July 22, 2024

Paper #GMD-2024-41 | Model evaluation paper**: 'Evaluation of CORDEX ERA5-forced 'NARCliM2.0' regional climate models over Australia using the Weather Research and Forecasting (WRF) model version 4.1.2'**

We thank the Editor and two reviewers for their constructive input, and for assessing this manuscript as suitable to publication following the opportunity to implement revisions.

As you will see from our point-by-point responses below, we have carefully gone through the reviewer comments and suggestions (please see Tables 1-2 on pp. 2-33 in this document).

We believe that the reviewer comments have helped strengthen this manuscript, and we are very grateful for their reviews.

Kind regards,

Giovanni Di Virgilio, Fei Ji, Eugene Tam, Jason Evans, Jatin Kala, Julia Andrys, Christopher Thomas, Dipayan Choudhury, Carlos Rocha, Yue Li, and Matthew Riley

Table 1. Reviewer 1, Sugata Narsey: Comments

#	Issue Description	Discussion	Revision (in re-submitted manuscript)
	Recommendation: Minor revisions This manuscript documents the evaluation of the NARCliM2.0 regional climate model (RCM) driven using ECMWF Reanalysis v5 (ERA5). The manuscript is well-written, and systematically addresses key aspects of the evaluation of their model. They go further than a basic evaluation, providing useful insights into the regional impacts of multiple parameterisation configurations of the model. They find that changing the physics choices in their model can have quite dramatic effects on regional climate biases for Australia. A nice addition to this study is their analysis of the relative sources of bias estimated by interchanging their RCM with the previous version of the NARCliM RCM, and also interchanging the driving ERA5 reanalysis with the ERA Interim reanalysis data previously used. By doing so they find that the RCM set-up appears to be a stronger influence on the mean state bias in their regional climate simulations compared to the choice of driving reanalysis data. This manuscript forms an important scientific basis for the production of a nationally significant projections dataset and is an important contribution to regional modelling for the Southern Hemisphere. I have some comments around evaluation choices, and around specific wording especially with regards to claims of improvement for precipitation since the biases over northern Australia appear significant. However overall, this manuscript is appropriate for publication, and my recommendation is for minor revisions.	We are very grateful to the reviewer for assessing our work, for their positive remarks and for recommending publication following Minor Revisions.	N/A
	Reviewer #1: Main Comments		

1	The distribution plots show nationally aggregated data, however I find this problematic since the map plots show that Tmax and precipitation in particular have opposing biases in the northern and southern regions (Fig 3 and Fig 7). Additionally, the bimodal distribution of Tmin in Fig 4 might be a function of mixing two climatically distinct regions. Why not split it into at least two regions? Then you can clearly state the improvements for the southern parts of Australia.	We agree - it is a good idea to stratify the daily distribution plots by region. The reviewer suggests doing this for 'at least two regions'; however, we regionalised the data according to the four Australian Natural Resource Management (NRM) region clusters which are broadly aligned with climatological boundaries (Fiddes et al., 2021) and with the IPCC reference regions (Iturbide et al., 2020). PDFs for all four NRM clusters for all three variables will be included in the revised manuscript Supplementary Material. Sample results are shown below for the PDFs of minimum temperature for the Southern Australia and Northern Australia NRM clusters, e.g. bimodality in modelled distributions is not apparent for Southern Australia.	See sample new PDF results below (pp. 12-15); all regionalised PDFs will be included in the revised manuscript Supplementary Material.
2	The evaluation conducted here focuses on rainfall and temperature, which I agree are the most important variables to consider. However, some investigation of the circulation state in the RCMs may be of use to help understand the systematic biases, for example over northern Australia (R3-7), and over SE Australia (R2-4).	We agree with the reviewer that temperature and precipitation are the most important variables to consider (e.g. because they are used in many climate impact studies). However, it is a good idea for future work to investigate RCM simulations of the circulation state via a future study aiming to explain the processes/mechanisms underlying the varying RCM skill profiles observed in this study. Such an analysis could form a component of a model intercomparison study of the various CMIP6-forced	Great idea for a future study. No changes to main text.

		RCMs for this region. As we stated at the end of the Introduction, for this current work our aim is: "*Here, our focus is on evaluating the performances of the different RCM generations, with an investigation of the mechanisms underlying the varying model performances to be the subject of future work.*"	
3	The statements around general improvements in precipitation are not well-founded in my view, since the dry biases over northern Australia are large compared to NARCliM1.5 runs. I would prefer if the claims were either made specific to the inner domain, or else more carefully described in this manuscript. Alternatively, the authors can show whether the biases in the NARCliM2.0 runs (especially for northern Australia) are actually smaller as a percentage of annual mean climatological rain.	Agreed, we have revised the text to more carefully explain the RCM performance for mean and extreme precipitation as shown right:	Original text (at lines 11-12 in Abstract): "ERA5-RCM precipitation simulations show lower bias magnitudes versus ERA-Interim-RCMs, though dry biases remain over monsoonal northern Australia and extreme precipitation simulation improvements are principally evident at convection-permitting 4 km resolution." Revised text in Abstract: "At 20 km resolution, improvements in the mean and extreme state precipitation of the ERA5-RCMs versus ERA-Interim RCMs are principally evident over south-eastern Australia, whereas strong biases remain over northern Australia. At convection-permitting scale over south-eastern Australia, mean absolute biases for mean and extreme precipitation for the ERA5-RCM ensemble are around 79% and 10%

			smaller versus the ERA-Interim RCMs that simulate for this region.” Also, the text is revised at lines 479-486 (see qualifiers added in **bold**): “Overall, CORDEX-CMIP6 ERA5-RCMs confer improvements in the simulation of mean precipitation **over south-eastern Australia** relative to the CORDEX-CMIP5 ERA-Interim RCMs, with two ERA5 RCMs in particular (R3, R4) showing considerable improvements **over this region**.”
4	It is outside the scope of this study, however it would be interesting to know if the different physics configurations and their associated regional climate biases have much bearing on the future change signal in the model when holding the driving global model data constant. Similarly, it would be interesting to know if the evaluation of the ERA5 runs with different physics configurations translates in an evaluation of the CMIP6 historical scenario runs	This is a great idea for a future study, thank you for suggesting this. For now, as the reviewer mentioned it is outside of the scope of this study, however, we shall mention this idea for future work in the Conclusion (see new text right).	We will state some of these ideas for future investigation in a new paragraph added to the Conclusion in the revised manuscript (see new text below): “Our present focus was to evaluate the performances of the different RCM generations assessed here. Future work will explore other topics, such as the potential influences of the different RCM physics configurations and their associated biases on the nature of the future change signals in subsequent CMIP6 GCM-forced simulations, e.g. when holding the driving GCM data constant. Additionally, future model-intercomparison studies that compare biases between the different RCMs contribution to CORDEX-Australasia will be highly valuable.”

5	Also outside the scope of this study, but it would be interesting to intercompare the dry-bias tendency over northern Australia in the NARCliM2.0 runs with other regional simulations using different models that introduced similar dry biases for the Australian monsoon. Although such a bias may be undesirable, there is a real opportunity here to shed light on some fundamental characteristics of the dynamics Australian monsoon, in particular the feedback mechanisms associated with land surface behaviour, convection, and large-scale circulation.	Thank you for suggesting these interesting ideas for future study – we see a lot of value in such future investigations, for now though, as the reviewer mentions, it is outside of the scope of this study. As above though, we shall briefly mention this idea for future work in the Conclusion (see above for #4).	Please see above.
	Reviewer #1: Specific comments		
6	L63-64: It’s now May 2024 and this statement is outdated; I believe the BARPA paper is now published (https://gmd.copernicus.org/articles/17/731/2024/gmd-17-731-2024-discussion.html) and there may be others by now. Might be worth a quick search.	Noted – this study is now acknowledged in the revised manuscript as shown right.	“Previous work to dynamically downscale ERA5 over CORDEX Australasia includes the BARPA-R (Bureau of Meteorology Atmospheric Regional Projections for Australia) regional climate model which simulates over CORDEX Australasia at 17 km resolution (Howard et al., 2024). Evaluation of BARPA-R’s skill in simulating the Australian climate observed good performance overall, including a 1°C cold bias in daily maximum temperatures and wet biases of up to 25 mm/month over inland Australia.”
7	L205-208: Did you follow the same experiment design in all other respects except for run length? Fig 13 shows the inner domain. Are these ERA5 swapped with ERA Interim sensitivity experiments conducted at the fine-scale for the inner domain? Maybe it’s specified somewhere but I couldn’t see it. Worth clarifying here.	Here, the experiment designs for the CORDEX-CMIP6 RCMs and the CORDEX-CMIP5 RCMs are identical in all other respects except for switching the driving data and running the simulations for 14 months instead of 30 years (the latter owing to compute and time constraints). For instance, the ERA5-RCMs CORDEX-CMIP6 (NARCliM2.0)	Revised text (changes in **bold**): “2) fourteen-month simulations are performed where otherwise identically parameterised **and configured** CORDEX-CMIP6 NARCliM2.0 R1-R7 RCMs are forced by ERA-Interim as opposed to ERA5, and similarly the WRFJ-K-L RCMs from the CORDEX-CMIP5 era are forced with ERA5

		RCMs are run for the same 4 km convection permitting domain using the same physics options and model setups with the only changes being to swap ERA5 for ERA-Interim and running for 14 months. The text has been revised to make this clearer (please see right).	instead of ERA-Interim. **For instance, the ERA5-RCMs CORDEX-CMIP6 (NARCliM2.0) RCMs are run for the same 4 km convection permitting domain using the same physics options and model setups with the only changes being to swap ERA5 for ERA-Interim and running for 14 months**.”
8	L210-211: The short periods are probably fine, but why not just do a quick bootstrap check to see how representative 14-month periods are for rainfall in the longer run period using either AGCD or your simulations?	Owing to the large computational costs involved, we could only simulate for 14 months. Given this hard constraint, it was important that the target year in question had experienced as broad a variety of climatic patterns as possible (i.e. being neither predominantly/ consistently dry-warm v wet-cold). 2016 (plus the preceding two months as spin-up) was hence an ideal choice if one is restricted by model resource constraints, because it is renowned as a year starting with a strong El Niño event and several months of hot dry weather but then from approximately the middle of the year shifting to a pattern of frequent heavy rainfall until approximately October (Trewin, 2017). This has now been stated in the revised text (it was not mentioned before and should have been stated). However, for our future model development work where we may alter the period being simulated, a test	Addition to the text: “Australia experienced a range of weather extremes during 2016 driven by a range of climatic influences making 2016 a suitable target year (Bureau of Meteorology, 2017).”

		procedure like that outlined by the reviewer is something we will trial.	
9	L231-233: Based on Fig 3 it seems Fig 2 might obscure some compensating biases between north and south. Is this the case?	Correct. As the reviewer suggested, we have prepared regionalised PDFs for all variables. In this case (for maximum temperature), the ERA5-RCMs (e.g. R5) showing warm mean annual max. temperature biases mapped in Fig. 3 over northern Australia also overestimate daily occurrences of warmer than average maximum temperatures (see new PDF for Northern Australia below – pp. 14-15). Conversely, the mapped cold biases over southern Australia in Fig. 3 correspond with overestimations over cooler than average temperatures in the PDF for southern Australia (see new below).	The new regionalised PDFs are included in the revised Supporting Information accompanying the revised manuscript, and they are briefly referred to in the main text. See example PDFs for maximum temperature for northern and southern Australia below (pp. 14-15).
10	Fig 3: The stippling is difficult to see. Can you improve somehow?	Agreed – the stippling and other elements of all map-based figures have been improved for clarity as shown in the example below (see p. 33).	Map based figures will be revised in the main text as suggested (and as implemented in the example figure provided here p. 33).
11	L270 and Fig 4: Is the bimodality due to mixing different climate zones?	Based on the reviewer's suggestion above (#1) and the subsequent geographically-stratified PDF analyses we have conducted for the revised manuscript, overall, yes, we believe the biomodality present for the continental-scale analyses is due to mixing of different climate zones (see also sample PDFs shown below). The revised manuscript includes the geographically stratified PDFs and also the main text is	"Observed PDFs at the continental scale show a slight bimodality that is captured by ERA5-R1, ERA5-R4, ERAI-WFJ, ERAI-SWWA and ERAI-CCLM. However, this bimodality is generally not present in PDFs stratified for specific NRM regions (Figures SX-SX)."

		revised/supplemented as in the example shown right.	
12	Fig 6: Why not show log(P)? Might be easier to see differences and similarities.	We had trialled log(P) early on when conducting these analyses for daily precipitation distributions and concluded this did not confer much improvement to the clarity of the PDFs.	No changes to main text.
13	Fig 7: Would the biases over northern Australia look this dramatic if you showed it as a percent of AGCD climatology? It's hard to know for example which absolute bias is more concerning between runs, since a small absolute bias in the dry regions may matter more than a large absolute bias in the monsoon region.	As requested, we prepared a similar plot to Fig. 7, but showing relative biases (please see below p. 16). Some observations include: - The magnitude of some of the absolute biases for annual mean precipitation over northern Australia do not appear as dramatic for the relative bias as compared to the correspoding absolute bias. For instance, when examining relative biases for R3, the relative bias magnitudes along the east and south coasts become more similar in magnitude to the relative bias magnitude over the north. Conversely, for the original plot (Fig. 7 in main text) showing absolute biases, the northern bias magnitude appears larger than the absolute bias magnitudes along the south and east coasts. - Overall, the relative performance ranking of the ERA5-RCM	No changes to main text.

		simulations using relative bias or absolute bias is similar.	
14	Fig 13: If this is not 4km explicit convection runs than perhaps show larger domain? Otherwise, see comment for L205-208 above.	Please see response to comment #7 above/in this table.	Addressed by #7 above.
15	L459: The claim of general improvement in precipitation and even max temperature is not quite true in my opinion. The bias over northern Australia appears large and systematic. I think it is appropriate to claim general improvement over the inner domain though. See main comment 3.	Please see response to comment #3 above/in this table.	Addressed by #3 above.
16	Section 4.1: you note the dry bias vs wet bias may relate to microphysics scheme. Looking at fig 7 the three runs (R2-4) with MYNN2 boundary layer scheme are all wet biased over SE Australia. Is this a coincidence?	It is an interesting observation, though it might need more exploration via future study e.g. because some ERA5-RCMs that do not use MYNN2 for PBL also show wet biases over SE Australia e.g. R1 (YSU) and R5 (ACM2).	No changes to main text.
17	L492: suggest "especially over northern Australia *where all other runs contain a systematic dry-bias*".	The reviewer's suggestion is added to the text.	Added to the main text: "For both mean and extreme precipitation, ERA5 R1 and R2 are notable in that they are more wet-biased than the other ERA5 RCMs, especially over northern Australia where all other ERA5-RCMs contain a systematic dry-bias."
18	L568: Again, I don't agree with this claim of general improvement for precipitation.	Please see response to comment #3 above/in this table.	Addressed by #3 above.
19	L577: It also appears important here at coarser scales when precipitation is parameterised, based on Fig 7.	Text revised accordingly – please see column right.	Revised text: "Nevertheless, our results for the CORDEX-Australasia domain suggest that the choice of microphysics scheme is important, especially for precipitation extremes."

20	L584-585: Potentially also in the rainfall biases, especially where dynamical feedbacks are known to occur in the real world such as over northern Australia during the summer monsoon season.	We agree with the reviewer that the change in land surface scheme could change the land surface feedback (via soil moisture) to the precipitation biases. This as a possibility that requires more extensive analysis to investigate. Revised text to be revised as shown right:	Revised text: “The different land surface schemes (e.g. Noah-Unified versus Noah-MP) likely play a role in RCM skill in simulating maximum temperature, as well as changing the land surface feedback (via soil moisture) to the simulation of precipitation – these possibilities require more extensive analysis to investigate.”

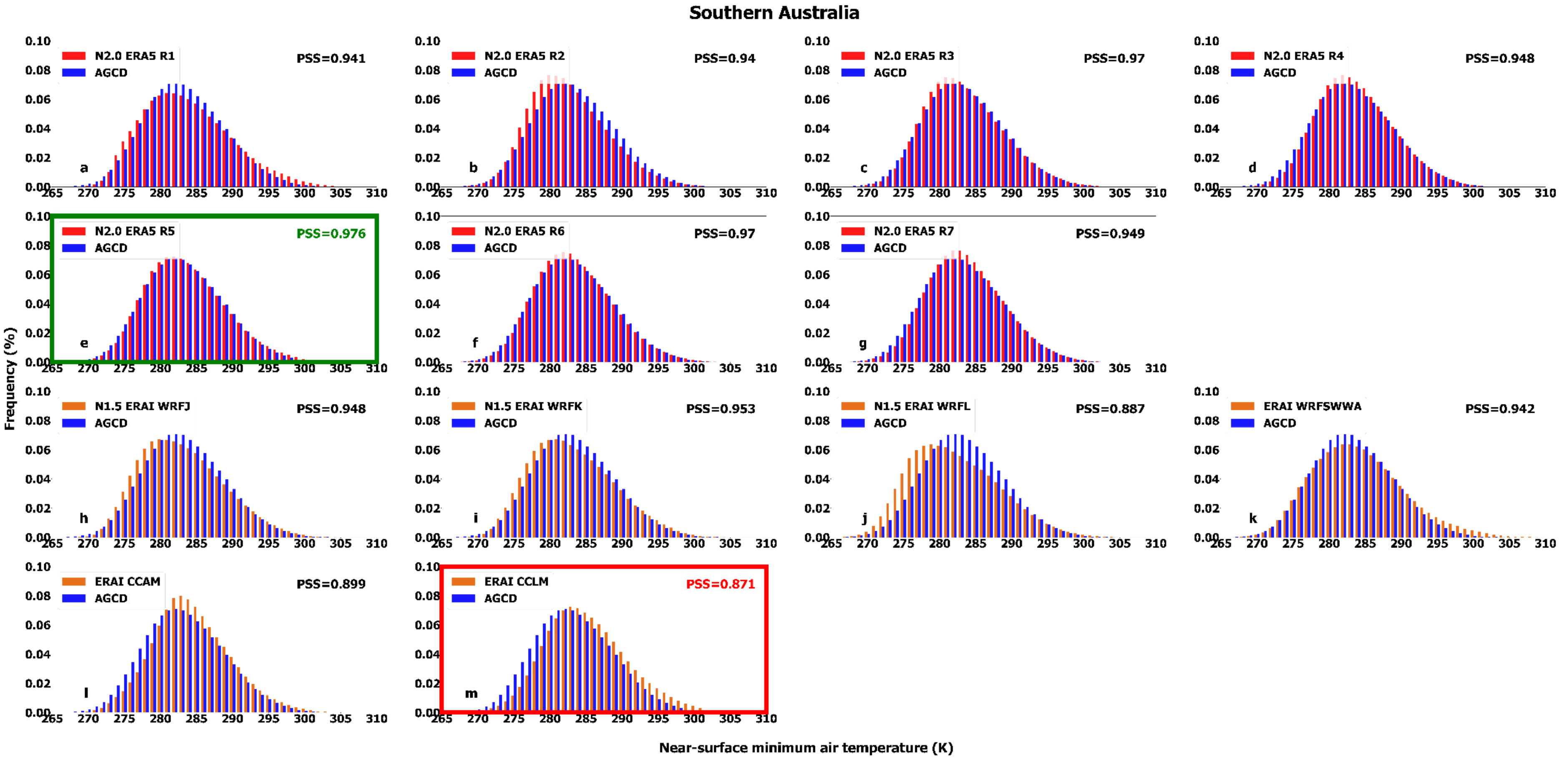

Figure S1. Probability density functions of mean daily maximum near-surface air temperature (K) with bin width of 1 K over the Southern Australia natural resource management region.

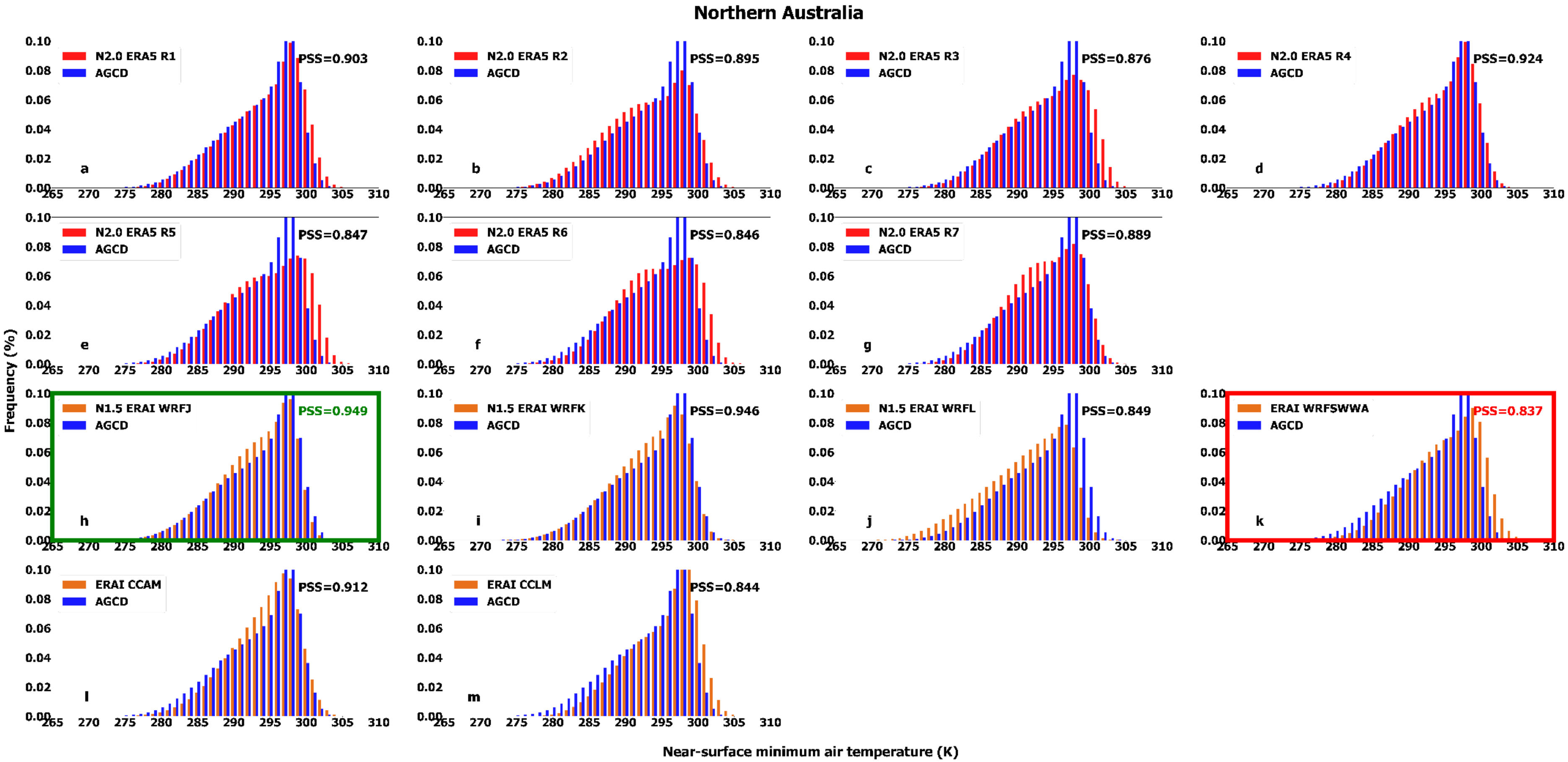

Figure S2. Probability density functions of mean daily minimum near-surface air temperature (K) with bin width of 1 K over the Northern Australia natural resource management region.

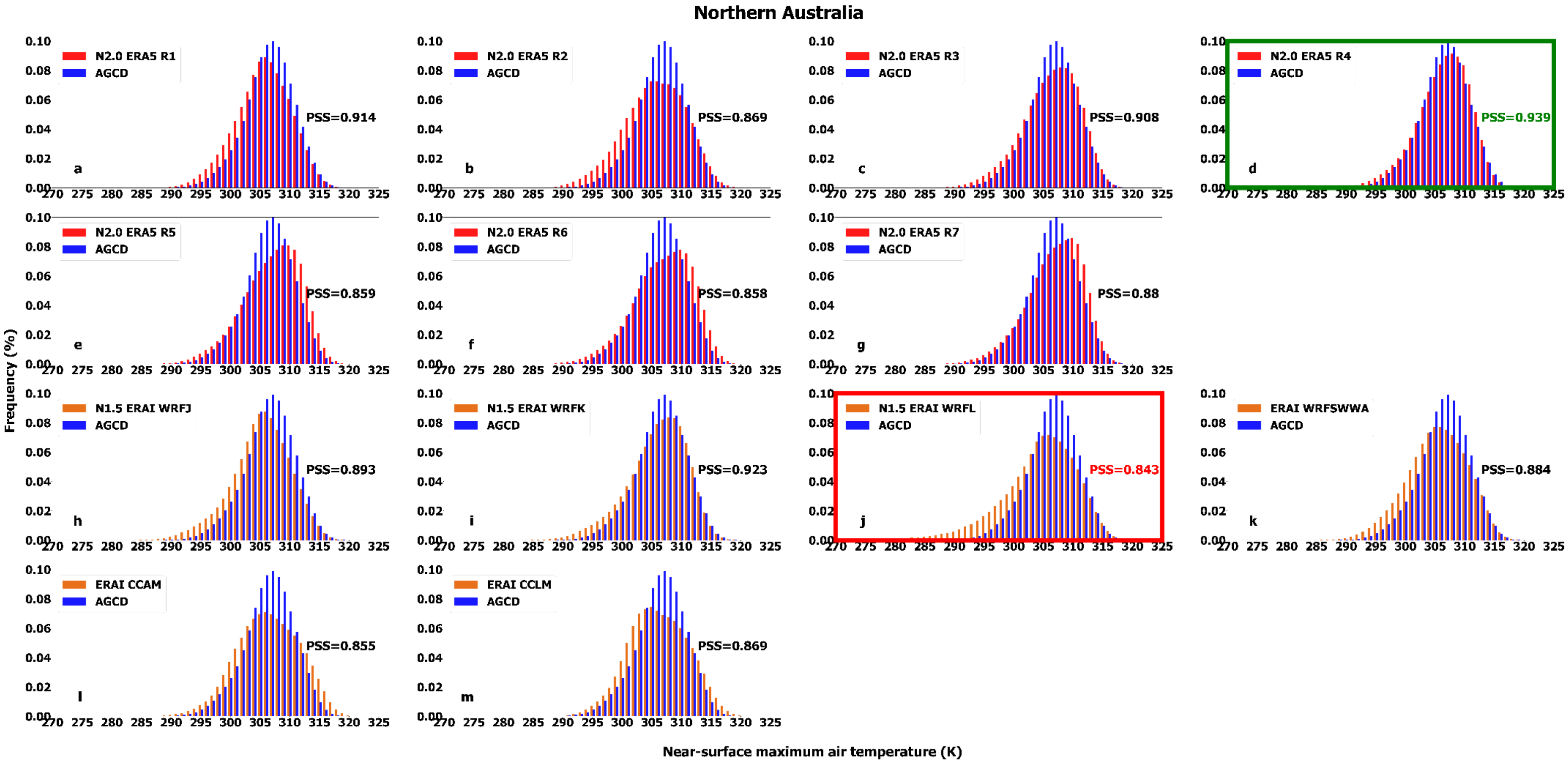

Figure S3. Probability density functions of mean daily maximum near-surface air temperature (K) with bin width of 1 K over the Northern Australia natural resource management region.

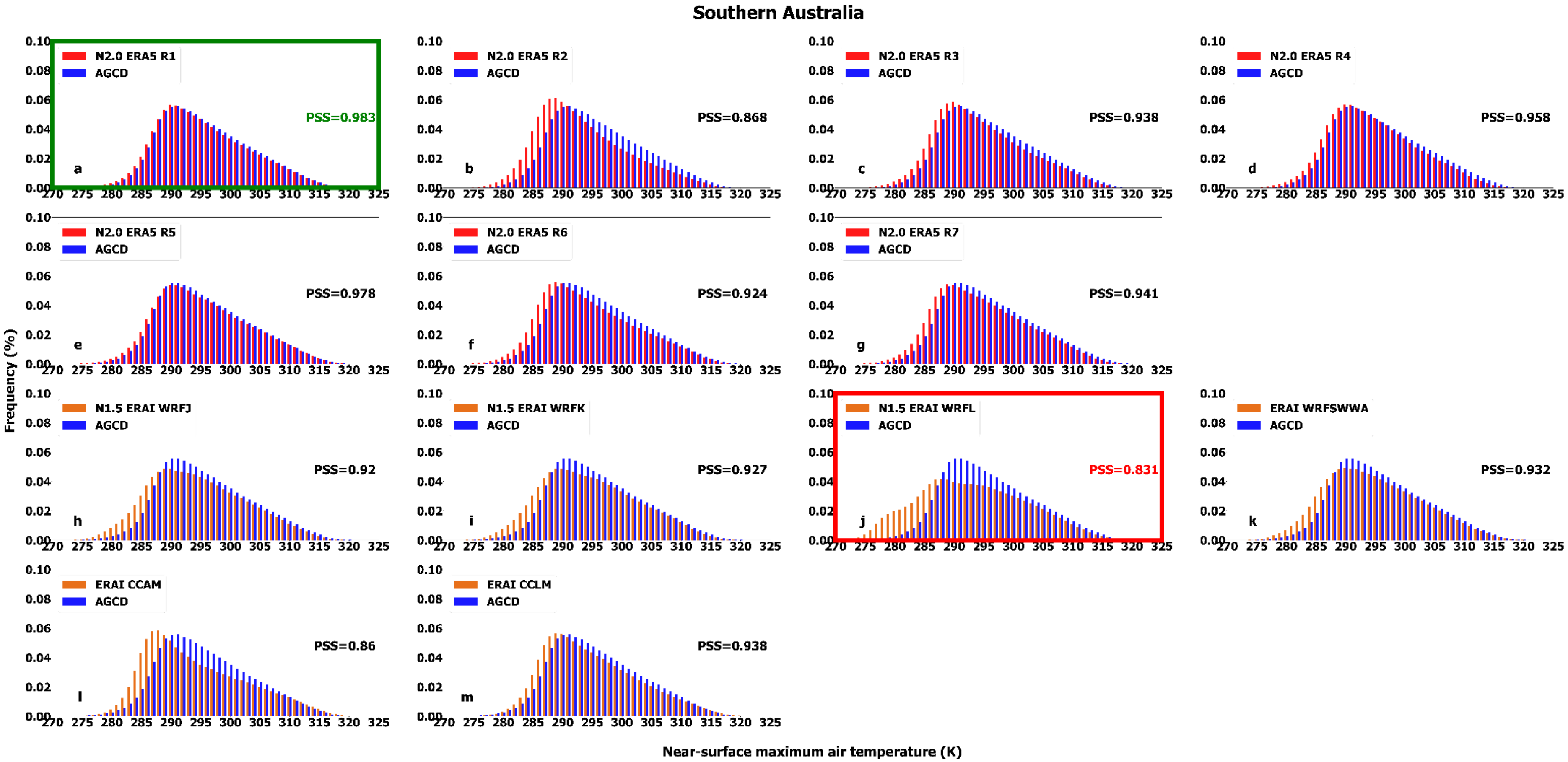

Figure S4. Probability density functions of mean daily maximum near-surface air temperature (K) with bin width of 1 K over the Southern Australia natural resource management region.

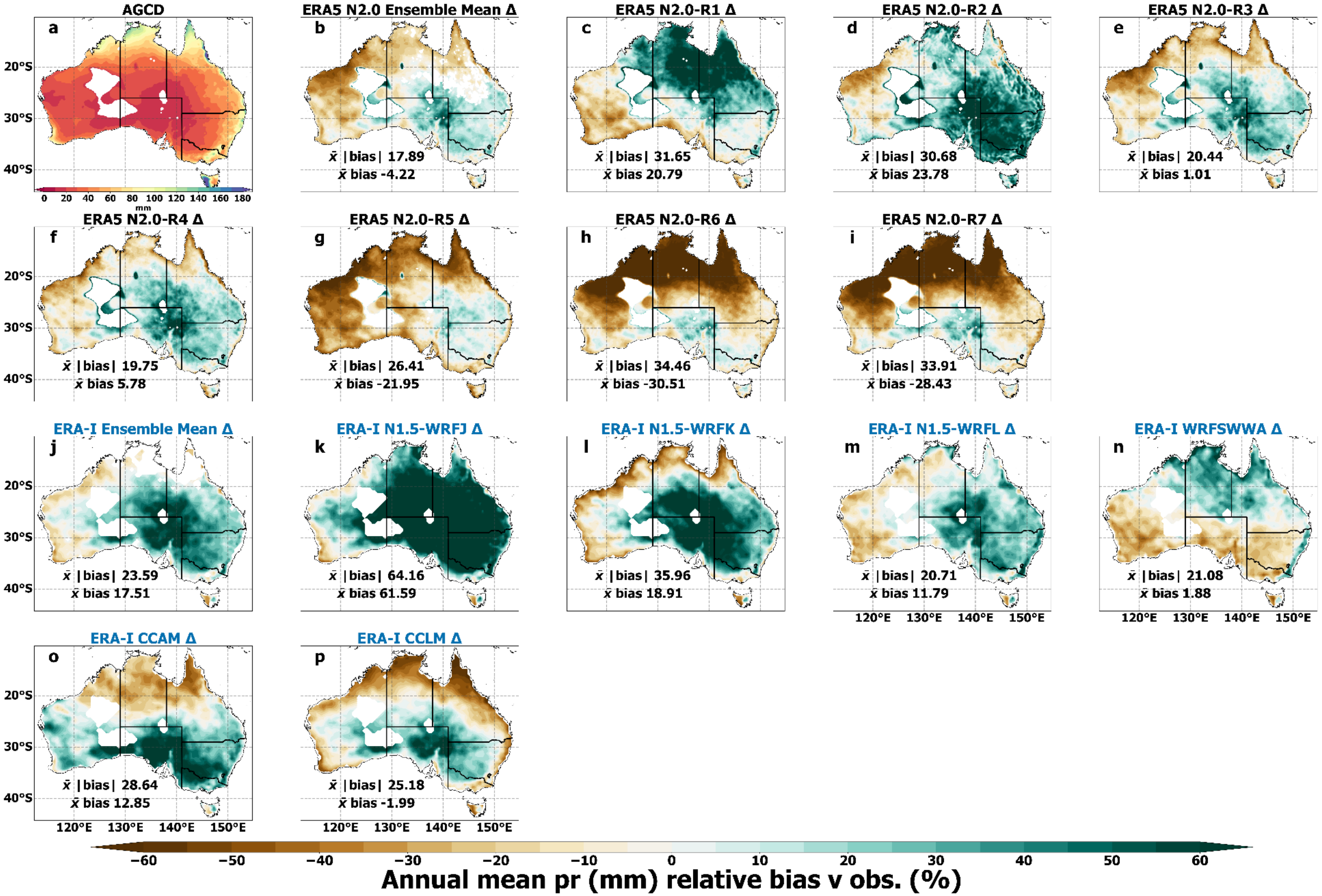

Fig. 7A Annual mean precipitation relative bias with respect to gridded observations for the RCMs for 1981-2010.

Table 2. Anonymous Referee #2 (RC #2) Comments

#	Comment Description	Discussion	Revision (in re-submitted manuscript)
	Reviewer #2 "This paper provides an evaluation of the representation of precipitation and diurnal screen-level temperatures from a set of 7 model configurations of the NARCLIM2.0 regional climate model. By benchmarking model performance against a previous version of NARCLIM and repeating the analysis of a previous paper, the authors follow an objective, pre-determined framework, which is to be commended. NARCLIM2.0 is shown to have a reduction in outlier model configurations with large temperature biases in excess of 2K. Their results highlight model dependence, particularly of precipitation, on the choice of parametrisation schemes and identify a pervasive dry bias in northern Australia. I have comments around the description and justifications of model configuration choices and some more minor comments on the presentation of the results. Overall, this is an important and well written manuscript suitable for publication following these revisions."	We are grateful to the reviewer for reviewing the manuscript, for their positive remarks on this work, and for recommending publication following their suggested revisions.	N/A
	Reviewer #2 General Comments:		
1	There are a large number of models, statistics and maps presented in this paper which makes it difficult to form an overall view of the improvement in model performance across generations. I would suggest you include a summary table of the mean absolute error, bias magnitudes and Perkins Skill	Thank you for this suggestion, we have created a table showing the mean absolute biases for climate means, climate extremes and Perkins Skill Scores for the majority of the analyses performed. Please see Table 1 below (pp. 31-32).	This table will be included in the revised manuscript.

	Scores reported across the paper and supplementary materials.		
2	The text at lines 137-140 suggests that the NARCLIM2.0 model configurations have been selected based on empirical performance during a single year, and that compatibility between parametrisation schemes or recommendations from the WRF model developers may not have been considered. Please add some text to provide assurance that these for each of the 7 selected configurations, the combination of parametrisation schemes is physically sensible. For example, have these combinations been used and recommendations by separate studies, developed and tested in combination, or at least not contain schemes developed specifically for use with a different setup or combinations precluded in the WRF user guide? Are the PBL schemes compatible with the surface schemes, and is shallow convection appropriately dealt with by the combination of PBL and convection schemes?	Thanks for pointing this out – this text will be re-phrased in the revised manuscript as shown in the column right, because the WRF parameterisation options were selected based on their performance in previous studies (e.g. via literature review). Most parameterizations were tested independently, such as those for the planetary boundary layer (PBL) or microphysics. However, some were tested in combination, such as PBL schemes combined with microphysics schemes, and they all performed well. When building the different configurations of the WRF model for test simulations, we were aware of certain physical parameterizations that are not compatible with each other due to overlapping functionalities or specific design constraints. For instance: • The Kain-Fritsch (KF) cumulus scheme and the Thompson microphysics scheme should not be used together. • The Yonsei University (YSU) PBL scheme should not be used with	"The seven ERA5 WRF configurations were selected from an ensemble of seventy-eight structurally different WRF RCMs. Each of these seventy-eight RCMs used different parameterisations for planetary boundary layer, microphysics, cumulus, radiation, and land surface model, where parameterisation options were selected via literature review and recommendations from WRF model developers. These test RCMs were run for an entire annual cycle (2016 with a two-month spin-up period commencing 1 November 2015). The seven ERA5 WRF configurations were selected from this larger ensemble based on their skill in simulating the south-eastern Australian climate, whilst retaining as much independent information as possible (Evans et al. 2014; Di Virgilio et al. in review)."

		the Monin-Obukhov surface layer scheme. - The Noah land surface model (LSM) should not be used with the Pleim-Xiu (PX) PBL scheme. - The Rapid Radiative Transfer Model for Global Circulation Models (RRTMG) longwave and shortwave schemes should not be used with other radiation schemes like Dudhia or Goddard. We avoided these incompatible combinations when setting up physics tests to assess RCM skill in simulating the south-east Australian climate for the NARCliM2.0 project. After this test/model development process, we also sought advice from WRF developers on the physics parameterisations being used for NARCliM2.0.	

3	More details on common aspects of the experimental design would be welcome: how are ozone and aerosols represented in these models? How frequently does the SST update? What datasets have been used as static inputs to the land-surface schemes (vegetation fraction etc)?	• SST is updated daily • Aersols are used the default WRF option aer_opt=0, therefore aerosols were not applied. • Ozone is default for radiation models used • Static inputs for land-surface schemes are also WRF defaults	
4	As the differences between the parametrisation schemes forms a large component of this paper, please provide references for the schemes. Some explanation of the dynamic vegetation scheme would also be welcome.	The revised manuscript will include the text shown column right providing high-level explanation of the dynamic vegetation cover option that is used with the Noah-MP land surface model. Also, we have included citations for the physics parametrisations in Table 1, which is where these schemes are listed. However, in more detail on Noah-MP's dynamic vegetation cover option: The dynamic vegetation cover option allows for the prognostic representation of plant phenology, leaf area index (LAI), and canopy stomatal resistance. Vegetation dynamics in the Noah-MP modelling system encompass plant	"Noah-MP provides a 'dynamic vegetation cover' model option (referred to as dynamic vegetation in the WRF users' guide) (Niu et al., 2011). When deactivated (the default), monthly leaf area index (LAI) is prescribed for various vegetation types and the greenness vegetation fraction (GVF) comes from monthly GVF climatological values. Conversely, when dynamic vegetation cover is activated, LAI and GVF are calculated using a dynamic leaf model. We clarify here that dominant plant-functional types do not change when using this option, but only the LAI and GVF, i.e. only the amount of green cover changes."

		photosynthesis, respiration, and the partitioning of assimilated carbon among various plant parts, including leaves, roots, and wood. This system can represent both seasonal and long-term changes in vegetation phenology and carbon exchanges over the land surface (Ise et al., 2010; De Kauwe et al., 2017; Gim et al., 2017). The incorporation of vegetation dynamics and photosynthesis-based stomatal resistance in the Noah-MP LSM enables the exploration of carbon partitioning within plant compartments (e.g., leaves, roots, and stems) and provides a prognostic representation of vegetation growth and senescence through canopy states, such as LAI (Hosseini et al. 2022). De Kauwe, M. G., Medlyn, B. E., Walker, A. P., Zaehle, S., Asao, S., Guenet, B., et al. (2017). Challenging terrestrial biosphere models with data from the long-term multifactor Prairie Heating and CO 2 Enrichment experiment. *Global Change Biol.* 23, 3623–3645. doi: 10.1111/gcb.13643 Gim, H. J., Park, S. K., Kang, M., Thakuri, B. M., Kim, J., and Ho, C. H. (2017). An improved parameterization of the allocation of assimilated carbon to plant parts in vegetation dynamics for N oah-	

		MP. *J. Adv. Model. Earth Sys.* 9, 1776–1794. doi: 10.1002/2016MS000890 Ise, T., Litton, C. M., Giardina, C. P., and Ito, A. (2010). Comparison of modeling approaches for carbon partitioning: impact on estimates of global net primary production and equilibrium biomass of woody vegetation from MODIS GPP. *J. Geophys. Res.: Biogeosci.* 4, 115. doi: 10.1029/2010JG001326 Hosseini A, Mocko DM, Brunsell NA, Kumar SV, Mahanama S, Arsenault K and Roundy JK (2022) Understanding the impact of vegetation dynamics on the water cycle in the Noah-MP model. Front. Water 4:925852. doi: 10.3389/frwa.2022.925852	
5	The selection of RCMs for this study comes across as ad-hoc and incomplete: HadRM3P, RegCM4 and REMO2015 also contributed ERA-interim driven simulations to CORDEX-CMIP5 Australasia but have not been evaluated. Additionally, three additionally, three ERA5-driven CORDEX-CMIP6 Australasia simulations appear to have also been recently published before the submission date. While including extra models at this stage may be out of scope, the paper may sit better in the literature if it focuses purely on NARCLIM/WRF-based models.	The primary focus of this paper is to evaluate the performance of the ERA5 driven NARCliM2.0 (WRF) simulations. Comparison with previous ERA-Interim driven WRF simulations is done to elucidate the generational improvement in climate simulation. A comprehensive assessment of all ERA-Interim driven simulations is outside the scope of this paper. Further studies that evaluate different RCMs downscaling ERA-Interim/ERA5 would be useful.	The text below mentioning another ERA5-driven evaluation RCM also contributing to CORDEX-CMIP6 has been added to the literature review section of the revised Introduction: "Previous work to dynamically downscale ERA5 over CORDEX Australasia includes the BARPA-R (Bureau of Meteorology Atmospheric Regional Projections for Australia) regional climate model which simulates over CORDEX Australasia at 17

		Additionally, in the revised version of the Introduction for this manuscript, we have also cited a recent evaluation of CORDEX CMIP6 RCMs (i.e. BARPA-R; Howard et al. 2024; GMD) – please see added text in column right.	km resolution (Howard et al., 2024). Evaluation of BARPA-R’s skill in simulating the Australian climate observed good performance overall, including a 1°C cold bias in daily maximum temperatures and wet biases of up to 25 mm/month over inland Australia.”
6	On a similar note, you may consider acknowledging that NARCLIM2.0 will contribute to an ensemble of downscaled climate projections for Australasia. (e.g. https://www.sciencedirect.com/science/article/pii/S2405880723000298)	Agreed, in the revised manuscript, we acknowledge this fact (please see text to right).	NARCliM2.0 comprises one of several RCM ensembles contributing to dynamically downscaled climate projections for Australasia (Grose et al. 2023). Citation: Grose, M., 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.

7	Map quality: stippling is hard to see, while coastlines and state boundaries show up as inconsistently rendered, adding to confusion. Can these be improved? Perhaps the figures would be easier to read if the stippling density was decreased and line thicknesses increased.	Good suggestions – we have modified the figures to improve clarity as suggested by decreasing stippling density and increasing their marker size, and also ensuring a clearer/more consistent representation of state/jurisdictional boundaries – please see example of one of the revised figures below (p. 33). All map-based figures will be revised accordingly in a revised version of the main text.	Please see comment left. All map-based figures will be revised as per the reviewer's suggestions and shown in the sample revised plot on p.33.
8	Can you provide a recommendation of which of R1-R7 you would recommend to be used in downscaling GCMs going forward?	The revised manuscript will include the new section of text shown in column right providing a suggestion for which of R1-R7 might be prioritised for subsequent CMIP6-forced dynamical downscaling.	New text added to the revised manuscript: "Although a single 'all-round' best-performing ERA5-RCM configuration cannot be selected, the model performances for the climate variables and statistics assessed here yield some insights if selecting a subset of ERA5-RCM configurations for subsequent CMIP6-forced downscaling. Overall, ERA5-R1 provides a good simulation of both mean and extreme maximum temperature and is broadly comparable to the other ERA5 RCM configurations with respect to minimum temperature. However, its simulation of mean and extreme precipitation is relatively poor as compared to some of the other ERA5 RCMs. ERA5-R2 has an unusual performance profile relative to the other ERA5-RCMs. Although this RCM shows generally good

			performance for minimum temperature, extreme maximum temperature and precipitation, it shows poor performance for mean maximum temperature in that is considerably more cold-biased than the other ERA5 ensemble members. ERA5-R2 is the only ERA5-forced RCM configuration in this ensemble to use Kain-Fritsch cumulus physics, and it showed mean maximum temperature biases of roughly similar magnitude and spatial pattern as the ERA-Interim WRFJ and WRFK RCMs which also used the same scheme. However, ERA5-R2 also generated a strong mean maximum temperature cold bias over south-eastern Australia at the 4 km convection-permitting scale which did not use cumulus parameterisation. ERA5-R3 shows good performance for mean minimum temperature and mean precipitation and reasonable performance for mean maximum temperature. The performance of R4 is broadly similar to R3, but it has substantially inferior performance versus R3 for maximum and minimum temperature extremes. R5 shows consistently good performance for maximum temperature. Its performance is more mediocre for precipitation and minimum temperature. Both R6 and R7 frequently showed the strongest biases, typically over large regions such as eastern Australia for both temperature variables, and over northern Australia for

			precipitation. As such, they are the poorest performers overall in the ERA5 ensemble, with performance for extreme minimum temperature often being particularly poor. From the specific perspective of the ERA5 RCM performances, and based on the present evaluations, overall R3 and R5 may be considered favourable RCM configurations for CMIP6-forced dynamical downscaling. However, as noted, some other ERA5 RCM configurations show good performance for specific variables and statistics, and thus could warrant inclusion in a larger ensemble and/or one adopting a sparse matrix approach (Christensen and Kjellström, 2020).”
	Reviewer #2 Line Specific Comments:		
9	Line 10: Please be more explicit about what these statistics (0.54K; 0.81K) are. They seem to be from R5 but I'm not sure why (R1 has a lower mean absolute error for the p99).	Agreed – please see column right for the revised text that is included in the revised manuscript.	New text added to revised manuscript: “ERA5-RCMs substantially reduce cold biases for mean and extreme maximum temperature versus ERA-Interim-RCMs, with the best-performing RCMs showing small mean absolute biases (ERA5-R5: 0.54K; ERA5-R1: 0.81K, respectively), but produce no improvements for minimum temperature.”

10	Lines 11-12 and lines 479-486: I can't see systematic improvement in mean state precipitation of the 7 CORDEX-CMIP6 RCMs over the 6 CORDEX-CMIP5 RCMs. Certainly, WRFJ has a very large wet bias, however the performance of WRFL is comparable to R3 and R4.	Agreed – the text should be much more specific because for the 20km outer domain these performance improvements are principally evident specifically over south-eastern Australia, and also for the convection-permitting (4 km) inner domain over south-east Australia. Text revised accordingly in both locations as shown right.	Original text (at lines 11-12 in Abstract): "ERA5-RCM precipitation simulations show lower bias magnitudes versus ERA-Interim-RCMs, though dry biases remain over monsoonal northern Australia and extreme precipitation simulation improvements are principally evident at convection-permitting 4 km resolution." Revised text in Abstract: "At 20 km resolution, improvements in the mean and extreme state precipitation of the ERA5-RCMs versus ERA-Interim RCMs are principally evident over south-eastern Australia, whereas strong biases remain over northern Australia. At convection-permitting scale over south-eastern Australia, mean absolute biases for mean and extreme precipitation for the ERA5-RCM ensemble are around 79% and 10% smaller versus the ERA-Interim RCMs that simulate for this region." Also, revised text at lines 479-486 (see qualifiers added in **bold**): "Overall, CORDEX-CMIP6 ERA5-RCMs confer improvements in the simulation of

			mean precipitation **over south-eastern Australia** relative to the CORDEX-CMIP5 ERA-Interim RCMs, with two ERA5 RCMs in particular (R3, R4) showing considerable improvements **over this region**.”
11	Line 194: Please specify the bin width used when calculating the Perkins Score.	The bin width used for precipitation is 0.5 mm. The bin width used for temperature variables is 1 K. This information is added to the relevant figure captions in the revised text.	Figure captions in the revised text now state the relevant bin widths (Please see examples for new regionalised PDFs requested above pp. 12-15)
12	Lines 380-383: Please review the meaning in this paragraph as it's confusing. In the first sentence you say the ERA5-driven and ERA-interim driven simulations are similar, in the second you say that the ERA5-driven show large reductions in biases.	Agreed – text is revised to improve clarity/meaning as shown right.	Previous text: “Without switching the driving reanalyses, ERA5-forced CORDEX-CMIP6 ‘NARCliM2.0’ RCMs and ERA-Interim CORDEX-CMIP5 RCMs simulate annual mean maximum temperature over the inner domains (Fig. 8) in a similar manner as compared to over Australia (Fig. 3). That is, the ERA5-NARCliM2.0 RCMs show large reductions in the marked cold biases (Fig. 8b-i) that characterise the ERA-Interim-forced RCMs (Fig. 8j-m), with ensemble mean |biases| of 1.09K and 2.46K, respectively.” Revised text: “Prior to switching their driving reanalyses, the ERA5-NARCliM2.0 RCMs show large reductions in cold bias (Fig. 8b-i) relative to the ERA-Interim-forced RCMs (Fig. 8j-m),

			with ensemble mean bias magnitudes of 1.09K and 2.46K, respectively.”
13	Lines 341 and elsewhere: consider saying 'bias magnitudes' (or mean absolute error) over |biases| in the text.	Phrasing revised throughout.	Phrasing modified throughout the revised manuscript.

14	Lines 488-490: I don't agree that the convection-permitting P99s from R3-R7 are markedly improved over WRFK and WRFL: perhaps a little along the coast but it's fairly marginal.	The text is revised to more accurately state the nature of differences in performance. That is, at convection permitting scale of 4 km, whilst two of the ERA5-RCMs perform poorly for P99 precipitation (R1 and R2), the best-performing ERA5-RCM for P99 precipitation (R3) shows an area-averaged bias of 8.08 mm versus 9 mm to 14.33 mm for the three ERA-I-RCMs simulating at 10 km. Hence, overall, we observe a small performance improvement, but it's not a marked one. Please see previous version of the text versus the revised text in column right.	Previous text, lines 488-490: "However, over the convection-permitting domain, many ERA5-RCMs show enhanced simulation of extreme precipitation relative to the ERA-Interim RCMs, except ERA5-R1 and R2 which are strongly wet-biased." Revised text: "However, at convection-permitting scale, some ERA5-RCMs show small improvements of around 10% in the simulation of extreme precipitation relative to the ERA-Interim RCMs, except ERA5-R1 and R2 which are strongly wet-biased"
15	Figure 8-13: are you able to include cutouts of the 20km outer domains of ERA5 R1-R7 in these figures?	We appreciate the rationale for this suggestion; however, we are also concerned that this might constitute including a lot of plot panels in Figure 8-13, which be too much for some readers.	

Table 1. Diagnostics for seven (R1-R7) ERA5-forced regional climate models (RCMs) and six ERA-Interim-forced RCMs and their respective ensemble means for 1981-2010 with Australian Gridded Climate Data as reference data. Mean absolute biases are shown for annual and seasonal mean maximum and minimum temperature and precipitation, for annual extreme maximum and minimum temperature and precipitation, as well as Perkins Skill Scores (PSS) for the daily distributions of these variables for the CORDEX-Australasia domain over Australia (20 km resolution). Diagnostics for climate means and extremes are shown for the nested inner domain over southeast Australia (4 km resolution). Bold values indicate which of the ERA5-RCMs R1-R7 has the best diagnostic score from this set.

			CORDEX Australasia (20 km)					**Southeast Australia (4 km)**			
			Climate Means: Mean |bias|			**Climate Extremes: Mean |bias|**	**PSS**	**Climate Means: Mean |bias|**			**Climate Extremes: Mean |bias|**
Variable	**Generation**	**RCM**	**Annual**	**DJF**	**JJA**	**Annual**		**Annual**	**DJF**	**JJA**	**Annual**
tasmax (K)	ERA5-RCMs	Ensemble	0.85	0.81	1.34	0.86	N/A	1.09	0.67	1.68	0.84
		R1	0.83	**0.68**	1.10	**0.73**	0.957	1.11	0.91	1.35	0.86
		R2	1.61	1.23	2.30	1.02	0.917	1.85	1.17	2.45	0.88
		R3	0.90	0.80	1.35	0.88	0.950	1.29	0.85	1.80	0.95
		R4	0.92	1.03	1.01	1.26	**0.958**	1.29	1.33	1.56	1.32
		R5	**0.54**	1.09	**0.84**	0.81	0.942	**0.51**	0.62	**1.26**	**0.64**
		R6	0.85	1.18	1.58	0.86	0.922	0.87	**0.55**	1.76	0.70
		R7	0.85	0.99	1.39	1.08	0.938	0.86	0.64	1.58	1.17
	ERAI-RCMs	Ensemble	1.33	0.80	2.24	0.91	N/A	2.46	1.52	4.09	1.81
		WRFJ	1.58	1.29	2.26	1.56	0.940	2.01	1.76	2.64	2.14
		WRFK	1.37	1.06	2.02	1.32	0.945	1.82	1.36	2.44	1.80
		WRFL	2.67	0.99	5.67	1.11	0.880	3.57	1.47	7.19	1.52
		WRFSWWA	1.07	0.92	1.33	0.89	0.952				
		CCAM	0.98	0.97	1.57	1.44	0.904				
		CCLM	0.92	0.94	1.21	1.37	0.946				
tasmin (K)	ERA5-RCMs	Ensemble	0.73	0.89	0.96	1.48	N/A	0.80	0.69	1.05	1.60
		R1	0.95	1.12	0.85	1.30	0.943	0.96	1.11	0.83	1.35
		R2	**0.77**	1.03	**0.70**	**1.02**	0.935	**0.68**	0.89	**0.62**	**1.13**
		R3	**0.77**	1.02	0.96	1.47	0.938	0.75	0.71	1.01	1.51
		R4	0.81	**0.73**	1.23	1.90	**0.944**	0.92	**0.56**	1.47	2.16
		R5	0.93	1.22	1.07	1.55	0.937	0.96	0.93	1.12	1.57
		R6	0.89	1.23	1.24	1.69	0.933	0.93	0.84	1.21	1.71
		R7	0.89	0.99	1.41	1.97	0.930	1.11	0.68	1.61	2.29
	ERAI-RCMs	Ensemble	0.73	0.69	0.76	1.01	N/A	0.62	0.73	1.05	1.07
		WRFJ	0.63	0.69	0.76	0.96	0.976	0.49	0.73	0.75	0.98
		WRFK	0.70	0.72	0.78	0.96	0.975	0.50	0.68	0.75	1.06
		WRFL	1.47	0.78	2.80	2.86	0.915	1.44	0.83	2.70	2.62
		WRFSWWA	1.75	1.78	1.68	2.15	0.912				
		CCAM	1.07	0.59	1.82	1.50	0.945				
		CCLM	2.25	1.75	2.75	3.33	0.900				

pr (mm)	ERA5-RCMs	Ensemble	7.28	18.42	4.31	8.64	N/A	3.97	7.99	7.58	11.40
		R1	13.48	27.82	5.17	20.02	0.773	6.75	12.90	**4.49**	22.47
		R2	11.33	22.79	5.06	14.83	**0.817**	6.17	8.90	8.67	20.90
		R3	8.31	19.72	**5.02**	9.80	0.805	5.65	8.47	8.74	**8.08**
		R4	**7.46**	**16.33**	5.67	**9.21**	0.801	5.96	**6.36**	9.94	9.51
		R5	12.59	33.93	5.21	11.40	0.814	**4.63**	17.69	5.47	8.95
		R6	16.29	49.29	6.16	10.25	0.787	4.68	16.40	10.32	8.22
		R7	15.92	46.43	6.23	9.91	0.787	4.83	12.54	10.76	8.19
	ERAI-RCMs	Ensemble	7.48	12.73	5.96	7.60	N/A	18.96	23.85	14.08	10.35
		WRFJ	20.65	31.54	12.38	8.75	0.798	28.14	39.36	17.82	14.33
		WRFK	12.86	23.31	9.83	11.06	0.770	17.05	12.25	16.32	9.69
		WRFL	7.81	15.96	7.63	9.45	0.678	12.20	21.67	9.55	9.00
		WRFSWWA	9.81	16.82	7.75	20.94	0.806				
		CCAM	10.39	22.85	9.17	15.77	0.837				
		CCLM	11.66	24.05	5.61	17.69	0.798				

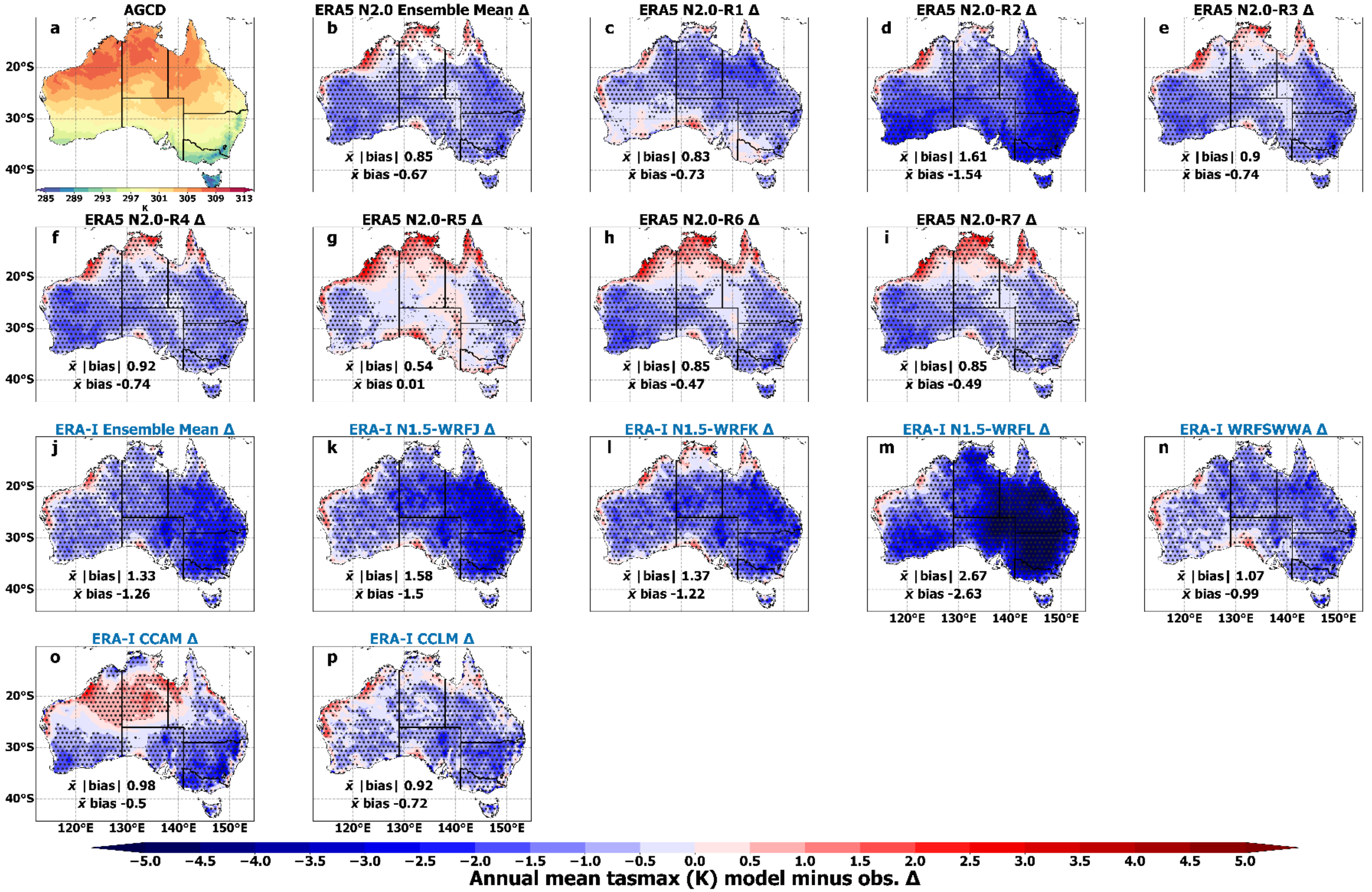

Figure 3 (Revision): As per Figure 3 in the first submission of this manuscript ("Annual mean near-surface atmospheric maximum temperature bias with respect to Australian Gridded Climate Data (AGCD) observations for 1981-2010. …") but with stippling density decrease and marker size increased, and clearer/more consistent representation of state/jurisdictional boundaries.