal spatial design for groundwater level
768 monitoring networks, Water Resour. Res., 55, 9454–9473, https://doi.org/10.1029/2019WR025728,
769 2019.
770 Penna, D., Mantese, N., Hopp, L., Dalla Fontana, G., and Borga, M.: Spatio-temporal variability of
771 piezometric response on two steep alpine hillslopes, Hydrol. Process., 29, 198–211,
772 https://doi.org/10.1002/hyp.10140, 2015.
773 Penna, D., Tromp-van Meerveld, H. J., Gobbi, A., Borga, M., and Dalla Fontana, G.: The influence of soil
774 moisture on threshold runoff generation processes in an alpine headwater catchment, Hydrol. Earth
775 Syst. Sci., 15, 689–702, https://doi.org/10.5194/hess-15-689-2011, 2011.
776 Pirastru, M., Iovino, M., Marrosu, R., Di Prima, S., Giadrossich, F., and Awada, H.: Large-scale lateral
777 saturated soil hydraulic conductivity as a metric for the connectivity of subsurface flow paths at
778 hillslope scale, Hydrol. Process., 36, e14649, https://doi.org/10.1002/hyp.14649, 2022.
779 Rinderer, M., van Meerveld, H. J., Stähli, M., and Seibert, J.: Is groundwater response timing in a pre-
780 alpine catchment controlled more by topography or by rainfall?, Hydrol. Process., 30, 1036–1051,
781 https://doi.org/10.1002/hyp.10634, 2016.
782 Ross, C. A., Ali, G. A., Spence, C., and Courchesne, F.: Evaluating the ubiquity of thresholds in rainfall-
783 runoff response across contrasting environments, Water Resour. Res., 57, e2020WR027498,
784 https://doi.org/10.1029/2020WR027498, 2021.
785 Salve, R., Rempe, D. M., and Dietrich, W. E.: Rain, rock moisture dynamics, and the rapid response of
786 perched groundwater in weathered, fractured argillite underlying a steep hillslope, Water Resour. Res.,
787 48, W11528, https://doi.org/10.1029/2012WR012583, 2012.

https://doi.org/10.5194/egusphere-2026-7190
Preprint. Discussion started: 25 November 2026
© Author(s) 2026. CC BY 4.0 License.



788 Scaife, C. I., Singh, N. K., Emanuel, R. E., Miniat, C. F., and Band, L. E.: Non-linear quickflow response
789 as indicators of runoff generation mechanisms, Hydrol. Process., 34, 2949–2964,
790 https://doi.org/10.1002/hyp.13780, 2020.
791 Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-
792 storm analysis of the Panola hillslope, Water Resour. Res., 42, W02410,
793 https://doi.org/10.1029/2004WR003778, 2006a.
794 Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill
795 and spill hypothesis, Water Resour. Res., 42, W02411, https://doi.org/10.1029/2004WR003800,
796 2006b.
797 Tudaji, M., Nan, Y., and Tian, F.: Assessing the value of high-resolution data and parameter transferability
798 across temporal scales in hydrological modeling: a case study in northern China, Hydrol. Earth Syst.
799 Sci., 29, 2633–2654, https://doi.org/10.5194/hess-29-2633-2025, 2025.
800 Zhao, R. J.: The Xinanjiang model applied in China, J. Hydrol., 135, 371–381,
801 https://doi.org/10.1016/0022-1694(92)90096-E, 1992.