Response to RC2 (Marie-Claire ten Veldhuis)

Thank you for your review of this study. We generally agree with your comments, are happy to address them, and believe this will substantially improve the paper. In particular, to address a main overarching comment, we intend to significantly shorten the figure descriptions in the Results section as requested, and also reduce the number of panels in some figures, in cases where not all panels are needed to support the information conveyed. On the methodology, we will expand the explanations in response to the comments, but please note that some of the requested information was already present in the Data section (this will be moved to Methods).
Please find below our responses to the individual detailed comments.

Abstract:

- 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 will expand the interpretation of the results in the Abstract.
- 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.
  In the IPCC Glossary (IPCC, 2021), "projections" are defined as follows: "*A potential future evolution of a quantity or set of quantities, often computed with the aid of a model. Unlike predictions, projections are conditional on assumptions concerning, for example, future socio-economic and technological developments that may or may not be realized.*".
  Our use of "projections" meets the IPCC definition: in this case, the *model* is a combination of climate models and simple models of drought characteristics. By analogy, if instead of SPI and SPEI we'd estimated drought conditions with two hydrological models, one of which ignored evaporative losses and one of which included them (without making them dependent on moisture availability), it would be valid to say that the future drought projections were sensitive to the choice of hydrological model.
  We propose to address this by being clear and more specific in our wording where this phrasing occurs, e.g. specifying that we mean the projected changes *to drought characteristics.*
- 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.
  See previous.
  Proposed rephrasing L14: "In general, far greater increases in drought frequency and extent are found when using SPEI for drought quantification than when using SPI". (Alternatively: replace "results" by "implies").
- 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.
  See previous.
  Although it is already specified here that this concerns the projected changes *in the distribution of drought durations*, an alternative wording could be: "the quantification of projected changes […] depends on …"

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).
Proposed solution: "[…] reference crop (A 1998), in which a fixed role of vegetation and fixed high moisture availability are assumed, such that only the effect of the atmospheric variables is left in the spatiotemporal variation of the resulting reference evapotranspiration."

- P3, l65: same phrasing issue as in Abstract.
See earlier comment. Alternative phrasing: How sensitive are quantifications of projected changes in drought characteristics to […]"

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.
  Thank you for pointing this out. We will add a paragraph and equation to the methods section to explain the standardised indicators (including specific methodological choices such as the fitted distributions used in the SI calculation, see your later comment), so that readers who aren't already very familiar with these indicators aren't required to refer back to the cited literature.

- P5, L151: "observation-based calibration": this needs clarification. How was this calibration done, this is currently not explained.
  Proposed wording change: *[…] was used to fit the distributions for the SPI and SPEI calculation. This observation-based calibration was also applied […].* This, combined with the added SPI/SPEI explanation (see above), will hopefully clarify the statement.

- 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?
  Thank you, we will rewrite this to make it clearer.
  The characterisation in space came in 3 forms: (1) each grid cell separately (for frequency, as visualised in maps); (2) UK-averages (extent expressed as fraction of the surface in drought, plus the summary heatmaps); (3) averages by UK administrative regions (for duration of individual events and investigating the seasonal contributions).
  The characterisation in time was done in the following ways: (1) frequency of exceeding SI thresholds (% of time); (2) duration and counts of individual events, which are defined as continuously negative SI; (3) seasonal contributions, in which we compared the SI12 for each year to the SI6 representing the October-March and April-September periods making up that year. So in this case, the term "seasonal" is used to indicate hydrological winter (SI6 for March, which represents the anomaly for October-March) or hydrological summer (SI6 for September, which represents the anomaly for April-September).

- P6, L162: please clarify definition of 'extreme drought'. At present, the choice of SI<-2 sounds arbitrary.
  This is the threshold used to separate "extreme" drought in the paper that originally proposed the SPI and in many studies that apply standardised indicators of drought. It is also used as the threshold for extreme drought in some drought monitoring

systems (e.g. https://eip.ceh.ac.uk/hydrology/water-resources/). As you rightfully point out, it is a bit arbitrary. A few other studies have used other threshold levels intended to better reflect impacts, however the threshold values linked to other drought types/impacts vary spatially across the UK (as shown by Parsons et al., 2019), so to keep it simple we opted for the thresholds of -2 and -1 standard deviations, to stay consistent with what is conventionally used in much of the literature.
We are happy to add a concise clarification of the “drought” and “extreme drought” threshold choices to the text.

- P6, L177: “a distribution fitted to the relatively short times series”. This needs explanation: what distributions were fitted, how exactly?
  This will definitely be clarified by the expansion of the SPI/SPEI explanation in Methods (see earlier comment). These distributions are fitted as part of the standardisation process for the SP(E)I calculation.

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.
It is already explained in the final paragraph of the section on the UKCP18 regional climate projections (L113-118):

*“As comparison to observations revealed significant bias in the simulation of both precipitation and PET, **these variables were statistically post-processed using the ISIMIP3b change preserving bias adjustment method (Lange, 2019) version 2.4.1 (Lange, 2020)**. 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.”*

We will move the bias adjustment paragraph to a dedicated subsection in “Methods” instead, and also expand the explanation further in response to a comment from 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.
Thank you for pointing this out, we will clarify this. Indeed, our “reference” period does not represent the pre-industrial level. The UKCP18 Derived Projections report on which we based our time slice selection documents the timings of crossing 2 and 4 °C of warming relative to pre-industrial levels, which refers to the period of 1850-1900 (Gohar et al., 2018). Our “REF” time slice of 1981-2005 was chosen as the UKCP18 RCM simulations are available from 1 December 1980 to 30 November 2080, with RCP8.5 affecting emissions starting 2006, and the timings of passing 2 and 4 °C of warming are also based on 25 year centred running means.

- 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.

We will clarify “%” and “number of events” to communicate more clearly what they mean and how they are distinct units. % is used for the percent of the time the SI values are below certain thresholds (-1 or -2), while “number of events” is used to count the number of individual drought events with different durations (defined as runs of consecutive negative values).

- 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.

We will reduce the descriptions to enhance the most essential information.

- 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!

Thank you for pointing this out, this will be clear when we add the definition and more detailed explanation for SPI and SPEI calculation (see first Methods comment).

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

This was explained in the data section on UKCP18-RCM (second to last paragraph):

While AED increases with rising temperatures, changes in humidity, net radiation and wind speed can also play a significant role. Therefore, we represented AED by PET calculated using Penman-Monteith, which includes the effect of all these variables. This method leads to a more robust correlation between the resulting SPEI and soil moisture under a warming climate compared to using the temperature-only Thornthwaite method (Feng et al., 2017) and is recommended over simpler temperature-based methods (e.g. Dewes et al., 2017), however it is still subject to significant limitations (Milly and Dunne, 2016; Greve et al., 2019). The calculation of PET for the UKCP18-RCM follows the same variant of the Penman-Monteith method used by Robinson et al. (2017), to ensure consistency with CHESS-PE. It uses these variables simulated by the UKCP18-RCM ensemble: specific humidity, pressure at sea level, net downwelling longwave radiation, net downwelling shortwave radiation, wind speed at 10m and daily average surface air temperature. PET was set to zero wherever a calculated value was negative (which occurred for less than 1% of the values overall and, when split by ensemble member and month, also less than 1% for all cases except December in ensemble member 1 with 1.2% of negative values). ***To investigate the influence of the projected temperature trend on changes in SPEI-based droughts and the deviation of SPEI from SPI, we also computed an alternate version of projected SPEI (SPEIdtr−tas) using a detrended version of UKCP18-RCM temperature. For this****, a linear trend was fitted to, and subsequently subtracted from, the simulated temperature time series for each grid cell and month separately. This detrended temperature dataset was used to compute PET as described above, resulting in a PETdtr−tas variable in which any trend left is due to trends in other variables (specific humidity, radiation, wind speed and pressure) or in interactions between variables. As these variables are closely intertwined in the climate models, this unavoidably introduces a physical discrepancy between temperature and the other variables used in the PET calculation. This is taken into account in the interpretation.*

Same as for the comment on bias adjustment, we will move this from the UKCP18 Data subsection to a dedicated separate Methods subsection, and clearly signpost the explanation of detrending with a subheader in a new subsection on potential evapotranspiration.

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.
We will rephrase this as it is indeed confusing. “Purely temperature-based PET” refers to some quantification methods of PET that rely on only temperature data, e.g. Thornthwaite. In our results, temperature detrending has a large impact because it reduces the saturated vapour pressure, which in combination with unchanged specific humidity leads to lower vapour pressure deficit (relative humidity increases), thus reducing PET.

Figure 6: this is first time Observations are shown in any of the results graphs! Why only now and not in the earlier graphs?
Due to bias correction, RCM-derived reference period statistics for regional averages or single grid cells lie close to the observations, so we decided not to show them in the seasonal cycle and aridity figures, as they would have little added value. In Figure 6, spatial co-occurrence becomes important, which was not explicitly considered in the bias correction, and as such we found it interesting to show observations here.
However, we can see that it could still be good for the reader to see the observations earlier. We will add dots for the observations to the scatter plot in Figure 4, showing their position with respect to the UKCP18-RCM reference period SPI6 and SPEI6. We will also add observations to Figures 2 and 3, which would also help address a comment by Reviewer 1 regarding the comparison of the observations and bias corrected simulations.

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.
We will look into amending the colours and/or other line properties to improve readability. The colour scale had the following reasoning behind it: green hues were avoided for the reference period in order to not imply that the ‘81-‘05 reference period is not already affected by climate change. A gradual colour scheme was chosen for intuitive interpretation of the progression from reference period to +2 and +4 °C.

Note: I stopped reading here. 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.
We agree that our results, discussion and key messages can be reported significantly more concisely, and we will reduce the text accordingly. While the point stated in the title is indeed not entirely unexpected, we think the magnitude of the differences between these popular drought indicators may commonly be underappreciated (e.g. in cases where only 1 drought indicator is used to represent overall “drought”), and as such we analysed and demonstrated the differences in great depth across different characteristics of drought.

After a critical re-examination of the text and figures, we will significantly condense the text in Sections 5 and 6 in order to better support the key messages in a concise way, including shorter figure descriptions. With regards to figures: we think all figures have valuable information to add, although we are considering leaving out the overall aridity figure (Figure 2). The figures that show results for all 12 regions (Figures 2, 8, 9, 10 and 11) could convey the essential information when showing results for only 4 regions. We will select 4 regions that best represent the regional variability of the responses, and only keep these for the main text. This will automatically merge Figures 8 and 9 (seasonality) into one, and will allow us to place SPI- and SPEI-based results side by side in a merged version of Figures 10 and

11 (durations), thus effectively decreasing the number of figures in the main text by 2. As some readers might still be interested in all or in specific regions, we will include the results for the other regions in the Supplementary materials.

References

Gohar, L., Bernie, D., Good, P., & Lowe, J. A. (2018). UKCP18 derived projections of future climate over the UK. Met Office, Exeter.

Parsons, D. J., Rey, D., Tanguy, M., & Holman, I. P. (2019). Regional variations in the link between drought indices and reported agricultural impacts of drought. Agricultural systems, 173, 119-129.

IPCC, 2021: Annex VII: Glossary [Matthews, J.B.R., V. Möller, R. van Diemen, J.S. Fuglestvedt, V. Masson-Delmotte, C. Méndez, S. Semenov, A. Reisinger (eds.)]. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 2215–2256, doi:10.1017/9781009157896.022