Dear authors,

Your manuscript has now received two external referee reports, to which you have responded in the open discussion. Both referees seem to agree in their assessment that the manuscript is of interest, and can potentially make a valuable contribution to the literature. However they also indicate that the manuscript need to be improved on several aspects. Given the nature of the comments, I classify these as revisions, reflecting the need for considerable changes to the text (for instance the length of the Discussion as indicated by both referees). This also means that I will return a revised manuscript to the referees. You can use your previous replies as a starting point for a revision. Please contact me in case you have any questions. Looking forward to receiving a revised version of your work!

Best regards

Ryan Teuling

Response:
Thank you for your decision and for your time handling this manuscript. We have revised it following the reviewer comments and our initial responses, and believe these changes have significantly improved the manuscript. The Discussion text was updated and almost halved in length, and overall the revised manuscript is about 7 pages shorter compared to the initial submission. The Introduction was re-written to better convey the motivation and added value of our study, including explanation of our UK focus. The Methods section has been expanded with the requested additional explanations (e.g. additional detail on bias correction and the SP(E)I calculation).

In the point-by-point responses below, reviewer comments are included in black**, author responses in** blue**, and manuscript excerpts in** *italic blue***.**
In some cases, we refer to the relevant manuscript sections instead of including them here, in particular when addressing comments led to substantial modifications throughout entire sections.

There are also some grammar/wording changes not directly addressing reviewer comments, which were not documented below but can be inspected in the tracked changes-document. We also included additional relevant references in the Introduction (justifying the relevance of UK-focussed drought research), Methods (aridity index), Discussion (primarily in 6.3, including work published since the initial manuscript submission), and Conclusions (crediting some of the hydrological modelling work being undertaken using the UKCP18 projections for the UK).

Response to RC1 (anonymous)

This paper examines projected changes in drought frequency, extent, seasonality and duration for Great Britain using the latest UKCP18 projections. It analyses the differences between two atmospheric drought indicators (SPI and SPEI) and shows that the choice of drought indicator can have a big impact on the derived drought indicator.

Overall, this is an interesting study that is well written and well presented. The analyses are extensive, well thought out and well executed. I believe the paper will have appeal to a wide readership, however, there are some core issues that need to be resolved to enhance the key messages of the paper. Firstly I think the motivation of the study needs to be more clearly defined and more detail is required on the choice of climate projections and bias correction. Secondly, I strongly encourage the authors to take a critical read of the discussion and shorten it to the core messages – it is currently very long and your interesting results are getting lost.

More detailed comments are provided below that I hope the authors find useful.

Thank you for your time and your review. We believe addressing your comments has substantially improved the quality of the paper. In particular, the core issues on motivation and the choice of climate projections have been addressed in the introduction and methods sections respectively, and the discussion has been drastically shortened to about half its original length.

Comments

1. **Motivation.** The motivation for the study needs to be clearer. Currently the introduction reads like a series of definitions rather than a compelling story of why this study should be undertaken. There are two core areas where I think this could be improved:
    1. There have been quite a few studies that have used or compared drought indices (many are cited in your discussion) so what does this study offer that is novel and different?
    2. Why focus on Great Britain (which often isn't thought of as a country that experiences lots of droughts!) and how are your research questions relevant to this region?

We have modified the introduction section throughout to more clearly convey the motivation of the study, including two paragraphs responding to the two core areas mentioned above:

1. Novelty: "*Although previous studies have compared historical and projected changes using these SI in different regions of the world (e.g. Stagge et al., 2017; Chiang et al., 2021), this study adds a new level of detail by an in-depth analysis of different drought characteristics and attention to within-GB regional differences, and is the first to use UKCP18 with these SI to assess projected changes in drought characteristics for GB.*"

2. UK focus: *"This study focuses on Great Britain (GB) to compare projected drought changes based on the SPI and SPEI. Despite not typically being thought of as a particularly drought-prone area, GB has experienced several droughts in the past which lead to widespread impacts, including impacts on ecosystems (including algal blooms and fish kills), agriculture and domestic water supply (Rodda and March, 2011; Kendon et al., 2013; Turner et al., 2018). The impacts of climate change on future droughts in the UK is therefore a key concern for stakeholders including water managers and farmers (e.g. Watts et al., 2015)"*

New references for this paragraph:

Kendon, M., Marsh, T., and Parry, S.: The 2010–2012 Drought in England and Wales, Weather, 68, 88–95, https://doi.org/10.1002/wea.2101,2013.

Rodda, J. and March, T.: The 1975/76 Drought – a Contemporary and Retrospective View, Tech. rep., Centre for Ecology & Hydrology, http://nora.nerc.ac.uk/id/eprint/15011/1/CEH_1975-76_Drought_Report_Rodda_and_Marsh.pdf, 2011.

Turner, S., Barker, L. J., Hannaford, J., Muchan, K., Parry, S., and Sefton, C.: The 2018/2019 Drought in the UK: A Hydrological Appraisal, Weather, https://doi.org/10.1002/wea.4003, 2018

Watts, G., Battarbee, R. W., Bloomfield, J. P., Crossman, J., Daccache, A., Durance, I., Elliott, J. A., Garner, G., Hannaford, J., Hannah, D. M., Hess, T., Jackson, C. R., Kay, A. L., Kernan, M., Knox, J., Mackay, J., Monteith, D. T., Ormerod, S. J., Rance, J., Stuart, M. E., Wade, A. J., Wade, S. D., Weatherhead, K., Whitehead, P. G., and Wilby, R. L.: Climate Change and Water in the UK - Past Changes and Future Prospects:, Progress in Physical Geography, https://doi.org/10.1177/0309133314542957, 2015

2. **Use of UKCP18.** There needs to be better justification for the use of the regional projections from UKCP18 in this study – why use this set of projections instead of the local UKCP18 projections for example? Or use a set of climate projections that encompasses different GCMs or RCP scenarios (for example)?

The following paragraphs were added in the Data section to further justify our use of the UKCP18 regional simulations:

"UKCP18 is the most recent set of national climate projections for the UK and have been produced by the Met Office Hadley Centre with the aim of providing a range of storylines to support adaptation efforts in the UK (Murphy et al., 2018)."

…

"The ensemble thus does not sample GCM-RCM structural uncertainty, only parameter uncertainty, and was designed to cover a range of possible futures. While multiple GCM-RCM structures would add another interesting dimension to the study, expanding the ensemble was outside the scope and capacity of the study. The horizontal resolution of the RCM simulations is 12km over GB (available on OSGB36 grid projection). As droughts tend to be more spread out in space and time, we judged that the 12km daily resolution of the UKCP18 RCM pose a better trade-off between practicality and spatiotemporal detail than the higher-resolution convective permitting simulations for this study."

3. **Bias Correction.** At the end of Section 2.2 there is a section on bias correction which needs more detail. Why did you choose these bias correction methods and how are they applicable to the types of biases you observe between the climate projections and observational data? It would be helpful to add some plots in the supplementary information showing these biases to help the reader understand the nature of the biases and how they were corrected. For example, you note in section 3.2 that a direct comparison of the results between climate model ensemble members and observations is only possible because their distributions are similar – it would be helpful to see evidence of this.

The section on bias adjustment was expanded with additional explanation and moved to a dedicated Methods subsection:

"3.2 Bias adjustment
As comparison to observations revealed significant bias in the simulation of both precipitation and PET (see Figs. S1 and S2), these variables were statistically post-processed using the ISIMIP3b change preserving bias adjustment method (Lange, 2019) version 2.4.1 (Lange, 2020). The biases we observed for different quantiles were not equal to the biases observed in the mean, which is why we selected a bias adjustment method that took this into account. Similarly, biases also varied between months and locations, so the bias adjustment needed to be specific for each month and grid cell. The ISIMIP3b bias adjustment method is based on quantile mapping, but also preserves projected changes in the variables being corrected, and enables separate adjustment of the frequency of dry days – a desirable feature for drought research. For precipitation, the gamma distribution and mixed additive/multiplicative per-quantile change preservation were used. For PET and PETdtr−tas, the Weibull distribution, detrending and mixed additive/multiplicative per-quantile change preservation were used. A dry threshold of 0.1 mm day-1 was selected below which there is considered to be no precipitation or PET. In what follows, UKCP18-RCM indicates the bias adjusted data."

To show the nature of the biases before adjustment, maps of ensemble-averaged precipitation and potential evapotranspiration bias (in three sections of their distributions) were included in the supplementary material. As the biases were well-corrected, similar post-correction bias maps would take up two pages for little added information, and were thus not included. Instead, to give evidence for the good match between observations and reference period simulation statistics after bias adjustment, we have added statistics of the observations to the following plots:

- Fig. 2 and Fig. S3 (seasonal cycles for the different regions)
- Fig. 3 (aridity map)
- Fig. 4 (observations included in scatter plot for two time periods: the 50 year SI calibration period and the 25 year reference period)

4. **Discussion.** The discussion section is extremely long and as a result, a lot of your interesting results get a little lost amid all the discussion. The authors need to have a critical read of the discussion and carefully consider if all the text is needed. As a suggestion, I would significantly shorten section 6.4 as this tends to be a review of the literature, rather than a discussion of your results with context from the literature (you could just use the summary paragraph – you don't really need all the preceding text).

In the revision, the discussion was almost halved in text length, which we think has benefited readability. For Section 6.4 in particular, we have kept the intro and summary paragraphs with slight modifications, and removed about 2 pages of text in total. Sections 6.1, 6.2 and 6.3 were also significantly shortened.
Section 6.3 (Role of AED) was largely re-written, to give it more structure for the reader. We also included some additional relevant literature, including interesting works on evaporation published since the initial submission of this paper.

Minor/Technical Corrections

P2 L45. 'Drought indices that only rely on atmospheric data are a popular choice due to data availability and **propagating model uncertainties**.' I don't understand this sentence – why are they a popular choice due to propagating model uncertainties?

Changed to be clearer as follows: *"While indicators exist for variables relevant to different drought types, drought indices that only rely on atmospheric data are a popular choice due to (historical) data availability and due to their ease of use (they do not require the deployment of an impact model, such as a hydrological model)."*

P3 L65. You could add into the third research question the uncertainty in the RCM as a lot of your results focus on the differences between ensemble members. E.g. How sensitive are the projected changes in drought characteristics to the choice of atmosphere-based drought indicator and parametric uncertainty in regional climate models?

We amended the research question as follows, to include this as well as the GMWL as a source of uncertainty: *"How sensitive are the quantified projected changes in drought characteristics to the choice of atmosphere-based drought indicator, and how does it compare to other sources of uncertainty?"*

P3 Section 2.1. It may be useful to state the time-period you used from each observational dataset.

Added sentence in Section 2.1: "…, using the following time periods: 1961-2010 for the SI calibration (see Section 3.4), 1981-2010 for the bias correction, and 1981-2005 for comparison to the reference period UKCP18-data in this study."

P5 L141-142. Why did you include aridity as well? What does it add to the story? The motivation could be a little clearer.

Added to Section 3.4*: "While drought refers to a period of below-normal water availability for a given context, aridity refers to the climatic average moisture availability (Dai, 2011). This is included in this study in order to help establish an understanding of the mean climatic changes projected for precipitation and PET in UKCP18-RCM, before proceeding to assessing projected changes in drought characteristics. To this end, the aridity index (AI) was calculated as the annual average ratio of precipitation to PET (e.g. UNEP, 1992; Feng and Fu, 2013; Greve et al., 2019), which is more intuitive to interpret than the standardized indicators."*

P6 L168. The area for London is small, but it is a central hub which contains around 14% of the population of GB and is likely to be significantly affected by droughts in the future. Consequently, leaving out the figures for London because the area is small seems an odd choice, given the significant impacts changes in droughts will have in this small part of GB. Potentially a better reason would be because the results are very similar the South-East region or East of England?

The results are indeed very similar to those of South-East England, but as we have now reduced the number of figures to be shown in the main text to four, we have also included the results for London in the supplementary materials along with the rest of the regions (Figs. S3-S8).

Figure 1. It is difficult to see the labels for North and West Scotland – can these be moved or made clearer?

Label readability has been improved and the selected "main text" regions were highlighted on the map.

P9 L203-205. This sentence is a little difficult to understand – can it be rewritten?

In the end, this sentence disappeared in the compression of Section 5.1 (see RC2 and responses), because the section now focusses more on the key take-aways from the discussed figures in favour of individually describing the projections for each SI, GMWL and drought category.

P9 L209. 'For extreme meteorological drought, all ensemble members **project multiples** of the reference period frequency by +4 ◦C.' I don't understand what you mean here.

This sentence also disappeared in the compression of Section 5.1.

P10 L233 'due to SPEI6 occurrences beginning to saturate when they have already become quite frequent'. What do you mean by 'beginning to saturate'?

This sentence also disappeared in the compression of Section 5.1.

P28 L666. I think you are missing some key references from the brackets?

This sentence disappeared in the compression of Section 6.4. The references that could have been meant to go here (Touma et al., 2015 or Lee et al., 2019) were included in Section 6.3 instead (L463), for similar reasons as why they would've been cited in the disappeared sentence.

Code and Data Availability – Great to see that the data you produced are available but this section needs to be a full description of all the data used in the study, including links to all the data you used for analysis (i.e. for the CHESS-PE, HadUK and UKCP18 data). I would also reword to 'The SPEI and SPI data **produced in this study** are available on Zenodo (doi:10.5281/zenodo.6123020) (Reyniers et al., 2022b) alongside the bias adjusted UKCP18- based PET (doi:10.5281/zenodo.6320707) (Reyniers et al., 2022a).'

Thanks again for pointing this out, this section now reads:

"The SPEI and SPI data produced in this study are available on Zenodo (doi:10.5281/zenodo.6123020) (Reyniers et al., 2022b) alongside the bias adjusted UKCP18- based PET (doi:10.5281/zenodo.6320707) (Reyniers et al., 2022a). Python code for the computations and analyses is available upon reasonable request. The CHESS-PE data used in this study was obtained from the UK CEH Environmental Information Data Centre (https://doi.org/10.5285/9116e565-2c0a-455b-9c68-558fdd9179ad) (Robinson et al., 2020). HadUK-Grid data was obtained from the Centre for Environmental Data Analysis (http://dx.doi.org/10.5285/d1343358 (Met Office et al., 2019), as well as the raw UKCP18-RCM simulations (https://catalogue.ceda.ac.uk/uuid/589211abeb844070a95d061c8cc7f6565 (Met Office Hadley Centre, 2018)."

Response to RC2 (Marie-Claire ten Veldhuis)

In this paper changes in drought characteristics are evaluated for GB, for 2 future climate scenarios. Two drought indices are used to characterize drought severity, SPI and SPEI, for various space and time-scales. The study finds increase in most drought characteristics (frequency, extent, duration etc) for future climate conditions, not entirely unexpected.

In particular, the authors emphasize that the choice of drought index influences the quantitative assessments of projected drought changes.

Given this perspective, it is particularly important to document not only the indices used but also the full range of methods applied to reach their conclusions. This is the main problem I see with this paper: many aspects of the methods used for analysis are not clearly explained. Secondly, the Results and Discussion sections are very long and need to be strongly condensed to convey only the essential information. To give an example, section 5.1 covers almost 2 pages to describe a single figure, followed by 2 more pages for a second figure. That's a lot of descriptive information that can be drastically shortened, based on a critical reassessment of what pieces of information are really important and worth for the reader to know.

[concluding reviewer comment moved here as it also addresses the manuscript in general:]

Note: I stopped reading here [after Section 5.1 and Fig. 6]. Sections 5 and 6 are very lengthy and many of the results point in the same direction. Are all these figures and subsections really needed to make the point stated in the title, that "Projected changes in droughts are strongly influenced by the choice of drought index"? I strongly recommend that the authors take a critical view of their results and make a selection of the materials that most strongly support their conclusions. Then report these clearly and concisely.

Thank you for your time and your comments, we believe that the paper has substantially improved after addressing them. The Methods section was expanded and structured to address all comments. The Discussion and Results sections were shortened drastically, both by condensing the text and reducing the volume of Figures shown. Specifically, Section 5.1 was reduced from 92 to 48 lines, and the results and discussion sections were both about halved in text length. The number of figures was reduced by selecting only four regions to show in the main text, whose response is representative of the responses seen across all the analysed GB regions for the seasonal cycle, drought seasonality and duration distribution. These regions were highlighted on the map in Fig. 1, and results for the other regions were shown in the supplementary materials. In total, the paper reduced in length by about 7 pages.

Detailed comments:

Abstract:

1. General comment: the summary of results presented here is quite superficial, i.e descriptive rather than interpretive. Deeper interpretation of the results would make the Abstract a lot more appealing.

We re-wrote the abstract to better convey the relevance and context of our results:

"*Abstract. Droughts cause enormous ecological, economical and societal damage, and are already undergoing changes due to anthropogenic climate change. The issue of defining and quantifying droughts has long been a substantial source of uncertainty in understanding*

observed and projected trends. Atmospheric-based drought indicators, such as the Standardised Precipitation Index (SPI) and the Standardised Precipitation Evapotranspiration Index (SPEI), are often used to quantify drought characteristics and their changes, sometimes as the sole metric representing drought. This study presents a detailed systematic analysis of SPI- and SPEI-based drought projections and their differences for Great Britain, derived from the most recent set of regional climate projections for the UK. We show that the choice of drought indicator has a decisive influence on projected changes in drought frequency, extent, duration and seasonality by 2 ◦C and 4 ◦C above pre-industrial levels. The increases projected in drought frequency and extent are far greater based on the SPEI than based on the SPI. Importantly, compared to droughts of all intensities, isolated extreme droughts are projected to increase far more in frequency and extent, and show more pronounced changes in the distribution of their event durations. Further, projected intensification of the seasonal cycle is reflected in an increasing occurrence of years with (extremely) dry summers combined with wetter than average winters. Increasing summer droughts also form the main contribution to increases in annual droughts, especially using SPEI. These results show that the choice of atmospheric drought index strongly influences the drought characteristics inferred from climate change projections, comparable to the uncertainty from the climate model parameters or the warming level, and therefore potential users of these indices should carefully consider the importance of potential evapotranspiration in their intended context. The stark differences between SPI- and SPEI-based projections highlight the need to better understand the interplay between increasing atmospheric evaporative demand, moisture availability and drought impacts under a changing climate. The region-dependent projected changes in drought characteristics by two warming levels have important implications for adaptation efforts in GB, and further stress the need for rapid mitigation."

2. Check phrasing here: the phrasing suggests that projected changes are sensitive to the choice of drought index (L5). However, projected changes are simply what the simulated climate scenarios tell us, how they are summarized in quantitative metrics is where the differences in interpretation come in.

Following our original response to this comment clarifying that our use of projections (i.e., including the quantification step using the SPI and SPEI, which can be considered simple models) follows the IPCC definition, we addressed the issue of potentially confusing uses of "projection" by being clear about which variables are indicated in phrasings such as "projected changes", and by being explicit about the SI being included in the term "projections" by using phrasings such as "quantified using SPI", "SPEI-based projections" or "inferred using SPI". The sentence indicated in this specific comment was not included in the re-written abstract."

3. Same confusing phrasing is used throughout, e.g. (L14) "SPEI results in greater increases in drought frequency and extent". Obviously the drought characteristics do not change, only how the indices are computed.

This specific sentence disappeared in rewriting.

4. L16: "projected changes (..) depend on the drought index, (..)". Again, reasoning is flawed: projected changes are the same, the indices are different, not the other way around.

This specific sentence disappeared in rewriting.

Introduction:

- P2, L 39: it is suggested here that evapotranspiration only depends on atmospheric variables, but strictly speaking vegetation also plays a role (stomatal conductance).

This was amended as proposed, but then the sentence was left out in re-writing the introduction.

- P3, l65: same phrasing issue as in Abstract.

Revised version: *"How sensitive are the quantified projected changes in drought characteristics to the choice of atmosphere-based drought indicator, and how does it compare to other sources of uncertainty?"*

Methods:

- P5: it would be helpful to provide the definitions (and/or the equations) of the indices that are used in the paper (AI, SPI, SPEI), so the reader doesn't need to search back in the literature.

For aridity, context and the equation were added to the AI definition paragraph:

While drought refers to a period of below-normal water availability for a given context, aridity refers to the climatic average moisture availability (Dai, 2011). This is included in this study in order to help establish an understanding of the mean climatic changes projected for precipitation and PET in UKCP18-RCM, before proceeding to assessing projected changes in drought characteristics. To this end, the aridity index (AI) was calculated as the annual average ratio of precipitation to PET (e.g. UNEP, 1992; Feng and Fu, 2013; Greve et al., 2019), which is more intuitive to interpret than the standardized indicators. For time slices of 25 years, this gives:

$$AI = \frac{1}{25}\sum_{y=1}^{25}\frac{Precipitation_y}{PET_y}$$

For SPI and SPEI, the common general equation was added, and more detail was provided on their calculation so that the reader indeed doesn't need to refer back to the cited literature:

"The drought indices compared in this study are SPI and SPEI. Both are widely used in the literature to quantify droughts, and they imply contrasting assumptions of the surface water balance: for SPI, no evaporation takes place, while for SPEI, evaporation takes place and is not limited by moisture availability. Multi-scalar standardized climate indicators such as these allow for comparison of unusually dry (or wet) periods across locations with different climates. The SI are calculated as follows. First, the time series of a variable D (precipitation for SPI, precipitation minus PET for SPEI) is aggregated using a specified accumulation period length of n months, such that the value for each month in the resulting time series is the average of that month and the n preceding months. Then, a suitable distribution F D for that variable is fit to the aggregated time series, for each month and location. The SI value for an accumulation period length n at a time step t is then defined as follows:

$$SIn_t = \phi^{-1}(F_D(D_{n,t}))$$

with Dn,t indicating D accumulated over the n time steps preceding t (inclusive), and φ the standard normal distribution. Monthly values of SPI and SPEI are calculated using n of 3 to 24 months. Following recommendations provided by Stagge et al. (2015b), the two-parameter gamma distribution was used for calculating SPI and the generalized extreme value (GEV) distribution was used for calculating SPEI. For shorter SPI accumulation periods (1-3 months) and further into the future in the UKCP18-RCM simulations (with drying

summers), there may be occurrences of zero accumulated precipitation for grid cells in drier regions. To take this possibility into account, the SPI values corresponding to the probability of zero accumulated precipitation were calculated separately following the method proposed by Stagge et al. (2015b), which avoids the mean SPI becoming larger than 0. A 50-year period (1961-2010) of observation-based data (regridded HadUK-Grid and CHESS-PE) was used to fit the distributions for the SPI and SPEI calculation. This observation-based calibration was also applied to the UKCP18-RCM data to allow a direct comparison of the results between climate model ensemble members and observations. This is appropriate because the bias adjustment brings the distributions of the reference period climate model data close to the observed distributions."

- P5, L151: "observation-based calibration": this needs clarification. How was this calibration done, this is currently not explained.

  This refers to fitting the distributions to the observation data: *"A 50-year period (1961-2010) of observation-based data (regridded HadUK-Grid and CHESS-PE) was used to fit the distributions for the SPI and SPEI calculation. This observation-based calibration was also applied to the UKCP18-RCM data to allow a direct comparison of the results between climate model ensemble members and observations."*

- P6, section 3.3, Drought characterization: it is stated that spatiotemporal characterization is important - agreed. Unfortunately, the authors do not specify the space and time scales used in their characterization. What is "regional", "seasonal", what range of space and time scales did they investigate?

  This has been clarified by phrasing changes and additions throughout the corresponding paragraphs:

  "*Given the importance of both space (e.g. extent, spatial connectivity, local vulnerability) and time (e.g. seasonal timing, duration) for drought impacts, the spatiotemporal characterisation of droughts is an important element of any drought study. It is approached here in three ways. First, the frequency (fraction of the time in drought) of dry and extremely dry conditions was computed for each individual grid cell of GB separately, for each ensemble member and the observations. Second, the drought area extent was quantified as the fraction of the total GB area simultaneously in (extreme) drought. We then compute the frequency with which different drought extents are exceeded (fraction of time). Third, regionally averaged SI values were used to investigate drought seasonality and duration. For computing these regional averages, we used the UKCP18 administrative regions (ukcp18 data, 2021) shown in Fig. 1, as they represented a decent trade-off between the sizes of the regions, number of regions to compare and relevant differences in climatology, projected changes and societal relevance. For investigating the seasonal contributions to longer-term deficits (seasonality), we compared the 6-month aggregated regionally averaged SI (SI6) for March and September for each year to represent the winter and summer contributions to that year's overall dryness (SI12). Durations of individual drought events are defined as periods of continuously negative regionally averaged SI values reaching a threshold value of -1 or lower, following the theory of runs (Yevjevich, 1967). Each event is then assigned to the time slice (reference period, +2 ◦C or +4 ◦C) that contains its central time step, and the number of occurrences of droughts with different duration categories is assessed. Extreme droughts are identified as events that have a peak (i.e. minimum) SI value below -2.*"

- P6, L162: please clarify definition of ‘extreme drought’. At present, the choice of SI<-2 sounds arbitrary.

This has been elaborated as follows: *“In order to compare changes in overall drought conditions to changes in more extremely dry conditions, we consider a category of "all/total drought" covering all SI of -1 and lower, and a category of "extreme" drought covering SI values of -2 and lower. These threshold values are a subset of the classification originally introduced by McKee et al. (1993), which has been extensively used in studies using standardised drought indicators.”*

- P6, L177: “a distribution fitted to the relatively short times series”. This needs explanation: what distributions were fitted, how exactly?

This sentence now includes “*(see Section 3.4).”*, which is where the calculation of the SI (including the distribution fitting) is now explained in more detail.

4. Projected climate changes:

- In the caption of Figure 3 it is mentioned that “after bias adjustment using change preserving quantile mapping” is applied to the ensemble members.
This is not the right place to mention such a data processing step! Please explain adequately in the main text.

The corresponding Data paragraph has been moved to its own dedicated Methods subsection, and this new bias adjustment subsection has also been expanded with more detail in response to this comment as well as a comment by Reviewer 1.

5. Projected changes in drought characteristics:
   - L204: the authors refer to “2C above pre-industrial”, but as far as I understand their reference scenario is 1981-2005. That’s not exactly pre-industrial.. Please clarify or correct.

The “time slice selection” Methods section has been amended to clarify that the +2 and +4 are relative to pre-industrial levels (1850-1900), not the fixed reference period:

“Therefore, to assess the impact of climate change on drought characteristics in scenarios with lower climate sensitivity and more mitigation (resulting in lower warming levels above pre-industrial times), a time slice approach was implemented to investigate changes at two specific global mean warming levels. A common fixed reference period (1981-2005) was used for all ensemble members to compare to these future time slices and observations. For each ensemble member, a time slice was selected from 12 years before to 12 years after the year in which the centred 25-year rolling mean global temperature exceeds + 2 ◦C and + 4 ◦C above pre-industrial levels (defined as 1850-1900) in the driving global model (see Table 2 in Gohar et al. (2018)).”

The exact sentence this comment referred to disappeared in re-writing the results section.

- Figure 4: the use of % as a unit for frequency is very confusing here. If I understand correctly the % is calculated based on number of years (in 25 year climate period) that index values are below a given threshold. This is a guess, it is not clearly explained.
Much later, in Figure 10, the authors use “number of events” instead - a much more straightforward type of unit. I recommend using this unit throughout.

The figure caption now explains this as follows: "*Spatially averaged projections of drought frequency, expressed as the fraction of time SI is below the threshold, for each […].*"

Furthermore, the distinction between the "number of events" unit used for drought duration and the "% of time" unit used in Figure 4 has been clarified in the drought characterisation Methods section.

- LL 199-241: this is a very extensive description of a single figure (see earlier comment). Please reflect critically: what pieces of information are really worth mentioning?
- LL 242-290: same here, figure description is far too lengthy.

The section with these figure descriptions (5.1) has been almost halved in length, to better convey the key messages learned from both figures.

- L246: "the fit of the gamma and GEV distributions used in the calculation of SPI and SPEI". So gamma and GEV distributions were fitted apparently..? This should have been explained in the Methods Section!

See expanded explanation of SI calculation in the Methods section above.

- L266: "detrended temperature simulations". Again, please explain his properly in the Methods section – how was the detrending done, for what purpose exactly?

The corresponding paragraph has been moved from the Data section to its own dedicated Methods subsection.

L272: "purely temperature-based PET": this seems to suggest that temperature has a strong influence on PET, yet the influence of Radiation is much stronger (linear relationship with PET in Penman equation). Please check the reasoning here, it seems flawed.

Rephrased to clarify what we meant: *"PET calculation methods which only rely on temperature (e.g. Thornthwaite)"*

Figure 6: this is first time Observations are shown in any of the results graphs! Why only now and not in the earlier graphs?

In the initial submission, they were judged to be more interesting to show in Figures 6 and 7. Observations were now also added to Figures 2, 3 and 4 as well, demonstrating a good match with the reference period simulations for the seasonal cycles of P and PET, aridity index and SPI/SPEI. The latter shows the sensitivity of SPI and SPEI to the period used to calibrate / fit the distributions in their calculation, through the slight difference between the dots showing the 1981-2005 observations and 1961-2010 observations.

Also in Figure 6: a gradual color scale is applied here which makes it impossible to distinguish clearly between the 3 scenarios. Note that in the current representation their seems to be no significant difference between the Reference and +2C scenario.

The colour scheme for this figure has been adjusted for enhanced contrast, while still following the same reasoning behind the original colour scale which led to picking a gradual colour scale. Additionally, we re-arranged the order in which the line plots are overlaid, so that the thick lines representing the ensemble means are on top of the thin lines representing individual ensemble members. These changes make it now easier to see the differences between the reference and +2°C scenario for SPI, although the smaller

magnitude of the differences (compared to SPEI or +4°C) of course also makes them more difficult to distinguish.

All figures in the study were checked for colourblind-friendliness with https://www.color-blindness.com/coblis-color-blindness-simulator/.