GMD-2023-07 Anonymous Reviewer 1:

The work to produce these projections represents a mammoth effort with a large interdisciplinary team. This project/data set will have a lasting legacy with a large number of potential applications, and I hope the potential benefits of using this data set are realised by the community so as to reward the authors on their hard work. The projections have been well described and evaluated extensively. The choice of GCMs and downscaling methods has been done with care and are appropriate to the outcomes of the project. Finally, the projections themselves are of great interest to the hydrologic (among others) community. Please see my comments below which at first may seem extensive but are relatively minor and can often be treated as suggestions rather than being prescriptive.

Thank you for this comment. It is very much appreciated.

General comments:

Some of the figure legends/axes were a bit small and hard to read.

Many of the figures have been updated and/or redrafted. In particular, some figures have been removed from the main figure and placed in a supplementary figures section. This applies to Figures 2, 3, 8, 11, 12 and 13. In addition, high resolution EPS format figures have been supplied for publication.

The abstract describes succinctly the product development and how it was performed. I wonder if a sentence at the start of the abstract on the need for the product might help with context.

We have modified the second sentence to: "The NHP aimed to provide nationally consistent hydrological projections across jurisdictional boundaries to support planning of water-dependent industries. NHP is complementary to those previously produced by federal and state governments, universities, and other organisations, for limited geographical domains."

I know it was mentioned, but it just wasn't quite clear to me how the 9am-9am data from Australia was matched to the GCM data (which I assume is 12am-12am).

The GCM data are simulations that are not based on observations for any given day in particular, such that the timing of an individual event in the simulations (such as a rain-bearing storm, an El Nino event, etc.) is not intended to relate to the timing of a similar individual event in observations. However, the climatology (e.g., average statistics based on long-term data aggregation) can be matched between the GCM and observations data, as is done in this study, as this doesn't rely on an individual day in the model data being matched to an individual day in the observations. This means that for this study's purpose, it doesn't matter what time of day the daily rainfall observations are based on, as long as the model and observations data both use daily aggregation, which is the case.

We tried to explain how we reconciled the timing of the observations with that of the projections in Section 3.1. Since we are dealing with *projections* (as opposed to predictions), the temporal synchronicity (at least at daily timescale) between the model and of observations is of little

relevance. However, the following sentence (from our paper), hopefully explains it best: "Note that, while observations are made from 9 am to 9 am, the bias corrections are calculated by calendar day."

Page 5, Line 15: Clarification for me please - Is it usual to only use SSTs as the forcing from the GCM in CCAM? I understand CCAM doesn't have lateral boundary conditions making it quite unique – is my understanding correct?

You are correct, the CCAM model has a stretched grid with a focus over a region of interest, so is more an actual GCM, rather than an RCM, which uses as its forcing the output of the GCM for its lateral boundary conditions.

CCAM can be run in two major configurations. The first uses spectral nudging of the winds to force CCAM back to those generated by the host GCM and in that configuration tends to follow the projections from the host GCM more closely. In the configuration used for NHP, the forcing from the GCM comes from bias correcting the SST's and CCAM is allowed to develop its own climate and the subsequent projections can be quite different from the host GCM.

We have modified the text (P.5, L8-12):

" CCAM can be forced via two methods, one using bias-corrected sea surface temperatures (SST's) and another using spectral nudging to update the circulation to that of the host GCM. It was the former configuration used for NHP, whereby the mean and variance SSTs of the host GCM are bias corrected to provide boundary conditions for CCAM to produce 50 km resolution projections of the atmospheric state over the Australian continent (Clarke et al., 2019; Hoffmann et al., 2016)."

Section 2.2: I don't think the authors should change their text, but as a comment, it felt the GCM selection was given 1-2 lines of attention on Page 4 and then two pages of attention was given to how the GCM projections fit within the ensemble of GCMS. This felt a little unbalanced to me. I understand it is important to show the spread of possible futures and how this ensemble covers it, but some text sounds like the authors justifying that 'only' four GCMs are sufficient. In particular on Page 7, Lines 9-15 almost seem to defensive to me, and I don't see a reason why the authors need to defend 'only' four GCMs when they do in fact represent a good range plausible future. Moreover, I think the authors analysis is superior for the fact that they considered the best GCMs for Australia (rather than using all GCMs blindly).

Thank you for this comment. We tried to emphasise that GCMs spanned a broad range of the phase space of the GCMs recommended by CCiA and moreover, the complete range of CMIP5 GCMs.

Page 5, Line 21: It feels odd to state the method not used was spectral nudging when the method that was used wasn't stated?

Please see the answer to the previous comment (Page 5, line15) above.

Page 7, Line 9: “uncertainties are underestimated” – which uncertainties? Should this be “uncertainties in the GCM choice are underestimated”?

We have modified the sentence to:

" Nevertheless, the spread of the NHP ensemble is less than that of CCiA, suggesting that the NHP GCM ensemble spread may be less than the CCiA GCM recommendations."

Page 7, Line 25: “calibrate the GCM output” I would prefer the word calibrate to not be used, also calibrating data doesn’t quite make sense. Can this be reworded please?

"Calibration" changed to "bias correct".

Page 8, Line 20. Up to the authors if they want to keep this sentence, but AWAP is gridded and will by definition underestimate point data. It is true also if one looks at catchment averages for extremes AWAP is slightly biased down but (to me anyway) the differences aren’t great. See Figure 5 in Nathan, R., Jordan, P., Scorah, M., Lang, S., Kuczera, G., Schaefer, M., Weinmann, E., 2016. Estimating the exceedance probability of extreme rainfalls up to the probable maximum precipitation. J. Hydrol. 543, 706–720. https://doi.org/10.1016/j.jhydrol.2016.10.044

We wanted to point out some of the biases of the AWAP data set. We modified the sentence: " Furthermore, in areas of steep topography…" to " Furthermore, in areas of steep topography and a sparse gauge network,…", to emphasise that the AWAP gridded analysis is subject to increasing disparity between point observations in data-sparse regions and topographically complex regions.

Page 9, Line 25: The description sounds more like downscaling “modify coast-scale GCM projections at a finer scale” rather than bias correction. Maybe some rewording in this paragraph would be appropriate?

Anonymous Reviewer 3 made the point that using bias correction techniques for down scaling should be a key point. However, we are hesitant to use the term "downscaling". Please refer to our response to their point (page 3, line 24). The bias correction does provide provide fine-scale information compared to the GCM/RCM input, so we think this is appropriate terminology.

Page 12, Line 24: When you say decreases the warming signal it sounds like it has decreased the trend, but to me to the trend before and after bias correction of the CCAM data (brown and blue lines) is identical? So maybe some rewording here is necessary. See the comment below.

Sentence modified to: "…the application of ISIMIP2b has generally decreased the warming bias…"

Page 12, Line 27: Maybe I am taking exception with the word “signal”. That implies to me some sort of temporal trend, but here you just talking about the GCM being wetter, which isn’t a signal, it’s just bias. Apologies about the long comment – maybe just changing the word from signal to bias would be beneficial?

Agreed, we have changed "signal" to "bias" (in this section)

Figure 8: Because you summarise seasonal results in Figure 9 and Figure 10, Figure 8 could just have the annual results only to make the figure more manageable? I know Vogel et al., 2022 has an extensive evaluation of the bias correction, but I think one figure just for one variable (say runoff) with all the bias correction methods would really be beneficial (can just be for one GCM) – given the amount of time spent outlining the bias correction methods (and their potential impact on the results).

We have removed the seasonal plots from this figure and placed them in the supplementary figures section.

Page 14, Line 32 (and elsewhere): Sometimes precipitation is pr, and sometimes it is Pr (in italics).

All instances of Pr changed to "pr". The same has been done for qtot, e0, etot, so, ss and sd, where sometimes we had used capitals with subscripts.

Figure 11: Bottom panel missing units on the y-axis?

The units have been included for the bottom panel.

Page 16, Line 19: Did you use the wet and dry season? I think Figures 13 and 14 just use the regular DJF etc seasons?

We didn't show the wet-dry season as this NRM (SSWF) is in a temperate climate zone and subject to the four seasons. We used the wet/dry season analysis to construct analyses for the Assessment Reports (https://awo.bom.gov.au/about/overview/assessment-reports) described in Section 8.3.

Page 16, Line 9: Not sure, but I know of work that found that the bias correction method was the greater contributor to the ensemble spread. Not sure if the authors have comments on why the different results? See Wasko, C., Guo, D., Ho, M., Nathan, R., Vogel, E., 2023. Diverging projections for flood and rainfall frequency curves. J. Hydrol. 620, 129403. https://doi.org/10.1016/j.jhydrol.2023.129403

Was that because that evaluation was for extremes? We have shown for the mean state.

Figure 15: Given that AWRA-L is a water balance model, has it been evaluated for extremes and if not can a comment be made on its applicability for this purpose. The above manuscript and the following found a possible underestimation of extremes or the change signal in changes for extreme events. Ho, M., Nathan, R., Wasko, C., Vogel, E., Sharma, A., 2022. Projecting changes in flood event runoff coefficients under climate change. J. Hydrol. 615, 128689. https://doi.org/10.1016/j.jhydrol.2022.128689

Yes, an evaluation of the ability of the AWRA-L model's ability to simulate extremes and climate variability has been undertaken, coming to the conclusion that AWRA-L is reliable and accurate enough to be able to simulate the wide range of plausible projected outcomes (Azarnivand et al., 2022).

Section 7.3: I wonder if the “maps” came first (Section 7.3 was Section 7.1), then it would make an easier transition to Section 7.1 and Section 7.2. Looking at the maps, you see the strongest change in SSWF and then you can drill down on the results for that region. My other concern with just focussing on the JJA season. Most rainfall occurs in the summer in the tropics so the results presented here aren’t as meaningful as they could be – I guess I would prefer these maps to be annual – and to be the first item displayed in Section 7. This would also follow better as again Section 7.4 focuses on SSWF.

We followed the structure that was presented in the Assessment Reports, in particular Section 4 of the reports (e.g. https://awo.bom.gov.au/about/overview/assessment-reports#regionsandreports). We appreciate the suggestion of the reviewer, however, we wanted to demonstrate the lines of reasoning that were followed in constructing the assessment reports for the Australian Water Outlook.

Section 8: Am I right in saying that temperature projections are not available as part of the Australian Water Outlook Service but are available on NCI? I feel temperature is an important variable for example when calculating fire risk, and one that many other users would be interested in.

Yes, that is correct. The AWO service was designed with water availability its central focus. However, we of course hope that the bias corrected climate variables will be used by others for differing impact studies. Temperature and other climate variables are available as application-ready datasets as per NCI Data Collection.

Section 8.3: Line 27 confused me a bit – would it be better to have a link to the reports here (instead of the end of Section 8.3)?

We tried to emphasise how the assessment reports build on the foundational work performed in CCiA. We think that the confusion is due to the footnote in line 27 being to the CCiA reports. First we describe the construction of the assessment reports and then provide a link as a footnote. As a compromise, we have made the link to the assessment reports explicit in the text rather than providing them as a footnote.

Section 9: The first paragraph could almost be removed, and the section relabelled “Limitations”.

Thank you for this comment. We have deleted the first paragraph (which was a rehash of the uncertainties of the NHP data set) and relabelled this section "Limitations of the hydrological projections".

Page 22, Line 7: I wonder if “due to time and personnel constraints” could be rephrased with “due to the large spatial domain…” it is clear (to me anyway) that you couldn’t be expected to use more GCMs than you already have due to the large domain and sheer scale of the project.

Yes, although it was a factor (time and personnel constraints) we like the recommendation of the reviewer. Furthermore, another reviewer (number 2) mentioned a similar issue.

Editorial:

Page 2, Line 11: "…south-east with changes in streamflow typically…" might read better.

Changed as recommended.

Page 3, Line 5: missing a space after the reference.

Modified.

Page 5, Line 24: Doesn't have to be bold and can be part of the paragraph.

Thanks for spotting. Text formatted as "normal" rather than "heading 2".

Page 7, Line 9: extra new line.

Modified.

Page 19, Line 22: Change from 3rd person to 1st person with "we". Could revert to be consistent with the rest of manuscript.

Modified to use "we" as recommended.

Page 20, Line 23: I think there is a track changes mark under the apostrophe in "model's".

This paragraph (and the rogue track changes mark) was deleted as per the reviewer's previous recommendation. (See #Section 9 comment).

Page 21, Line 12: CO2 (subscript the 2)

Modified.

References:

Azarnivand, A., Sharples, W., Bende-michl, U., Shokri, A. and Srikanthan, S.: Analysing the uncertainty of modelling hydrologic states of AWRA-L – understanding impacts from parameter uncertainty for the National Hydrological Projections., 2022.

Clarke, J., Grose, M., Thatcher, M., Hernaman, V., Heady, C., Round, V., Rafter, T., Trenham, C. and Wilson, L.: Victorian Climate Projections 2019 Technical Report., 2019.

Hoffmann, P., Katzfey, J. J., McGregor, J. L. and Thatcher, M.: Bias and variance correction of sea surface temperatures used for dynamical downscaling, J. Geophys. Res. Atmos., 121(21), 12,877-12,890, doi:10.1002/2016JD025383, 2016.