**Responses to RC2:** GMD-2023-162
Stefan Rahimi et al.

The reviewer comments are presented followed by underlined author responses.

In this manuscript, the authors investigated the performance of dynamically downscaled simulations of climate conditions over the western U.S. from 1980 to 2100. The simulations were conducted using the WRF model driven by sixteen selected CMIP6 GCMs for various SSP scenarios. The WRF model was run on two nested domains with resolutions of 45 km and 9 km, yearly initiated with a one-month spin-up for the land surface, spanning from August 1 to September 1 of the next year. The authors have addressed many aspects related to the dynamical downscaling technique applied in climate change projection. The manuscript is very interesting, and it's worth publishing in the Geoscientific Model Development. However, I would like to request the authors clarify the following points.

1. Since the WRF model was initiated yearly, soil moisture data for the land model were provided by GCMs. These soil moisture data may be significant differences compared to those in case continuous running of the model. Have the authors conducted tests to assess the impact of this difference on the model outputs?

In short, yes – tests were conducted to assess the impact of different spin-up periods.

The assumption that one month of soil spin-up is sufficient for climate change simulations is a massive one and deserves scrutiny (see comment by RC1), and the assumption should be regarded as a substantial limitation of WUS-D3. From the outset, we committed to some type of parallelization strategy to reduce integration times similar to other studies (e.g., the previous works of Zobel et al. below). However, we were not initially sure how much spin-up time was to be used nor how to parallelize. We eventually justified a one-month spin-up in reanalysis-direven tests in Rahimi et al. (2022), Here, we conduycted two year-long test experiments for water year 2010. In case 1, we used a single month of spin-up, and in test 2, four years of spin-up were integrated. Broadly speaking, we found there to be minimal differences in simulated soil moisture, soil temperature, surface air temperature, and precipitation between the two cases.

We are wary of the spin-up issues and resulting discontinuities in land-surface variables. For example, snow in WUS-D3 simulations is generally far too wet over the historical period, a feature common to different GCMs. By the end of each simulated year (31 August), snow does not completely melt out at all locations, leading to a discontinuity in the snow fields between 31 August and 1 September. Across these areas, this results in discontinuities in surface energy fluxes as the reviewer suggests. We have thus added Sec. 2.6 to the manuscript cautioning end-users about this issue:

'     Despite one month of spin-up in parallelized yearly WRF experiments, our adopted spin-up strategy neglects high-resolution soil memory on time scales greater than one month. This assumption may be particularly problematic across regions where a transient simulation is necessary to equilibrate the soil conditions to a state which properly resolves the local-scale land-atmospheric coupling. For instance, some grid points do not see complete meltout of snow by 31 August 1993, but since data is retained from 1 September 1993 onwards, there are instances where discontinuities in surface snow coverage exist. This leads to discontinuities in surface energy variables (e.g., sensible heating; not shown). We encourage end-users of WUS-D3 to be wary of this pitfall. To alleviate this discontinuity, we propose that the atmospheric temperature, precipitation, surface radiative fluxes, winds, and specific humidity from WRF be used to drive offline calibrated hydrology models that are time-continuous and can be integrated much more rapidly (e.g., Bass et al., 2023). We acknowledge that this approach is inadequate across regions with a strong land-atmosphere coupling.

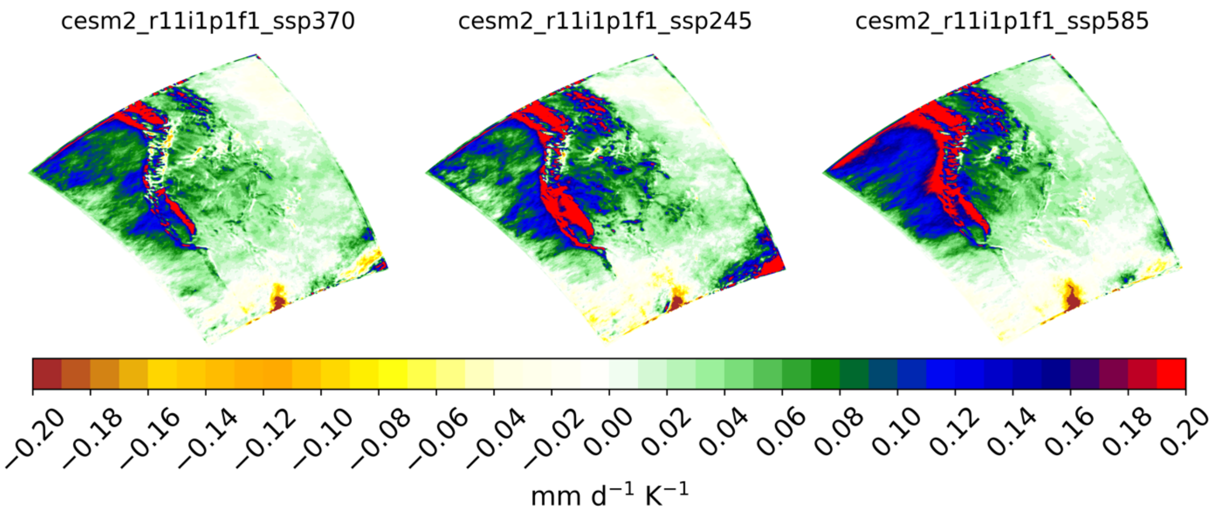
2. Why did the authors used one-way nesting from the parent domain (45-km) to the inner domain (9-km) instead of employing two-way nesting to gather feedback on local features that could benefit from a finer resolution?

This choice was made for pratical and technical reasons.

Tehnically, given our choice (determined via testing) to spectrally nudge the 45-km grid's large-scale temperature and horizontal winds above the boundary layer to prevent model drift, we were concerned about how any such simulated feedbacks may be obfuscated by the nudging. We did explore the option of two-way nesting however, which led to a practical limitation. Specifically, we were ultimately downscaling to 3 km (not the subject of this manuscript), and continual crashes were found at the grid interfaces (45 with 9 and 9 with 3) which in some locations bifurcated complex terrain. In short, nudging choices and grid location led us to choose a one-way nesting approach.

3. I'm uncertain about the method the authors used to determine the projected changes in temperature (K $K^{-1}$) and precipitation (mm $d^{-1}$ $K^{-1}$) per degree of global warming in Figure 7. Just to clarify, are these changes being calculated only for the end of the 21$^{st}$ century? Please let me know if I understood correctly. If so, is the calculation of "global warming" based on the ensemble mean derived from all sixteen GCMs for the entire globe?

Thank you for the opportunity to clarify. The changes you see in Fig. 7 are being computed for the 2070-2100 period (end-century) relative to the historical period (1980-2010). In the instance of precipitation changes, we compute the ensemble-mean precipitation change [mm/d] and divide by the ensemble-mean global temperature change [K]. We compute anthropogenic response this way for consistency with the IPCC and because this approach effectively eliminates scenario uncertainty. For instance, considering the CESM2 experiments only, if we compute the change in precipitation normalized by the amount of global warming for mid-century (2030-2060), we will get a similar plot to another version in which we compute the change in precipitation for end-century normalized by the end-century global warming. We also see similar maps of of the global mean temperature-normalized precipitation response between the SSP2,3, and 5 emission trajectories of the CESM2 experiments ofr the end-century period:




**Non-text references**

Bass, B., Rahimi, S., Goldenson, N., Hall, A., Norris, J., and Lebow, Z. J.: Achieving Realistic Runoff in the Western United States with a Land Surface Model Forced by Dynamically Downscaled Meteorology, Journal of Hydrometeorology, 24, 269–283, https://doi.org/10.1175/JHM-D-22-0047.1, 2023.

Zobel, Z., Wang, J., Wuebbles, D. J., and Kotamarthi, V. R.: High-Resolution Dynamical Downscaling Ensemble Projections of Future Extreme Temperature Distributions for the United States, Earth's Future, 5, 1234–1251, https://doi.org/10.1002/2017EF000642, 2017.

Zobel, Z., Wang, J., Wuebbles, D. J., and Kotamarthi, V. R.: Evaluations of high-resolution dynamically downscaled ensembles over the contiguous United States, Clim Dyn, 50, 863–884, https://doi.org/10.1007/s00382-017-3645-6, 2018.

Rahimi, S., Krantz, W., Lin, Y.-H., Bass, B., Goldenson, N., Hall, A., Lebo, Z. J., and Norris, J.: Evaluation of a Reanalysis-Driven Configuration of WRF4 Over the Western United States From 1980 to 2020, Journal of Geophysical Research: Atmospheres, 127, e2021JD035699, https://doi.org/10.1029/2021JD035699, 2022.