MS No.: egusphere-2026-973:
Answers to the Editor and Referees

Original Referee comments are in italics, **authors' response is in bold**.
All revisions described in this response have been fully implemented in the clean revised manuscript, and all line-number references correspond exclusively to the clean revised manuscript.

Referee #2: Dr. Jitao Zhang

Authors' response: We appreciate the referee's comment and thank them for their constructive feedback. We have carefully addressed the points raised and revised the manuscript to enhance its clarity and overall quality. All suggested changes will be included in the revised version.

Major comments:

The drought threshold reference period must be clarified. It is unclear whether drought thresholds are calculated separately for each dataset and period, or whether historical thresholds are fixed and then applied to future simulations. This is a critical methodological point. If future thresholds are recalculated using the future climate period, the analysis identifies droughts relative to future climatology rather than changes relative to present-day conditions. The authors should explicitly state the threshold strategy and justify it.

Authors' response #1: We thank the reviewer for raising this important point regarding the choice of drought threshold reference period. We agree that the drought threshold reference period must be stated explicitly, because the interpretation of future drought changes depends on whether thresholds are recalculated for the future period or kept fixed at the historical baseline. We have clarified this in Section 2.3, lines 181–186. Specifically, we now state that values were standardized relative to the 1971–2000 reference period and that the percentile threshold (τ) used for drought identification was calculated from the corresponding normalized historical baseline. The same historical threshold was then applied to both the historical and future periods for event identification. We also added the justification that this fixed historical-threshold strategy was chosen because the objective of the study is to quantify changes in drought characteristics relative to historical climatological conditions. Thus, the projected drought changes reported in the manuscript are interpreted relative to the historical baseline rather than as drought anomalies recalculated relative to future climatology.

The drought classification scheme is not sufficiently reproducible. The seven drought types are central to the paper, but the classification rules are not described in enough detail. The manuscript should provide a clear algorithm, ideally as pseudocode or a decision table. Important details include the definition of "proximal rainfall deficit," *the six-pentad window, the treatment of overlapping conditions, the priority order among drought types, the calculation of SWE anomalies, and the handling of grid cells with no snow or very limited SWE.*

Authors' response #2: We thank the reviewer for this important comment. We agree that the seven drought types are central to the study and that the classification procedure should be described in a more reproducible way. We have therefore revised Section 2.4 by adding a stepwise decision table that translates the drought-typology workflow shown in Fig. 2 into explicit event-level classification rules. The added decision table specifies how each runoff drought event is assigned to one of the seven drought-generation types. In particular, we now define a proximal rainfall deficit as a precipitation-deficit event occurring within a six-pentad moving window, including the current pentad and the preceding five pentads, during the first half of the runoff drought event. The table also clarifies the priority order among drought types: rainfall-driven mechanisms are evaluated first, and if no proximal rainfall deficit is detected, the event is then evaluated using the snow-related criteria.
The revised table further specifies the seasonal and physical criteria used in the classification, including June–September for dry-season overlap and summer rainfall deficits, September–October sub-zero temperature conditions for rain-to-snow transitions, and November–March for snow-season runoff deficits. We also clarified the SWE-related criteria, including the use of SWE<1 mm to define absent or negligible snow storage and the use of prior snow availability followed by melt to distinguish warm-snow-season droughts from composite droughts. These additions explain how overlapping conditions and grid cells or events with limited SWE are handled, and make the drought classification scheme more transparent and reproducible. The analysis code used for drought-event identification and classification into the seven drought-generation types is publicly available from Zenodo: `https://doi.org/10.5281/zenodo.18479544`. The repository provides the implemented workflow and can be used by interested researchers to understand, reproduce, and apply the drought-classification scheme proposed in this study.

The parameter selection for drought identification appears subjective. The selected values of mit, ml, and tau are said to be chosen by inspecting sensitivity analyses. This is not sufficiently objective. The authors should explain what criterion defines "stable event characteristics" and show whether the main conclusions are robust to alternative parameter choices. The runoff minimum length of eight pentads, approximately 40 days, is especially important because it may exclude shorter but hydrologically meaningful drought events.

Authors' response #3: We thank the reviewer for this important comment. We agree that the selection of (mit), (ml), and (τ) should be described more objectively. We have therefore revised Section 2.3, lines 165–178, to clarify the purpose and interpretation of the sensitivity analysis and to define what we mean by "stable event characteristics".
In the revised manuscript, we now state that the sensitivity analysis was used to evaluate how drought-event characteristics depend on the percentile threshold (τ), the minimum inter-event time (mit), and the

minimum length (*ml*). For each variable separately, we assessed how the number of drought events (N) and mean event duration changed across neighbouring parameter combinations, as shown in Supplementary Figs. S2 and S3. We now define stable event characteristics as parameter ranges in which these statistics changed gradually rather than abruptly with adjacent parameter values, while preserving consistent regional patterns across MED, NEU, and WCE.

We also specifically clarified the rationale for the runoff minimum length of (*ml*=8) pentads. Because runoff (Q) responds to catchment storage and routing processes, very short threshold crossings may represent transient runoff fluctuations rather than sustained hydrological drought events. The sensitivity analysis showed that short minimum lengths retained many short runoff-deficit events, whereas values around (*ml*=7–9) pentads produced comparatively stable regional event characteristics. Therefore, (*ml*=8) pentads, approximately 40 days, were selected as a compromise that focuses on sustained runoff droughts while avoiding excessive fragmentation of dependent drought periods or inclusion of short-lived runoff fluctuations. We have clarified this rationale in Section 2.3 and now refer explicitly to Supplementary Figs. S2 and S3.

4.Historical model biases are large and require stronger treatment. The GCM-driven simulations substantially underestimate event counts for several drought types, especially snowmelt and composite droughts. The argument that future-minus-historical changes are robust because both periods share the same bias structure is not fully convincing. The authors should show model spread, model agreement, and possibly robustness categories, rather than relying mainly on the ensemble mean.

Authors' response #4: We thank the reviewer for this comment. We agree that historical model biases and inter-model uncertainty require stronger consideration. We have revised the analysis by adding GCM robustness (Fig. R1) and inter-model spread assessments (Fig. R2) in addition to the ensemble mean changes. The robustness analysis identifies areas where at least 4 of 5 GCMs agree on the direction of projected change, while the spread analysis quantifies uncertainty in the magnitude of changes using inter-model variability. We have also revised the Discussion to emphasize that future-minus-historical comparisons allow robust assessment of relative changes but cannot fully account for structural model biases, especially for snow-related and composite drought types.

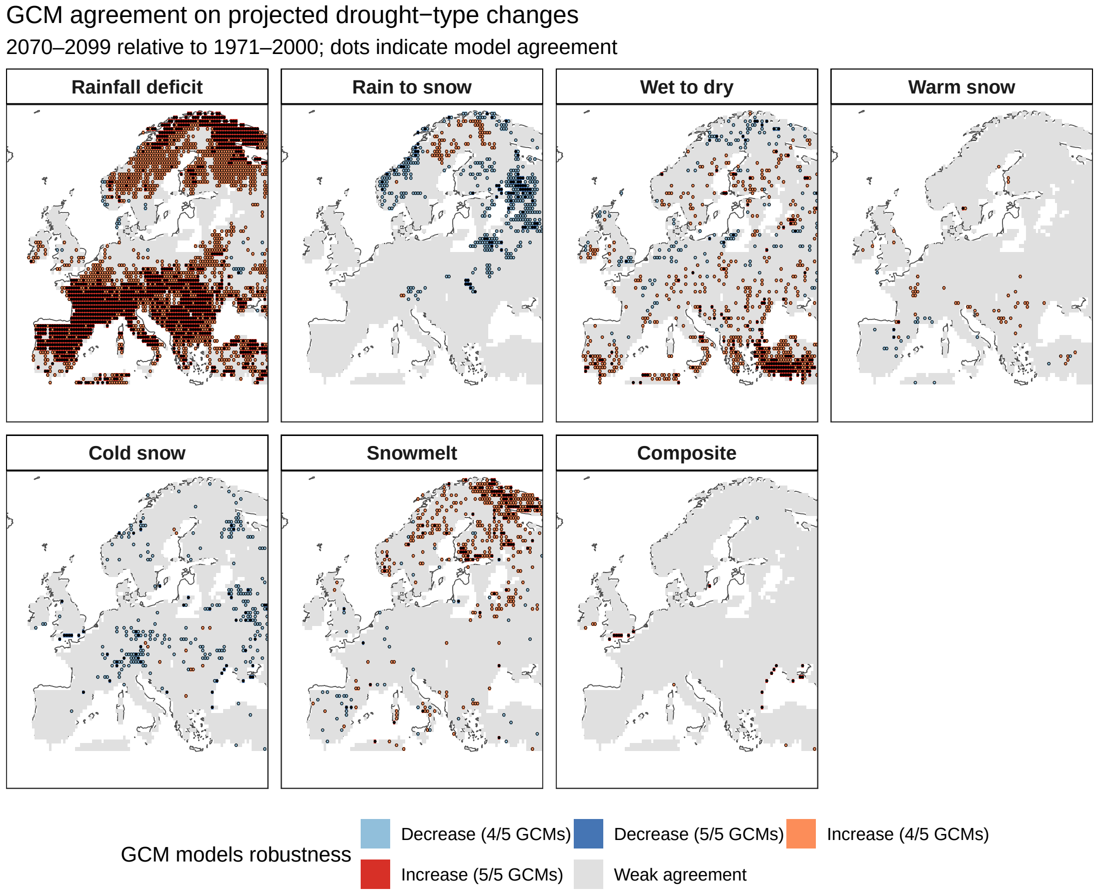

Figure R1: Spatial distribution of GCM agreement in projected drought-type changes. Colours indicate the direction of projected change, with robust increases/decreases defined as agreement among at least four of five GCMs. Dark shades represent agreement among all five GCMs, light shades represent agreement among four GCMs, and grey areas indicate weak agreement (<4 GCMs).

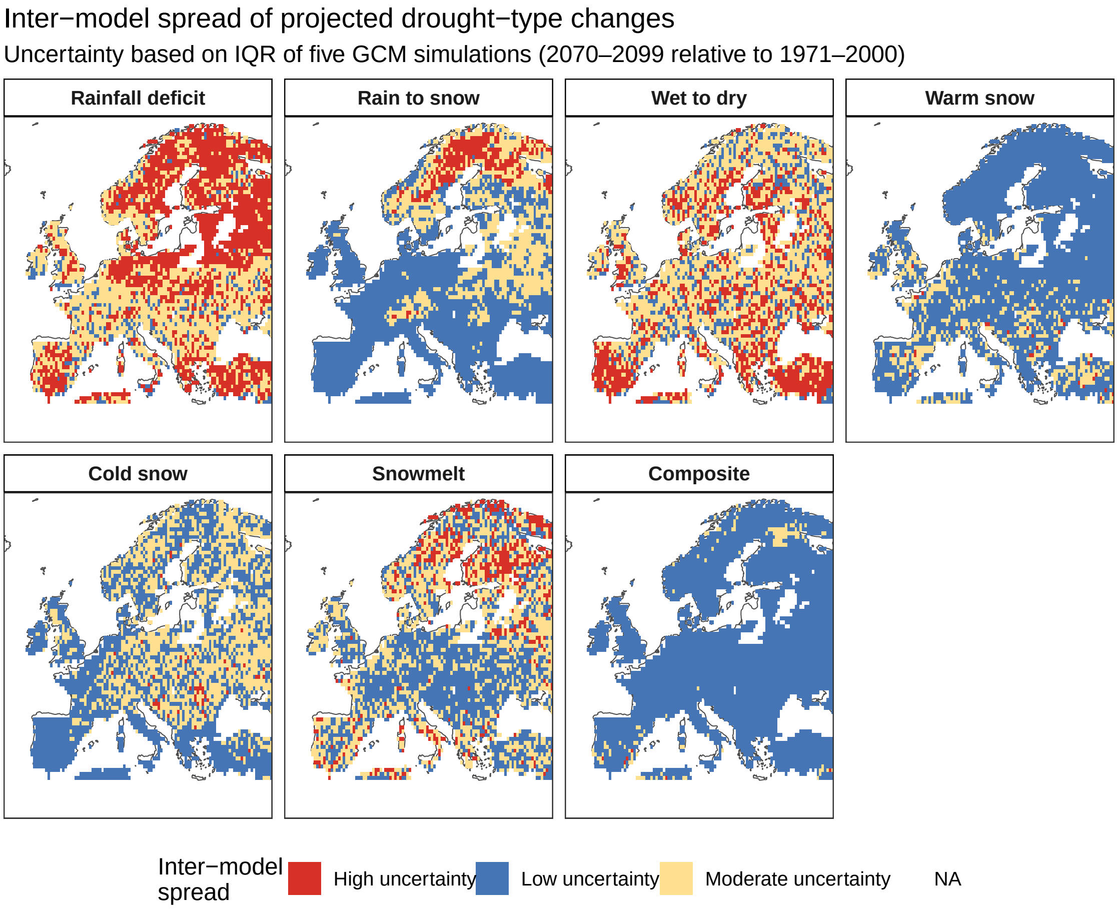

Figure R2: Spatial distribution of inter-model spread in projected drought-type changes. Colors indicate the magnitude of uncertainty among the five GCMs based on the interquartile range (IQR) of projected changes between 2070–2099 and 1971–2000. The uncertainty is classified based on the IQR of model projections: low uncertainty ($\mathrm{IQR} < 25$), moderate uncertainty ($25 \leq \mathrm{IQR} < 75$), and high uncertainty ($\mathrm{IQR} \geq 75$). Blue areas indicate low model spread, yellow areas indicate moderate spread, and red areas indicate high spread, reflecting greater differences among GCM responses. Higher spread indicates larger uncertainty in the magnitude of projected drought-type changes.

The meaning and units of drought severity need clarification. The methods define severity as cumulative deficit, but the results frequently discuss mean runoff deficit in ($\mathrm{mm\ day}^{-1}$)*. The manuscript should clearly distinguish between cumulative severity, mean intensity, runoff deficit anomaly, and absolute change. The sign convention should also be standardized. Presenting deficits as positive magnitudes may reduce confusion.*

Authors' response #5: We thank the reviewer for highlighting this important issue regarding the definition and interpretation of drought severity. We agree that the distinction between cumulative severity, mean runoff deficit, and runoff deficit anomalies was not sufficiently clear in the previous version.

We have revised Section 2.3 to clarify the terminology, units, and sign

convention. In the revised manuscript, drought severity is defined based on the runoff deficit relative to the threshold: $D_Q(t) = Q(t) - Q_\tau(t)$ where $Q(t)$ is the simulated runoff and $Q_\tau(t)$ is the corresponding threshold value. Because drought conditions occur when runoff falls below the threshold, $D_Q(t)$ has negative values, with more negative values representing stronger deficits. Event severity is summarized as the mean runoff deficit ($\overline{D}_Q$), expressed in mm day^{-1}, rather than cumulative deficit.

We have revised the Results and figure captions to consistently refer to this metric as "mean runoff deficit" and clarified that it represents deficit intensity rather than cumulative drought severity. The sign convention has also been standardized throughout the manuscript: negative values indicate stronger runoff deficits, while positive changes indicate weaker deficits relative to the reference period.

Some conclusions appear stronger than the evidence supports. For example, the manuscript states that temperature-driven processes, especially rain-to-snow droughts, show the most pronounced projected changes. However, Table 4 shows large inter-model disagreement for rain-to-snow events, including both decreases and increases. The authors should moderate this claim or support it with model-agreement statistics.

Authors' response #6: Yes, we agree that the previous statement about temperature-driven processes, especially rain-to-snow droughts, was too strong given the inter-model disagreement shown in Table 4. We have revised the manuscript to moderate this conclusion and to distinguish more clearly between individual-model event counts in Table 4 and ensemble-mean drought characteristics shown in Figs. 4–6.

Table 4 reports total event counts for each GCM and shows that rain-to-snow drought counts have mixed responses across models. Therefore, we now describe projected changes in rain-to-snow event frequency as model-dependent and uncertain. We also revised the Results and Conclusions to emphasize robust signals only where model agreement is stronger, such as increasing rainfall-deficit and wet-to-dry droughts and decreasing cold-snow-season droughts. Ensemble-mean changes in mean runoff deficit, duration, and mean number of events are now interpreted together with GCM agreement and inter-model spread (see response to Comment #4) rather than used alone to support strong conclusions.

The Getis-Ord hotspot analysis requires more methodological detail. The manuscript should specify the spatial weights matrix, distance threshold, neighborhood definition, significance levels, treatment of edge effects, and whether multiple-testing correction was applied. Without this information, claims about statistically robust hotspots are difficult to evaluate.

Authors' response #7: We thank the reviewer for this helpful comment. We have revised Appendix A to provide additional methodological details on the Getis–Ord hotspot analysis. Specifically, we now describe the spatial weights matrix, neighborhood definition, absence of a fixed-distance threshold, significance levels, treatment of edge and coastal grid cells, and multiple-testing correction.

Spatial neighborhoods were defined using a symmetric k-nearest-neighbor spatial weights matrix based on grid-cell centroids, with $k = 5$ nearest

neighbors and an adaptive value of $k = \max(2, \min(5, n-1))$ for drought-type groups with fewer valid grid cells. The neighbor list was symmetrized and spatial weights were row-standardized. Thus, no fixed-distance threshold was used. Edge and coastal cells were retained, and their neighborhoods were defined using the nearest available valid land-grid cells without artificial padding outside the European land mask.

Statistical significance was evaluated using Getis–Ord z-scores, and p-values were adjusted using the Benjamini–Hochberg false discovery rate procedure (1) to account for multiple local tests. We also clarified the interpretation of hotspot classes for runoff deficit. Because runoff deficit is expressed as negative values, negative coldspots indicate spatial clusters of increasing runoff-deficit severity, whereas positive hotspots indicate spatial clusters of decreasing runoff-deficit severity. These clarifications have been added to Appendix A of the revised manuscript.

Spatial resolution and regridding procedures need to be described.EOBS is at 0.25°, ISIMIP2b is at 0.5°, and mHM may operate at another resolution. The final analysis grid, interpolation/regridding method, land mask, area weighting, and regional aggregation procedure should be clearly stated

Authors' response #8: Thank you for your suggestion. We have revised Sections 2.1 and 2.5 of the Methods to clarify the spatial resolution, regridding procedure, final analysis grid, land-mask treatment, and regional aggregation approach (lines 110–112 and 224–225). All drought analyses were conducted on a common $0.5^\circ \times 0.5^\circ$ latitude–longitude grid. Although the native EOBS product is available at $0.25^\circ \times 0.25^\circ$, the EOBS precipitation and temperature forcing used in this study were prepared at $0.5^\circ \times 0.5^\circ$, consistent with the ISIMIP2b GCM forcing. The mHM simulations were then forced with these $0.5^\circ \times 0.5^\circ$ precipitation and temperature inputs, and the resulting runoff (Q) and snow water equivalent (SWE) outputs were extracted and analyzed on the same $0.5^\circ \times 0.5^\circ$ grid. Thus, EOBS-driven, historical GCM-driven, and future GCM-driven mHM outputs were all compared on identical grid-cell coordinates.

The final analysis grid consisted of the common European land grid cells available across EOBS-driven and GCM-driven mHM simulations. Ocean cells and grid cells with missing values were excluded. No additional spatial interpolation was applied after the mHM simulations or after drought-event classification; all drought metrics, bias estimates, projected changes, and hotspot analyses were calculated directly at the common $0.5^\circ \times 0.5^\circ$ grid-cell level.

Regional aggregation was performed by assigning each $0.5^\circ \times 0.5^\circ$ land grid cell to one of the three IPCC AR6 European reference regions used in the study: Northern Europe, Western-Central Europe, and the Mediterranean. Regional summaries were then calculated from the grid-cell-level drought metrics within each region. These clarifications have been added to the revised manuscript.

The use of RCP4.5/CMIP5 needs more careful framing.The framework is useful, but the manuscript should avoid overgeneralizing from one scenario and five CMIP5 models. The authors should frame the results as a mid-range CMIP5/ISIMIP2b case study and discuss how CMIP6, ISIMIP3b, or EURO-CORDEX could affect regional *signals.*

Authors' response #9: We appreciate the reviewer's insightful comments. In response, we have clarified the interpretation and scope of our results and added a paragraph to the Discussion section discussing uncertainties arising from the chosen scenario and model ensemble.

Revised statements L453–467: While the focus on RCP4.5 provides a plausible mid-range climate-change pathway, the projected changes should be interpreted as a CMIP5/ISIMIP2b case study based on five GCMs rather than as a full exploration of scenario and model uncertainty. Lower- or higher-emission pathways could alter both the magnitude and spatial extent of drought-type changes, and more severe forcing scenarios may lead to nonlinear hydrological responses not captured by the present ensemble (2). Similarly, CMIP6/ISIMIP3b simulations may modify the magnitude and spatial distribution of regional drought signals because of differences in climate sensitivity, large-scale circulation responses, precipitation projections, and bias-adjustment methods relative to CMIP5/ISIMIP2b. Higher-resolution regional climate model ensembles, such as EURO-CORDEX, may further refine the spatial expression of drought mechanisms, particularly in mountainous, snow-influenced, and coastal regions where topography, snow accumulation and melt, and precipitation-phase transitions are important (3). Thus, the quantitative projections reported here are specific to the selected CMIP5/ISIMIP2b RCP4.5 ensemble, whereas the process-based drought-typology framework is transferable to other scenario and model ensembles. In addition, the framework can be applied to outputs from alternative hydrological and land-surface models, enabling assessment of uncertainties arising from differences in model structure, parameterization, and process representation, including runoff generation and snow dynamics.Finally, the exclusion of anthropogenic influences such as land use change and water withdrawals may limit process realism at regional scales.

Suggestions to result section. 1.Present model uncertainty more explicitly. For each drought type and region, show the ensemble median, interquartile range, and number of models agreeing on the sign of change. 2.Add a robustness criterion, for example: "robust increase" where at least four out of five GCMs agree on the sign of change. 3.Report both absolute and relative changes. Absolute changes are useful for impacts, while relative changes help compare rare and common drought types. 4.Include regional summary tables for MED, WCE, and NEU. These tables could report changes in mean deficit, duration, frequency, and model agreement by drought type. 5.Avoid interpreting ensemble-mean maps alone where individual models differ strongly. For rain-to-snow droughts in particular, model spread should be shown. 6.For drought timing, directly quantify changes in onset month, termination month, and duration rather than inferring duration from heatmap patterns alone.

Authors' response to the suggestion regarding result section
We thank the reviewer for these valuable suggestions. We have revised the analysis to better represent uncertainty and model agreement in future drought projections.
For suggestion 1, we now include ensemble statistics for each drought type and region, including the ensemble median, interquartile range (Q25–Q75), and the number of GCMs agreeing on the direction of projected change. This provides a clearer representation of the spread among mod-

els rather than relying only on the ensemble mean (see response to Comment #4).

For suggestion 2, we have added a robustness classification based on inter-model agreement. A change is considered a “robust increase” or “robust decrease” when at least four out of five GCMs agree on the sign of change. Cases with lower agreement are classified as “mixed or weak agreement”. This criterion is now included in the regional summary analysis (see response to Comment #4).

For suggestion 3, we now report both absolute and relative changes. Absolute changes describe the magnitude of projected changes, while relative changes allow comparison between drought types with different historical frequencies and magnitudes.

For suggestion 4, we have added a regional summary table for the MED region (refer to Table R1) and equivalent tables for the NEU and WCE regions in the revised manuscript. These tables summarize projected changes in drought frequency, runoff deficit magnitude, drought duration, and model agreement for each drought type. The frequency metric is calculated as the mean number of affected grid cells normalized by the total number of grid cells within each region. The runoff deficit magnitude is reported in mm day^{-1}, while drought duration is expressed in days.

For suggestion 5, we have revised the interpretation of the results by considering model agreement and uncertainty rather than relying only on ensemble-average results. In particular, rain-to-snow droughts are now discussed with caution because they show larger inter-model variability and weaker agreement among GCMs (see response to Comment #4 and #6).

For suggestion 6, we have revised the drought timing analysis by directly quantifying changes in drought onset and termination months. An additional column has been added to the heatmap figure showing the difference between GCM_fut (2070–2099) and GCM_hist (1971–2000) for drought onset and termination timing. We have also added text in the manuscript describing these timing changes and their implications for drought persistence.

Modified statement L373-384: To quantify changes in drought timing, we calculated the difference in mean onset and termination months between GCM_fut (2070–2099) and GCM_hist (1971–2000) (Fig. R3, last column). The results show that projected changes are generally small, indicating that climate change mainly affects drought persistence rather than causing major shifts in seasonal timing. The largest timing shifts occur for snow-related droughts, with earlier occurrence linked to warming-driven changes in snow processes, while rainfall-deficit droughts largely maintain their seasonal pattern but extend over longer periods. Overall, the temporal analysis indicates that the seasonal occurrence patterns of European drought events remain broadly consistent under future climate conditions. However, the change analysis reveals regional- and drought-type-dependent shifts in mean drought onset and termination timing, with the strongest responses occurring for temperature-sensitive snow-related droughts. These timing changes, together with projected increases in drought duration and intensity, indicate a tendency toward longer drought persistence under future climate conditions.

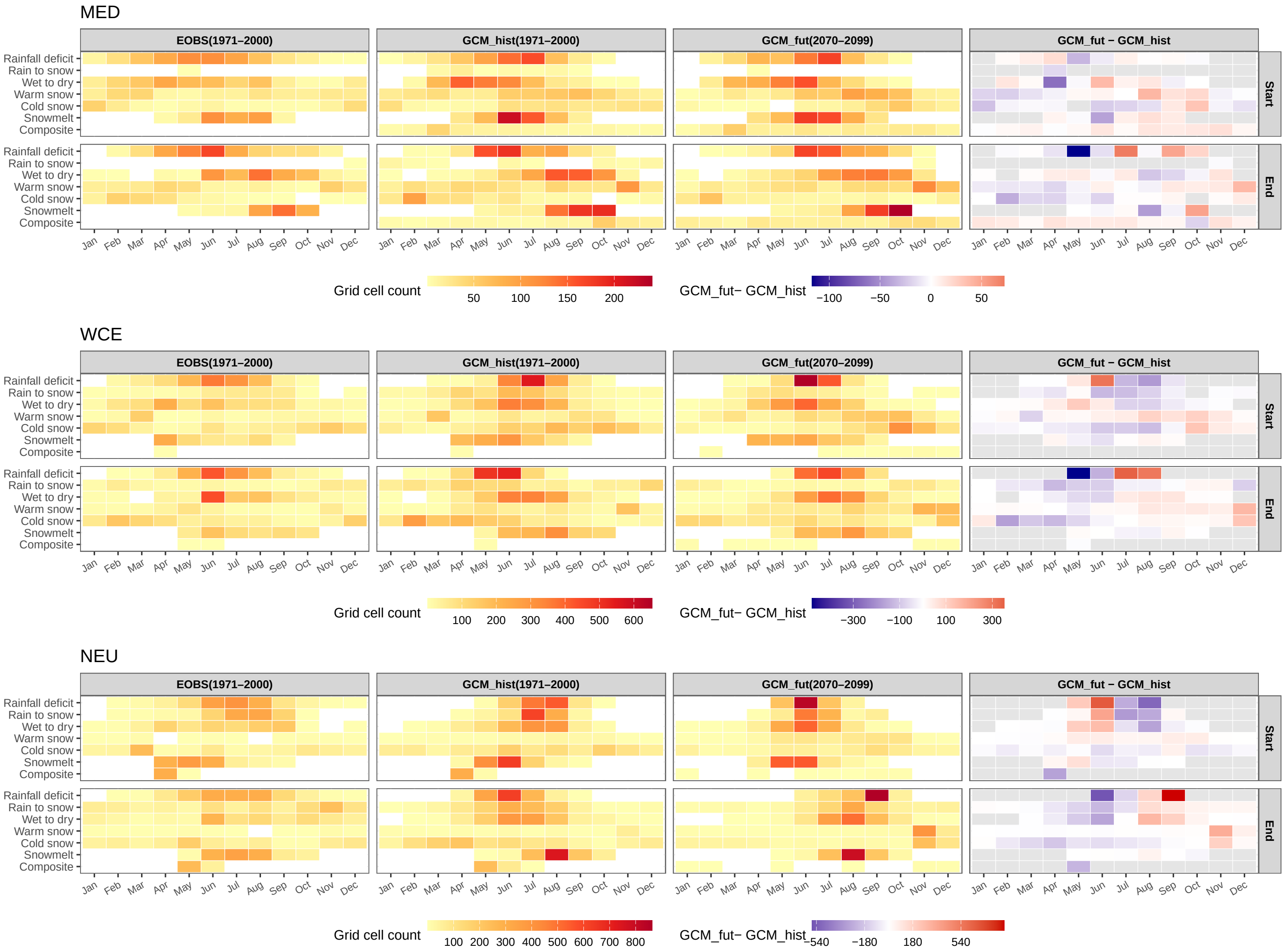

Figure R3: Heatmap representation of (a) drought start and (b) ends across reference (1971–2000) and future (2070–2099) periods under RCP4.5. The additional column shows the projected timing change diff(GCM_fut, GCM_hist) for drought onset and termination. Each tile shows the grid cell count corresponding to the average start or end month for a given drought event type.

Table R1: Historical and future drought typology metrics and projected changes for MED region

Region	Drought type	Metric	Hist. Median	Hist. Q25	Hist. Q75	Future Median	Future Q25	Future Q75
Abs. Change	Rel. Change (%)	Signal						
MED	Rainfall deficit	Mean number of events	0.94	0.92	0.94	0.94	0.90	0.95
0.01	0.00	+						
MED	Rainfall deficit	Mean runoff deficit (mm day^{-1})	-4.31	-4.32	-3.76	-8.96	-13.10	-7.43
-4.64	107.89	-						
MED	Rainfall deficit	Mean duration (days)	132.52	125.36	135.90	205.27	171.17	225.27
69.37	54.90	+						
MED	Rain to snow	Mean number of events	0.00	0.00	0.01	0.00	0.00	0.00
0.00	–	-						
MED	Rain to snow	Mean runoff deficit (mm day^{-1})	-6.54	-8.36	-4.93	-18.94	-24.81	-16.59
-9.53	189.60	±						
MED	Rain to snow	Mean duration (days)	208.89	207.50	232.00	284.00	232.50	292.00
83.87	35.96	±						
MED	Wet to dry	Mean number of events	0.78	0.76	0.87	0.82	0.80	0.84
0.02	5.13	+						
MED	Wet to dry	Mean runoff deficit (mm day^{-1})	-8.65	-8.78	-7.08	-13.21	-18.37	-10.18
-4.56	52.72	-						
MED	Wet to dry	Mean duration (days)	187.20	182.73	206.37	227.27	225.40	243.21
40.08	21.40	+						
MED	Warm snow	Mean number of events	0.26	0.22	0.27	0.28	0.23	0.32
0.02	7.69	+						
MED	Warm snow	Mean runoff deficit (mm day^{-1})	-1.37	-1.64	-1.20	-2.17	-2.30	-1.60
-0.24	58.39	-						
MED	Warm snow	Mean duration (days)	89.27	87.77	100.84	97.47	93.21	112.61
3.94	9.19	+						
MED	Cold snow	Mean number of events	0.17	0.13	0.17	0.11	0.06	0.12
-0.06	-35.29	-						

Region	Drought type	Metric	Hist. Median	Hist. Q25	Hist. Q75	Future Median	Future Q25	Future Q75
Abs. Change	Rel. Change (%)	Signal						
MED	Cold snow	Mean runoff deficit (mm day^{-1})	-1.49	-1.67	-1.47	-1.71	-2.06	-1.61
-0.45	14.77	-						
MED	Cold snow	Mean duration (days)	83.19	82.51	87.91	102.72	101.33	119.23
19.53	23.48	+						
MED	Snowmelt	Mean number of events	0.42	0.42	0.43	0.39	0.37	0.41
-0.01	-7.14	-						
MED	Snowmelt	Mean runoff deficit (mm day^{-1})	-1.87	-2.11	-1.58	-2.23	-2.38	-1.95
-0.47	19.25	-						
MED	Snowmelt	Mean duration (days)	88.83	86.05	100.04	88.30	81.57	102.72
2.25	-0.60	±						
MED	Composite	Mean number of events	0.07	0.00	0.13	0.13	0.00	0.18
0.05	85.71	±						
MED	Composite	Mean runoff deficit (mm day^{-1})	-4.01	-4.94	-2.61	-5.34	-5.58	-3.37
-0.19	33.17	±						
MED	Composite	Mean duration (days)	139.31	116.00	140.43	128.11	108.47	149.87
-3.86	-8.04	±						

Suggestions to discussion section. 1.Discuss more explicitly how threshold choice affects future drought interpretation. This should be a central limitation because threshold methodology can strongly influence projected frequency and duration. 2.Moderate claims about "resilience" or "improvement" in Northern and Western-Central Europe. A decrease in one drought type does not necessarily imply lower overall water-resource risk. 3.Better connect the process-based results to adaptation implications. For example, Mediterranean rainfall-deficit and wet-to-dry droughts imply different management needs than snowmelt or rain-to-snow droughts in northern regions. 4.Discuss the implications of using naturalized/modelled runoff without anthropogenic water withdrawals, reservoirs, irrigation, or land-use change. 5.Expand the discussion of PET uncertainty. Since the manuscript interprets some changes through evaporative demand, sensitivity to PET formulation should be treated as more than a minor limitation. 6.Distinguish clearly between robust findings and exploratory findings. The Mediterranean intensification appears well supported; changes in rare or snow-related drought types are more uncertain and should be discussed accordingly. 7.Reframe the contribution more precisely. Rather than claiming a fully definitive continental-scale assessment, the paper is strongest as a transferable framework for process-based drought typology applied to CMIP5/ISIMIP2b projections over Europe.

Authors' response to the suggestion regarding discussion:
We thank the reviewer for these insightful and constructive suggestions regarding the discussion section. We agree that several of these aspects warranted a more explicit and balanced view in the manuscript. In the revised version, we have expanded the discussion and clarified how threshold choice may influence the interpretation of future drought frequency, duration, and drought-type transitions.

Furthermore, we have strengthened the connection between the identified drought-generating processes and their potential adaptation and management implications across different European regions. In particular, we discuss how drought types driven by rainfall deficits and increased evaporative demand may require different management responses than droughts associated with snow-related processes.

We also agree that the use of naturalized hydrological simulations represents an important limitation. The revised discussion now more explicitly acknowledges the absence of anthropogenic influences such as water withdrawals, reservoir operations, irrigation, and land-use change, and discusses how these factors may modify future drought characteristics. At the same time, we note that the use of naturalized hydrological simulations provides a controlled framework for isolating and assessing the sensitivity of drought characteristics to climate forcing alone. This approach allows us to attribute projected changes primarily to climatic drivers while avoiding the additional uncertainties – and often unknown future trajectory – associated with human water management effects. Nevertheless, we recognize that future drought evolution will likely emerge from complex interactions between climatic and anthropogenic factors, which are not explicitly represented in the current framework. We have clarified this distinction and highlighted it as an important avenue for future research.

In addition, we have expanded the discussion of uncertainties associated with potential evapotranspiration (PET) estimates, particularly given the

role of evaporative demand in the interpretation of several projected drought changes. To provide additional context, we draw upon insights from our recent comprehensive modelling assessment that evaluated the influence of different PET formulations on hydrological simulations (4). While a detailed sensitivity analysis is beyond the scope of the present study, the findings from this work help contextualize the extent to which PET-related uncertainties may affect the simulated drought responses. Finally, we appreciate the reviewer's suggestion regarding the framing of the study. We have revised the concluding discussion to more clearly position the contribution as a transferable process-based drought typology framework applied to CMIP5/ISIMIP2b projections over Europe, while avoiding language that could be interpreted as implying a definitive continental-scale assessment.

References

[1] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).

[2] Lehner, F. *et al.* Projected drought risk in 1.5 c and 2 c warmer climates. *Geophysical Research Letters* **44**, 7419–7428 (2017).

[3] Jacob, D. *et al.* Euro-cordex: new high-resolution climate change projections for european impact research. *Regional environmental change* **14**, 563–578 (2014).

[4] Thakur, V. *et al.* Unveiling the impact of potential evapotranspiration method selection on trends in hydrological cycle components across europe. *Hydrology and Earth System Sciences* **29**, 4395–4416 (2025). URL `https://hess.copernicus.org/articles/29/4395/2025/`.