This article introduces the development of a national hydrological projections (NHP) service for Australia, including the choice of GCMs and RCM, application and evaluation of three bias correction methods, and driving the Bureau's landscape water balance hydrological model (AWRA-L) to produce hydrological projections. This national hydrological service provides valuable information on the impact of climate change on hydrological cycles over Australia to end users. The overall structure of the manuscript is coherent while wordsmithing is necessary, especially in the first half of the article. Besides, I have a few comments and suggestions for authors to consider.

Thank you for the recognition of the NHP service.

Specific comments

Page 2 Line 22-26: It is mentioned here that Australian states may prefer to use their own downscaled projection products. Key issues are that data are too heterogenous for use across intersect jurisdictional boundaries, and clear instructions are not provided. These issues are addressed in the NHP service, but are users in these states now tend to use your products rather than use state operated ones? Could you give some insights into this point?

Thank you for this question, it is an important one. We don't consider the NHP projections to be a replacement for those produced by states (or other jurisdictions) but rather a supplement. As we explained in the paper, the current projections are too heterogeneous for use across boundaries and the current projections often stop at those boundaries.

There has been uptake of the NHP data sets, in particular the Energy Sector Climate Initiative (ESCI), which was initiated to provide information on energy security by the Australian Energy Market Operator (AEMO), CSIRO and The Australian Bureau of Meteorology. Two case studies were investigated by ESCI: projections of runoff for hydroelectricity production (Hydro Generation (climatechangeinaustralia.gov.au)) and projections of soil moisture for infrastructure (Soil moisture & infrastructure (climatechangeinaustralia.gov.au)). These are examples where state projections would have not been able to supply the necessary data, since the National Energy Market (NEM) operates in Queensland, New South Wales, Victoria, Australian Capital Territory, South Australia and Tasmania (see AEMO | National Electricity Market (NEM)).

The NHP data was also used by the Western Australian government, who wanted to investigate projections of runoff in the Pilbara region. The Western Australian government does not have local hydrological projections and so utilised the NHP data sets. In consultation with water managers, conditions for "drought" and "wet" years were defined and the NHP data used to identify changes at the catchment scale. An example is shown in the figure below:

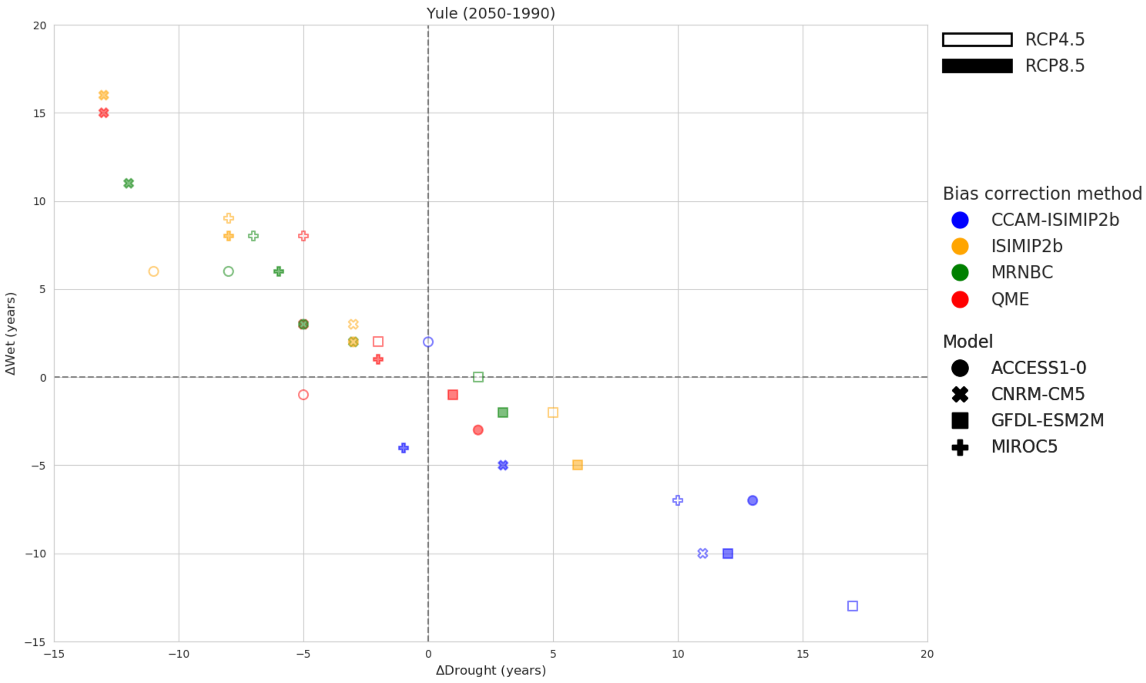

Figure 1: The NHP based assessment of change in drought frequency vs change in wet year frequency for each model (markers), bias correction method (colours), and emissions pathway (marker fill) by 2050 for the Yule catchment.

Using this information, it was shown that changes in drought frequency and changes in wet year frequency are highly correlated at the catchment scale.

The above are two examples of how the NHP data has filled the gap that would not be enabled by the current state-based projections. However, we hope that users who have access to state-based hydrological projections will use the NHP data sets to expand their projections.

Page 3 Line 1: I could not fully understand what transient projection means. Please give a clearer definition and/or example. Also, repeated word 'applied' in footnote 1.

Previous projections (using scaling methods) have only been provided for a discrete time corresponding to the length of the observational baseline period (for example 30 years), rather than continuous projections up to 2100. We have replaced "transient" with "continuous".

Repeated word "applied" has been deleted.

Page 3 Line 24: According to the context, you simply interpolate GCM and RCM to 5km spatial resolution before applying the bias corrections. I reckon using bias correction techniques for downscaling should be the key point here.

We are hesitant to use the word "downscaling" when applying bias correction. Although the term is used, we consider downscaling to either be statistical or dynamical. In the former, statistical relationships between the large-scale GCM output (e.g. 500 hPa winds) and the historical observations are derived and subsequently applied to the GCM projections to downscale the GCM projections to a local scale. Dynamical downscaling is the use of an RCM to produce the finer scale projections. It is true that the bias correction has provided finer-scale resolution and picked up important features, in particular, cooler temperatures in elevated topography in the Great Dividing range and various coastal rainfall features. However, the output of RCMs will still have biases (as demonstrated by the CCAM output for NHP). The first part of this point (referring to downscaling) was in reference to the use of CCAM, not the use of bias correction a downscaling technique.

Another point I am interested in is whether you have tried only applying simple mean (additive or multiplicative) correction to the GCM outputs to drive hydrological model, and using sophisticated methods to correct hydrological outputs. What is your rationale of bias correcting climate outputs prior to driving the hydrological model? Even though the multivariate bias correction accounts for the inter-variable, temporal and spatial structure of the model outputs, the bias adjustment process may have changed temporal features of the model series.

It is well documented that a large proportion of the biases along the projection impact modelling chain come from the GCM data itself (Azarnivand et al., 2022; Bosshard et al., 2013; Dobler et al., 2012; Giuntoli et al., 2015; Joseph et al., 2018). Therefore, it makes sense to bias correct GCM data rather than impact model outputs. Secondly, a full (both spatially and temporally) land surface dataset is needed to bias correct the hydrological outputs which is not available for Australia. Thirdly, the available options for hydrological reanalysis across Australia to bias correct the outputs do not match in the AWRA-L model historical dataset both in accuracy and reliability (Frost et al., 2018). Even so, we have evaluated the AWRA-L model's ability to simulate extremes and climate variability, 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).

With regard to changing the temporal characteristics (for instance, wet-dry spell length), we are currently examining the NHP data to investigate how the bias correction has modified these. We hope that others examine these features, especially those related to "extremes" (by which we mean the upper percentiles) to determine how the bias correction has modified these.

Page 4 Line 8: This is the first time AWRA-L model is mentioned in the introduction. I think more descriptions of AWRA-L are needed in this section because the choice/development of hydrological model is definitely an important part of the NHP project.

The introduction served to outline the motivations for initiating and developing the NHP project. In the dot points prior, we outline the major decisions made, in a sequence that we developed and applied during the various stages of the NHP project. We did consider the use of other hydrological models (e.g. GR4J), however, the AWRA-L was settled on, primarily because the expertise was available at the Bureau to run it and it also the Bureau's operational model. We think that outlining the major decisions (as in the dot points) and then describing them in detail is appropriate.

Page 4 Line 21-26: In line 24, what does ‘variation between CMIP5 models’ mean here? The temporal variance and climatological mean? Regarding the GCM selection, I would like the authors to explain more about how you narrow down the selection from 8 to 4 GCMs. You mentioned that all required variable data are available among 47 CMIP5 models, and CCiA recommended 8 models. What are your criteria to choose these four CMIP5 models out of eight. In addition, I am curious why you include an RCM to increase the ensemble range, and four RCM simulations are only corrected using one bias correction technique. Why not simply include other four GCMs recommended by CCiA?

"Variation between CMIP5 models" means those that were considered to range between hot/cool and wet/dry projections. We have modified the sentence to read:

"… as well as to provide a reasonable representation of the wet/dry and cool/warm variation between CMIP5 models…".

Figures 2 and 3 are too small. Please consider redo them into a 4 rows × 4 cols plot. Figure 8 is also too small. Please consider split it into two or more plots.

We have modified Figures 2 and 3, to only show RCP8.5 and moved the RCP4.5 figures to Supplementary Figures.

Page 6 Line 26: What is your rationale of calling these four GCMs a 'reasonable' subsample of the CCiA ensemble? Please specify.

We have removed the word "reasonable".

Page 7 Line 25: The bias correction methods, ISIMIP2b and MRNBC, are trained over 1976-2005. Is it because the wind speed observations start from 1975? The QME method is trained over 1975-2017, which is 13 years longer. Please clarify and comment on to what extent the use of different training period would affect the bias corrected climate variables, and further the hydrological projections.

One of the main aims for producing hydrological projections in this study was to use a 30-yr time period to help capture more detail of the long-term variability than a 20-yr time period would. The starting year of 1976 was chosen as some studies have shown there was a climate shift that occurred in Australia (particularly in the south-west of the continent) in the mid-70s (e.g. Hope et al., 2010).

The time-period of 1975-2017 for calibration of the QME was used to capture some recent extreme meteorological events (e.g., the Black Saturday bushfires in 2009 and the Queensland floods in 2010/2011) including with this method originally being designed to have a key focus on details for extreme cases. A thorough evaluation of these contrasting bias correction methods and a ranking against several criteria has recently been published by Vogel et al. (2022). For instance, the QME performed better than the other two methods when evaluated against 5-year maxima and extreme percentiles (but less good in some cases for other metrics such as mean runoff), which may relate to several aspects of the QME algorithm and time periods used. For example, the QME method was applied for 3-month seasons as part of its aims around maximising the sample size for extremes, as a complementary approach to the application for individual months for the other methods used in this study). Further details about the choice of the 1976-2005 reference time period can be found in the Australian Water Outlook FAQ page (https://awo.bom.gov.au/faqs/projections).

Page 8 Line 25: Before 1990, daily climatological averages (for each day of the year) are used for solar radiation. How did this affect the training of the bias correction models as the 'true' values are not recorded?

The use of the daily climatological values was a pragmatic decision made due to the non-availability of measurements before this time. It is difficult to know exactly how the use of the climatological values has affected the overall bias correction methods. However, we note that the evaluation of Vogel et al. (2022 see Table 3), indicated that the MRNBC (a multivariate method) performed particularly well when evaluated against hydrological output metrics. This indicates that a multivariate technique that considers the joint marginal dependencies of the distributions (including solar radiation) has performed better than the univariate methods, further indicative that the use of climatological solar radiation values has not been detrimental to the projected hydrological outputs.

Page 12 Line 18-19: This statement could be moved to before Section 4.1, where the data required for the bias correction is introduced.

We have moved this as suggested (now near the beginning of page 10):

"Three statistical bias correction methods were applied to the GCM output and one (ISIMIP2b)[1] to the CCAM output (see **Error! Reference source not found.**). In addition, the original GCM data, spatially interpolated to the resolution of the AWAP, was archived to provide data before the application of bias correction; we refer to this data as "NOBC" and "NOBC-CCAM" for the GCM and CCAM data, respectively."

Page 13 Line 27: Do you train bias correction model using the period from 1976, and apply the trained model to correct climate model simulations from 1960 to 2099? Please clarify this in the text.

Yes, that is what we did. The sentence has been modified:

"To produce the historical AWRA-L hydrological data for the reference period (1976–2005), it was necessary to use historical GCM simulations (bias corrected to the reference period) beginning in 1960."

In Figure 8c, despite small absolute biases, the relative biases for root zone soil moisture over all four seasons are very large compared to other variables. Will the large relative biases of the soil moisture lead to inaccurate information for the community?

Please see our response to the comment below. This was a mistake in the drafting of the figure on our part, where we mixed panels (b) and (c) of Figure 8.

Page 14 Line 17: Figures 9 and 10 are plotted without any interpretations. Please give some comments/explanations on these results. From my understanding, part of Figure 9 shows area-averaged relative biases presented in Figure 8. However, the results of MRNBC-ACCESS1-0 in Figure 9c contradicts those in Figure 8c, where the averaged relative biases should be at least <-10%. Moreover, in these figures, these relative biases or bias values may not be representative because the negative and positive values may be cancelled out in the area averages. I suggest plotting averaged absolute biases without signs across Australia and NRM regions.

This was a mistake in the drafting of the figure. Thank you for spotting this error in the review. The numbers are now consistent, however, there are slight differences. This is due to the different ways the biases were calculated in the spatial plots (Figure 8) compared to the bar plots (Figures 9 and 10). In the former, area-averaged biases were calculated by first calculating the relative bias (%) for each grid cell and then averaging it over Australia. For the bar plots, the Australia-wide mean was calculated first and then the relative bias (%) calculated from the two regionally averaged values.

The manuscript has been modified to reflect this (page 14, lines 20-32 and page 15, lines 1-3).

Page 14: It would be better to show the bar charts for NRM regions (similar to Figure 9 and 10), at least SSWF, in the manuscript or in the supplementary material.

Figures S7 and S8, which are for the SSWF NRM have been included in the supplementary material.

Page 15 Line 23-24: Why not showing the ensemble statistics using the yearly averaged data instead of 30-year running mean? I suspect the range of 10th and 90th percentiles over time will not be too messy.

We show the 30-yr running mean to extract any trends in the data and, in particular, to show differences between the RCP8.5 and RCP4.5 scenarios. We do include and example of one model in the relevant Figure (Figure 11) to illustrate one possible future and the role of year-to-year

[1] Due to time constraints, the only bias correction algorithm applied to the CCAM output was the ISIMIP2b method.

variability. The analysis produced was designed to have correspondence with that of CCiA (CSIRO and Bureau of Meteorology, 2015; see Figure B.6.2.4, Pg. 84). The model ensemble can be a good representation of the projected multi-decadal trends, however, is of limited value to explain yearly variability.

Page 16 Line 9-10: Do the results imply that perhaps only one best performing bias correction technique is needed for your application? Perhaps more GCMs and/or RCMs should be included to better gauge the uncertainty of the future projections.

That may indeed be the case. For instance we found that QME performed best when measured against extremes, while the MRNBC performed the "best" overall when ranked across a range of metrics (Vogel et al., 2023). It would be ideal to include a full suite of CMIP models, or at least, a subset based on benchmarking criteria. That opportunity was not available to us for NHP (as explained in Section 2). In any case, the frequentist approach may not be the best way in which to represent uncertainty, particularly for impact assessment studies (Shepherd, 2021).

There are emerging opportunities however for the Australian Climate Service (ACS) to produce similar data sets based on a carefully selected subset of CMIP6 models (Grose et al., 2023). One of their key findings is that: "The projections cannot be considered a probabilistic or balanced estimate of uncertainty given the limited ensemble size and underlying epistemic uncertainties. The ensemble can however be used in a 'climate futures' or 'storyline' approach to illustrate plausible future climates that broadly span the range of possibilities suggested by CMIP6, while producing added value at the regional scale." The assessment reports produced for the NHP (https://awo.bom.gov.au/about/overview/assessment-reports) were also structured with a storyline approach, which we document in Section 8.3.

Figure 14: Blockings are apparently seen in the precipitation (Figure 14a) and root-zone soil moisture (Figure 14c) maps compared to runoff results (Figure 14b). Is it because low-res GCMs are spatially interpolated into fine resolutions followed by statistical bias corrections? Why runoff results are smoother? Why not much spatial variability is seen in the PET plot (Figure 14d)? Furthermore, in previous spatial maps, the data-sparse regions are masked out. I suggest doing the same masking for this figure.

Yes, that is the reason the blockings are seen. The interpolation used was a "conservative remapping" as opposed to a "bilinear interpolation", which may reduce some of the "blocking" in the visual representation of the results. We are currently investigating which interpolation method is more appropriate and how it may affect communication of the ensuing results.

Page 17 Line 26: You may want to show the same analysis for the historical period to be confident about the performance of GCM-driven hydrological projections in simulating the extreme events.

The GCMs do not represent a perfect historical analysis - they are set up with boundary conditions but because of parameterisations and lack of fine scale resolution, the GCMs are not a digital twin of Earth. Thus, the reference period of change needs to be the GCM historical period, as the actual historical period of observations or reanalysis has no bearing on the GCM simulated historical period. However, this is why we bias correct the output from the GCM, to bias correct for climate variability, to make sure that the extremes fit within the distribution of the observations. In addition, (Vogel et al., 2023) has performed a thorough evaluation of the GCM variables and the GCM-driven hydrological projections, especially with regard to extremes (5-year maximum and 99, 99.5 and 99.9 percentiles).

Editorial

Page 1 Line 23: It is hard to understand this sentence without reading the main content. I suggest replacing 'one to output from a regional climate model forced by…' with 'one regional climate model (RCM) that is forced by …'

The suggested modification would not convey what was achieved, however, we do appreciate the difficulty in interpretation. We have modified it to now read (italicised words are additions):

"Three bias correction techniques were applied to *all* four CMIP5 global climate models (GCMs) and one *method* to a regional climate model (RCM) forced by the same four GCMs, resulting in a 16-member ensemble of bias-corrected GCM data for each emission scenario."

Careful proofreading is required throughout the article. Typical issues are:

- Missing commas: For example, in Page 3 Line 11-13, there should be a comma after 'To address these deficits in hydrological projections'.
  - Modified as suggested.
- Sentences too long to read: For example, in Page 6 Line 1-4, it is better to break this sentence into two or more before the words 'hence' and 'nevertheless'.
  - Split into two sentences. It now reads:

  "Due to Australia's large size and geographical location, the climate of Australia varies markedly from the tropical north to the temperate south. As such, nationally-averaged precipitation may not provide meaningful insight from a climatological perspective, nevertheless, it does impart interpretation of how our choice of GCMs occupies the phase space spanned by the CMIP5 models."
- Duplicate descriptions: For example, the sentence in Page 10 Line 19-21 is a duplicate description of Page 10 Line 25. The first sentence in Section 6 is also mentioned before.
  - Page 10, line 25 (now page 10, line 28) has been modified to: "Steps (1) and (2) have different implementations depending on whether the correction to be applied is additive or multiplicative."

Furthermore, we have carefully gone through the manuscript, to shorten lengthy sentences and pay careful attention to grammar. This has been aided using grammar checking software. We hope the reviewer finds the new version satisfactory.

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.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M. and Schär, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, Water Resour. Res., 49(3), 1523–1536, doi:10.1029/2011WR011533, 2013.

CSIRO and Bureau of Meteorology: Climate Change in Australia Projections for Australia's Natural Resource Management Regions: Technical Report., 2015.

Dobler, C., Hagemann, S., Wilby, R. L. and StÃtter, J.: Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed, Hydrol. Earth Syst. Sci., 16(11), 4343–4360, doi:10.5194/HESS-16-4343-2012, 2012.

Frost, A. J., Ramchurn A. and Smith A.: The Australian Landscape Water Balance model (AWRA-L v6) Technical Description of the Australian Water Resources Assessment Landscape model version 6. [online] Available from: www.bom.gov.au/other/copyright.shtml (Accessed 1 June 2021), 2018.

Giuntoli, I., Vidal, J. P., Prudhomme, C. and Hannah, D. M.: Future hydrological extremes: The uncertainty from multiple global climate and global hydrological models, Earth Syst. Dyn., 6(1), 267–285, doi:10.5194/ESD-6-267-2015, 2015.

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, Clim. Serv., 30, 100368, doi:10.1016/J.CLISER.2023.100368, 2023.

Hope, P., Timbal, B. and Fawcett, R.: Associations between rainfall variability in the southwest and southeast of Australia and their evolution through time, Int. J. Climatol., 30(9), 1360–1371, doi:10.1002/joc.1964, 2010.

Joseph, J., Ghosh, S., Pathak, A. and Sahai, A. K.: Hydrologic impacts of climate change: Comparisons between hydrological parameter uncertainty and climate model uncertainty, J. Hydrol., 566, 1–22, doi:10.1016/J.JHYDROL.2018.08.080, 2018.

Shepherd, T. G.: Bringing physical reasoning into statistical practice in climate-change science, Clim. Change, 169(1–2), doi:10.1007/S10584-021-03226-6, 2021.

Vogel, E., Johnson, F., Marshall, L., Bende-Michl, U., Wilson, L., Peter, J. R., Wasko, C., Srikanthan, S., Sharples, W., Dowdy, A., Hope, P., Khan, Z., Mehrotra, R., Sharma, A., Matic, V., Oke, A., Turner, M., Thomas, S., Donnelly, C. and Duong, V. C.: Downscaling and bias correction for climate change impact studies, J. Hydrol. (in prep.), 1–41, 2022.

Vogel, E., Johnson, F., Marshall, L., Bende-Michl, U., Wilson, L., Peter, J. R., Wasko, C., Srikanthan, S., Sharples, W., Dowdy, A., Hope, P., Khan, Z., Mehrotra, R., Sharma, A., Matic, V., Oke, A., Turner, M., Thomas, S., Donnelly, C. and Duong, V. C.: An evaluation framework for downscaling and bias correction in climate change impact studies, J. Hydrol., 622, 129693, doi:10.1016/J.JHYDROL.2023.129693, 2023.