egusphere-2025-2982: Author response to the Editor and Referee #1 (minor revision)

Dear Dr. Kheyrollah Pour,

Thank you for the positive decision of publish subject to minor revisions and for forwarding the second-round evaluation from Referee #1. We are grateful to the referee for a careful reading on our study. We have addressed every point, and we summarize the principal changes below. A point-by-point response follows, with the referee's comments in black and our responses in blue.

On behalf of all authors,

Abolfazl Jalali Shahrood

egusphere-2025-2982: Responses to Anonymous Referee #1 20 July 2026

We thank the referee again for the time and care taken with our manuscript. In the following, reviewer comments are in black and author responses are in blue. The principal changes are as follows. (i) We state explicitly that each water year is named for its starting calendar year. (ii) We clarify the DVD connectivity rule, add the selection/tie convention, and confirm by a per-year check that the discharge-stability interval is a single continuous period in all 59 water years. (iii) We correct the treatment of the DSO threshold sensitivity: DSE and its earlier trend are robust, but DSO, DSP, and all DSO-derived quantities are now treated as method-dependent and interpreted qualitatively, and the DSP trend is reported as indicative because it does not persist under an alternative variability scale. (iv) We document the snow choices (trace recoded to zero; presence threshold; interpolation) and correct the number of interpolated days to 35, and we show that the one significant snow trend is insensitive to these choices. (v) We expand the statistical workflow, add Benjamini–Hochberg false-discovery-rate control with reported q-values, and treat multiple tests and early–late correlation differences (a permutation test). (vi) Figure S2 is now presented as operational context, not validation. The underlying dataset and detection algorithms are unchanged. The link in "Appendix A" pointed to a review-only repository that was included inadvertently and is not a permanent public archive. We have removed Appendix A from the manuscript; the repository remains accessible via the same link so that the referee can consult it during this review round if needed. As the link was included in the preprint in error, we intend to restrict access to the repository once the review is complete, subject to the editor's and referee's agreement. The availability statement has accordingly been revised into a "Code and data availability" section: the raw air-temperature, snow-depth, and discharge records remain publicly available from the FMI and SYKE portals (as cited), while the RiTiCE analysis code and the processed water-year input series are available from the corresponding author upon reasonable request.

Point 1 - Water-year labelling

RC1: p 5, lines 123–125: The record is described as 1 October 1966 to 30 September 2025 but labelled WY1966–WY2024. Please state explicitly that each water year is named for its starting calendar year (for example, WY1966 = 1 October 1966 to 30 September 1967).

Response: We have added this statement to Section 2.1 Line: 109-110: "Each water year is named for its starting calendar year; for example, WY1966 spans 1 October 1966 to 30 September 1967, and the full record WY1966–WY2024 covers 1 October 1966 to 30 September 2025."

Point 2 - Connectivity rule for DSO and DSE

RC1: p 10, lines 196–209; p 18–19, lines 335–340: As written, there is no requirement that the second run reaches DSE, overlaps the first run, or represents the same continuous period whose end was used to define DSE … the interval calculated as DSP = DSE − DSO could contain one or more unstable gaps … Please clarify the actual connectivity rule used in the implementation … including how ties between candidate runs are handled. Please also provide pseudocode or a worked water-year example showing the two thresholds, the resulting binary masks, the selected runs, and the final DSO and DSE. If no connectivity constraint is imposed, DSP should instead be described as an interval between two independently detected transition markers, and the interpretation of DSP and its derived analyses should be reconsidered accordingly.

Response: The referee is correct that the implementation does not impose connectivity as a hard constraint: DSE is the end of the longest stable run under the full-water-year threshold, and DSO is the start of the longest stable run under the stricter pre-DSE threshold, detected separately. We therefore tested whether the two runs are in fact connected, documented the tie rule and the algorithm, and reconsidered the interpretation of the DSO-derived quantities (Point 3). Specifically:

(a) Tie handling. When several runs share the maximum length, the earliest run is selected, because the run detector updates its choice only for a strictly longer run. This is now stated in Section 2.2 Line: 169-170.

(b) Connectivity is empirically exact in this record. Because the pre-DSE variability scale is stricter than the full-water-year scale, any day-to-day change that is stable under the pre-DSE threshold is also stable under the full-water-year threshold. We checked, for every water year, whether any day-to-day change between DSO and DSE exceeds the full-water-year threshold. In 0 of 59 water years is there any intervening unstable day. DSO therefore always lies within the same contiguous full-water-year stable run whose end defines DSE, so DSP = DSE − DSO is a single continuous discharge-stability period throughout this record, which is consistent with the description of DSO and DSE in Table 1.

(c) This connectivity is now confirmed diagnostically (a per-year counting check, added to the detection code and referenced in Section 2.2) rather than imposed as a constraint; the detector cannot silently fail on a disjoint year, and no disjoint year occurs here.

(d) Pseudocode. To keep the manuscript methods concise, we did not add the pseudocode to the paper, but we provide it here for the reviewer. std is the population standard deviation and longestRun returns the earliest longest run of stable days:

```
Algorithm 1  DVD detection of DSO and DSE for one water year
Input : daily discharge Q[1..365]
dQ[t] = Q[t+1] - Q[t]                  for t = 1..364     # forward differences
sigma_full = std(dQ)                                      # full-WY variability scale
mask_full[t] = 1 if |dQ[t]| <= sigma_full else 0
(rs, re) = longestRun(mask_full)                          # earliest longest run of 1s
DSE = re + 1                                              # end day of full-WY stable run
dQpre = dQ[1..DSE-1]                                      # differences before DSE
sigma_pre = std(dQpre)                                    # freshet-excluded scale
mask_pre[t] = 1 if |dQpre[t]| <= sigma_pre else 0
(ps, pe) = longestRun(mask_pre)
DSO = ps + 1                                              # start day of pre-DSE stable run
DSP = DSE - DSO
Output: DSO, DSE, DSP
Connectivity diagnostic: n_unstable = count(|dQ[DSO..DSE-1]| > sigma_full) = 0 in all
59 years.
```

(e) Worked example (WY1966, the year shown in Figure 2), giving the two thresholds, masks, and selected runs the referee requested. The full-water-year variability scale is $\sigma\_full = 7.0$ m^3 s^{-1}; the longest full-water-year stable run spans day-to-day changes 55–222, so DSE = 223. Restricting to the pre-DSE differences gives the stricter scale $\sigma\_pre = 2.6$ m^3 s^{-1}; the longest pre-DSE stable run begins at difference 57, so DSO = 58. Hence DSP = 165 days, and [DSO, DSE] lies entirely within the single full-water-year stable run. Figure 2 shows the binary masks and selected runs for this year.

(f) In the manuscript we added the reader-necessary items only: the reason two thresholds are used (the full-water-year variability is dominated by the spring freshet, so a single threshold can place DSO implausibly early), and the connectivity rule with tie handling (Section 2.2). The freshet reasoning is also the mechanistic explanation for the DSO scale sensitivity addressed next.

Point 3 - DSO threshold sensitivity and its propagation

RC1: p 37, lines 650–660: The newly reported threshold sensitivity is too large to be treated only as uncertainty in the absolute timing of DSO … The manuscript should first clarify whether the reported 91- and 7-day displacements are medians of signed differences, median absolute differences, or another quantity, and should show their distributions across the 59 water years. … DSO is used directly to calculate DSP, the DSO–CSO and DSO–CCT CVIs, onset-group ordering frequencies, early–late comparisons, and several PCT and IVI correlations … Robustness of DSE therefore cannot, by itself, demonstrate robustness of the reported DSP shortening. Similarly, a median DSE displacement of 7 days does not establish that its Sen slope and PW-MK significance remain unchanged.

Response: We accept this in full and have restructured the treatment.

The 91 and 7 days are medians of signed differences (IQR/1.349 minus SD). Across the 59 water years, DSO shifts later in 98% of years (median +91 d, mean +86 d, range −45 to +179 d) and DSE shifts earlier in 100% of years (median −7 d, mean −10.5 d, range −1 to −42 d). As the displacements are almost unidirectional, the signed and absolute medians coincide (both 91 and 7 d). The full distributions are shown in Figure R1 below (provided for the referee rather than added to the paper).

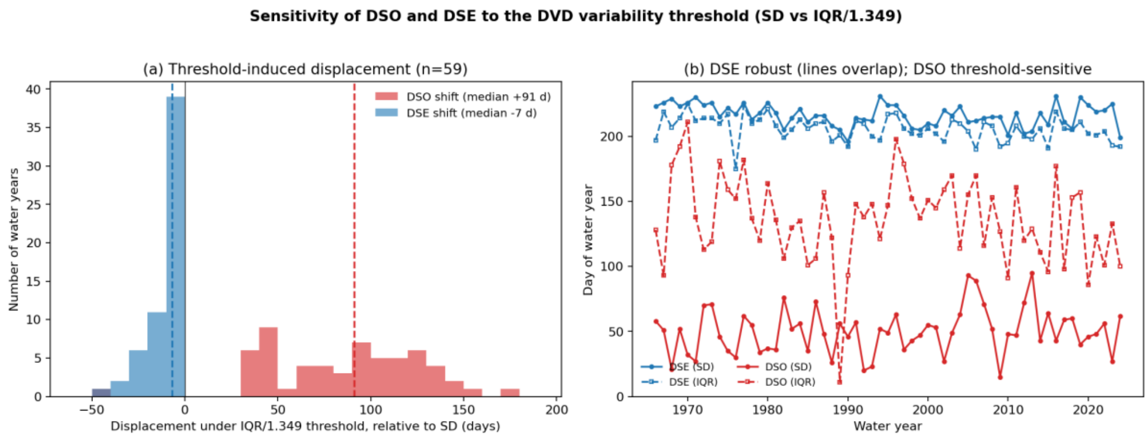

Figure R1. Sensitivity of the discharge timings to the DVD variability scale. (Left) Distributions across the 59 water years of the signed displacement (IQR/1.349 minus SD) in DSO and DSE. (Right) Per-water-year DSO and DSE under the SD-based and IQR-based thresholds.

Trends under the alternative scale. We now report the Sen slope and PW-MK significance of DSO, DSE, and DSP under both scales:

Series	SD threshold: Sen (d yr^{-1}), p	IQR/1.349: Sen (d yr^{-1}), p	Assessment
DSE	−0.182, p=0.023	−0.235, p=0.001	Robust (significant, same sign, both scales)
DSO	+0.154, p=0.281	−0.393, p=0.133	Not significant under either scale; sign not stable
DSP	−0.333, p=0.041	+0.200, p=0.574	Not robust: sign flips and loses significance

Consequently, we no longer claim the sensitivity "has little bearing on our conclusions," and we no longer defend DSP shortening on the basis of DSE robustness. In the revised manuscript we (i) treat DSE and its earlier trend as robust; (ii) withdraw DSP as a quantitative trend and report it as indicative (Sections 3.4 and 4); and (iii) flag all DSO-derived quantities - the DSO–CSO and DSO–CCT offsets, the DSP-window discharge average, the onset-group ordering frequencies involving DSO, and the DSO-based correlations - as method-dependent, to be read qualitatively. The revised limitation text reads: "DSE and its earlier trend are robust to this choice, whereas

DSO, DSP, and all DSO-derived quantities are method-dependent and are interpreted qualitatively rather than as quantitative trend estimates; in particular, the DSP trend does not persist under the alternative scale." Lines: 517-519.

What does not depend on the scale, and is retained is that the qualitative onset-group result (DSO, CSO, and CCT have flexible interannual ordering with no coherent onset trend) and the release-group invariance (CWT→DSE→CSE in all 59 years).

Point 4 - Snow metric: presence threshold, accuracy, trace, interpolation

RC1: The snow metric has a related sensitivity problem: a 1 cm presence threshold is used for observations with stated accuracy of ±2 cm, while isolated missing days are interpolated and trace values are set to zero.

Response: We have documented these choices and shown that the one significant snow trend is insensitive to them (Section 2.5).

Please note that the FMI records snow depth in whole centimeters, so the values are integers and the resolution of the data is 1 cm. The ±2 cm is the measurement error bar, which means how close a reported value is to the true depth. For example, a recorded value of 10 cm has ±2 cm error bar.

Trace (−1 → 0). The FMI code −1 denotes a trace: snow present but below the 1 cm measurement resolution. Trace occurs on 9316 days (about 43% of the record), outside the winter snowpack season. For a continuous snow cover metric, recording trace to 0 is the conservative choice; assigning a positive depth would spuriously extend the continuous snow period.

Interpolation counts corrected. The previous text stated "12 missing days." On re-checking the raw FMI series against a complete daily calendar, the correct figure is 35 days - 12 days with missing values on recorded dates plus 23 absent calendar dates - over the WY1966–WY2024 window. The text now reports 35 and the breakdown. We additionally corrected a single spurious sub-1 cm interpolated snow value (WY2013) so that snow presence ($S \neq 0$) matches the 1 cm reporting resolution; this shifts only CSO and CSP derived quantities by less than 0.01 units and changes no conclusion. We thank the referee, as the earlier number understated the gap-filling.

CSO and CSE are defined by the longest continuous run of snow presence and are therefore governed by the sustained winter snowpack (tens of centimetres), not by ±2 cm noise around the 1 cm mark. Raising the presence threshold from 1 cm to 2 or 3 cm keeps the only significant snow trend - the earlier shift in CSE - significant and shifts the detected dates by a median of 0–1 day, while CSO remains trendless:

Presence threshold	CSE: Sen (d yr^{-1}), p	CSO: Sen (d yr^{-1}), p
1 cm	−0.118, p=0.038	+0.128, p=0.247
2 cm	−0.118, p=0.044	+0.057, p=0.381
3 cm	−0.133, p=0.035	+0.106, p=0.191

Point 5 - Full documentation of the statistical workflow

RC1: The statistical workflow also needs fuller documentation and a more cautious treatment of multiple tests and early–late differences in correlation.

Response: Section 2.6 now states, in order: the trend test (Mann–Kendall), the AR(1) pre-whitening (PW-MK; Von Storch, 1999), the significance level, the slope estimator (Sen's slope; Theil, 1950; Sen, 1968), the correlation method (Spearman ρ on paired water-year values with p-values), the early/late split with sample sizes, the Benjamini–Hochberg false-discovery-rate control (Point 6), and a permutation test for early–late correlation differences (Point 7).

Point 6 - Multiple testing

RC1: … a more cautious treatment of multiple tests …

Response: We now compute Benjamini–Hochberg q-values within predefined families (the six PCT trends, the three IVI trends, the six CVI trends, the three seasonal air-temperature trends, the degree-day trends, and the hydroclimatic extreme-value trends) and report the q-value alongside the uncorrected p-value. Our main conclusions rest on the trends that are individually strong and survive correction; trends that are nominally significant (uncorrected $p < 0.05$) but fall just above the corrected threshold ($q > 0.05$) are reported with their q-values and described as nominal. The within-family outcome is: No extreme-value metric survives the correction (the strongest, annual minimum air temperature, has uncorrected $p=0.0498$ and $q=0.356$).

Thus the backbone of the study - the later CCT and earlier CWT, the contracting CDP, the earlier DSE, and the significant warming in all three seasonal windows - survives correction. The nominal trends (CSE, the two CVI offsets, DSP, and TDD start–CWT) are reported with their q-values and interpreted with due caution; DSP is additionally withdrawn on the independent grounds of its threshold sensitivity (Point 3). We note that the CVI offsets do not survive within the six-member CVI family, so the widening/narrowing offsets are presented as nominal rather than as established trends. We additionally applied the Benjamini–Hochberg correction within each Spearman correlation matrix; every correlation interpreted in the study remains significant after correction, so the reported correlation structure is not an artefact of multiple testing.

Point 7 - Early–late correlation differences

RC1: … a more cautious treatment of … early–late differences in correlation.

Response: We tested every early–late correlation difference with a permutation test: the 59 paired water-year values were pooled and randomly reassigned to the early and late periods 20,000 times to build the null distribution of the difference in Spearman ρ. With about 29–30 water years per sub-period these comparisons have low power, and none of the 10 correlation-change claims is significant (the closest, CSO–CCT, gives $p \approx 0.07$). We have removed the "strengthening," "weakening," and "structural decoupling" language from Sections 3.3 and 4 and now present the full-record correlation structure. We describe the early–late differences as descriptive only. The DSO–DSE late-period negative correlation is reported as nominal and is no longer interpreted as decoupling.

Point 8 - Figure S2 as operational context

RC1: The supplementary material is useful, especially the example years, but Figure S2 should be presented strictly as operational context rather than validation.

Response: We agree. Section 2.5 now describes the ice-thickness and water-temperature records as operational context, explicitly "not as validation of the proxy timings," and the Figure S2 caption now ends: "These records provide operational context only and are not a validation of the detected PCTs." The Discussion already framed them as "plausibility checks rather than validation," which is retained.

Point 9 - Residual freeze-up / break-up terminology

RC1: … the text now generally avoids equating DSO and DSE with visually observed freeze-up and break-up …

Response: We thank the referee for acknowledging the improved terminology. As a final check we swept the manuscript for any remaining equation of DSO/DSE (or CSO/CSE) with visually observed freeze-up/break-up. Every remaining mention of freeze-up/break-up is either a literature reference or an explicit statement that the proxy timings are not those events; Table 1 and the opening of Section 2.2 make the distinction explicit. No further wording change was required.

We believe these revisions resolve the robustness and documentation concerns transparently while preserving the findings that are genuinely robust - the asymmetric, release-group-coherent shift driven by CCT and CWT, the contracting CDP, the earlier DSE, and the qualitative onset-group flexibility - and appropriately down-weighting DSO, DSP, and the early–late correlation differences. We thank the referee for comments that have measurably improved the manuscript.

On behalf of all authors,

Abolfazl Jalali Shahrood