Memo on WRF Downscaling with Bias Corrected Boundary Conditions

The motivation, development and evaluation of four “new” GCMs

Prepared by Stefan Rahimi, Dave Pierce, Dan Cayan and Julie Kalansky
University of Wyoming & UCLA, Scripps Institution of Oceanography, UCSD
August, 2024

- This research is funded by the California Energy Commission (CEC) through its Electric Program Investment Charge (EPIC) Program, which invests in scientific and technological research to accelerate the transformation of the electricity sector to meet the state’s energy and climate goals.
- The research project, EPC-20-006, will integrate the latest downscaling approaches applied to the recently produced global climate models (GCMs) with an engagement process to develop a robust, usable, set of climate projections applicable for California.
- This memo and data here within are being shared to support transparent and timely consideration of interim deliverables that are relevant for energy stakeholders and all those interested in California’s next generation of climate projections.

This memorandum is submitted to the CEC by UC San Diego’s Scripps Institution of Oceanography. The report meets deliverable requirements under Task 4 of the California Energy Commission’s Project EPC-20-006: Development of Climate Projections for California and Identification of Priority Projections.

Synopsis

In this document, we present four new dynamically downscaled global climate model (GCM) projections, building on the original four GCMs under of EPC-20-006, Task 4. Specifically, new GCMs from the 6th Coupled Model Intercomparison Project (CMIP6) are downscaled here using the Weather Research and Forecasting (WRF) model. Information on the development of the original four dynamically downscaled projections can be found in the Memos on (a) the Development and Availability of Dynamically Downscaled Projections Using WRF[1] and (b) the Evaluation of Downscaled GCMs using WRF[2] - Task 4 deliverables of the California Energy Commission's Project EPC-20-006.

The four new GCMs that were dynamically downscaled were, like the original four projections, preselected based on their abilities to simulate the atmospheric circulation (both in the Northern Hemisphere and across California), extremes, and California's temperature and precipitation; as well as availability of data at 6-hour intervals on native GCM levels to drive WRF. However, the need for photovoltaic (PV) and hub-height winds by climate data end-users required the production of these four new simulations. In this memo we discuss the motivations behind their selection as well as the choice to bias correct the GCM boundary condition data *before* (*a priori* to) dynamical downscaling. Further, we present comparisons of the original four projections to the four new *a priori* bias corrected projections.

1. Introduction: Bias Correction and downscaling

Global climate models use millions of calculations and estimations to represent interactions in the earth system. This information and data are extremely useful for understanding the climate system and predicting its future change. However, all models have errors. Systematic model errors are called biases. Examples of climate model biases include overly wet winters or heat waves that are not as extreme as observed. The process of artificially reducing model biases is called bias correction.

Bias correction (BC) of Global Climate Model (GCM) data, downscaled or otherwise, is the process by which biases in their historical-era meteorological fields are quantified relative to another observationally constrained dataset and removed. The process of BC is to-date an unfortunate necessity since numerical models predict solutions that can have unrealistic biases. For climate data users interested in impacts modeling, these biases can be problematic.

Take electricity utility planners, for example, who use downscaled temperature outputs to project peak load and electricity grid strain in a warming climate. Their in-house electricity load models are built on historically observed demand, which is related to historically observed temperature, as more buildings turn on (off) their air conditioners on hot (cold) days. If these electric utility load models are driven directly with non-bias corrected downscaled GCM data which are, say, 3°C too hot on average and/or misrepresent the variability and intensity of 3-day heatwaves, the models can easily return projections of future peak load that are nonsensical. Worse, utility companies may invest too much or too little capital in system resiliency upgrades, ultimately leading to punishing consequences to taxpayers through increased taxes or rolling blackouts. In either case, planners using downscaled data in their impacts models must be confident in the fidelity of historical load/meteorology relationships in order to place faith in their future projections, thus motivating the use of BC.

In dynamical downscaling (DD), Regional Climate Models (RCMs) are driven at their boundaries by data from reanalyses (e.g., Rahimi et al., 2022; Rasmussen et al., 2011; Rasmussen et al., 2023) or GCMs (e.g., Bukovsky & Karoly, 2011; Rahimi et al., 2024a,b; Wang & Kotamarthi, 2015) across a limited area of the planet over which computational power can be

[1] https://www.energy.ca.gov/media/7265
[2] https://cal-adapt.org/files/01_Memo_Evaluation_of_Downscaled_GCMs_Using_WRF_CEC_final.pdf

focused into resolving complex terrain and coastline geometry, as well as otherwise high-resolution land and atmospheric processes. RCMs like the Weather Research and Forecasting (WRF) model (Skamarock et al., 2019; Skamarock & Klemp, 2008) used in EPC-2006, solve the primitive governing equations subject to the imposed boundary conditions, meaning that inherited biases from the boundary condition data, either from GCMs or reanalyses, may lead to biases in downscaled precipitation, temperature, and other variables. These biases in downscaled fields may be enormous, necessitating BC of these data before delivery to stakeholders. Specifically in EPC-20-006, the research team found there to be California-wide mean biases of -1.71°C and +1.86 mm d^{-1} for our original experiments (downscaled to 3 km x 3 km) relative to observations for surface air temperature and precipitation, respectively, over the 1981-2010 period. Further, local biases were much larger, particularly over the Sierra Nevada (Figure 1). Conversely, a well-calibrated dynamically downscaled reanalysis (ERA5; henceforth ERA5-WRF; Rahimi et al., 2022; Hersbach et al., 2020), showcased much smaller biases of +0.13°C and +0.19 mm d^{-1} for these variables, respectively. Temperature and precipitation biases in all in the original 4 dynamical downscaled models of these products were removed from many of these WRF data, as their statistics were forced to match the Livneh gridded dataset (Livneh et al., 2013) via an *a posteriori* (after downscaling) BC using PresRat.

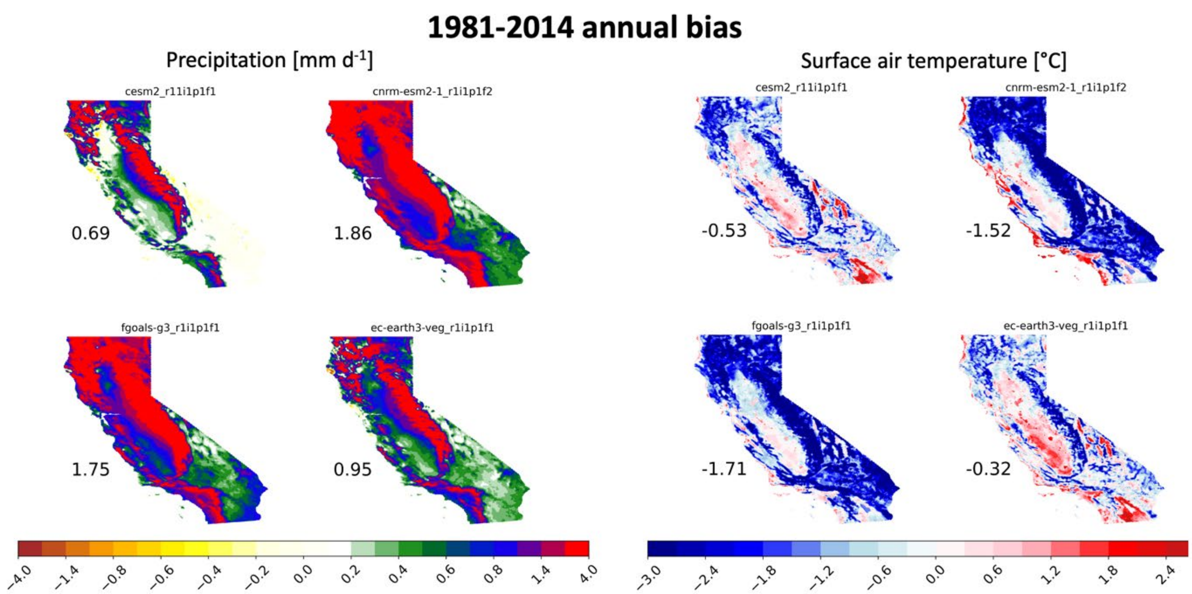

Figure 1. Historical-era (1981-2014) annual mean biases in (left) precipitation [mm d^{-1}] and (right) surface air temperature [°C] for the four original downscaled GCMs.

Despite a rigorous GCM selection process, the large bias disparity between ERA5-WRF and the 4 original dynamically downscaled GCM projections (see blue-tabled GCMs in Table 1) indicated that the transmission of GCM biases to WRF was to blame for the relatively large biases in dynamically downscaled precipitation and temperature. This raised the following questions:

1. Were the GCM inputs to WRF so biased that the local-scale long-term climate trends distorted in unphysical ways? This is important since PresRat preserves WRF trends.
2. Is there a BC that can be applied to the GCM boundary conditions *a priori* to DD that can lead to downscaled GCM precipitation and temperature comparable to WRF-ERA5 and observations, effectively lowering the intensiveness of BC via PresRat *a posteriori* to DD?

Below we address these two questions, as we justify and describe the choice to implement *a priori* BC to the newest four dynamically downscaled GCMs. A listing of all CEC GCM experiments is provided in Table 1. The original four GCM projections do not have their boundary condition data bias corrected prior to DD (henceforth referred to as no-BC experiments), while the newest four projections do have their boundary condition data bias corrected prior to DD (henceforth referred to as the w/BC experiments). We also provide a fifth 3-km w/BC experiment, EC-Earth3-Veg[3], which is identical to the no-BC version.

Table 1. GCMs delivered as part of EPC-20-006, classified by the use of bias correction (BC). GCM nicknames are given in parentheses. Note, EC-Earth3-Veg does not contain the new PV and wind variables of the other bias corrected experiments.

GCM	Country
Initial four projections: No BC of GCM boundary conditions	
1. CESM2 r11i1p1f1 (CESM2)	United States
2. CNRM-ESM2-1 r1i1p1f2 (CNRM)	France
3. FGOALS-g3 r1i1p1f1 (FGOALS)	China
4. EC-Earth3-Veg r1i1p1f1 (ECE)	Sweden
Newest projections: BC of GCM boundary conditions before downscaling	
1. MIROC6 r11i1p1f1 (MIROC)	Japan
2. TaiESM1 r1i1p1f1 (TaiESM)	Taiwan
3. EC-Earth3 r1i1p1f1 (ECEnoVeg)	Sweden
4. MPI-ESM1-2-HR r3i1p1f1 (MPI)	Germany
5. EC-Earth3-Veg r1i1p1f1 (ECEnew)	Sweden

2. Bias Correction of the GCM boundary conditions

2.1. BC method

Since the completion of the no-BC simulations (referred to as the initial four projections), the UCLA and Scripps team have been performing tests to assess if GCM biases that are transmitted to WRF can affect climate trends. Specifically, we tested if GCM biases in atmospheric temperature, horizontal winds, and moisture, as well as sea-surface temperatures (SSTs) could explain the very different precipitation, snow, and temperature bias profiles between the no-BC and WRF-ERA5 experiments.

We impose a BC method that is minimally aggressive and has physically identifiable effects in WRF downscaling. By ‘minimally aggressive’, we mean that we want as much of the original GCM signal to be preserved as possible after BC. Thus we implement the BC procedure adopted in Bruyère et al. (2014) and Holland et al. (2010), which decomposes the above GCM variables (x) into the sums of their monthly climatological mean state ($x_{GCM,0}$), averaged from 1980-2014, and the residuals (x'_{GCM}) which contain the synoptic variability and climate trends:

$$x_{GCM} = x_{GCM,0} + x'_{GCM}. \quad (1)$$

Next, we do the same for a reference dataset used in BC, in this case, native ERA5 (not ERA5-WRF):

$$x_{ERA5} = x_{ERA5,0} + x'_{ERA5}. \quad (2)$$

[3] This was an early test simulation to assess the effectiveness of BC. Thus, we do not provide wind and PV variables for this simulation. ECE and ECEnew have an identical phasing of internal variability.

We then define the mean-state bias as $\Delta = x_{GCM,0} - x_{ERA5,0}$, and subtract this from the original GCM signal to arrive at our bias corrected signal (x_{BC}) and use this signal to drive WRF:

$$x_{BC} = x_{GCM,0} + x'_{GCM} - \Delta, \tag{3}$$
$$x_{BC} = x_{ERA5,0} + x'_{GCM}. \tag{4}$$

This BC procedure is applied monthly and at every grid point to the entire 1980-2100 time series. Time stationarity is assumed in Δ, which is unsampled to 6-hourly, the boundary condition update interval in WRF. In the no-BC experiments, WRF is driven by x_{GCM}.

We chose this BC technique for the following reasons:

1. BC may break/violate the physical relationships linking the variables used to drive WRF. (e.g., mass continuity, energy conservation). While WRF does not conserve mass or energy anyways and integrates its own internal physically consistent solution, a more aggressive BC procedure (e.g., by quantile) may create imbalances in WRF that may manifest as unintended and unphysical artifacts in DD.
2. Despite the assumption of time stationarity in the BC, the above method preserves the original GCMs' trends and variability.
3. A gridpoint-specific versus a spatial-mean BC was chosen since WRF's large-scale temperature and horizontal winds are spectrally nudged to parent GCMs in downscaling. The spatial geometry of these biases can be linked to biases in dynamically downscaled precipitation, temperature, and snow (Rahimi et al., 2024).

2.2. Quality checks of the BC

A two-stage check for the success of the implemented BC was conducted. First, we checked that the historical means (1980-2014) of the bias corrected fields (horizontal winds and temperature above the boundary layer and SSTs) matched those from ERA5-WRF across the 45-km intermediate nest; here ERA5-WRF should match native ERA5 above 1000 meters altitude due to our implemented spectral nudging approach in WRF. As can be seen, the w/BC experiments exhibit near-zero biases (Figure 2) relative to ERA5-WRF, while the no-BC experiments can be quite biased, especially for temperature fields. Second, we compared biases in dynamically downscaled temperature, precipitation, and snow water equivalent (SWE) across the western United States (U.S.) relative to PRISM (Daly et al., 1994) and SNOw TELemetry (SNOTEL; Serreze et al., 1999) on our intermediate 9-km x 9-km grid. Biases in these variables were generally much smaller in the w/BC experiments relative to the no-BC experiments when considering an identical 9-GCM ensemble mean (Fig. 3; including GCMs not delivered in EPC-20-006). Specifically, biases in temperature (precipitation) averaged across 11 western U.S. states were -1.48°C and -0.08°C (0.38 mm d^{-1} and 0.07 mm d^{-1}) in the no-BC and w/BC experiments, respectively, relative to PRISM. At higher elevations, biases were compared to SNOTEL given the strong correlation with biases and elevation. Spatially averaged at more than 700 SNOTEL sites, near-zero and +20% (-12% and +25%) biases in precipitation (snow) were simulated in w/BC and no-BC experiments, respectively. Further, the simulated spread in bias is much larger in the no-BC relative to the w/BC experiments. Bias comparisons are summarized in Rahimi et al. (2024b) and reveal that biases in extreme precipitation are smaller in the w/BC relative to the no-BC experiments.

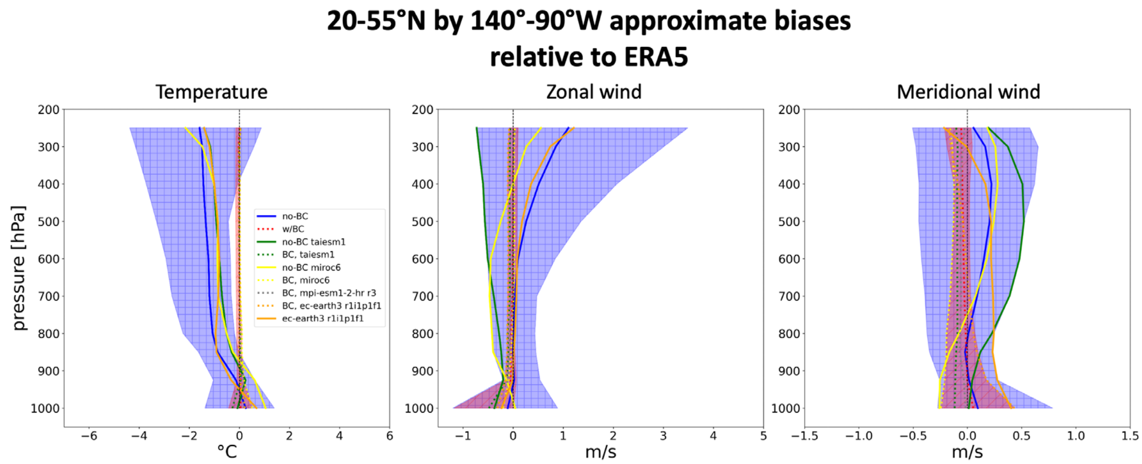

Figure 2. Regional-mean approximate downscaled GCM bias profiles from w/BC (dashed) and no-BC (solid) experiments across the Northeast Pacific and the western U.S. for (left) temperature [°C], (center) zonal wind [m s^{-1}], and (right) meridional wind [m s^{-1}]. Here, these are approximate biases because the reference data product used here is ERA5-WRF (see LOCA Version 2 – Training Data Adoption Justification Memo). Since we are using spectral nudging in the 45-km WRF experiments, these fields are approximately conserved in downscaling across large spatial scales. Here, no-BC and w/BC are constructed using an identical 11-GCM mean with their spreads given by blue and red shading, respectively. We also include individual curves for new GCMs reported to CEC (all bias corrected; dashed curves) and where possible, the no-BC pair (not reported to CEC).

2.3. Effects of BC on future climate trends

From the beginning of DD, BC of the GCM boundary conditions was avoided due to its potentially distorting effects on climate trends and subsequently increased uncertainty in future projections. Subsequent analyses of no-BC and w/BC experiments from the Western U.S. Dynamically Downscaled Dataset (WUS-D3) have revealed that, across the western U.S., *a priori* BC introduces relatively little uncertainty into climate projections when compared to the irreducible uncertainties attributed to internal climate variability and GCM choice (Risser et al., 2024). Locally, however, trends in surface air temperature can be modified by as much as 1.5°C per century (Figure 4). Across the Sierra Nevada in California, for instance, w/BC experiments simulate stronger warming than their no-BC counterparts in excess of 1°C per century.

This difference in trends is related to wet precipitation (Fig. 3c,e) and snow biases (Fig. 3f) in the no-BC experiments; this wet behavior of the no-BC relative to the w/BC experiments continues from the historical into the future era (through 2100; not shown). With more snow on the ground from year to year in no-BC experiments, bare, darker ground will be exposed more gradually than in the w/BC experiments as regional transient warming facilitates snow melt via latent heating at the expense of sensible heating for a given energy input to the land surface. BC of the GCM boundary conditions dries the WRF solution comparatively, reducing seasonal precipitation and snow while warming surface air temperatures as more energy is diverted into sensible heating of the overlying atmosphere, leading to an environment more prone to snow loss and warmer surface air temperature trends.

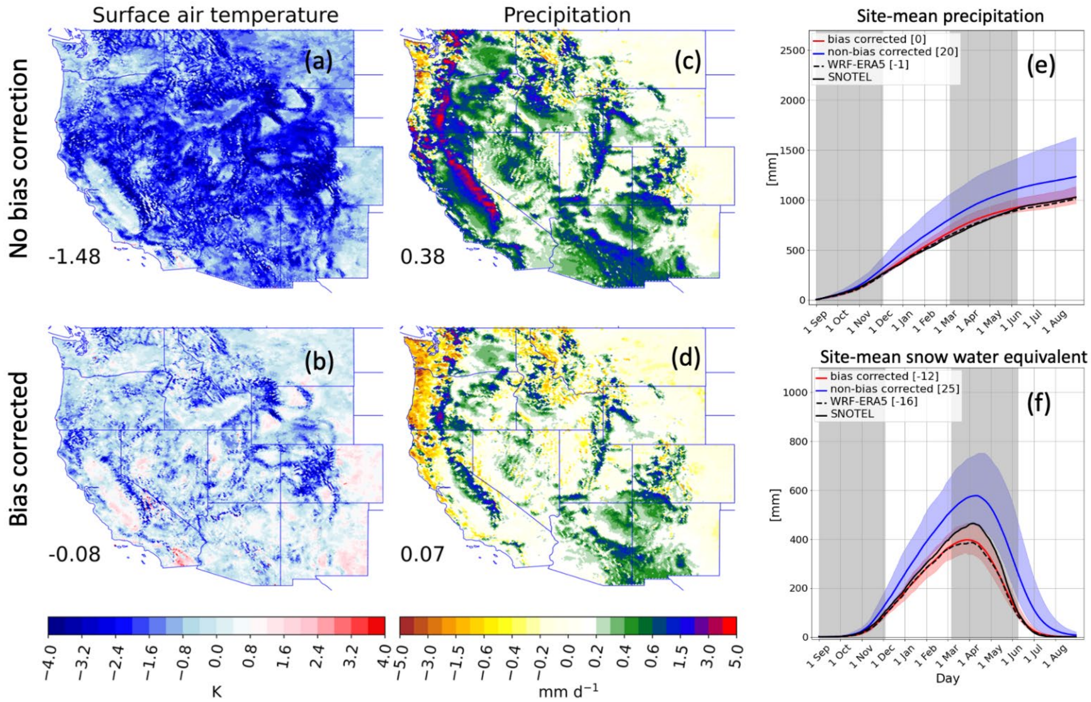

Figure 3. Relative to PRISM, ensemble-mean climatological (1981-2010) bias patterns in annual-mean (a,b) temperature and (c,d) precipitation, with 11-state-mean biases indicated in the lower-left of each subpanel. (e, f) time series of cumulative annual precipitation and snow water equivalent (SWE), respectively [mm], across 703 western U.S. SNOTEL sites. Shading denotes the range in w/BC (red) and no-BC (blue) experiments. Observations (black solid curve) and WRF-ERA5 (black dashed curve) are also presented.

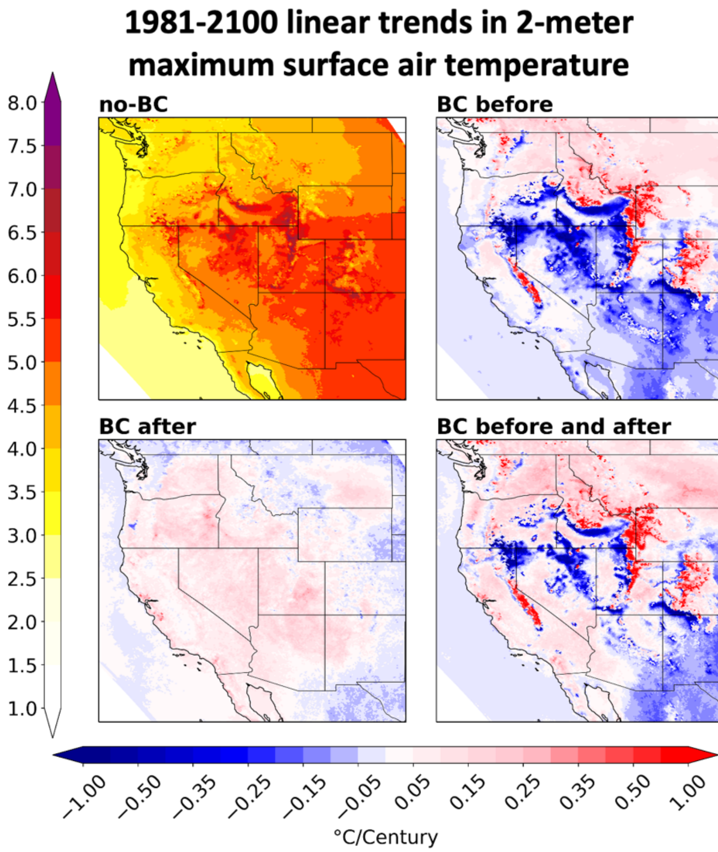

Figure 4. The upper-left panel shows downscaled 9-GCM-mean linear trends in springtime 2-meter maximum daily air temperature [°C/Century]. The upper-right panel depicts alterations in linear trends if BC is applied prior to downscaling for identical GCMs (subtracting the no-BC from the w/BC experiments). The bottom-left figure is the same as in the upper-right, but for linear trend alterations when BC is applied after downscaling only using PresRat. The bottom-right figure is the same as in the upper-right, but for linear trend alterations when BC is applied both before and after downscaling.

2.4. Why does BC dry the WRF solution?

As discussed previously, in the experiments, transmitted GCM biases are leading to no-BC solutions that are relatively wetter and generally unrealistic compared to those in w/BC. For a more holistic review of transmitted GCM biases and their effects in DD, refer to Rahimi et al. (2024b). However, we found three mean GCM biases that generally explained the unrealistically wet behavior of the no-BC experiments:

1. An overly cold troposphere (Fig. 2a): This increases the likelihood of heterogenous (ice+liquid) precipitation processes in clouds, increasing precipitation efficiency. Further, for parcels lifted adiabatically, saturation and subsequent precipitation formation processes will occur more readily given lower saturation vapor pressures.
2. An overly unstable lower troposphere (Fig. 2a): Too much thermodynamic instability in the lower levels of the atmosphere tends to increase convective available potential energy (CAPE). CAPE is positively correlated with vertical velocity, which itself is positively correlated with a faster cloud droplet nucleation (and subsequent precipitation) rate.
3. An overly cyclonic atmosphere just upstream of western North America (Figure 5): This feature is coupled to the cold bias (2) via the thermal wind in GCMs, which is maximized in intensity just upstream of the western U.S. The presence of this feature can be diagnostically linked to excessive upward vertical velocities using quasigeostrophic

theory that, when transmitted to WRF, manifests as too much precipitation. This bias cannot be removed if a spatial-mean BC were implemented.

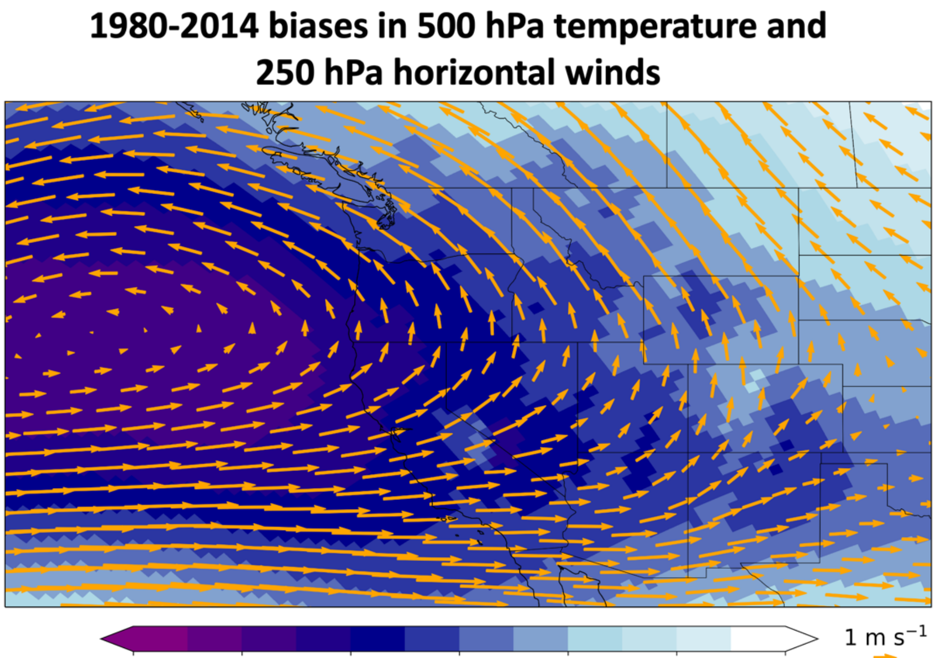

Figure 5. 1980-2014 biases in 9-GCM-mean 500 hPa temperature [°C] and 250 hPa horizontal winds [m s^{-1}].

2.5. Overview of BC used in EPC-20-006

Since different types of BC are used for different datasets in EPC-20-006, we provide a general summary of these data by BC type in Table 2.

All simulations have the raw 6-hourly, hourly, and daily datastreams saved within an open data bucket on Amazon S3. Principally, there are two sets of bias corrected WRF outputs for use in this project.

1. **Bias correction *a priori***: The full WRF datastreams for these experiments can be regarded as bias corrected, since the GCM boundary conditions were bias corrected to ERA5 before downscaling. This bias correction was only performed to TaiESM, MIROC, MPI, and ECEnoVeg. These data are best for use in process studies, downscaling to higher resolution, or for when variables not provided in LOCA2 hydrid downscaling are needed in analysis.
2. **Bias correction *a posteriori:*** This bias correction procedure, conducted using PresRat, is only applied to the no-BC WRF experiments (CESM2, CNRM, ECE, and FGOALS) for certain variables and temporal frequencies. Specifically, hourly simulated precipitation and temperature outputs were bias corrected station data at multiple sites across California. These outputs may be better suited for impact studies than the outputs in (1) given their minimal historical biases.

2.6. Why the w/BC GCMs were different from the no-BC GCMs

Once we decided to employ *a priori* BC in DD, we were then faced with the choice of whether to select new GCMs from the GCMs identified for downscaling over California (Krantz, et al, 2021). In the end, we chose a new set of GCMs for the w/BC experiments (Table 1) which, like the no-BC GCMs, were minimally biased. Further, the new GCMs, in the mean, were characterized by amounts of regional warming and precipitation change like the no-BC GCMs (Section 4). There are three fundamental reasons why we opted to downscale different GCMs in w/BC experiments:

1. While downscaling the original 4 GCMs, we found practical issues with some of the native GCM data despite a rigorous GCM identification and selection process. This is not uncommon to DD, but nonetheless gave us pause about re-downscaling the same GCMs. For example, FGOALS-g3 contains a western seaboard coastline that is offset ~2° of longitude relative to where it should be. Due to this, FGOALS-g3 was out of contention for w/BC experiments and we began to examine other minimally biased contenders for DD.
2. We wanted a broader sense of structural uncertainty. By downscaling different GCMs in w/BC experiments, our complete set of projections (considering both no-BC and w/BC experiments) trace out a larger swath of plausible regional climate responses.
3. By downscaling 4 new GCMs, each of which with their own unique phasing of internal climate variability, it was considered that these new simulations could also provide outputs of use in targeted case study downscaling (another task yet to be performed as part of EPC-20-006). For example, MPI-ESM1-2-HR has an intense drought in the early 2090s (not shown).

3. Historical performance of new 3-km x 3-km bias corrected WRF GCM projections

Next, we evaluate historical biases in precipitation and temperature in the w/BC experiments across California with respect to the original four GCMs (Table 1). We begin with a quantile-based state-mean examination of WRF-simulated versus PRISM-based precipitation, as well as mean, maximum, and minimum temperature in Figure 6. Beginning with precipitation (Fig. 6, upper-left), we can see that the no-BC (blue) experiments are generally much wetter than the w/BC experiments (red), which is consistent with the broader western U.S. precipitation bias profiles (Fig. 3). Despite the fact that biases grow with increasing quantile, w/BC experiments reduce the bias from +1.1 to 0.1 mm, +4.9 to +1.6 mm, and +4.7 to +2.6 mm for the 50th, 99th, and 99.9th percentiles, respectively. The substantial reduction in bias is evident by comparing ECE and ECEnew. Finally, although precipitation bias is evident for the most extreme quantiles examined here, these values are modulated by particularly high precipitation amounts simulated across data-sparse regions of complex terrain (e.g., the Sierra Nevada). Given large observational uncertainties across these regions, it may not be appropriate to classify WRF and PRISM differences as biases for higher quantiles (Lundquist et al., 2019). We also note a much-reduced simulated spread in the w/BC experiments comparatively which perhaps is an intuitive result given that GCM climate means of bias corrected fields were forced to match the mean-state climate of native ERA5; the w/BC experiments expectedly straddle the ERA5-WRF curve.

Next, we focus on surface air temperature. For mean surface air temperature (upper-right, Fig. 6), the cold bias is effectively removed in the w/BC experiments, with the original spread of the no-BC experiments nearly vanishing. We can also see that a large portion of the bias in no-BC-simulated mean air surface temperature is the result of highly biased maximum surface air temperature (lower-left, Fig. 6), whose bias and spread are lowered in the w/BC experiments. In terms of minimum surface air temperature, no-BC experiments are slightly less warm-biased compared to their w/BC counterparts (lower-right, Fig. 6), albeit with a larger model spread.

We believe the nature of the bias differences between w/BC and no-BC experiments in maximum and minimum temperature stems in part from low cloud coverage differences between the two sets of simulations. Specifically, no-BC experiments precipitate too much relative to the w/BC experiments and observations. If we use precipitation as a proxy for cloudiness, then no-BC experiments will tend to be cloudier, which would tend to cool maximum daily temperatures and warm nocturnal temperatures. During the overnight hours, cloudiness would tend to prevent thermal infrared emissions from escaping to space, thus limiting the degree to which minimum temperatures could be lowered. These simple arguments tend to explain the relatively warmer maximum temperatures of w/BC experiments in Figure 6, but not the warmer, more biased behavior of w/BC experiments in simulating minimum temperatures in which no-BC is nearly always cooler than w/BC experiments. Thus, we speculate that an overly snowy solution in no-BC experiments could be partially responsible for its colder representation of minimum daily temperatures relative to w/BC experiments. This could also partially explain why no-BC-simulated maximum surface air temperatures are colder than those in w/BC.

Finally, despite the largest biases presenting for the largest quantiles for precipitation, the power of bias correcting the dynamic fields for WRF *a priori* to reduce PRISM-relative biases in 50th percentile (median) downscaled precipitation and temperature is evident in Figure 7. Specifically, every w/BC experiment outperforms the best performing no-BC experiment (CESM2) in simulating median precipitation, while two w/BC experiments (MIROC and TaiESM) simulate median temperature biases comparable to the best-performing no-BC experiments (ECE and CESM2). Further, biases in w/BC simulations are closer than no-BC experiments to those in ERA5-WRF.

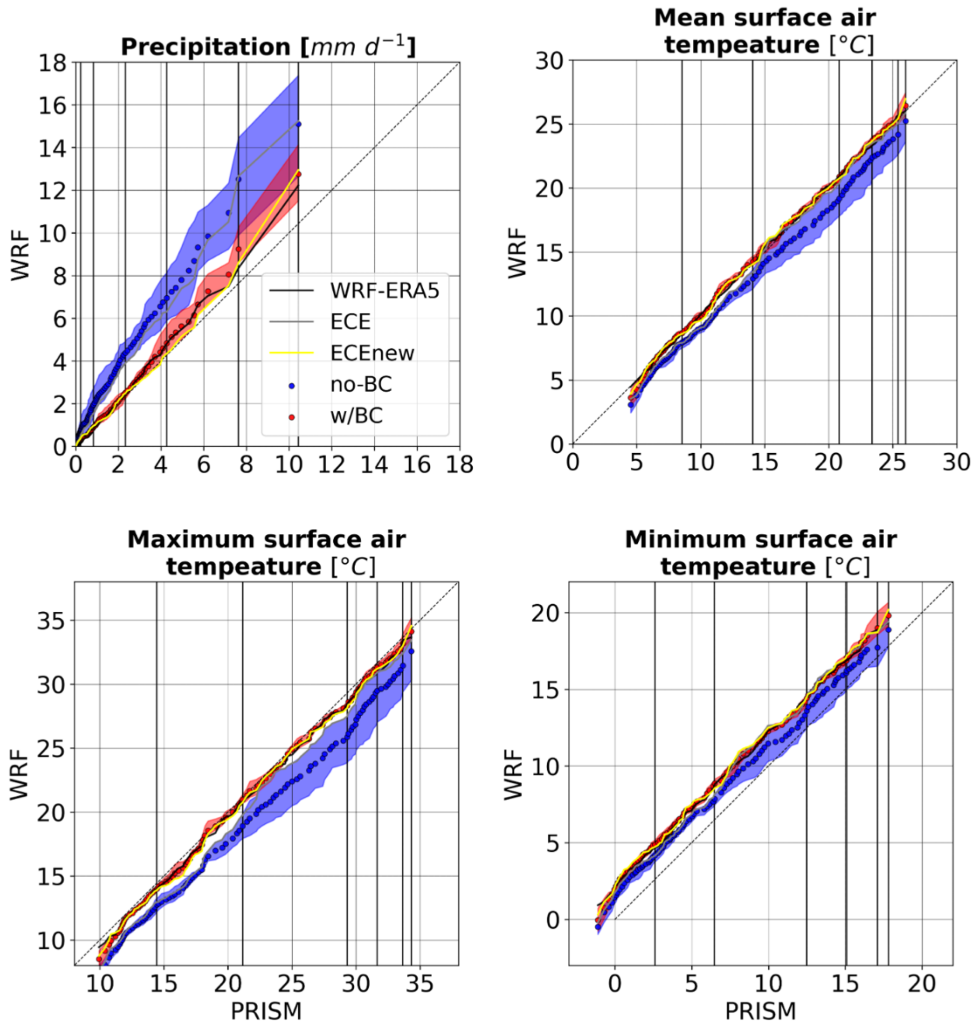

Figure 6. California-mean historical-era WRF-simulated variables scattered against reference data from PRISM by quantile for (upper-left) precipitation [mm d^{-1}], mean surface air temperature [°C], maximum surface air temperature [°C], and minimum surface air temperature [°C]. GCM means of the original 4 (no-BC) simulations comprise the blue scatter, while GCM means from the new 5 (w/BC) simulations are depicted by the red scatter. Red (blue) shading denotes the w/BC (no-BC) experimental spread, and the WRF-ERA5 is depicted by the black solid line. EC-Earth3-Veg is the only experiment that was duplicated in the w/BC and no-BC runs and is depicted by the light yellow and gray curves, respectively. Vertical bolded lines indicate the 25th, 50th (median), 75th, 90th, 99th, and 99.9th percentile values from PRISM.

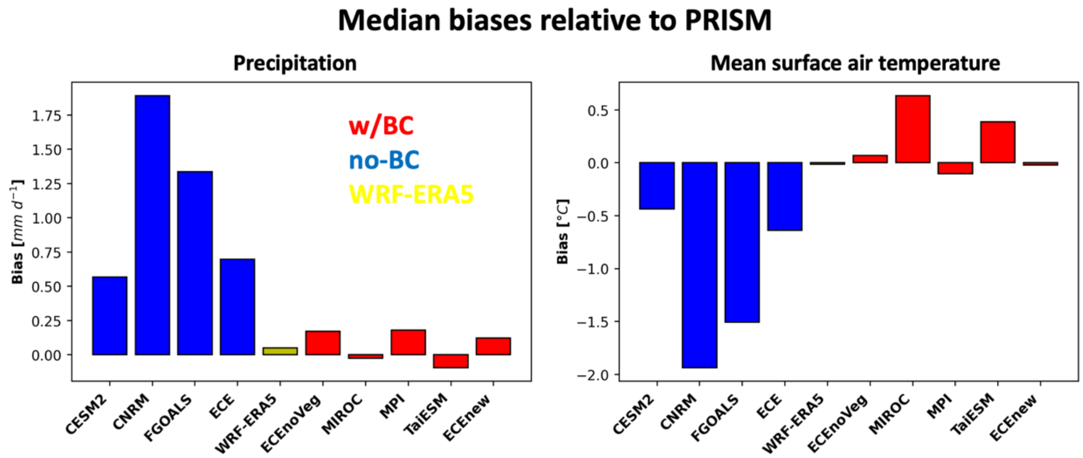

Figure 7. California-mean historical biases in median (left) precipitation [mm d^{-1}] and (right) daily mean surface air temperature [°C] from individual GCMs within no-BC (blue) and w/BC (red) simulations. WRF-ERA5 is provided for reference.

4. Comparison of w/BC and no-BC Future Projections

We restrict our examination of the w/BC GCMs' climate response to precipitation and temperature and compare these new w/BC projections to the original no-BC projections in Figure 8. As in our no-BC projections, the large-scale climate change pattern is preserved in downscaling (see Krantz et al, 2021).

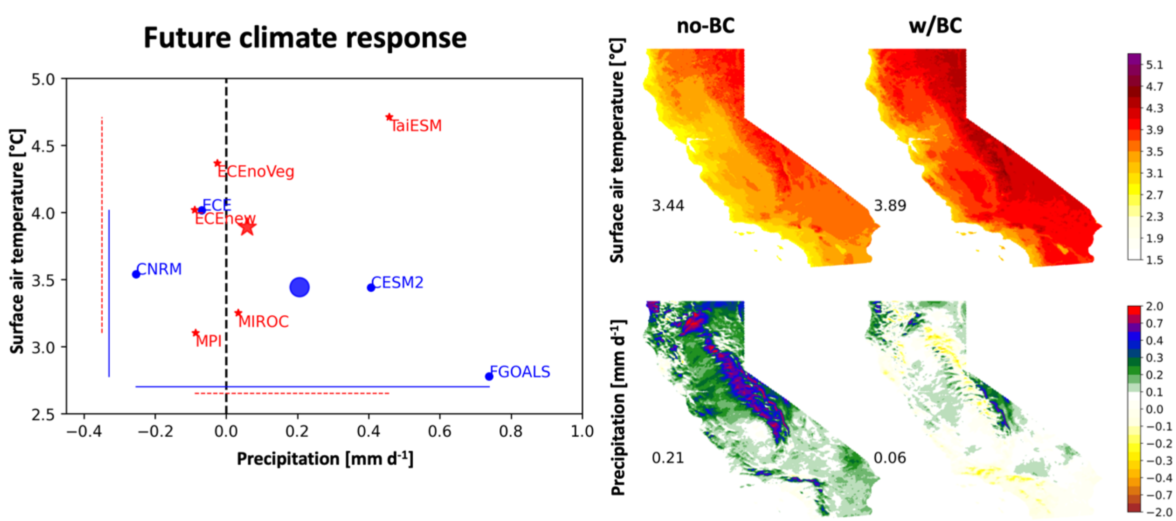

Figure 8. Future-era (2070-2099 mean) annual and California-mean changes in surface mean air temperature [°C] and precipitation [mm d^{-1}] relative to the 1981-2010 period. The left panel shows a scatter of future-era warming against future-era precipitation changes. Individual GCMs are shown as red stars (blue circles) for w/BC (no-BC) experiments with larger symbols indicating the ensemble mean changes. Red dashed (blue solid) horizontal lines showcase the precipitation spread. Vertical lines also denote spread but for temperature. Geographic figures on the right show the spatiality of (top) temperature and (bottom) precipitation changes for (left) no-BC and (right) w/BC experiments in their respective ensemble means. Numbers in the lower-left portion of each subpanel denote California-mean changes in either temperature [°C] or precipitation [mm d^{-1}].

Relative to no-BC experiments, the climate response of w/BC experiments is warmer and drier in the ensemble mean. Further, w/BC experiments have a smaller future precipitation spread and a broader warming spread than their no-BC counterparts (as denoted by the red dashed lines in Fig. 8, left panel). Geographically, w/BC experiments do not simulate the future-era wetting along the spine of the Sierra Nevada as in the no-BC experiments, although annual-mean wetting is predicted in both ensembles in the lee of the Sierra Nevada.

5. References

Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dynamics*, *43*(7), 1847–1856. https://doi.org/10.1007/s00382-013-2011-6

Bukovsky, M. S., & Karoly, D. J. (2011). A Regional Modeling Study of Climate Change Impacts on Warm-Season Precipitation in the Central United States. *Journal of Climate*, *24*(7), 1985–2002. https://doi.org/10.1175/2010JCLI3447.1

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal of Applied Meteorology and Climatology*, *33*(2), 140–158. https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holland, G., Done, J., Bruyere, C., Cooper, C. K., & Suzuki, A. (2010). Model Investigations of the Effects of Climate Variability and Change on Future Gulf of Mexico Tropical Cyclone Activity. Presented at the Offshore Technology Conference, OnePetro. https://doi.org/10.4043/20690-MS

Krantz, W., Pierce, D., Goldenson, N., Cayan, D.: Memorandum on Evaluating Global Climate Models for Studying Regional Climate Change in California, Memo to California Energy Commission as part of EPC-20-006. https://www.energy.ca.gov/sites/default/files/2022-09/20220907_CDAWG_MemoEvaluating_GCMs_EPC-20-006_Nov2021-ADA.pdf, 2021.

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., et al. (2013). A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *Journal of Climate*, *26*(23), 9384–9392. https://doi.org/10.1175/JCLI-D-12-00508.1

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, *100*(12), 2473–2490. https://doi.org/10.1175/BAMS-D-19-0001.1

Rahimi, S., Krantz, W., Lin, Y.-H., Bass, B., Goldenson, N., Hall, A., et al. (2022). Evaluation of a Reanalysis-Driven Configuration of WRF4 Over the Western United States From 1980 to 2020. *Journal of Geophysical Research: Atmospheres*, *127*(4), e2021JD035699. https://doi.org/10.1029/2021JD035699

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Krantz, W., Bass, B.,, et al. (2024a). An Overview of the Western United States Dynamically Downscaled Dataset (WUS-D3). *Geoscientific Model Development* 17, no. 6: 2265–86. https://doi.org/10.5194/gmd-17-2265-2024.

Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Risser, M., et al. (2024b). Understanding the Cascade: Removing GCM Biases Improves Dynamically Downscaled Climate Projections. *Geophysical Research Letters*, *51*(9), e2023GL106264. https://doi.org/10.1029/2023GL106264

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., et al. (2011). High-Resolution Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *Journal of Climate*, *24*(12), 3015–3048. https://doi.org/10.1175/2010JCLI3985.1

Rasmussen, R. M., Chen, F., Liu, C. H., Ikeda, K., Prein, A., Kim, J., et al. (2023). CONUS404: The NCAR-USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bulletin of the American Meteorological Society*, *1*(aop). https://doi.org/10.1175/BAMS-D-21-0326.1

Risser, M. D., Rahimi, S., Goldenson, N., Hall, A., Lebo, Z. J., & Feldman, D. R. (2024). Is Bias Correction in Dynamical Downscaling Defensible? *Geophysical Research Letters*, *51*(10), e2023GL105979. https://doi.org/10.1029/2023GL105979

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, *35*(7), 2145–2160. https://doi.org/10.1029/1999WR900090

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., et al. (2019). A Description of the Advanced Research WRF Model Version 4. https://doi.org/10.5065/1dfh-6p97

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, *227*(7), 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037

Wang, J., & Kotamarthi, V. R. (2015). High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America. *Earth's Future*, *3*(7), 268–288. https://doi.org/10.1002/2015EF000304