Responses to CC1: GMD-2023-162
Stefan Rahimi et al.

The reviewer comments are presented followed by underlined author responses.

The paper presents a comprehensive study introducing downscaling work to a 9 km resolution for 16 CMIP6 GCM experiments using the WRF model. The manuscript well describes the methodology and the WUS-D3 dataset. I recommend publishing the paper after some minor revisions as follows.

Comments:
1. GCM selection: the authors outlined 6 processes considered in the evaluation and selection of GCMs. While they refer to the ranking methodology in a technical note (Krantz et al. 2021) and a paper currently under revision (Goldenson et al. 2023, in revisions), I recommend providing more information on two key aspects. Firstly, elaborate on the process selection – explain why these 6 processes were chosen; why extreme precipitation across California is included among the selected processes, given the coarse resolution of GCMs and the fact that this diagnostic variable might not play a role in the ICBC of the downscaling framework. Secondly, provide more details on the ranking methodology: clarify how these 6 processes are considered in the final ranking; are they equally weighted? How do temporal and spatial patterns contribute to the selection process?

Happy to provide additional context and clarification here. For the first question, the processes were chosen based on our team's experience in processes that are of regional relevance. For California specifically, we actually considered a set of metrics for biases within just this category. Specifically, we considered the following, conditioned on days when GCM precipitation exceeded the 95th percentile (biases relative to ERA5): integrated water vapor, sea level pressure, and the 250 hPa zonal wind. We also used the third empirical orthogonal function (EOF) of the 500 hPa geopotential, whose spatial pattern is strongly correlated with extreme precipitation (> 99th percentile) across the region (See Chen et al., 2021). Additionally, we also included a metric for large-scale circulations from the GCM that may affect Los Angeles extreme precipitation.

To answer the first question, the metrics of GCM bias were of hemispheric, Pacific Ocean, western U.S., and California scope. Since California is so latitudinally expansive however, and because the landfalling atmospheric rivers that bring the state extreme precipitation generally provide abundant rainfall downstream to the western U.S. interior, we believe that including biases across California should be regarded as regional versus point biases. Further, since the large-scale patterns of temperature and horizontal winds are preserved above the boundary layer via spectral nudging on the 45-km WRF grid, we believe that consideration of biases in the large-scale dynamic fields associated with extreme precipitation events across California, have regional-scale and western U.S. consequences.

Regarding the second question, spatial patterns and temporal patterns are considered in the selection process. For instance, the time-variability of ENSO and high-frequency synoptic variability of landfalling waves are considered, while the spatial variability of the California precipitation mode is considered via the identification of where the geopotential anomalies exist upstream of the region. Additionally, our metrics per Simpson et al., (2020) consider jet stream landfall position bias, accounting for spatial bias. More generally however, these processes were not considered equally in the finalized GCM selection process. First, metric redundancy was addressed by computing a set of EOFs from the metrics, and only retaining a subset of EOFs that capture most of the variation between models. The result is a reduced set of linear combinations of metrics that efficiently captures nearly all of the variance across GCMs; this process constituted a weighting of the metrics themselves based on redundancy with other metrics. It was found that the first 6 EOFs described 91% of the variance amongst models, with only a subset of metrics explaining most of the variability between GCMs. After the EOF decomposition, and overall score was computed for each GCM. The least biased (highest scoring) GCMs were then generally selected.

We have rewritten Sec. 2.2 to read: 'Prioritizing SSP3-7.0 with an end-of-century radiative forcing of 7 W m^{-2}, we selected 14 GCMs (Table 1) based on three criteria: (i) their skill in simulating important processes that govern western North American climate over the historical (1980-2010) period, (ii) their collective representativeness of the broader CMIP6 ensemble spread in future temperature and precipitation responses, and (iii) data availability. Aspects considered in the GCM evaluation included:

1. Large-scale meteorology associated with Santa Ana and Diablo winds – important for extreme wind and fire risk across the southwestern U.S. We use this metric to minimize the usage of GCMs which simulate a distorted portrayal of the Pacific High.
2. The El Niño Southern Oscillation (ENSO) – well-known to modulate the interannual variability of precipitation and temperature across the western U.S. We use this metric to prioritize GCMs which adequately capture the ENSO-Western U.S. teleconnection.
3. Northern Hemisphere blocking and circulation (Simpson et al., 2020) – Wave characteristics, both over climate and synoptic time scales, are directly related to the variability of precipitation across the Western U.S. We use this metric, for instance, to ensure that GCMs are down-selected if they are too progressive in their simulation of mid-latitude waves.
4. Landfalling jet characteristics – Atmospheric rivers are responsible for a majority of West-Coast precipitation. As such, we only select GCMs that demonstrate superior performance in their landfalling position and tilt.
5. GCM-simulated surface air temperature and precipitation – while these variables can be incorrectly simulated in GCMs despite the more-or-less correct treatment of their local driving processes, which may be more important for driving a regional climate model, we include these variables to account for the relationships between the GCM-simulated processes and GCM-simulated surface temperature/precipitation profiles.
6. Extreme precipitation across California – Generally, extreme precipitation events in California are driven by large-scale synoptic events (described by column water vapor, 500 hPa geopotential, and upper tropospheric wind speeds). These

large-scale patterns can have ramifications for weather and climate as they propagate downstream, hence we include an evaluation of bias in these fields for our GCM selection.

7. Regional wind shear – Wind shear helps to moduleate the lifetime of precipitation systems through storm-scale organization and is a measure for the larger-scale background baroclinicity which is important for storm tracks. We thus evaluate its bias.

The ranking system is described in Krantz et al. (2021), and the process of choosing GCMs to downscale based on end-user needs and locally relevant atmospheric processes is described in Goldenson et al. (2023). To emphasize, being subject to these selection processes, the GCMs downscaled in this study span the range of future changes in temperature and precipitation from CMIP6 across the WUS.

For more details on the GCM selection process, we refer readers to Krantz et al., (2021). However, we highlight that temporal and spatial variability was considered in ranking a preferred set of GCMs to downscale. Specifically, the time-variability of ENSO and high-frequency synoptic variability of landfalling waves are considered, while the spatial variability of the California precipitation mode Chen et al. (2021) was factored into our analyses via the identification of where the geopotential anomalies exist upstream of the western U.S. on extreme precipitation days. Additionally, our metrics per Simpson et al., (2020) consider jet stream landfall position bias. Finally, Krantz et al. (2021) performed a variance decomposition using empirical orthogonal functions to reduce the effects of metric redundancy, weighting them accordingly in the final rankings of GCMs.'

2. L330: The authors stated that "Interestingly, downscaling generally reduces warming (leftward pointing arrows)" and hypothetically attributed it to the reduced snow albedo feedback with downscaling. I recommend that the authors prove this hypothesis by comparing the snow outputs of both WRF and GCMs.

The suggestion of a stronger warming response in the GCMs relative to WRF is intended only as a hypothesis and thus needs to be tested. Our group plans to look at this in another paper (in preparation).

3. The authors conducted a more in-depth analysis of the changes in rx1day and tmax99. However, there is no explanation as to why only these two indices, among many possible extreme indices, were selected. Furthermore, why did the authors opt for the absolute index (rx1day) when analyzing rainfall, while choosing the percentile index for temperature.

The purpose of this manuscript was to present the dataset rather than conduct extensive scientific process studies and analyses of extremes. However, we wanted to inspire the community to use the dataset, so we conducted initial analyses to examine mean changes in mean temperature and precipitation, as well as rx1day precipitation and Tmax99. We looked at these common extreme metrics (and mean changes) to showcase how the footprint of topography is represented in the climate response, a feature not characteristic of the GCMs. Rx1day is a common metric for extreme precipitation while Tmax99 is also common in extreme heat analyses, so for an overview of the dataset, we thought that presenting these two metrics alone would be enough to inspire community analysis. We certainly acknowledge that this analysis was by no means comprehensive, and future studies using WUS-D3 should include a more expansive set of metrics.

Minor comments:
1. L210, 215 should refer to Table 1's last column. The caption of Table 1 should also provide an explanation of the last column (SST mode)

Done!

2. Please add the names of locations mentioned in the text to Figure 1, such as California's Central Valley, Sierra Nevada, and state names, …

This is a great idea and has been done for state names. Regarding the mountain ranges, this may be difficult since we list 5-6 over a large geographic region. Indicating these regions in Fig. 1 may clutter the figure. Thus for now, we only include state names.

References

Krantz, W., Pierce, D., Goldenson, N., and Cayan, D.: Memorandum on Evaluating Global Climate Models for Studying Regional Climate Change in California, *The California Energy Commision,* https://www.energy.ca.gov/sites/default/files/2022-09/20220907_CDAWG_MemoEvaluating_GCMs_EPC-20-006_Nov2021-ADA.pdf, 2021.

Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., Lauritzen, P. H., Marsh, D. R., Mills, M. J., Medeiros, B., and Richter, J. H.: An Evaluation of the Large-Scale Atmospheric Circulation and Its Variability in CESM2 and Other CMIP Models, Journal of Geophysical Research: Atmospheres, 125, e2020JD032835, https://doi.org/10.1029/2020JD032835, 2020.