Review of “Performance and process-based evaluation of the BARPA-R Australasian regional climate model version 1” for GMD

General comments

The authors introduce BARPA-R, a regional model used for downscaling reanalysis and GCMs at ~17km resolution over Australia. The paper focuses first on evaluation of the ERA5 reanalysis driven simulations, this is useful for isolating regional model biases. The evaluation is very comprehensive and extends far beyond the standard climatological metrics, to include aspects of large-scale atmospheric circulation and feature tracking, as well as lagged metrics to consider land atmosphere feedbacks.

The authors have been careful in their experimental design to align with CORDEX requirements. The model and experimental design appropriately described and justified.

My comments below are mostly minor and relate to improving the presentation and discussion of certain results. Overall, I think this is an important well-written and comprehensive manuscript.

Specific comments

-should be “ERA5” not “ERA-5” check for consistency throughout

-Line 138, what version of ERA5 was used? Given the known issues with lower stratospheric temperature and humidity biases https://www.ecmwf.int/en/elibrary/81149-global-stratospheric-temperature-bias-and-other-stratospheric-aspects-era5-and . I doubt this matters much for downscaling, but useful to document

-Line 95, include additional details about what atmospheric variables from ERA5 are used to drive BARPA, and at what levels etc.

-Is AGCD the same as AWAP? Perhaps discuss further. Related to this, are there papers that have tried to quantify the observational uncertainty from this product? This is likely important for some biases. See additional comments below.

-Lines 140 onwards, I agree with the idea that similar biases in BARPA-R and ERA5 should be interpreted as good (or at least acceptable) for the reasons discussed. But I think this needs to be highlighted in the abstract and/or conclusion – as it is fundamental to the results presented here and this point could be missed by readers.

-Lines 155, suggest explaining additive and multiplicative biases more thoroughly

-Figure 2 (and others). The number in the panel, what is this exactly? Is it the mean bias? This should be detailed in the figure or caption. It would be useful to include MAE (mean absolute bias) also, so that this isn’t contaminated by cancellation of errors of different signs which clearly is present in some of the results. At other times you present additive and multiplicative biases so I was confused.

-Figure 3 and 4 – this is very comprehensive – but there is a lot packed into these figures. I wonder if the number of comparisons can be reduced (different bars) so that the figure is easier to read (or broken into a separate figure). The 3 separate y-axis seems somewhat excessive in my opinion.

-Figure 5 – Nice figure. Further discussion around what might be contributing to these differences in trends seems warranted. For temperature, does ERA5 assimilate observations in a way that is temporally consistent (i.e. so that local trends are expected to be realistic)?

-Figure 6, can you be sure that AGCD is appropriate to use for trend analysis of rainfall – i.e. considering station inhomogeneity and the effects of interpolation? For example, the following paper found highly inconsistent trends for precipitation indices over the US in observational products (including different in situ gridded products):

Gibson, P. B., D. E. Waliser, H. Lee, B. Tian, and E. Massoud, 2019: Climate Model Evaluation in the Presence of Observational Uncertainty: Precipitation Indices over the Contiguous United States. J. Hydrometeor., 20, 1339–1357, https://doi.org/10.1175/JHM-D-18-0230.1.

-Lines 280, Suggested discussion point - has the BARPA-R wet bias been shown to be a general bias seen in other UM regional models at this approximate resolution?

-Line 301- there is a “not shown, TODO” left in that needs updating :)

-Figure 9-10. Given the BARPA-R is forced by ERA5, it would be quite odd if large scale features like the climatology of the subtropical jet diverged much. So, the agreement is not that surprising. Smaller scale convective features are where we would expect more divergence, as you show. Perhaps worth discussing this point more.

-Figure 11. This is very interesting and well presented – a valuable contribution to the paper