September 27, 2024

Paper #GMD-2024-87 | Model experiment description paper**: 'Design, evaluation and future projections of the NARCliM2.0 CORDEX-CMIP6 Australasia regional climate ensemble'**

Author Comments (ACs) – Referee 1

Table 1. Anonymous Referee 1 (RC1) Comments

#	Issue Description	Discussion	Revision (in re-submitted manuscript)
	Referee #1: General Comments		
1	The authors have compared the experimental designs and results across three generations of NARCliM RCMs. The latest iteration, NARCliM 2.0, features enhanced spatial resolution and utilizes CMIP6 experiment outputs as large-scale forcing data, representing advancements over earlier phases. The ensemble simulations of NARCliM 2.0 were conducted after a rigorous evaluation and selection process involving CMIP6 models and various physics configurations of the WRF model. This approach has the potential to provide more robust projections of regional climate over Australia. The ensemble simulations, incorporating diverse GCM-RCM combinations, make significant contributions to CORDEX. Therefore, I recommend acceptance pending minor revisions, including clarifications, correction, and reorganization in certain sections. Specific comments are outlined below:	We are very grateful to the reviewer for reviewing our work, for their positive remarks on this work and manuscript, and for recommending acceptance following Minor Revisions.	Please see point-by-point responses below.
	Referee #1: Specific comments		

2	L108: Please replace "NARCliM2.0" with "NARCliM 2.0 (NARCliM 1.5)".	Agreed.	The naming of NARCliM is changed throughout the revised manuscript as suggested by the reviewer.
3	Section 3.2.1: It is unclear which variables were evaluated to assess CMIP6 GCM performance. Note that precipitation, daily maximum and minimum surface air temperatures do not serve as boundary conditions for driving the RCM. It would be preferable to evaluate U, V, T, Q, Z, SST, PSL for dynamical downscaling purposes. This issue should be properly addressed or discussed.	In this study, we evaluated the performance of CMIP6 GCMs by analysing mean climate, including annual and seasonal climatology of maximum and minimum temperatures, and precipitation; climate extremes, such as the 99th percentiles of daily maximum temperature and precipitation, and the 1st percentile of minimum temperature; as well as the teleconnections of ENSO, IOD, SAM, and their influence on Australian regional rainfall. The focus on temperature and rainfall is due to them being the best observed climate variables that provide the most direct comparison to observations (i.e. being gridded observational products). However, we also acknowledge the reviewer's suggestion of using variables such as U, V, T, Q, Z, SST, and PSL, which serve as initial and boundary conditions for driving the RCMs. If we want to evaluate U, V, T, Q, Z, PSL, etc, we would have to use re-analysis as the "surrogate truth/observations". This would be a useful thing to do, i.e., comparing CMIP6 against re-analysis for these variables, but it's a different exercise. These variables / this approach will be incorporated in future studies, and this is acknowledged in a revised version of the manuscript (please see text right).	The revised main text includes the statements below explaining the benefits of focusing on gridded observations of temperature and rainfall in the GCM evaluation, as well as acknowledging the reviewer's suggestion that variables such as U, V, T, Q, etc can be included in future GCM evaluation studies: 1) "Temperature and precipitation variables are chosen for evaluation because they are well-represented in high-quality gridded observational data sets for the Australian continent (King et al. 2013)." 2) "Variables such as winds (U, V), air temperature (T), water mixing ratio (Q), geopotential height (Z), sea surface temperature (SST), and sea level pressure (PSL) could be incorporated into future GCM evaluation studies as these variables serve as boundary conditions for driving RCMs. Evaluating such variables would require use of re-analysis data as surrogate observations."
4	Table 2: Please clarify how many GCM-RCM runs were conducted for CORDEX-CMIP6 NARCliM 2.0. Specify the combinations used. Were all five GCMs downscaled by seven RCMs each? Presenting this	The CORDEX-CMIP6 NARCliM 2.0 regional climate projections are a 10-member ensemble comprising two configurations of the WRF RCM dynamically downscaling the five shortlisted GCMs under three SSPs for 20 km and 4 km (i.e.	Text revised as follows: 1. The text preceding / introducing Table 2 is now revised to add mention that the five CMIP6 GCMs are used to force two,

	information in a table format would aid readers in quickly accessing these details.	convection-permitting scale). Although statements of this sort had been made at various points in the text of the submitted manuscript (please see example below), we agree with the reviewer that this key point can be further clarified (see changes in revised manuscript in column right). The tremendous compute (financial) requirement to run these simulations necessitated us to be selective in the number of RCM configurations chosen to dynamically downscale the shortlisted CMIP6 GCMs. For instance, the ultimate outcome of the CORDEX ERA5-forced NARCliM 2.0 simulations and their evaluation was the selection of the two definitive RCM configurations R3 and R5 to run the CMIP6-forced phase of NARCliM 2.0. An example of existing text in submitted manuscript (see lines 898-901); added text in revised manuscript shown in column right: "In summary, the CORDEX-CMIP6 NARCliM 2.0 regional climate projections are a 10-member ensemble comprising two configurations of the WRF RCM dynamically downscaling five GCMs under three SSPs at 20 km resolution over CORDEX-Australasia and at 4 km convection-permitting resolution over south-east Australia"	definitive RCMs comprising NARCliM 2.0 CORDEX-CMIP6: "As a result of the above process, the five CMIP6 GCMs listed in Table 2 are selected to force each of the two definitive NARCliM 2.0 RCMs selected via the RCM physics testing and ERA5 evaluation processes." 2. The caption for Table 2 is also revised accordingly: "**Table 2.** Basic details of the CMIP6 GCMs used to force two RCMs comprising the NARCliM 2.0 CORDEX-CMIP6 ensemble."
5	L423-424: The authors employed a cold restart for the SSP experiments. Did the authors examined the duration required for deep soil spin-up? Why not use soil moisture from a historical RCM run in 2014	Ideally, we would complete the long-term historical simulation first and use the final restart file from this simulation to initialize the first SSP simulation. However, due to time constraints we had to run historical and SSP simulations	Following text added to the revised manuscript:

	or ERA5 reanalysis as initial conditions for the SSP experiments?	concurrently, using a one-year spin-up period. In this study, we conducted a cold restart for the historical simulation in 2014 and used the final restart files from 2014 to initialize the first SSP simulation in 2015. We also evaluated the time needed for deep soil spin-up, which is approximately 3 to 6 months for different Australia regions. To account for this, we used a 12-month spin-up period, which is sufficient to minimize the impact of the cold restart.	"We tested the time duration required for soil moisture to equilibrate from the cold start and found that 1 year is sufficient."
6	Section 4 Evaluation methods: these evaluation methods were already used in previous sections. It would improve clarity to present this section earlier in the manuscript.	Thanks for this suggestion: we agree it is better to swap section 3 and 4 and make some changes accordingly.	Main text to be revised as suggested.
7	L453-456: RMSE and PSS are typically used to assess model performance in simulating individual variables. However, it remains unclear how overall RCM performance in simulating multiple variables is determined. Did the authors normalized the biases/RMSEs when sum them together? Otherwise, the biases/RMSEs are in different order of magnitude. The authors may consider employing the Model Climate Performance Index (Gleckler et al., 2008) or multivariable integrated skill score (Zhang et al., 2021) for a comprehensive assessment in terms of the model performance in simulating multiple variables.	There are several methods to evaluate the overall performance of RCMs. In this study, we ranked the RCMs individually based on their bias, RMSE, and PSS for maximum temperature, minimum temperature, and precipitation. Each variable was ranked separately for each metric. The ranks were then summed to determine the overall ranking for each RCM. Thank you for suggesting these references; in particular, in future studies we will try the approach of Zhang et al., (2021).	Text below added to the revised manuscript to provide more clarity on this matter: "There are several methods to evaluate the overall performance of RCMs. In this study, we ranked the RCMs individually based on their bias, RMSE, and PSS for maximum temperature, minimum temperature, and precipitation. Each variable was ranked separately for each metric. The ranks were then summed to determine the overall ranking for each RCM."
8	L699: Please replace "CMPI6" with "CMIP6".	Thank you for pointing that out – corrected.	"CMPI6" corrected to "CMIP6".
9	L707-712: Could you explain why projected changes in TAS exhibit distinct spatial patterns between NARCliM 2.0 and NARCliM 1.5/1.0?	Thanks for this comment. In this work, we looked at future projections of mean maximum temperature (TASMAX) rather than mean temperature (TAS). Given your comment, we compared differences in the spatial patterns of projected changes in both TAS and TASMAX	The manuscript had stated the need for further work in this space, noting our comments in the column left. Lines 913-916:

		between CMIP6 and CMIP5 GCMs (see Figure 1 below this table). Both GCM generations show broadly similar spatial patterns of change (at least qualitatively). However, there are clear differences in magnitude, e.g. whilst both CMIP5-6 show stronger warming changes across an east-west band of central Australia, the magnitude of change is larger for CMIP5, probably in large part to the differences in GHG assumptions (See Figure 1 below this table). Additionally, GCM skill in simulating observed TASMAX is fairly similar for both GCM generations (see Supporting Information Figure S7), noting though that the spatial patterns of bias are somewhat different (e.g. the CMIP6 ensemble mean is more cold biased over northern Australia than CMIP5; conversely the CMIP5 GCM ensemble mean is more cold biased over southern and eastern Australia). This topic requires an additional in-depth investigation to understand and explain which is out of scope for this paper. For example, TMAX is usually driven at the larger scale by changes in MSLP, e.g, the sub-tropical ridge and its intensification, this in turn probably affects changes in precip. and surface energy balance, so we would need to examine changes in potentially: MSLP, precip., soil moisture, sensible, latent heat fluxes etc. Our aim with this current work is to explain key model design processes and the basic performance characteristics of the NARCliM models, i.e. to lay a foundation for future work in this space. There might be several	"Other differences in the projections between NARCliM generations require further investigation in order to explain, such as NARCliM 1.5's latitudinal warming gradient for maximum temperature that contrasts with NARCliM 2.0's band of faster warming over central Australia relative to northern and southern regions."

		factors that underlie the different/distinct spatial patterns in projected temperature changes for NARCliM 2.0 and NARCliM 1.x. For instance, changes in model spatial resolution are one possible candidate, given that the resolution of CMIP6 GCMs is higher than CMIP5 GCMs, and the same applies to NARCliM 2.0 RCMs versus its predecessors. However, we expect that there will be other factors that explain the observed differences in NARCliM RCM behaviour.	
10	Fig.15: The quality of this figure appears low. Why do the stippling areas form very regular circles in the many subpanels, e.g., b, c, e, n, p, t, u, v? Consider presenting these figures as supplementary material and summarizing the statistics using a Taylor diagram.	We agree that the quality of the original Figure 15 was insufficient: this figure is now revised, e.g. with DPI increased from 300 to 600, stippling size increased, panel title font size increased, etc., – please see revised figure below this table.	Figure 15 revised (please see example below this table).
11	L804-816: These discussions are somewhat tangential to the study's main focus and could be shortened or omitted. Instead, further investigate/discuss the differences in projected changes in the surface air temperature and precipitation among the three generations of NARCliM. For example, explore why widespread wet biases observed in NARCliM 1.x are substantially reduced in NARCliM 2. Are these biases attributable to GCMs, RCMs, or both?	This study focuses on summarizing the improvements in the NARCliM2.0 design, including the incorporation of the Noah-MP land surface model, which has significantly reduced cold biases in both ERA5 and GCM-driven simulations. This section discusses the successful application of Noah-MP in other regions, which aligns with the results we achieved in our project. Additionally, we explore how Noah-MP performance in Australia can be further enhanced by selecting specific settings rather than relying on default ones for future regional climate modelling. We believe these discussions are relevant to the focus of the study. We also appreciate the reviewer's suggestion to address why the wet biases in NARCliM1.0 and NARCliM1.5 were reduced in NARCliM2.0. The	Revised text now includes the following additional text (based on that in column left): "Overall, the CMIP6 GCMs used to drive NARCliM 2.0 show marginally reduced wet biases relative to the CMIP5 GCMs used to drive NARCliM1.5 RCMs (e.g. area-averaged ensemble mean absolute biases are 7.13 mm and 8.89 mm, respectively; Supporting Information Figure S15). This suggests that the underlying nature of the CMIP6 driving data is not the principal factor underlying the observed improvements for NARCliM 2.0's simulation of mean precipitation. In fact, the RCMs appear to have a substantial influence on the reduced maximum

		main aims of the present paper are more focused on introducing the model design processes, and the basic performance profiles of the new models as compared to the previous generations, with more detailed explorations explaining differences in model skill etc to be the topics of future work. For instance, this topic is also being discussed in more detail in another paper, 'Three Generations of NARCliM: Model Evaluation and Future Projections over CORDEX Australia,' which is currently under review. That said, we can suggest initial explanations as to why widespread wet biases observed in NARCliM 1.x are substantially reduced in NARCliM 2.0: Overall, the CMIP6 GCMs used to drive NARCliM 2.0 show marginally reduced wet biases relative to the CMIP5 GCMs used to drive NARCliM1.5 RCMs (e.g. area-averaged ensemble mean absolute biases are 7.13 mm and 8.89 mm, respectively; Supporting Information Figure S15). This suggests that the underlying nature of the CMIP6 driving data is not the principal factor underlying the observed improvements for NARCliM 2.0's simulation of mean precipitation. In fact, the RCMs appear to have a substantial influence on the reduced maximum temperature biases. Conversely, in terms of RCM design features, the use of the Noah-MP LSM in the NARCliM 2.0 RCM physics tests conferred overall RCM skill improvements relative to RCMs using the Noah-Unified LSM for both mean max	temperature biases. Conversely, in terms of RCM design features, the use of the Noah-MP LSM in the NARCliM 2.0 RCM physics tests con-ferred overall RCM skill improvements relative to RCMs using the Noah-Unified LSM for both mean max temperature and precipitation. The developers of Noah-MP suggest that some limitations in the Noah-Unified LSM have been modified to better represent several parameters such as soil moisture and heat fluxes, leaf area-rainfall interaction, vegetation and canopy temperature distinction, drainage of soil, and runoff. The production NARCliM2.0 RCMs forced with CMIP6 GCMs used Noah-MP, whereas NARCliM1.x RCMs used Noah-Unified. Given these performance improvements observed for RCMs using Noah-MP versus RCMs using Noah-Unified, it's plausible that the different land surface schemes (i.e. Noah-MP for NARCliM 2.0 versus Noah-Unified for NARCliM 1.x) play a role in the improved NARCliM2.0 RCM skill in simulating mean precipitation (as well as max temp), for instance, via changing the land surface feedback (via soil moisture) to the simulation of precipitation. However, this possibility requires more extensive investigation via future studies."

		temperature and precipitation. The developers of Noah-MP suggest that some limitations in the Noah-Unified LSM have been modified to better represent several parameters such as soil moisture and heat fluxes, leaf area-rainfall interaction, vegetation and canopy temperature distinction, drainage of soil, and runoff. The production NARCliM2.0 RCMs forced with CMIP6 GCMs used Noah-MP, whereas NARCliM1.x RCMs used Noah-Unified. Given these performance improvements observed for RCMs using Noah-MP versus RCMs using Noah-Unified, it's plausible that the different land surface schemes (i.e. Noah-MP for NARCliM 2.0 versus Noah-Unified for NARCliM 1.x) play a role in the improved NARCliM2.0 RCM skill in simulating mean precipitation (as well as max temp), for instance, via changing the land surface feedback (via soil moisture) to the simulation of precipitation. However, this possibility requires more extensive investigation via future studies.	

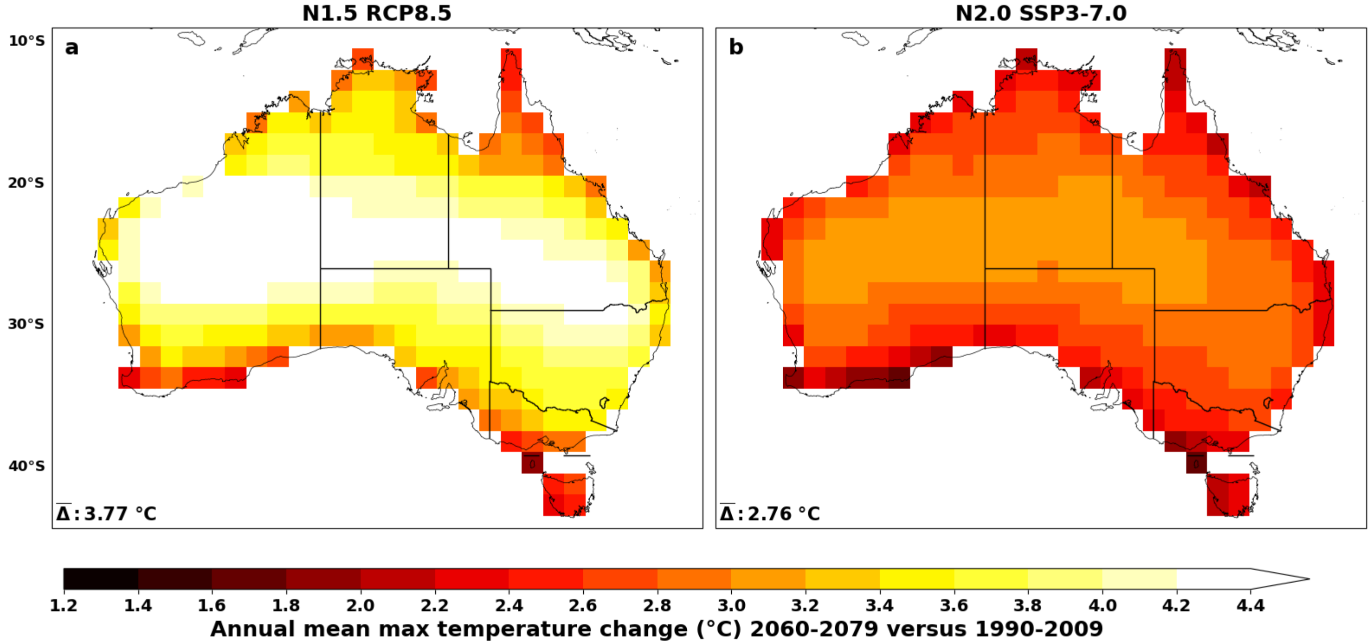

Figure 1 (New). Future projections of mean maximum temperature for the ensemble means of CMIP5 GCMs forcing NARCliM 1.5 and CMIP6 GCMs forcing NARCliM 2.0

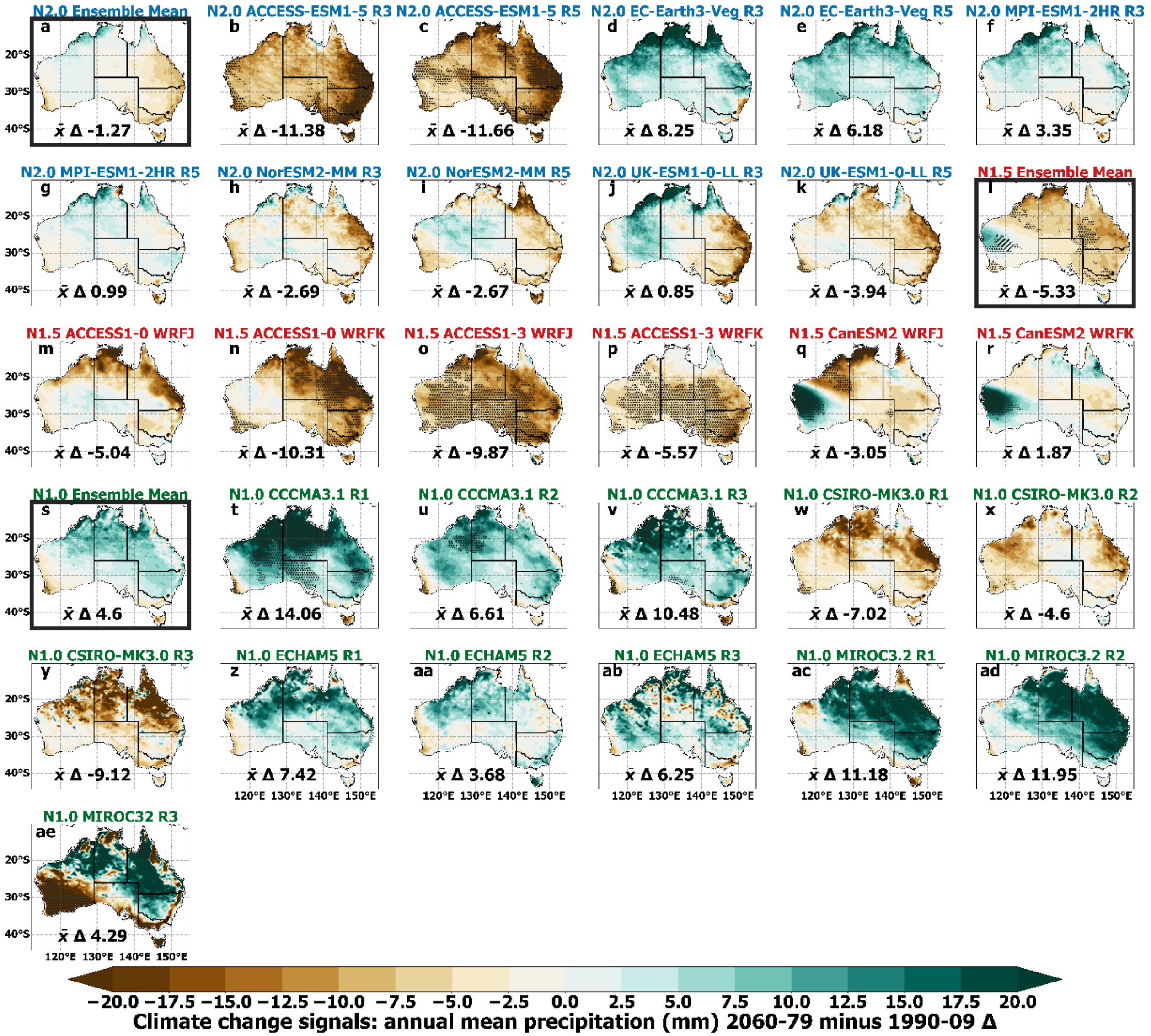

Figure 15: revised version