We thank reviewer 1 for their careful and constructive reviews. The comments will help us substantially improve the clarity of our methods and the framing of our contribution. Below, the original review comments are shown in **bold** and our responses to each comment use normal text. Line numbers refer to the reviewed preprint; specific revised text will be added in the tracked-changes manuscript if the editor is in support of a revised manuscript.

Reviewer 1

This manuscript addresses an important topic and provides a useful national-scale analysis of drought-to-flood and flood-to-drought streamflow transitions. I particularly appreciate the comprehensive sensitivity analysis and the discussion of implications for water resources management. However, several methodological details need substantial clarification. I also encourage the authors to sharpen the literature review and more explicitly state the scientific gap and contribution of the study.

We thank the reviewer for the positive assessment and for the detailed, actionable feedback, particularly on the methods section and framing. We will sharpen the introduction and make Sections 2.2 and 2.3 substantially clearer, as detailed below.

Major comments

1. The introduction does a good job motivating the management challenges associated with drought-to-flood and flood-to-drought transitions. However, the literature review on related work is currently limited. For example, the manuscript refers to "impactful transitions" (line 175) and the overlap between meteorological and hydrological whiplash (lines 178-179). First, meteorological whiplash and hydrological whiplash need to be defined more clearly. Second, they need to be situated within existing literature. It would help to explain where previous studies stand in identifying impactful transitions, quantifying meteorological versus hydrological whiplash, and assessing their overlap. In addition, the novelty currently reads mainly as the "exploration of both DtF and FtD transitions" (line 171), but the manuscript should more clearly state the specific knowledge gap and how this study fills it. For example, is the key contribution the national-scale streamflow-based characterization, the comparison of DtF and FtD asymmetry, the sensitivity to event definitions, or the linkage to meteorological transitions?

We agree and will revise the introduction in three ways. (1) We will define both terms explicitly on first use: *meteorological (hydroclimate) whiplash* as a rapid shift between extreme wet and dry conditions in atmospheric/water-balance indices (e.g., SPEI, precipitation), following Swain et al. (2025), and *hydrological whiplash* as the analogous rapid shift between extreme high- and low-flow conditions in the streamflow record, following Hammond et al. (2025). We will make clear that the two are related but distinct, and that watershed processes can decouple them. (2) We will expand the literature review to state where prior work stands: meteorological/precipitation whiplash is comparatively well characterized (He and Sheffield, 2020; Na and Najafi, 2024; Puxley and Martin, 2025; Swain et al., 2025; Mullens and Engström, 2025), whereas streamflow-based transition studies are more recent and have focused predominantly on the DtF direction (Götte and Brunner, 2024; Matanó et al., 2024; Anderson et al., 2025), on single directions or single regions, or on annual/multi-year indicators (Li et al., 2025; Maharjan et al., 2025). We will note that the overlap between meteorological and hydrological whiplash has been assessed largely at the level of individual extremes or coincidence statistics rather than direction-resolved transition co-occurrence. (3) We will rewrite the novelty statement (previously line 170–171) to state the specific gap and contributions concretely: this is, to our knowledge, the first national-scale, streamflow-based characterization that (i) treats DtF and FtD transitions symmetrically and quantifies their asymmetry, (ii) systematically tests the sensitivity of transition prevalence and rapidity to a broad matrix of event definitions, (iii) introduces an "impactful" transition threshold based on percentile-space magnitude, and (iv) directly quantifies direction-resolved co-occurrence with meteorological whiplash. We will state that the primary contribution is the comparative DtF/FtD asymmetry framework and its linkage to meteorological transitions, with the sensitivity analysis and national coverage supporting the robustness of those findings.

2. Section 2.2 is central to the study, but many of the definitions and implementation choices are difficult to follow. Several terms are introduced before they are defined. The authors should reorganize this section so that event-definition components are introduced first, followed by using the definitions and then the sensitivity settings.

We agree that Section 2.2 introduced several terms (inter-event period, minimum duration, pooling, flood separation) before defining them. We will reorganize the section into three clearly labeled parts: (2.2.1) definitions of the event-identification components (percentile threshold, inter-event period and pooling, minimum duration for droughts; percentile/POT threshold and flood separation for floods), with each term defined at first mention; (2.2.2) how those components are combined to identify events and derive transitions; and (2.2.3) the sensitivity matrix of tested definitions and the three representative (strict/study/lenient) definitions. This will remove the forward references and the duplicated material noted in 2d.

2a) The manuscript uses minimum drought durations of 5, 15, and 30 days. It is worth reconsidering whether 5-day or even 15-day low-flow periods should be called “droughts,” or whether “dry periods” would be more accurate. Also, the duration may represent different lengths under fixed and variable-threshold methods. In the variable-threshold approach, it seems each daily value represents a 7-day averaged condition, so it has to be clarified whether the effective duration equals the nominal duration under the two approaches.

2a. We agree that "drought" is a strong term for a 5- or even 15-day low-flow period and will add a clarifying statement that shorter minimum-duration definitions capture short "dry periods" rather than sustained hydrological droughts; we retain the shorter durations only as sensitivity endpoints and use a 15-day minimum in the main analysis. Regarding effective duration, the reviewer is correct that the nominal duration is not identical between approaches. In the variable-threshold approach each daily value reflects a 7-day moving-average condition, so a nominal *n*-day exceedance corresponds to an effective window of approximately *n*+6 days of underlying flow. We will state this explicitly and clarify that inter-event pooling and minimum-duration criteria are applied to the (smoothed) daily series in the variable case and to the daily series in the fixed case, and we will quantify the resulting difference.

2b) The phrase “the interevent period is [...] less than 10% of the drought volume” (line 241) is confusing because a period is being compared with a volume. The authors should define drought volume and explain how this criterion is calculated.

2b. We agree the phrasing compared incommensurate quantities. We will define drought volume as the cumulative deficit below the drought threshold, expressed in percentile-days (the sum, over drought days, of the difference between the threshold percentile and the observed percentile). The pooling criterion merges two events when the intervening above-threshold "surplus" (also in percentile-days) is less than 10% of the preceding drought's volume, following the inter-event volume criterion of WMO (2008). We will rewrite the sentence to make both the units and the comparison explicit.

2c) Line 242, does “pool drought events” mean if drought days have inter-event period <=5 days, those days are merged as one event? It is not until here that I understand why you need an inter-event period, which was introduced in line 223. Similarly, is the “15-day” for flood separation (lines 246-247) a similar concept of “inter-event period” used in drought? If so, why not introduce it earlier in the first paragraph in section 2.2?

2c. Correct, "pool drought events" means that drought periods separated by an inter-event gap at or below the threshold (<5 days, or below the volume criterion) are merged into a single event. We will introduce the inter-event period and its pooling role together in 2.2.1 so the concept is defined where it first appears. The 15-day flood separation is indeed the flood analogue of the inter-event period, we will introduce it alongside the drought inter-event period and label them consistently as "event-separation criteria" so the parallel is clear.

2d) Lines 240-247, aren't these the same as in lines 232-236? Again, it would be clearer to introduce all pooling, separation, and duration criteria before using the definitions.

2d. Correct, the former lines 240–247 largely restated the pooling/separation/duration criteria already listed in the sensitivity matrix (former 232–236). In the reorganized section, all pooling, separation, and duration criteria will be defined once in 2.2.1, and the study/strict/lenient settings will be given once in 2.2.3, eliminating the repetition.

3. The manuscript uses both fixed and variable percentile approaches, but the implementation is difficult to follow in several places.

3a) The zero-flow adjustment is described very briefly (lines 212-215). While reference is cited, the authors should explain how days with different numbers of preceding zero-flow days are combined with positive-flow days when assigning percentiles, and whether this procedure is applied before or after 7-day averaging in the variable approach. It is also unclear whether zero-flow duration is calculated using the full time series (e.g., fixed approach) or only within the comparison window for a given day-of-year (e.g., variable approach).

3a. We will expand the zero-flow description. We will state that zero-flow days are ranked below all positive-flow days, with ties among zero-flow days broken by the number of preceding consecutive zero-flow days (more preceding zeros → lower percentile), so that positive- and zero-flow days are placed on a single continuous ranking (the combined threshold-level / continuous-dry-period approach of Van Huijgevoort et al., 2012; Simeone et al., 2024). We will clarify that in the variable approach this ranking is performed after 7-day averaging, within the day-of-year comparison window, and that the preceding-zero-flow count used for tie-breaking is computed from the continuous time series (not reset within each day-of-year window), whereas in the fixed approach the ranking uses the full record. We will add this to Section 2.1.

3b) The authors should acknowledge that changing the reference period can change the percentile values themselves (line 210). This matters because percentiles are central to identifying droughts and transitions.

3b. We agree and will explicitly acknowledge that percentile values themselves depend on the reference period, and that we fix the reference period at 1981–2020 for both approaches to isolate transition behavior from reference-period effects, consistent with Hammond et al. (2022) and Simeone et al. (2024). We will note this as a caveat where percentiles are first introduced (former line 210).

3c) The main text appears to identify droughts using variable percentiles but then evaluate "impactful" transitions using a fixed full-record percentile space (line 266). This hybrid framework is difficult to follow.

3c. We recognize the hybrid framework was under-explained. Droughts are identified in variable-percentile space because day-of-year thresholds are needed to detect seasonally relative low-flow events. Floods are identified using fixed percentiles as flows over a static threshold may be considered may be considered a flood regardless of the time of year. Impactfulness is evaluated in fixed-percentile space because the fixed percentile measures movement through a site's full flow-duration curve and thus indexes absolute hydrologic-state change rather than seasonal re-ranking. We will add a short paragraph explaining why each space is appropriate for its respective purpose, and we will flag this design choice clearly rather than leaving it implicit.

3d) Line 329 states that the maximum increase and decrease in streamflow percentiles are calculated during each season and the preceding 90 days. Does this correspond to a 120-day window in total? However, the following sentence says these maximum changes are compared against the "seasons" where there were meteorological or hydrological transitions, which seems to imply a 30-day window. The wording is confusing and should be clarified.

3d. We agree the wording was confusing. "Each season and the preceding 90 days" was intended to describe a window comprising the transition season plus the 90 days before it, approximately 180 days, not 120. The subsequent comparison against "seasons" with transitions used the seasonal (~90-day) unit for defining *which* periods contained transitions, while the maximum percentile change was computed over the longer season-plus-90-day window to capture the full swing. We will rewrite the sentence to state both windows explicitly and to distinguish the detection unit from the change-computation window.

3e) The authors should also clarify how the maximum change in streamflow percentile is calculated. How do they ensure that the identified maximum change corresponds to the actual transitional event, rather than another large streamflow fluctuation occurring within the same window? An example would be very helpful to illustrate the procedure.

3e. The maximum change is the largest increase (or decrease) in streamflow percentile between any two days within the window. To ensure it reflects the transition of interest rather than an unrelated fluctuation, we constrained the paired days so that the direction of the maximum change matches the transition type (i.e., for a DtF window the maximum *increase* from a low-percentile day to a later high-percentile day). We acknowledge that in windows with multiple large swings this measure captures the dominant swing rather than a specifically labeled event. We will state this limitation and add a worked example (annotated on one case-study hydrograph) illustrating how the paired minimum and maximum are selected.

3f) The authors should also specify which event values are compared to calculate the 50-percentile shift, such as drought minimum to flood maximum, event mean to event peak, or event end to event start, etc.

3f. We will specify that the 50-percentile impactfulness shift is computed as the difference between the minimum fixed percentile during the initiating drought (for DtF) or the maximum fixed percentile during the initiating flood (for FtD) and the corresponding extreme of the terminating event. So, this is the full swing from the low point of the first event to the high point of the second (and vice versa for FtD). We will add this definition explicitly.

4. The manuscript compares hydrological DtF/FtD transitions with meteorological dry-to-wet and wet-to-dry transitions, but the temporal matching criteria need more explanation.

4a) The manuscript should clarify the timescale mismatch between hydrological and meteorological transitions. Hydrological transitions are currently identified using relatively short event-scale criteria (e.g., 5-, 15-, and 30-day for drought, and no stated duration for flood), while meteorological transitions are based on seasonal timestep (line 317; is it 3-month?). The co-occurrence rate likely depends on the selected timescale, and the timescale mismatch should be acknowledged.

4a. We agree and will explicitly acknowledge the timescale mismatch: hydrological transitions are identified at event scale (5–30-day drought minimums; POT-defined floods), whereas the Swain et al. (2025) meteorological transitions are defined at seasonal steps. We will state that the measured co-occurrence rate is conditional on this mismatch and on the alignment window, and we will add this caveat both in Section 2.4 and in the Discussion. This also motivates the sensitivity of co-occurrence to alignment choices reported via Table S1.

4b) The authors should clarify how event order is considered. For example, must a hydrological DtF transition occur after a meteorological dry-to-wet transition? If order is not required, then the manuscript should avoid causal language such as meteorological transitions "produce" hydrological transitions (line 575).

4b. Correct, we do not require event order: a hydrological transition is counted as co-occurring if the meteorological transition falls in the same, immediately preceding, or immediately following season (and vice versa). Because a meteorological transition occurring after the streamflow transition can still register as a match, our metric reflects temporal co-occurrence rather than a directed cause-and-effect relationship. We agree causal language is not warranted, though meteorology is the physical driver of hydrology, and will replace "produce" (line 575) with co-occurrence language (e.g., "co-occur with," "coincide with").

4c) Lines 325–326 state that the co-occurrence between hydrological DtF transitions and meteorological dry-to-wet transitions is evaluated during the same season or the immediately preceding or following season. Does this correspond to a total window of approximately 270 days? However, the caption of Figure 7 states that the co-occurrence is evaluated within one season. It was not until I reached Table S1 that I realized both metrics (same season only, and same or adjacent season) were analyzed. I suggest clarifying this distinction in the main text.

4c. The reviewer raises a fair point about ambiguity. Our intended meaning of "within one season" was "within one season's distance", that is, the same season plus the immediately preceding or following season. You are correct that this is a total window of 270 days. We will change the language throughout from "within one season" (6 instances) to "in the same or adjacent season" to reduce ambiguity.

4d) Lines 320–321: The authors may also want to clarify the rationale for the different timing choices used throughout the analysis. For the co-occurrence analysis, meteorological and hydrological transitions are assigned to the season in which the transition ends, whereas the rose plots use the month in which the first event ends. Is there a reason for using different temporal reference points in these analyses? A brief explanation would help readers understand the methodology.

4d. Yes, there is a deliberate reason for the two reference points. For most of our analysis, including the rose plots, we use the month in which the first event ends, because we are most interested in extreme-event transitions from the perspective of already being in an extreme: this framing is the most applicable for understanding transition risk and, eventually, for forecasting these transitions as they begin to unfold. For the co-occurrence analysis, however, we assign each transition to the season in which it ends (the timing of the terminating extreme), for two reasons: the completed shift is the event being matched between the two systems, and, importantly, Swain et al. (2025) define their meteorological transitions by the terminating extreme, so we adopt the same convention here to eliminate unnecessary differences between our approaches when comparing the two datasets. We will add a brief sentence explaining this rationale, so the two reference points are no longer unexplained.

5. Several statements should be interpreted more carefully. For example, when the manuscript reports the fraction of transitions occurring within 90 days or 30 days (Lines 431-433), this is more of a measure of the prevalence, not transition speed itself. Claims that DtF transitions are "faster" (line 431) should be supported by direct comparison of transition-time, such as Fig3 c, h.

Similarly, where the manuscript compares percentages across regions or groups, the term "fraction" may be more accurate than "frequency" (line 517), unless actual event counts are being compared. Small differences, such as 21.9% versus 20.8% (line 519), should not be overinterpreted as indicating that transitions are meaningfully more common in one region than another.

We agree on both counts. We will distinguish *prevalence* (the fraction of transitions falling under a duration threshold) from *speed* (the transition-time distribution itself). Where we claim DtF transitions are faster, we will support it with the median transition times in Fig. 3c,h rather than relying solely on the sub-30/90-day fractions. We will replace "frequency" with "fraction" wherever we report percentages rather than counts (including former line 517), and we will soften comparisons of small differences (e.g., 21.9% vs. 20.8%) so they are described as comparable rather than meaningfully different, reserving stronger language for larger, consistent gaps.

6. Several watershed characteristics need clearer definitions and data sources: PET method/source (line 292), define PET before first use (line 305), SWE data source (line 293), whether the "more than 25%" snow criterion is a long-term average, and the calculation of "input seasonality" (line 309). There is also inconsistency in the use of "concentration."

We will revise Section 2.3 to add these details. We will specify the PET method and source (gridMET-derived PET; Abatzoglou, 2013) and define PET (potential evapotranspiration) at first use, which will precede the wet/dry classification. We will give the SWE data source and clarify that the "peak SWE / total precipitation > 25%" snow criterion is based on the long-term mean of annual values over the reference period. We will describe input seasonality explicitly as the concentration index of daily surface-water input (rainfall, rain-on-snow, and snowmelt) within the year from Hammond (2024), with the concentration/event-dominated split at 0.5. To resolve the "concentration" inconsistency, we will use a single consistent term, "input concentration," throughout, defined as the temporal concentration of surface-water inputs (rainfall plus snowmelt), and will correct the former "flow concentration" wording (line 308) and the "rainfall and snowmelt concentration" wording (lines 533, 536) to match.

Minor comments

We thank the reviewer for the detailed line by line edits. Responses to each are below; typographical and reference fixes will all be made.

Line 200: The 95% completeness threshold still allows ~18 missing days per year... check whether missing days are clustered.

We plan to check the distribution of missing days. We believe that across selected gages, missing days are predominantly isolated or in short runs rather than clustered into extended gaps.

Line 201: Clarify whether rolling decades are updated every year or every 10 years.

We will clarify that the decades are fixed calendar decades (1981–1990, 1991–2000, ..., 2011–2020), not annually rolling windows, and will revise the wording accordingly.

Lines 225–227: Clarify whether flood percentiles use daily percentiles or 7-day percentiles like drought identification.

Flood percentiles/POT thresholds are computed on daily (1-day) flow, not the 7-day series, because flood peaks are short-duration. We will state this explicitly to contrast with the 7-day smoothing used for drought (variable) identification.

Line 252: The manuscript uses both "transition duration" and "transition time." Please use one term consistently.

Agreed. We will use "transition time" consistently throughout.

Line 319: "not" should be "note."

This will be corrected.

Line 352: Please specify what the transition time here refers to. It is FtD, not DtF. Right?

Correct, the 85–100-day transition times in that passage are from flood back to drought (FtD). We will correct the label.

Line 364: "Fig. A13" should be "Fig. S13."

This will be corrected here and wherever "A#" appears for supplementary figures.

Line 374: The sentence appears to be missing a word.

This will be corrected, the sentence will read "...the rate of change is much faster **with** lenient drought and flood definitions (13 days)..."

Line 451: Figure 6 and Figure 7 show DtF under 30 days and FtD under 90 days. What about DtF under 90 days and FtD under 30 days?

We note this appears to reference Fig. S6 and Fig. S7. We will add 2 figures to the supplemental information to include regional rose plots for DtF transitions under 90 days and FtD transitions under 30 days.

Line 502: Figure 4 appears to indicate prevalence rather than timing unless seasonality is shown using the combined methods.

Agreed. Figure 4 shows prevalence (event counts and the fraction of transitions under 90 days), not seasonality. We will correct the text to refer to prevalence rather than timing.

Line 517: these are not only "sub-seasonal DtF" but also impactful DtF that a 50% percentile shift is required. Same thing throughout this section.

Thank you. You are right that Fig. 6 reports the impactful fraction (sub-seasonal *and* ≥50-percentile shift). We will correct the wording throughout Section 3.3 to say "impactful DtF/FtD transitions" rather than "sub-seasonal," and will state the impactfulness criterion explicitly where these numbers are introduced.

Line 522: "Fig. 5a" should likely be "Fig. 6a."

This will be corrected to Fig. 6a.

Lines 549–552: this sentence is very long and difficult to read.

We will split this into two shorter sentences.

Line 569: is the 24.0% a national median?

It is the overall fraction of DtF transitions (across all sites, n=15,393) with a co-occurring dry-to-wet meteorological transition in the same or adjacent season, not a site median. We will clarify the wording.

Lines 667–668: Please clarify how Figure 6 and Figure S10 support the claim that FtD transitions within 30 days are mostly impactful. Fig. 6a,b,c shows the fraction of impactful events as a function of watershed climates, land use, and hydrologic regions, and there is no classification of transition time.

The reviewer is correct that Fig. 6 does not resolve transition time for impactful events. We additionally used impactful in this sentence in the general context rather than specifically as we have defined it related to transitions in this manuscript. We will remove the overreaching claim and use clearer language: "Although few FtD transitions occur within 30 days (Fig. 3f), sub-seasonal FtD transitions often coincide with large percentile declines (Fig. S10) and are likely disruptive for water management."

Line 701: Remove redundant "may."

This will be corrected ("may may be" → "may be").

Lines 738–740: Please briefly describe the Huge Creek watershed for readers unfamiliar with it, such as whether it is urbanized or experienced major modifications in early 2009.

We will add a brief description of Huge Creek near Wauna, WA, a small, partially developed coastal Puget Sound basin, and note that we are not aware of a discrete major modification in early 2009 beyond the direct impacts of these hydrological extremes. The rapid FtD behavior reflects its flashy, low-storage urban character rather than a specific event.

Lines 824–825: clarify how observational case studies "align with projections," since case studies are historical while projections refer to the future.

We will rephrase to avoid conflating the two: the historical case studies are *consistent with the direction of change* anticipated by projections of more frequent compound extremes, rather than "aligning with" future projections directly.

Lines 827–830: Are you comparing DtF (dry-to-wet) versus WtD (wet-to-dry)? The current sentence reads like comparing fixed vs variable percentile, or correlation between FtD and WtD.

The intended comparison is across transition types. DtF/dry-to-wet transitions show the most consistent streamflow-percentile responses, whereas FtD and meteorological wet-to-dry transitions correspond weakly, especially in fixed-percentile space. We will rewrite the sentence to make the transition-type comparison explicit and to remove the ambiguity with the fixed/variable distinction.

Line 895: Be more specific about which event characteristics are most sensitive to event definitions.

We will revise this to specify that absolute event and transition counts are highly sensitive to definitions, whereas the DtF-over-FtD asymmetry in the fraction of short-term (<30-day) transitions is robust across definitions.

Figure 1b: Add a description of the blue and red shading, similar to the Figure 2 caption.

We will add this. The caption will identify the blue and red lines as the drought and flood thresholds, respectively, consistent with Fig. 2.

Figure 2: Clarify the use of "rapid and sub-seasonal transitions." Some events have transition times >14 but <30 days (e.g., 20–54 days), which are short-term.

Agreed. We will use "short-term and sub-seasonal" in the Fig. 2 caption to accommodate transitions in the 14–30-day (short-term) and ≤90-day (sub-seasonal) ranges, consistent with our defined categories.

Figure 2: Some panels appear to show drought and flood overlapping at the same time. Explain how the same flow value can be classified as both.

The apparent overlap arises because drought is identified in variable (day-of-year) percentile space while floods are identified from daily flow against a fixed/POT threshold, and because the lenient definitions use different smoothing; a single day can therefore fall below the seasonally variable drought threshold while a short daily peak exceeds the flood threshold. Pooling can additionally contribute as droughts or floods can be pooled across short term above threshold gaps of the opposite extreme, particularly when comparing lenient and severe event types. We will add a note to the caption and text explaining that these overlaps reflect the combination of variable-percentile drought and daily-flow flood criteria (this is the "complexity of combining variable and fixed measures" referenced at line 377) and the impacts of pooling.

Figure 2 / Lines 413–415: Clarify "only use drought percentile thresholds." Also explain why one site uses flood stage rather than flow, in the main text/methods.

"Only use drought percentile thresholds" means that for the Guadalupe River we plot the drought thresholds but do not apply the flood identification (POT) or the pooling/minimum-duration criteria, because the record is too short (2000–2025) for our 40-year framework. We will clarify this in the caption. We will also move the explanation that this site is shown in stage (ft) rather than flow, because the extreme July 2025 peak is best represented by the available stage record, into the main text/methods rather than only Fig. S13.

Figure 4b: If boxes are sorted by median, please state this.

We will add this. the caption will state that boxes are ordered by median value.

Figure 6: Some boxes appear truncated before the 1.5×IQR range, while the caption says only outliers outside 1.5×IQR are omitted. Please check. Same in Figures S8–S9.

The plotting window limit is set to 1.5 times the IQR range of all regions/groupings, so individual groupings with more extreme values than the average across all groups extend beyond these limits. Outliers outside this range are omitted and whiskers are truncated at this limit. We will modify the captions of these figures to represent this clipping.

Figure 6: Consider using abbreviations for hydrologic regions or introducing region numbers/names in the main text.

We will introduce the twelve region numbers and names in the main text (Section 2.3) and will add abbreviated region labels to Fig. 6 so readers need not consult the supplement.

Figure 7 caption: Browner or bluer colors indicate sign, not necessarily more/smaller extreme changes. Also, if negative values indicate drier-than-average conditions, the color should be brown rather than blue.

The color scale encodes the magnitude of the percentile change relative to normal, not the sign of the flow percentile change. The direction of change is fixed by the transition type in each panel (an increase for dry-to-wet and DtF; a decrease for wet-to-dry and FtD), and the color shows how large that change is relative to the site's median change in that direction across all periods: browner values indicate a more extreme change than normal, bluer values a smaller change. For a wet-to-dry panel, for example, a browner point denotes a more extreme decrease than the site typically experiences, not a wetter condition. We agree the caption was unclear and will revise it to state this explicitly, and we will switch to a diverging palette without drought/flood color connotations to avoid the brown/blue scheme being read as dry/wet.

Additional caption text: The diverging color scale shows the magnitude of each transition's percentile change relative to the site's median change across all periods ([color A] = more extreme than normal, [color B] = smaller than normal); the direction of change is set by the transition type and color does not indicate wetter or drier conditions.

Figure S1: Make the figure-panel order consistent across percentile thresholds.

This will be corrected. Panel ordering will be made consistent (starting from the same inter-event/minimum-duration combination) across the 10%, 15%, and 20% thresholds.

Figure S2 caption: The first semicolon should be a colon, and "food" should be "flood."

Both will be corrected.

Figure S5: Add a space in "Fig.3definitions." "Fig. A4" should be "Fig. S4." Same issue for Figure S8 caption.

This will be corrected in both captions.

Figure S8/S9: The sentence referring to region names in Fig. A4 is redundant with the final caption sentence.

We will remove the redundant reference.

Figure S9: The red horizontal dotted lines are difficult to read.

We will increase spacing between boxes and thicken/recolor the median reference lines to improve legibility.

Figure S10: The phrase "transitions relative and 90 preceding days relative" appears broken.

This will be corrected. The caption will read "...during the transition season and the 90 preceding days, relative to the median percentile change across all periods."

Figure S11: In the third sentence, "x-axes" should be "y-axes."

This will be corrected.

Figure S13: Does this site satisfy the ≥8 of 10 complete years criterion (line 201)? Also the reference period 2000–2020 conflicts with the stated 1981–2020 (line 211).

The Guadalupe River gage (record beginning 2000) does not meet the 40-year selection criteria and is therefore *not* part of the 3,219-gage sample. It is shown only as an illustrative case study, using a 2000–2020 reference period. We will state this exception clearly in both the caption and at line 211, noting that all analyzed (sampled) gages use the 1981–2020 reference period, while this single illustrative site necessarily uses a shorter reference period.

Lines 1098–1100: The paper seems published in 2017 not 2021, and the doi is not found (should be 10.1038/nclimate3225).

Thank you for catching this. We will correct the citation year and DOI in the reference list and verify the reference against the published record.

Text S1: The sentence "Our flood identification using a peaks-over-threshold approach calibrated to yield approximately one event per year represents a middle ground

between overly restrictive and overly permissive flood definitions" is difficult to read. Consider breaking it into two shorter sentences.

We will split this into two shorter sentences for readability.