Response to Reviewer #1

This study examines changes in projections of meteorological, agricultural, and hydrological droughts in Australia while quantifying the contributions of different sources of uncertainty. The study uses CMIP5 climate projections from four global climate models (GCMs) under two carbon emission scenarios. Each of the climate projections is downscaled using four statistical and dynamical downscaling methods to drive the Australian Landscape Water Balance model (AWRA-L), to generate hydrological projections. The results indicate an increase in future droughts in Australia despite inherent uncertainties. Among the sources of uncertainty, the four GCMs contribute the most, followed by the downscaling methods and the emission scenarios. The study is well-designed, and the manuscript is well-written. The topic aligns well with the scope of the HESS journal, and I consider it suitable for publication, however, after addressing a few at least moderate comments:

We sincerely thank the reviewer for taking the time to review our manuscript and for the positive assessment.

1. Use of outdated climate projections: The study relies on hydrological projections driven by CMIP5 climate projections. While I understand that the study makes use of existing projections, it raises the question of whether these findings would hold with the latest CMIP6 climate projections, which represent the current state of knowledge on this topic. Although replicating this analysis with CMIP6 projections will be a major task, could the authors at least compare the CMIP6 and CMIP5 projections (without downscaling or running the hydrological model) from the same four models? This would provide insight into the similarities and differences between the two projections.

We acknowledge that newer climate projections are available through CMIP6 but the dataset used here is the most up-to-date national hydrological projections available for Australia and continues to be widely used by end users. The projections were not developed as part of this study but rather were a major effort by the Australian Bureau of Meteorology so it is not possible for us to replicate the simulations with CMIP6 (the CMIP6 version is in development but will not be available for the foreseeable future). As such, these projections are not "outdated" but remain actively used in Australia.

Ukkola et al. (2020) showed that while CMIP6 projections are slightly more robust for future meteorological drought (using similar metrics to this study), they are qualitatively very similar to CMIP5. This is the case both when comparing CMIP6 to equivalent CMIP5 models and the full CMIP5 ensemble. It is thus unlikely that the precipitation drought changes presented here would be substantively different using CMIP6. We also show below (comment #3) that AWRA-L evaporation projections follow precipitation closely so it is unlikely evaporation would be substantially different (and therefore runoff and soil moisture) despite the tendency for greater temperature increases in CMIP6 compared to CMIP5.

We also do not feel it would be particularly insightful for compare the "same" four CMIP6 models as in practice, there can be major differences across the equivalent models as the newer versions reflect multiple years of model development. Furthermore, the models used in the NHP projections were chosen to cover a range of climate futures (hot/dry/wet) as well as their skill in representing Australian conditions (as detailed on L164). The equivalent CMIP6 models would likely not fulfill the same criteria and other models would be chosen instead (as has been done in recent CMIP6-based projections over Australia, e.g. Grose et al., 2023).

2. Novelty of results: The main motivation of this study is to address the existing uncertainties in future drought projections in Australia that "have remained stubbornly uncertain due to a lack of model agreement in projected precipitation changes in most regions". To underscore the study's novelty, it may be helpful to compare the state of knowledge on future droughts in Australia both before and after this study. Including this comparison, perhaps in the discussion section, would enhance the manuscript's contribution to the field. Additionally, discussing the implications of this additional knowledge on decision-making regarding future droughts in Australia would further strengthen the study's relevance and novelty.

We do not want to over-state our results as the NHP projections are only one source of future projections with their own uncertainties. We have made it clearer in the revised manuscript that our results are broadly consistent with previous studies (e.g. Kirono et al., 2020):

> L569: "*These results broadly agree with previous studies that have shown increasing meteorological and agricultural droughts in southern and eastern Australia but the specific regions where projections are robust differ across studies (Eccles et al., 2024; Kirono et al., 2020; Spinoni et al., 2020).*"

Where more robust future changes are identified, these should nevertheless be treated with caution as one additional data point and considered in conjunction with other studies and data sources. As such we do not want to claim that our understanding of future changes in these regions is now clearer based solely on these projections.

Our study is novel in several ways. Firstly, it is to our knowledge the first study that provides insights into future drought across Australia by using a continental-scale hydrological model forced with downscaled and bias-corrected high resolution climate projections data. Previous studies of future drought have primarily focused on GCM outputs, such as changes in rainfall, used high resolution regional climate model outputs or concentrated on specific regions.

Secondly, it is the first study investigating drought metrics in the Bureau of Meteorology's NHP projections that are widely used by end users across Australia. Previous studies using this dataset have largely concentrated on longer-term averages of hydrological variables rather than focusing specifically on dry extremes. We also use a more comprehensive set of drought indicators than most past studies (precipitation, runoff and soil moisture) and analyse these using consistent metrics, allowing a direct comparison across drought types.

Finally, the NHP dataset provides a unique opportunity to identify how different methodological choices influence drought projections to inform future research. This is enabled by the combination of different GCMs, bias correction/downscaling methods and hydrological modelling available in the dataset. These points are discussed in the introduction which we have further revised:

L96: “*Here we investigate future changes in Australian droughts and consider some of these key sources of uncertainty in future projections. We use an ensemble of 32 simulations from the National Hydrological Projections (NHP) collection (Peter et al., 2024; Srikanthan et al., 2022; Wilson et al., 2022) covering the period 1960-2099. The projections were created using the Australian Landscape Water Balance model (AWRA-L; Frost et al., 2018) forced with GCM outputs bias-corrected and downscaled using alternative methods. Using this ensemble provides an opportunity to quantify the uncertainties arising from the choice of GCM, bias correction and downscaling methods and how these propagate into drought projections. We also consider three types of drought (meteorological, hydrological and agricultural) and calculate future changes in these using consistent methods to assess the robustness of future changes across different indicators of drought (time under drought, duration and intensity). These projections are widely used for adaptation and management purposes across Australia but past studies using NHP have largely focused on longer-term averages rather than dry extremes (e.g. Peter et al., 2023), making a thorough analysis of the projected future drought changes timely.*”

3. Diagnosis of uncertainties in hydrological projections: The changes in hydrological projections appear more widespread and severe than changes in precipitation. Further exploration of the causes of those differences would be beneficial. Is this discrepancy due to compounding effects of precipitation and temperature changes, or possibly intraseasonal changes (e.g., shifts in rainfall intensity)? Alternatively, could differences in downscaling methods, particularly their temporal disaggregation schemes, be contributing factors?

We agree with the reviewer that the robust runoff and soil moisture changes tend to be more widespread than changes in precipitation. One contributor is actual evapotranspiration (ET) changes accentuating precipitation-driven changes in runoff and soil moisture. While ET largely follow rainfall changes and is projected to decline in many regions, increases in the mean and variability of ET coincide with increased hydrological droughts in parts of the southeast (in particular eastern Victoria, Tasmania):

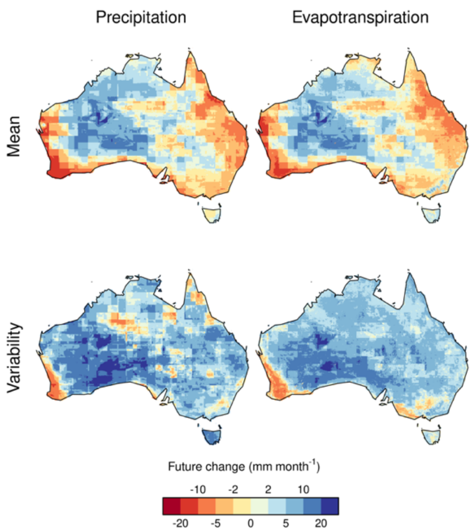

Figure R1: Future changes in mean (top row) and variability (bottom row) of 3- month running mean precipitation and actual evapotranspiration. Variability was quantified from the change in standard deviation of precipitation and ET.

We also found that a larger fraction of precipitation is projected to go into ET in the future in southeastern/eastern regions by quantifying changes in the mean ET/P ratio. This would act to amplify runoff/soil moisture droughts through a smaller proportion of precipitation going into these hydrological variables. The regions of increasing ET/P ratios are particularly evident in the southwestern and southeastern regions coinciding with regions where runoff/soil moisture droughts change more strongly than precipitation:

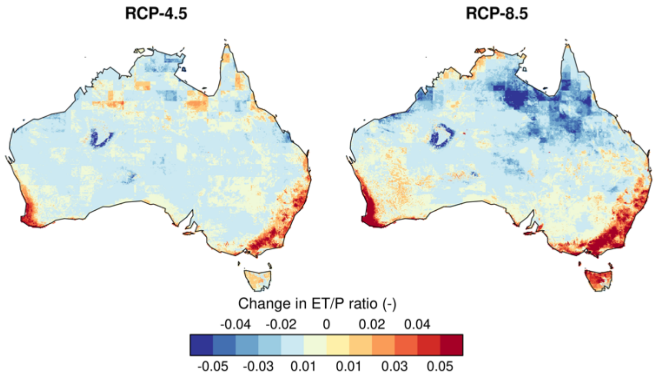

Figure R2: Future change in the evaporative ratio (ET/P). The historical ratio was calculated from the mean monthly ET and precipitation during 1970-2005 and the future ratio during 2064-2099.

Furthermore, because runoff is a smaller flux compared to precipitation and ET, given relative changes in ET and precipitation can yield larger relative changes in runoff (and similarly for soil moisture). As such, changes in ET and/or precipitation can be amplified in runoff and soil moisture. There are several observed examples of this in Australia e.g. from southwestern Australia where a 15-20% declines in rainfall coincided with >40% declines in dam inflows (Petrone et al., 2010). Ukkola et al. (2016) also found that fairly small changes in ET of ~6% led to much larger changes of 25-30% in streamflow.

The downscaling methods are unlikely to contribute as only the meteorological variables are downscaled and AWRA-L then run with these downscaled inputs, making the hydrological outputs consistent with the precipitation data. We also note that a full attribution of changes is not possible as the necessary outputs are not available (e.g. leaf area index that strongly controls ET). Individual ET components (i.e. soil evaporation, transpiration, interception evaporation) are also not available in the NHP collection that would allow us to infer some of the drivers. To reliably quantify the role of rainfall intensity shifts on runoff generation would require factorial simulations that are outside the scope of the study. Rainfall distribution is one of many factors controlling runoff generation and soil moisture recharge in AWRA-L and cannot be separated using the simulations at hand. CMIP6 models suggest a tendency towards higher intensity, lower frequency rainfall events in the future but how these influence runoff generation in the NHP simulations depends on AWRA-L mechanisms including the model's soil moisture state and properties so we cannot simply use the rainfall data to answer this question.

We have added text to the revised manuscript to explain the more widespread runoff drought changes and added Figures R1-2 to the supplementary information (Figs S4-S5 of the revised manuscript):

L334: "*While runoff drought changes show a smaller area of robust changes, they are of greater magnitude across most pixels compared to rainfall (16% points for runoff vs 8% points for rainfall on average across all robust pixels for time under drought; 1.6 months vs 0.5 months for duration and 3.1% points vs 2.2% points for intensity). Soil moisture changes similarly tend to be larger than rainfall changes for duration and intensity (0.8 vs 0.5 months for duration and 2.7 vs 2.2 % points for intensity). These stronger changes partly stem from increases in actual evapotranspiration (ET) (Fig. S4). While ET is largely projected to decline in the southern and eastern parts in line with rainfall, parts of the southeast coast and Tasmania that are projected to experience robust soil moisture and runoff drought increases also show increasing ET in the future. Furthermore, the ratio of ET to rainfall is projected to increase in the southeast where time under drought for runoff and soil moisture is projected to increase more strongly than for precipitation (Fig. S5). This will act to reduce runoff and/or soil moisture even when ET is declining. Both rainfall and ET variability are also projected to increase into the future, increasing the likelihood for drought (Fig. S4).*

Another likely factor in larger runoff drought changes is the amplification of rainfall and ET changes in runoff. Runoff is a smaller component of the water cycle compared to rainfall and ET across most of Australia (the fraction of rainfall partitioned to ET is >50% across >99% of the continent during the historical period). Any given relative changes in ET and rainfall are therefore amplified in runoff, yielding larger relative changes in runoff (and similarly for soil moisture). The relative future changes in mean monthly runoff exceed relative changes in mean monthly precipitation across 93% of pixels when averaged across the pixels showing robust future changes in time under hydrological drought. Similarly, runoff changes exceed those in ET across 99% of the pixels with robust changes. This amplification of runoff changes agrees with past observations of greater relative changes in runoff compared to precipitation (e.g. Petrone et al., 2010) and ET (e.g. Ukkola et al., 2016)."

4. Quantifying contributions of uncertainty in the abstract: It would be helpful to include the relative importance of each source of uncertainty in the abstract. For example, stating that GCMs contribute approximately X% to the uncertainties, and downscaling methods contribute Y%, would help more clearly summarize the results in the abstract. Additionally, it would be valuable to at least comment on the relative contributions of uncertainties from dynamical versus statistical downscaling methods, since statistical methods are more widely used, as they are numerically far less expensive.

We have modified the abstract to quantify the contributions and note that about half of the uncertainty in bias-correction/downscaling methods arises from the choice between dynamical and statistical downscaling:

L29: *GCMs represent the largest source of uncertainty (57-72% of the full range of projections) but the choice of DS-BC method is also important (~25-58%, with approximately half of this uncertainty arising from the choice between dynamical and statistical downscaling).*

As a separate issue, we would also like to note that minor errors were found in the soil moisture data in Figures 2 and 4 (data for the same model was repeated in some cases). These have been corrected leading to minor changes in the results. The skill scores in Figure 5 were also revised as data had also inadvertently been excluded when calculating these.

References

Grose, M. R., Narsey, S., Trancoso, R., Mackallah, C., Delage, F., Dowdy, A., Di Virgilio, G., Watterson, I., Dobrohotoff, P., Rashid, H. A., Rauniyar, S., Henley, B., Thatcher, M., Syktus, J., Abramowitz, G., Evans, J. P., Su, C.-H., and Takbash, A.: A CMIP6-based multi-model downscaling ensemble to underpin climate change services in Australia, Climate Services, 30, 100368, https://doi.org/10.1016/j.cliser.2023.100368, 2023.

Petrone, K. C., Hughes, J. D., Van Niel, T. G., and Silberstein, R. P.: Streamflow decline in southwestern Australia, 1950-2008, Geophys. Res. Lett., 37, L11401, https://doi.org/10.1029/2010GL043102, 2010.

Ukkola, A. M., Prentice, I. C., Keenan, T. F., van Dijk, A. I. J. M., Viney, N. R., Myneni, R. B., and Bi, J.: Reduced streamflow in water-stressed climates consistent with CO2 effects on vegetation, Nature Clim Change, 6, 75–78, https://doi.org/10.1038/nclimate2831, 2016.

Ukkola, A. M., De Kauwe, M. G., Roderick, M. L., Abramowitz, G., and Pitman, A. J.: Robust Future Changes in Meteorological Drought in CMIP6 Projections Despite Uncertainty in Precipitation, Geophys. Res. Lett., 47, https://doi.org/10.1029/2020GL087820, 2020.

Response to Reviewer #2

Ukkola et al. analyse an important topic, the impact of climate change on three drought types in Australia, including their seasonal changes. They use an ensemble of 32 members (4 x 4 x 2) based on the Australian Landscape Water Balance model (AWRA-L), forced with downscaled and bias-corrected 4 CMIP5 models. They used three statistical bias correction approaches and one combined downscaling and bias correction approach for two RCPs (4.5 and 8.5). Their results suggest an overall increase in all three drought types (meteorological, hydrological, and agricultural), particularly in winter and spring. They also attempt to quantify associated uncertainties. The topic of this manuscript is relevant to HESS readership and it is a nice contribution to the community. I find the current manuscript is very suitable for publication in HESS after addressing some minor comments listed below. The paper is mostly very clearly written and well-referenced.

We would like to thank the reviewer for taking the time to review the manuscript and for their positive assessment.

Here are my minor comments:

The abstract does not clearly address, which drought characteristics you quantify. You mention only one aspect in the abstract, L24: "time spent under drought", which could be better referred to as "drought duration"? From the abstract should be clear whether you also considered other characteristics, such as severity or spatial extent. If not, then the title of the manuscript should be adjusted accordingly. Also, L30 " future increases in drought" doesn't specify which aspect of drought is analysed. Also, in L98 in the Introduction, they say "across different indicators of drought", but they don't explicitly mention them. You mention them for the first time on L231ff. Results on drought intensity should be also mentioned in the abstract.

Thank you for pointing this out. We analyse three drought metrics (time under drought, duration and intensity) and have modified the abstract to make this clearer:

> L23: "*We show future increases for all three drought types, with largest increases projected in winter and spring. The sign of the changes is consistent across different drought metrics but projected changes are more robust for the time spent under drought than drought duration or intensity.*"

With this edit, intensity is now also mentioned as requested by the reviewer. We have also modified L98 to include the drought indicators:

> L101: "*We also consider three types of drought (meteorological, hydrological and agricultural) and calculate future changes in these using consistent methods to assess the robustness of future changes across different indicators of drought (time under drought, duration and intensity).*"

We have opted to not change L30 as this statement applies to all drought types and metrics analysed here. The details are provided earlier in the abstract and this statement is to provide an overview of the results.

We note that "time spend under drought" and "drought duration" are not identical metrics. Time under drought is the proportion of months in drought over a time period, whereas drought duration is the length of individual drought events, i.e. the number of consecutive months in drought (as detailed in section 2.2.3). Drought duration describes the sequencing of drought months and could be very different for the same number of drought months depending on how they cluster (e.g. lots of short droughts vs fewer longer droughts). Hence we consider both in the manuscript for completeness but mainly concentrate on time under drought as the key metric to keep the manuscript a manageable length.

Line 109: "model that is calibrated towards observed river streamflow, satellite soil moisture and evapotranspiration across the continent." It is not clear, whether it was done in this study, or they refer to any other previous work,

This was done as part of previous work (as we state on L166, the projections were obtained from the National Hydrological Projections dataset and not developed by us). We have reworded the sentence and added the relevant reference:

> L116: "*AWRA-L is a semi-distributed model that has been calibrated towards observed river streamflow, satellite soil moisture and evapotranspiration across the continent (Frost et al., 2018).*"

Lines 121-124: did you do any of these evaluations, which you could possibly include in the supporting information file?

These have been done as part of previous work with the relevant reference provided (Frost and Wright, 2018). We have provided an evaluation specifically for drought metrics (Figures 5-7) and now also add an evaluation against observed streamflow droughts (see comment below).

Section 2.2.1. on historical observations should include some of the model's evaluations. I fully understand that "gridded runoff and soil moisture observations are not available", but still, you could compare the routed runoff against observed streamflow observations to assess the credibility of your simulations in the historical period.

We have added an evaluation against observed streamflow using a newly-developed CAMELS-AUS v2 streamflow dataset (Fowler et al., 2024). This is the most comprehensive streamflow dataset of unimpaired catchments currently available for Australia. We used data over the period 1970-2020 for the evaluation as was done elsewhere in the paper (Figures 5-7). The data were screened to remove any catchments with >5% of the time steps missing during this period. Gap-filling has already been performed as part of the original dataset using output from a hydrological model. This left 216 catchments available for the evaluation out of the 561 catchments available in the full dataset. The location of these catchments is shown below in Figure R1, noting the uneven distribution of the catchments poses a limitation for this evaluation.

Figure R1: Location of CAMELS-AUS v2 catchments providing observed streamflow used for AWRA-L runoff evaluation.

Figure R2 below shows the evaluation against observed streamflow using the same method as used in Figure 5, splitting the AWRA-L historical outputs by GCMs and DS-BC methods. For duration, the results are consistent with Figure 5 which compares the GCM- and observations-forced AWRA-L simulations, albeit the skill scores are somewhat lower when comparing to observed streamflow. This is not surprising as the streamflow data is an independent constraint unlike the AWRA-L reference simulations used in Figures 5-7. The simulations broadly follow the observed density distribution but overestimate short duration and underestimate long duration hydrological droughts.

For intensity, the model simulations systematically underestimate observed values with a higher frequency of low intensities and a lower frequency of high drought intensity than observed. While this is also evident when comparing to the AWRA-L reference run (Figure S10), the biases are much larger when comparing to observed streamflow. The worse performance for drought intensity is not surprising as to successfully capture this metric, the model not only needs to simulate the occurrence but also the magnitude of events correctly. Similar underestimation of intensity has been found for CMIP5 models for meteorological drought (Ukkola et al., 2018). This implies that projected changes in hydrological drought intensity should be interpreted with caution.

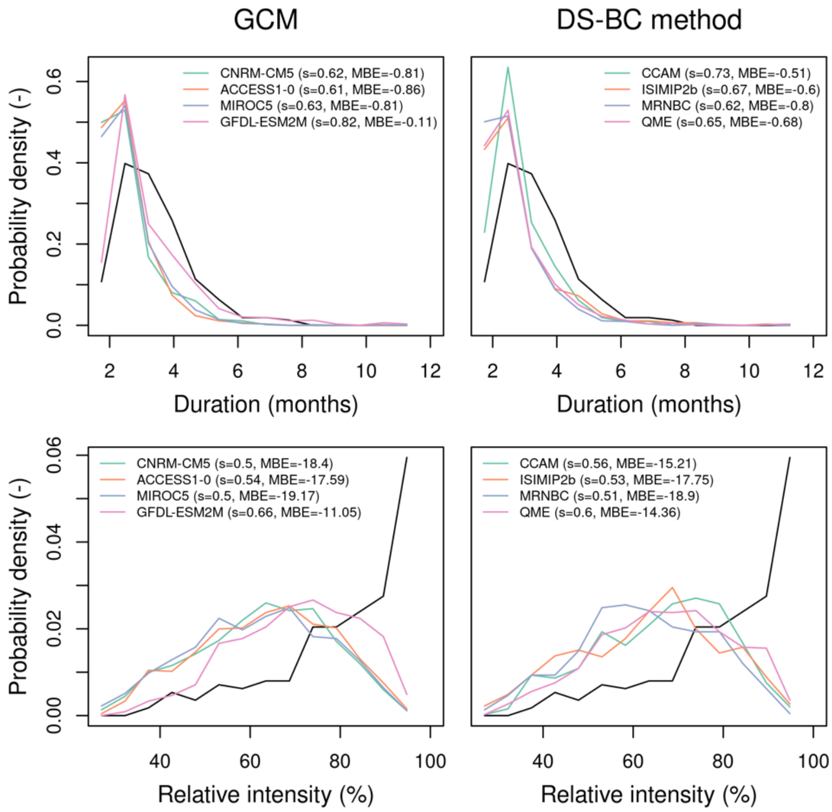

Figure R2: The density distribution of historical runoff drought duration (top row) and intensity (bottom row) across 216 river catchments. Observations from streamflow gauges are shown in black. For each GCM, the data were averaged across the four DS-BC members before plotting. For DS-BC methods, data were averaged across the four GCMs before plotting. Data for 1970-2020 were used to coincide with the observational data, with the historical model simulations extended using RCP4.5. The Perkins skill score (s) and mean bias error (MBE) are shown in the legend.

While these results suggest systematic biases in the average drought metrics, we have shown in a separate study recently submitted to HESS (Grant et al., in review) that the observationally forced AWRA-L reference simulation is able to capture the sign of observed trends in time under drought in the majority of catchments using the same CAMELS-AUS v2 dataset. For the 1981-2020 period, AWRA-L captures the correct sign of the trend (negative, positive, or zero) at 76% of the catchments, and 86% of catchments for the 1951-2020 trends despite lower agreement in the magnitude of trends (Figure R3). This suggests that despite biases in average metrics, the model is able to capture drought trends reasonably well.

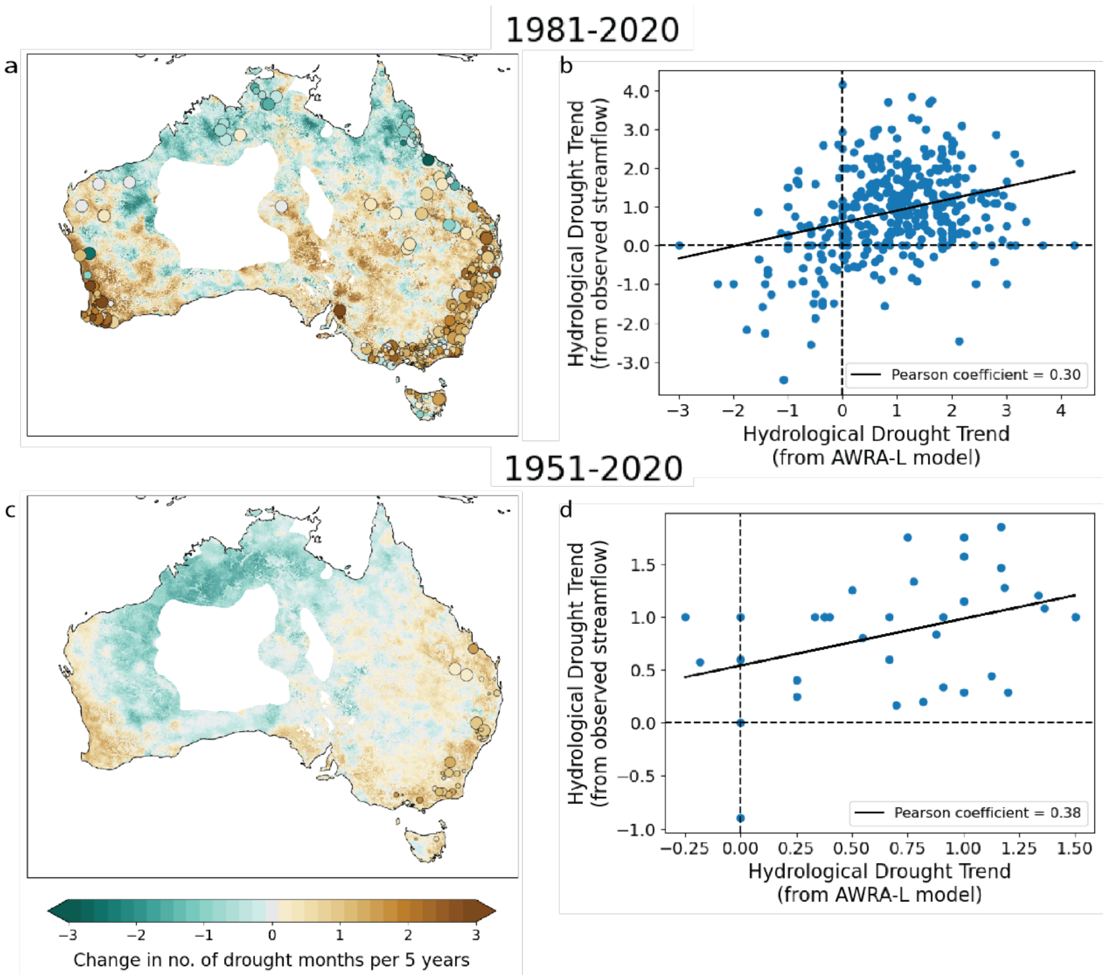

Figure R3: Evaluation of AWRA-L runoff against observed streamflow time under drought trends. Panels a and c show the observed streamflow time under drought trends at the catchments overlayed onto the AWRA-L runoff time under drought trends. Panels b and d show scatterplots of the AWRA-L runoff time under drought trends against the observed streamflow time under drought trends. Both types of plots are shown for 1981-2020 (a-b) and 1951-2020 (c-d) trends. The white spaces on a and c indicate the area masked out due to sparse observation network. (From Grant et al., in review)

We have added text to the Methods to detail the CAMELS-AUS dataset and its processing, as well as Figure R1 (Fig. S1 of the revised manuscript) to show the catchment locations:

> L157: "*We supplement these with streamflow observations from the version 2 of the Australian edition of the Catchment Attributes and Meteorology for Large-Sample Studies dataset (CAMELS-AUS v2; Fowler et al., 2024). The streamflow data were screened to remove any catchments with >5% of the time steps missing during 1970-2020. Gap-filling has already been performed as part of the original dataset using output from a hydrological model. This left 216 catchments available for the evaluation out of the 561 catchments available in the full dataset (see Fig. S1 for catchment locations).*"

We have also added Figure R2 to the supplementary information (Fig. S14 of the revised manuscript) and modified the results:

> L451: "*For runoff droughts, an evaluation against observed streamflow similarly shows that GFDL-ESM2 tends to simulate longer drought duration but it is in better agreement with observations than the other GCMs (Fig. S14).*"
>
> L460: "*This finding is consistent when comparing runoff against streamflow observations, albeit the skill scores are lower especially for MBE (Figure S14). This is not surprising given the streamflow observations are an independent constraint whereas AWRA-L underpins both the GCM-forced and reference runs in Figure 5.*"
>
> L471: "*All GCMs and DS-BC methods underestimate runoff drought intensity (except QME when compared to the AWRA-L reference run; Fig. S13 and S14). The Perkins skill score shows good agreement (≥0.85) with the AWRA-L reference run but is lower when compared to observed streamflow (0.5-0.66).*"

Lines 161-179: was this done by the authors, or taken from other authors, i.e., Peters et al.?

As stated on L160, we obtained the projections from the readily available NHP dataset. We have reworded L161-179 as follows:

> L176: "*We next describe how the NHP projections were constructed. Three statistical bias correction approaches and one combined downscaling and bias correction approach were first applied to the raw GCM data (Peter et al., 2024).*"

Line 197: are you sure that you are able to obtain steady-state conditions of your model's states in 10 years, given the very arid region? I guess in the dry regions, it can take longer… Did you check the time series of selected pixels? I can imagine the disagreement in results (grey in Fin 3) in central Australia could be driven by this factor of too short initialization.

While the GCM- forced runs analysed in this study begin in 1960, the initial conditions come from a much longer reference run forced with observed climatology (beginning 1911, effectively giving an extra ~50 years for spin-up). The GCM forcing has been bias-corrected to the same observations and thus should not result in large jumps in the forcing meteorology. As such we have only removed the first 10 years from the projections. Removing more years would result in a too short a time series to analyse historical drought metrics reliably. A key factor in model disagreement in central Australia is likely the highly stochastic nature of rainfall in these arid desert regions which would also be reflected in the runoff and soil moisture simulations (as we discuss L703 onwards).

Section 2.2.3 You run the model at a daily time step, and then drought analysis is done at a monthly time scale. I guess, you need to state this somewhere explicitly, possibly in this section. And then you apply 3 months averaging. This sequence needs to be stated clearly here in the section. Then, I would suggest moving L206-214 elsewhere because they are a bit distracting where they are. I would start directly with L203-205 and then move directly to L223 onwards. L205-208 could go to discussion.

We have opted to not move L206-214 as these detail the drought threshold used including why it was chosen and are an important part of the methods. This information also directly follows our choice to use percentiles to identify droughts and is thus appropriate in its current section. However, we have slightly shortened this section:

> L221: "*We use percentile thresholds to identify drought periods instead of commonly used metrics such as Standardised Precipitation Index (SPI; McKee et al., 1993) or Standardised Runoff Index (Shukla and Wood, 2008), as the percentile method does not involve assumptions about the data distribution. We use the 15^{th} percentile as the drought threshold such that months below this threshold are classified as drought. This corresponds approximately to the SPI threshold of −1 which is commonly used to characterise "moderate" droughts (McKee et al., 1993). We chose this threshold to identify events that are likely to lead to impacts whilst maintaining a sufficient number of drought events to reliably infer trends in the drought metrics (Ukkola et al., 2020). This method is also similar to that used by the Australian Bureau of Meteorology which uses percentiles in their drought reporting (http://www.bom.gov.au/climate/drought).*"

L216-219 were moved to the discussion (L571-575 of revised manuscript). L206-207 were kept in the methods as they again explain and justify our choice of methodology (i.e. why we have chosen to focus on seasonal droughts).

We also now state that daily variables were converted into a 3-month running mean time series:

> L232: "*To identify seasonal droughts, we first convert the daily precipitation, runoff and soil moisture time series into 3-month running means, such that each month's value is calculated as the mean of that and two preceding months.*"

L231: you could have also analysed spatial drought extent? Do you see distinct results for behaviours in duration and time under drought? If not, then I would suggest keeping just one.

As discussed above, time under drought and duration are not identical metrics and provide different information. Hence we have opted to present both. We do not think changes in spatial drought extent would provide substantially different information compared to the metrics already presented as it would largely follow the country-wide results presented in Figure 2 and the NRM region specific results in Figures 6-7 (we assume the reviewer means the commonly used "area under drought" metric). As the paper is already fairly long, we have opted to not add additional metrics.

In results, the results quantify the drought types. It would be interesting, which aspects lead to the runoff droughts, which seem to be by 20% longer, it's not only because of precipitation deficits, but surely from evaporative increases due to increased temperature? Also, in Fig.2, there are the reference values missing, to better relate the percentage increase to a reference. The 20% would be different from 2 months or from 4 months … ?

Figure 2 shows that the robustness of runoff droughts is similar to precipitation (e.g. 17% of land area shows robust runoff drought changes vs 20% for rainfall under RCP4.5) but the reviewer is correct that the magnitude of changes tend to be larger for runoff.

One contributor is actual evapotranspiration (ET) changes accentuating precipitation-driven changes in runoff and soil moisture. While ET largely follow rainfall changes and is projected to decline in many regions, increases in mean ET and ET variability coincide with increased hydrological droughts in parts of the southeast (in particular eastern Victoria, Tasmania):

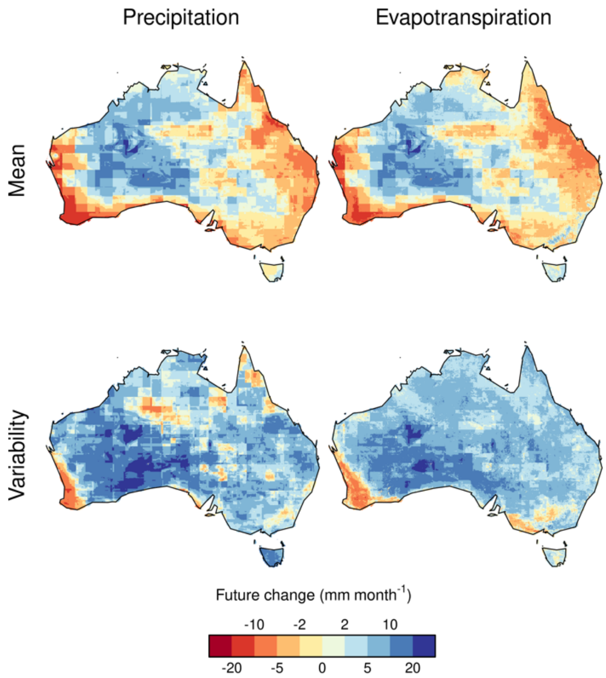

Figure R4: Future changes in mean (top row) and variability (bottom row) of 3-month running mean precipitation and actual evapotranspiration. The future change was calculated as the difference in mean/variability between 2064-2099 and 1970-2005. Variability was quantified as the standard deviation of monthly precipitation and ET.

We also found that a larger fraction of precipitation is projected to go into ET in the future in parts of Australia by quantifying future changes in the ET/P ratio. This would act to amplify runoff/soil moisture droughts through a smaller proportion of precipitation going into these hydrological variables even if ET is declining. The regions of increasing ET/P ratios are particularly evident in the southwestern and southeastern regions coinciding with regions where runoff/soil moisture droughts change more strongly than precipitation:

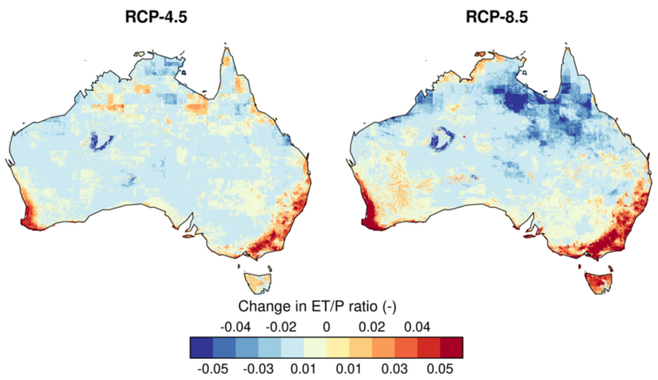

Figure R5: projected future change in the mean evaporative ratio (ET/P). The historical ratio was calculated from the mean monthly ET and precipitation during 1970-2005 and the future ratio during 2064-2099.

Furthermore, because runoff is a smaller flux compared to precipitation and ET, given relative changes in ET and precipitation can yield larger relative changes in runoff (and similarly for soil moisture). As such, changes in ET and/or precipitation can be amplified in runoff and soil moisture. There are several observed examples of this in Australia e.g. from southwestern Australia where a 15-20% declines in rainfall coincided with >40% declines in dam inflows (Petrone et al., 2010). Ukkola et al. (2016) also found that fairly small changes in ET of ~6% led to much larger changes of 25-30% in streamflow.

We note that a full attribution of runoff changes is not possible as the necessary outputs are not available (e.g. leaf area index that strongly controls ET). Individual ET components (i.e. soil evaporation, transpiration, interception evaporation) are also not available in the NHP collection that would allow us to infer some of the drivers.

We have added text to the revised manuscript to provide an explanation for the more widespread runoff drought changes and added Figures R4-5 to the supplementary information (Figures S4-S5 of the revised manuscript):

> L334: "*While runoff drought changes show a smaller area of robust changes, they are of greater magnitude across most pixels compared to rainfall (16% points for runoff vs 8% points for rainfall on average across all robust pixels for time under drought; 1.6 months vs 0.5 months for duration and 3.1% points vs 2.2% points for intensity). Soil moisture changes similarly tend to be larger than rainfall changes for duration and intensity (0.8 vs 0.5 months for duration and 2.7 vs 2.2 % points for intensity). These stronger changes partly stem from increases in actual evapotranspiration (ET) (Fig. S4). While ET is largely projected to decline in the southern and eastern parts in line with rainfall, parts of the southeast coast and Tasmania that are projected to experience robust soil moisture and runoff drought increases also show increasing ET in the future. Furthermore, the ratio of ET to rainfall is projected to increase in the southeast where time under drought for runoff and soil moisture is projected to increase more strongly than for precipitation (Fig. S5). This will act to reduce runoff and/or soil moisture even when ET is declining. Both rainfall and ET variability are also projected to increase into the future, increasing the likelihood for drought (Fig. S4).*

Another likely factor in larger runoff drought changes is the amplification of rainfall and ET changes in runoff. Runoff is a smaller component of the water cycle compared to rainfall and ET across most of Australia (the fraction of rainfall partitioned to ET is >50% across >99% of the continent during the historical period). Any given relative changes in ET and rainfall are therefore amplified in runoff, yielding larger relative changes in runoff (and similarly for soil moisture). The relative future changes in mean monthly runoff exceed relative changes in mean monthly precipitation across 93% of pixels when averaged across the pixels showing robust future changes in time under hydrological drought. Similarly, runoff changes exceed those in ET across 99% of the pixels with robust changes. This amplification of runoff changes agrees with past observations of greater relative changes in runoff compared to precipitation (e.g. Petrone et al., 2010) and ET (e.g. Ukkola et al., 2016)."

As for Figure 2 reference values, this figure shows the ensemble mean future change in time under drought relative to the historical baseline (i.e. the fraction of time under drought). As we state on L300, the reference value is 15% as per our definition of droughts as months below the 15th percentile:

"*Fig. 2 shows the ensemble mean future change in time under drought relative to the historical baseline (during which ~15% of the time is under drought as per our definition).*"

We have also added this information to the Figure 2 caption. Reference values for drought duration and intensity are presented together with the future changes in Figures S2-S3.

It might be useful to rearrange the sequence of the results. How about starting with 3.4, where observations are compared, and show basic characteristics of individual realization (Fig 5), then showing aggregated characteristics of drought for the full ensemble (Figs 2 and 3)…

We prefer keeping the current order as Figures 2 and 3 present the overall continent wide results. The subsequent results are then to explore the specific sources of uncertainty (e.g. GCMS vs bias correction/downscaling methods) and the reliability of these projections (through a comparison to observations). We feel presenting the aggregated results first gives necessary context to the subsequent sections and allows us to discuss model biases in the context of the future changes.

Why does the GFDL model stand so much apart? Is it because of precipitation or temperature differences?

Yes GFDL tends to project hotter and drier conditions in the future compared to the other GCMs. We discuss this in section 4.2:

L626: "*we found that ensemble members using GFDL-ESM2M as the forcing model were particularly anomalous compared to the rest of the NHP ensemble, indicating stronger increases in most regions. [..] The GFDL-ESM2M model projects greater future warming and drying over Australia than the other GCMs used here (Peter et al., 2023); our finding of larger drought increases in GFDL-ESM2M are consistent with this tendency*."

Nicely written discussion section, but could the conclusions be taken apart into one paragraph section at the end?

We have added a conclusions section to the revised manuscript:

> L715: "*We identify robust future increases in meteorological, hydrological and agricultural droughts across 20-29% of the Australian land surface under the RCP4.5 scenario and 38-48% under RCP8.5. The future changes are particularly robust in the highly populated and agricultural regions across southern and eastern Australia, with largest increases projected to occur in winter and spring. The ensemble members consisting of different combinations of GCMs and downscaling/bias correction methods broadly agree on increasing drought in southern and eastern Australia but the robustness of future projections vary by region. Uncertainty is particularly high in the monsoonal and tropical climates of northern Australia and the arid interior. Our results are in broad agreement with previous studies but future work is needed to systematically assess the differences in the magnitude and spatial extent of robust changes across different projections.*
>
> *When averaged across the broad NRM regions, the sign of future change is largely consistent across the different ensemble members but the magnitude of future change is strongly dependent on the choice of GCM and DS-BC method. This choice between statistical and dynamical downscaling represented the primary reason for differences among the DS-BC methods, with the choice of statistical bias-correction method (ISIMIP2b, QMR or MRNBC) only leading to relatively small differences in historical skill and future projections. Overall, we found that the GCMs are the dominant source of uncertainty but in some cases the choice of the DS-BC method can also change the sign of future change (Fig. 6). This suggests the choice of GCM is particularly important for adequately quantifying uncertainty in the future projections*."

As a separate issue, we would also like to note that minor errors were found in the soil moisture data in Figures 2 and 4 (data for the same model was repeated in some cases). These have been corrected leading to minor changes in the results. The skill scores in Figure 5 were also revised as data had also inadvertently been excluded when calculating these.

References

Fowler, K. J. A., Zhang, Z., and Hou, X.: CAMELS-AUS v2: updated hydrometeorological timeseries and landscape attributes for an enlarged set of catchments in Australia, Earth System Science Data Discussions, 1–21, https://doi.org/10.5194/essd-2024-263, 2024.

Ukkola, A. M., Pitman, A. J., De Kauwe, M. G., Abramowitz, G., Herger, N., Evans, J. P., and Decker, M.: Evaluating CMIP5 Model Agreement for Multiple Drought Metrics, Journal of Hydrometeorology, 19, 969–988, https://doi.org/10.1175/JHM-D-17-0099.1, 2018.

Grant, M. O., Ukkola, A. M., Vogel, E., Hobeichi, S., Pitman, A. J., Borowiak, A. R., and Fowler, K.,: Historical trends of seasonal droughts in Australia, *Hydrology and Earth System Science*, in review, https://doi.org/10.5194/egusphere-2024-4024.