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 #1:Emanuele Mombrini

Authors' response #1: 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:

I find the discussion of drought identification severely lacking, in a way that affects the legibility of the subsequent results. First, it is not clear to me how the sensitivity analysis for the drought identification parameters is conducted. The supplementary figures by themselves do not offer much explanation, and it is not clear to me in what sense event characteristics should be 'stable'. The minimum length chosen is 1, which is not a value present in the figures themselves, and it is not clear why an ml value outside the sensitivity analysis range would be chosen.

Authors' response #2: We thank the referee for this important comment. We agree that the original description of the sensitivity analysis was too brief and that the supplementary figures did not sufficiently explain how the final parameter values were selected. In the revised manuscript, we have expanded Section 2.3 and the Supplementary Material to clarify the purpose, implementation, and interpretation of the sensitivity analysis. We designed the sensitivity analysis to evaluate how the identified drought events depend on the threshold-level parameters: the percentile threshold τ, the minimum inter-event time (*mit*), and the minimum length (*ml*). For each variable separately, we assessed how key event statistics, especially the total number of detected events, mean duration, and mean severity, changed across parameter combinations. The purpose of the sensitivity analysis was to identify parameter settings that avoid artificial fragmentation of dependent drought events at very low *mit* and *ml* values, while also avoiding excessive merging or filtering of events at overly restrictive settings.
In the revised manuscript, we define "stable" event characteristics as parameter ranges where these statistics, particularly the number of detected events and their mean duration, do not show abrupt changes or systematic shifts when τ, *mit*, or *ml* are varied. The selected parameters, therefore, represent a compromise between retaining physically meaningful drought events and avoiding artefacts introduced by the event-pooling

procedure. This is consistent with the threshold-level approach, in which drought characteristics depend not only on the percentile threshold itself but also on the pooling and event-definition criteria used to derive independent events (1; 2).

We further agree that the presentation of $ml = 1$ for precipitation and SWE was not sufficiently transparent. For these variables, $ml = 1$ was retained because precipitation and snow-water-storage deficits are used here as process indicators rather than as the target drought variable itself, and short-lived deficits in precipitation or SWE can still be relevant for identifying drought-generation mechanisms and drought propagation. By contrast, runoff droughts are typically more persistent and therefore justify a stricter minimum-length criterion. This interpretation is consistent with process-based views of hydrological drought propagation and typology, in which short antecedent deficits in rainfall or snow conditions may still influence the development of runoff droughts (3; 4).

We also agree that this choice was not documented consistently in the original Supplement, because Figures S2–S3 only displayed ml values from 2 to 10. In the revised manuscript and Supplement, we will therefore explain the objective of the sensitivity analysis more explicitly, (ii) define the meaning of "stable event characteristics", and (iii) revise the Supplementary presentation so that the tested and selected parameter values are shown consistently.

It is not clear to me on which data the threshold values are calculated (i.e. if only on E-OBS or also the GMCs) and if the thresholds calculated through the chosen percentile are calculated for each time series. This is problematic since, given the use of percentile thresholds and the change of time period, this opens up questions about which reference period to use for such quantiles. Given the large time passing between 1970 and 2050, climate shifts which would change the underlying climate, and thus what constitutes a drought in statistical terms, are likely to happen. If the historical period is used as a reference for the future projection, this needs to be made explicit since it is fundamental in understanding the results themselves.

Authors' response #3: We thank the referee for highlighting this important point. We agree that the original manuscript did not describe the threshold calculation and the choice of reference period with sufficient clarity. In our analysis, drought thresholds were defined using a fixed historical reference period (1971–2000). Specifically, the variables were first standardized relative to the 1971–2000 baseline, and the drought threshold corresponding to the selected percentile (τ) was then calculated from the normalized values over the same 1971–2000 reference period. This historical threshold was subsequently applied to the full time series for drought-event identification, including the future simulations. Thus, future droughts were identified relative to the historical baseline rather than to a future-period climatology.

We agree that this point is fundamental for interpreting the projected changes and should have been stated explicitly in the original manuscript. In the revised Methods section, we will therefore clarify both the reference period and the threshold calculation logic. This choice follows the threshold-level framework in which seasonal thresholds are defined rela-

tive to a reference climatology and then used consistently for event detection and pooling (1; 2). In our case, retaining a fixed historical baseline allows projected future droughts to be interpreted as departures from historical hydroclimatic conditions rather than as deficits relative to an already shifted future mean state.
To avoid further ambiguity, we will also state directly in the Methods that the threshold was not recalculated from the future period. Instead, the historical 1971–2000 reference was retained when identifying future events, so that changes in drought frequency, duration, and severity can be interpreted against a common baseline across historical and projected conditions.
We added the following sentence for the revised Methods section - 2.3 Drought event identification and drought characteristics: "For each variable, values were standardized relative to the 1971–2000 reference period, and 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. 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. (lines 166–170)."

Another aspect lacking from the methodological discussion is how the timing of droughts is evaluated. Since this aspect is discussed at length in Section 3.4, a description of how this is evaluated is needed. I also find Figure 8 difficult to understand, as it often seems to show drought ending dates earlier than their starting ones (especially the (`GCM_hist`) rainfall deficit droughts). This doesn't seem to be reflected in the discussion of the figure itself, and based on Line 197 I think the start and end panels may have been switched.

Authors' response #4: We thank the referee for this important comment. We agree that the timing analysis was not described clearly enough in the original manuscript, although Section 3.4 discusses the results in detail. In the revised manuscript, we will explicitly state that, for each identified runoff drought event, the onset date was defined as the first pentad of the event and the termination date as the last pentad; the corresponding calendar months were then extracted and summarized by drought type, dataset, and subregion (MED, WCE, and NEU). We also clarify that Figure 8 does not compare paired individual start-end events. Instead, each tile shows the grid-cell count associated with the average onset month ("Start") or average termination month ("End") for a given drought type. Thus, the figure presents aggregated regional timing patterns, with warmer colors indicating months more frequently associated with average event onset or termination. We will revise the Methods text and Figure 8 caption accordingly. Suggested sentence for the revised Methods section: "For each identified runoff drought event, onset (start) and termination (end) were defined as the first and last pentads of the event, respectively. The corresponding calendar months were then extracted and summarized by drought type, dataset, and European subregion to characterize aggregated regional timing patterns."

Bias in the climate models is it discussed, with the assumption that although strong biases are found, comparison between historical and future projection climate data are unaffected by such biases due to both datasets sharing them. While I don't disagree with this argument, I do not feel confident in this assumption, and the manuscript doesn't provide either enough references or reasoning for it. A reference is given to Gurdmundsson and Seneviratne, 2016, at line 326, but it is not clear to me how their work supports the idea of "internally consistent relative changes". They in fact seem to confirm the results from climate models through comparisons with measured data, and write that when "the observational and the model based assessment is conflicting [...] results are inconclusive [.]". I would either find a more focused discussion in support of the author's assumption or make clearer the speculative nature of this argument.

Authors' response #5: We thank the referee for this careful and important comment. The original wording was too strong and could be interpreted as implying that projected changes are unaffected by climate-model biases. This was not our intention. We have revised the manuscript to make clear that comparing future and historical simulations within the same GCM-forcing chain reduces the influence of approximately time-invariant systematic biases, but does not eliminate the possible influence of model bias on the projected changes.
In the revised manuscript, we interpret the projected changes as internally consistent model responses rather than as bias-free model simulations. Future and historical simulations are compared within the same GCM, bias-correction framework, hydrological model, and drought classification procedure. This way, we avoid directly comparing biased future absolute values with observations, and it is therefore appropriate for assessing relative changes. However, we still assume that the bias-corrected GCMs provide a plausible climate-change signal for precipitation, temperature, snow-related processes, and runoff-relevant conditions. We now state this assumption explicitly and discuss it as a limitation, following previous work showing that bias correction cannot correct fundamental climate-model errors or implausible climate-change signals (5; 6).
We have also revised the referencing to (7). Rather than using this reference to justify the methodological assumption of internally consistent relative changes, we now use it only to place our regional drought-change patterns in the context of previous European drought-risk assessments. The relevant manuscript text has been revised as follows: "Although substantial biases remain in absolute drought characteristics, projected changes were assessed within each GCM by comparing future simulations with their corresponding historical simulations. This paired design reduces the influence of approximately time-invariant systematic biases and provides internally consistent modelled change estimates. However, it does not guarantee that projected changes are unaffected by model bias, particularly when biases are state-dependent or influence snow-related, threshold-dependent processes. We therefore interpret the projected changes as conditional model responses rather than bias-free predictions (5; 6)."

Minor comments:

On line 44 Brunner at al., 2021 is cited, but given that this paper cites only briefly the categorization of hydrological drought, Brunner et al., 2022 is a better reference. The same applies at line 81.

Authors' response #6: Thanks, the references were updated as per your suggestions.

The choice of threshold for drought identification is not described clearly: I think stating that the tau value indicates the non-exceedance probability would make this part easier to understand. I would also rephrase line 131 making clear that percentiles are in fact used in this work and not just "typically" used.

Authors' response #7: We have updated lines 131-134 in our manuscript: "In this study, the threshold was defined as a specific percentile (quantile) of the hydrometeorological variable, where (τ) denotes the corresponding non-exceedance probability. This percentile-based approach ensures that drought identification is relative to local climatological conditions rather than based on absolute values."

Figure 2 seems to contain a mistake, as according to it wet to dry season droughts result from deficits outside summer not transitioning to dry periods. It also lacks a connection between rainfall deficits droughts and the "transition to dry period" option. If these are not errors, then an explanation for why the 2022 categorization has been changed is needed.

Authors' response #8: We have modified the figure as per your suggestions.

In Table 3 I would add also the values for the ensemble climate model which biases are showed in Figure 3. The same applies to Table 4 and Figure 4.

Authors' response #9: We thank the reviewer for the suggestion. We have added the ensemble values to Tables 3 and 4. However, the tables and figures show different metrics: Tables 3 and 4 report total drought-event counts aggregated over the study domain/regions, whereas Figures 3 and 4 present gridded ensemble-based mean runoff deficit and its bias/change. We have clarified this distinction in the captions and text.

Table R1: Comparison of drought events between GCMs (RCP4.5) and EOBS (1971–2000), categorised by drought event type. *Number of events* refers to absolute counts in EOBS (N_{OBS}) or historical GCM simulations (N_{hist}). *Fraction of events* (%) indicates the proportion of each drought event type relative to the total event count per dataset. *Relative bias* (%) is calculated as $(N_{\text{hist}} - N_{\text{OBS}})/N_{\text{OBS}} \times 100$, representing the percentage difference between modelled and observed event counts. Negative values indicate underestimation by the model. The ensemble mean was calculated across the five GCMs.

Metrics	**Rainfall deficit**	**Rain to snow**	**Wet to dry**	**Warm snow**	**Cold snow**	**Snowmelt**	**Composite**
EOBS v25 observations							
Number of events	30993	3400	5273	957	2793	3952	475
Fraction of events (%)	64.7	7.1	11.0	2.0	5.8	8.2	0.9
MIROC							
Number of events	22222	2623	5243	682	1914	2110	121
Fraction of events (%)	63.6	7.5	15.0	2.0	5.4	6.4	0.3
Relative bias (%)	-28.3	-22.8	-0.6	-28.7	-31.5	-46.6	-74.6
GFDL							
Number of events	19842	3394	4414	476	1384	1725	85
Fraction of events (%)	63.4	10.8	14.1	1.5	4.4	5.5	0.3
Relative bias (%)	-36.0	-0.2	-16.3	-50.3	-50.4	-56.4	-82.1
HadGEM2							
Number of events	20992	2740	4681	561	2060	1952	96
Fraction of events (%)	63.5	8.3	14.1	1.7	6.2	5.9	0.3
Relative bias (%)	-32.3	-19.4	-11.2	-41.4	-26.2	-50.6	-79.8
IPSL							
Number of events	22502	2395	3848	572	1532	2663	251
Fraction of events (%)	66.6	7.1	11.4	1.7	4.5	7.9	0.7
Relative bias (%)	-27.4	-29.6	-27.1	-40.3	-45.1	-32.7	-47.2
NorESM							
Number of events	22412	1762	3933	458	1490	1674	316
Fraction of events (%)	69.9	5.5	12.3	1.4	4.6	5.2	0.9
Relative bias (%)	-27.7	-48.2	-25.4	-52.1	-46.7	-57.6	-33.5
Ensemble mean							
Number of events	21594	2583	4424	550	1676	2025	174
Fraction of events (%)	65.4	7.8	13.4	1.7	5.1	6.1	0.5
Relative bias (%)	−30.3	−24.0	−16.1	−42.5	−40.0	−48.8	−63.4

Table R2: Comparison of drought events between future GCM simulations (2070–2099, RCP4.5) and historical GCM simulations (1971–2000), categorised by drought event type. *Number of events* refers to absolute counts in future GCM simulations (N_{fut}). *Relative change* (%) is calculated as $(N_{fut} - N_{hist})/N_{hist} \times 100$, representing the percentage difference between future and historical event counts. Positive values indicate an increase and negative values indicate a decrease in event frequency under future climate conditions. The ensemble mean was calculated across the five GCMs, except for the composite category, where the MIROC value was not reported.

Metrics	**Rainfall Deficit**	**Rain to Snow**	**Wet to Dry**	**Warm Snow**	**Cold Snow**	**Snowmelt**	**Composite**
MIROC							
Number of events	29083	1122	7454	1033	312	3454	–
Relative change (%)	+31	-57	+42	+51	-84	+64	–
GFDL							
Number of events	22501	897	5211	651	825	1975	5
Relative change (%)	+13	-74	+18	+37	-40	+14	-94
HadGEM2							
Number of events	45400	2898	6328	1251	1475	3663	197
Relative change (%)	+116	+6	+35	+123	-28	+88	+105
IPSL							
Number of events	43787	1627	6887	940	385	3621	375
Relative change (%)	+95	-32	+79	+64	-75	+36	+49
NorESM							
Number of events	48232	4205	4259	1222	1050	3792	252
Relative change (%)	+115	+139	+8	+167	-30	+127	-20
Ensemble mean							
Number of events	37801	2150	6028	1019	809	3301	207
Relative change (%)	+75	−17	+36	+85	−52	+63	+11

The use of citations on lines 240 and 243 is confusing to me. The section presents results from the current study, and the citations seem to be put there to indicate consistency with those results. I would first present the results and then indicate that they are consistent with other studies such as those cited. This also applies to line 271–2.

Authors' response #10: We thank the referee for this helpful observation. We agree that the original placement of citations in this passage could make it unclear whether the cited studies were used to report the results of the present study or to provide context. In the revised manuscript, we have therefore retained the description of our results first and moved the citations to the end of the passage so that they clearly indicate consistency with previous studies. In the revised manuscript, we have rephrased the text around lines 240—243 and 271—272 so that our results are stated first, and the citations are added afterwards only to indicate consistency with previous studies.

Line 261 feels a bit misleading, as values above the third quartile are used for the Mediterranean region drought duration increase ("durations may lengthen by more than 200 days") compared to roughly the first and third quartile for the West-Central Europe region ("approx 25—75 days"). Given the highest values found in *the Mediterranean region are remarkable, I would compare similar values while also mentioning this fact.*

Authors' response #11: We thank the referee for this observation. We agree that the original sentence was unclear because it mixed projected absolute durations in the Mediterranean (Fig. 5b) with projected duration changes in Western-Central Europe (Fig. 5c–d). Our intention was to highlight both the very long-term drought durations in the Mediterranean and the more moderate change signal in Western-Central Europe, but these two quantities should not have been compared in the same phrasing. We have therefore revised the text to distinguish more clearly between projected future duration values and projected duration changes, as follows: "During the historical period, droughts in the Mediterranean were already longer than in Northern Europe and often exceeded 200 days in some areas. Projections for 2070–2099 (`GCM_fut`) indicate substantial further increases in drought duration in the Mediterranean, particularly for rainfall-deficit droughts. In many Mediterranean areas, projected drought durations remain above 200 days, and the strongest increases also exceed 200 days in the upper tail of the distribution, reflecting enhanced evapotranspiration and reduced precipitation."

References

[1] Tallaksen, L. M., Madsen, H. & Clausen, B. On the definition and modelling of streamflow drought duration and deficit volume. *Hydrological Sciences Journal* **42**, 15–33 (1997).

[2] Fleig, A. K., Tallaksen, L. M., Hisdal, H. & Demuth, S. A global evaluation of streamflow drought characteristics. *Hydrology and Earth System Sciences Discussions* **10**, 535–552 (2006).

[3] Van Loon, A. F. & Van Lanen, H. A. A process-based typology of hydrological drought. *Hydrology and Earth System Sciences* **16**, 1915–1946 (2012).

[4] Brunner, M. I., Van Loon, A. F. & Stahl, K. Moderate and severe hydrological droughts in europe differ in their hydrometeorological drivers. *Water Resources Research* **58**, e2022WR032871 (2022).

[5] Maraun, D. Bias correcting climate change simulations-a critical review. *Current Climate Change Reports* **2**, 211–220 (2016).

[6] Maraun, D. *et al.* Towards process-informed bias correction of climate change simulations. *Nature Climate Change* **7**, 764–773 (2017).

[7] Gudmundsson, L. & Seneviratne, S. I. Anthropogenic climate change affects meteorological drought risk in europe. *Environmental Research Letters* **11**, 044005 (2016).