An Overview of the Western United States Dynamically Downscaled Dataset (WUS-D3)

[1]Stefan Rahimi, [1]Lei Huang, [1]Jesse Norris, [1]Alex Hall, [1]Naomi Goldenson, [1]Will Krantz, [1]Benjamin Bass, [1]Chad Thackeray, [1]Henry Lin, [1]Di Chen, [1]Eli Dennis, [2]Ethan Collins, [3]Zachary J. Lebo, [1]Emily Slinskey, [1]Sara Graves, [1]Surabhi Biyani, [1]Bowen Wang, [1]Stephen Cropper, and the UCLA Center for Climate Science Team[+]

[1]Center for Climate Science, University of California Los Angeles, Los Angeles, California, 90095, U.S.A.
[2]Department of Atmospheric Science, University of Wyoming, Laramie, Wyoming, 82071, U.S.A.
[3]School of Meteorology, University of Oklahoma, Norman, Oklahoma 73019, U.S.A.
[+]A full list of Authors appears at the end of this manuscript

Correspondence to: Stefan Rahimi (srahimi@uwyo.edu)

Abstract. Predicting future climate change over a region of complex terrain, such as the western United States (U.S.), remains challenging due to the low resolution of global climate models (GCMs). Yet the climate extremes of recent years in this region, such as floods, wildfires, and drought, are likely to intensify further as climate warms, underscoring the need for high-quality and high-resolution predictions. Here, we present an ensemble of dynamically downscaled simulations over the western U.S. from 1980–2100 at 9-km grid spacing, driven by sixteen latest-generation GCMs. This dataset is titled the Western U.S. Dynamically Downscaled Dataset (WUS-D3).

We describe the challenges of producing WUS-D3, including GCM selection and technical issues, and we evaluate the simulations' realism by comparing historical results to temperature and precipitation observations. The future downscaled climate change signals are shaped in physically credible ways by the regional model's more realistic coastlines and topography: (1) The mean warming signals are heavily influenced by more realistic snowpack. (2) Mean precipitation changes are often consistent with wetting on the windward side of mountain complexes, as warmer, moister air masses are uplifted orographically during precipitation events. (3) There are large fractional precipitation increases on the lee side of mountain complexes, leading to potentially significant changes in water resources and ecology in these arid landscapes. (4) Increases in precipitation extremes are generally larger than in the GCMs, driven by locally intensified atmospheric updrafts tied to sharper, more realistic gradients in topography. (5) Changes in temperature extremes are different from what is expected by a shift in mean temperature and are shaped by local atmospheric dynamics and land surface feedbacks. Because of its high resolution, comprehensiveness, and representation of relevant physical processes, this dataset presents a unique opportunity to evaluate societally relevant future changes in western U.S. climate.

1 Introduction

Predicting climate change on a regional level is critical for assessing its societal impacts, such as changes to water resources, flooding, drought, heat waves, wildfire, and windstorms. Current-generation global climate models (GCMs) are ill-equipped for this task due to their coarse grid spacing (on the order of 1 degree longitude/latitude). This prevents GCMs from representing complex terrain and from resolving small-scale meteorological phenomena that define the local hydroclimate. To counter this limitation, a regional climate model (RCM) may be used to dynamically downscale the GCM projections over a limited area. The resulting high-resolution output allows us to study future weather- and climate-relevant processes that may unfold across a region of complex terrain and gain physical insights into the land–atmosphere drivers of regional climate change. Moreover, the output can be used to drive land-surface, hydrological, and fire models under future climate conditions.

The western United States (WUS) is a particularly complex natural laboratory for studying the heterogeneous patterns of historical climate and future climate change. It consists of major mountain ranges, deserts, shrublands, temperate forests, plains, and a complex coastline. It is affected by diverse atmospheric phenomena, such as extratropical cyclones, atmospheric rivers, persistent blocking highs, the North American Monsoon, summertime convective storms, wildfire-related downslope winds, and cooling coastal breezes. The complex interplay of these phenomena with local topography makes it impossible for GCMs to represent the diversity of microclimates within the WUS and how they may uniquely respond to larger-scale climate change. In general, GCMs project mid-latitude wetting to the north of the region and subtropical drying to the south, but with disagreement on where within the WUS the transition occurs (Meehl et al., 2007; Neelin et al., 2013). Moreover, intensified interannual swings

between extremely wet and extremely dry years (i.e., 'whiplash') are projected in parts of the region

65 (Swain et al., 2018; Chen et al., 2022). In recent years the WUS has experienced catastrophic

weather/climate events, such as the southwestern U.S. drought (Mankin et al., 2019.; White et al.,

2023), record-breaking floods in California in 2017 (White et al., 2019) and 2023, and the

unprecedented 2021 heatwave in the Pacific Northwest (White et al., 2023). In a warming climate, all of

these extreme events are likely to be intensified. Thus, dynamical downscaling of future GCM

70 projections over the WUS can provide a unique insight into how large-scale climate change may

interact with its complex terrain and diverse meteorological phenomena.

Direct dynamical downscaling of GCMs is far less common than that driven by historical

reanalyses (Liu et al., 2017, 2011; Rahimi et al., 2022; Rasmussen et al., 2011, 2014; Norris et al., 2019,

and many, many others) due to the fact that historical reanalyses tend to more reliably contain the

75 requisite data to drive RCMs (Bruyère et al., 2014; Coppola et al., 2020, 2021; Huang et al., 2020,

2021; Komurcu et al., 2018; Wang and Kotamarthi, 2015, 2013; Zobel et al., 2018, 2017; Bukovsky and

Karoly, 2011; Bukovsky et al., 2021; Mearns et al., 2012; Scalzitti et al., 2016). Further, since

dynamical downscaling uses the laws of physics to arrive at the high-resolution end-product, it can be

superior to other purely statistical-based downscaling methods. For example, dynamical downscaling

80 does not explicitly assume stationarity (Lanzante et al., 2018) in the creation of future projections, as

with other forms of downscaling (e.g., statistical); the parameterization choices within RCMs do contain

empirically-derived assumptions that are not completely free of time stationarity. Dynamical

downscaling can however be used to tie explicitly simulated extreme weather events to the governing

large-scale dynamics simulated within their driving GCMs. Additionally, RCMs can solve for the full

---

complement of physical quantities relevant to climate that are otherwise not available in statistical downscaling, which typically focus on a small set of variables. For example, statistically downscaled precipitation and temperature data products, even when obtained using multivariate relationships, may contain no information about water vapor content, surface pressure, cloud depth, etc. Finally, the use of

105   physics to arrive at the downscaled result means that feedbacks between the landscape and the overlying atmosphere, and other land and atmosphere processes, may be effectively simulated (e.g., the snow-albedo feedback).

There are three significant barriers to using RCMs to dynamically downscale GCMs: (1) RCMs require sub-daily three-dimensional variables as initial and boundary conditions, which are not typically

110   sufficiently archived in GCM databases; (2) RCM configurations may not be designed to ingest GCM data as boundary conditions; and (3) It is extremely computationally expensive. Because of these barriers, dynamical downscaling of full GCM ensembles (e.g., the Coupled Model Intercomparison Project Phase 6; CMIP6) at landscape-resolving (~10 km) grid spacings generally remain out of reach.

Despite these barriers, we present results from sixteen new dynamically downscaled CMIP6

115   simulations over 11 WUS states, including the whole of the Western Electricity Coordinating Council (WECC) region, comprising the Western U.S. Dynamically Downscaled Dataset (WUS-D3). These simulations span 1980-2100, combining the historical and Shared Socioeconomic Pathway (SSP) output for each GCM. Downscaling a wide variety of CMIP6 models yields a diverse suite of possible future climates over the WUS at a landscape-resolving scale (9-km grid spacing). In the following sections, we

120   present our methodology and technical challenges encountered, as well as a characterization of the historical performance and future change signals from our dataset.

130

2 Methodology

2.1 WRF Setup

We use the Weather Research and Forecasting (WRF) model version 4.1.3 (Skamarock et al., 2019) to dynamically downscale the simulations of 14 CMIP6 GCMs (Table 1) from 1980–2100. In

135 each simulation, historical forcings were applied up to 2014, and then the forcings associated with the SSP3-7.0 scenario thereafter. SSP3-7.0 is a high-emissions scenario in which greenhouse-gas emissions double by end-of-century (O'Neill et al., 2016). We also downscale one GCM's (CESM2) SSP2-4.5 and SSP5-8.5 projections. In these scenarios, emissions remain roughly constant until 2050 before falling thereafter, and triple by end-of-century, respectively.

140 WRF is configured as was documented for a related downscaling of the ECMWF Fifth Generation Reanalysis (ERA5; Hersbach et al., 2020) in Rahimi et al. (2022; WRF-ERA5). We downscale each GCM year separately and in parallel; at the beginning of each downscaling period (on August 1), the RCM is initialized to the driving GCM state. In this way, an N-year simulation can be completed in the same wall clock time as a 1-year experiment. For each year of integration, we choose

145 the beginning of the retained WRF output for analysis to coincide with minimum snowpack across the WUS (September 1). This approach produces one month of spin-up for the land surface. Thus, WRF is initialized on August 1 to surface and three-dimensional data from each GCM and integrated through September 1 of the following year (13 months, including the spin-up month) on 39 atmospheric levels. This approach is similar to that of Zobel et al. (2018, 2017), who also initialized WRF experiments at

yearly intervals, but only included 1 day of model spin-up. Despite our 1-month spin-up, soil moisture,

land surface fluxes, and streamflow may still suffer from biases due to imperfect soil texture categories

155 and their associated hydrophysical properties (Dennis and Berbery, 2021). However, because soil

texture is a necessary component of the land surface model, and these underlying datasets are imperfect,

these effects are somewhat unavoidable without massive regional calibration. WRF's parallelization

procedure, which is advantageous for executing simulations in weeks instead of years, is performed to

the detriment of time continuity in simulating the surface and subsurface runoff, as well as energy

160 fluxes, with high precision. The consequences of this choice will be expanded upon in Section 2.6.

Atmospheric carbon dioxide and methane concentrations vary yearly in our simulations based

on northern-hemispheric-mean values from input4MIPs (Durack et al., 2017). Prior to 2015, CMIP6

historical values are prescribed. From 2015 onward, these values are taken from the SSP3-7.0 scenario,

except for the alternate SSP CESM2 experiments. WRF's radiative code is modified to enable

165 concentrations to be manually inputted; this modification is no longer needed as of WRF version 4.4.2.

Because coupling WRF to an atmospheric chemistry model is 6-20 times more computationally

expensive, transient aerosol forcings were not explicitly considered in our study. Further, historical-era

21-category land-use/land-coverage (LULC) information from the Moderate Resolution Imaging

Spectrometer is used in all experiments. Since CMIP-projected LULC changes were not implemented in

170 WUS-D3, the anthropogenic forcings considered in this study stem directly from carbon dioxide and

methane concentrations, and indirectly from all greenhouse gas, aerosol, and LULC forcings in the

forcing GCMs at the lateral boundaries.

180   We dynamically downscale all GCMs to two grids of 45-km and 9-km grid spacing (Figure 1). On the parent 45-km grid, the horizontal winds, temperature, and geopotential height are relaxed (relaxation coefficient of 0.0003 s^{-1}) to their respective GCM-simulated fields above the planetary boundary layer via spectral nudging for wavelengths greater than 1,500 km. Smaller waveforms are allowed to evolve freely on the WRF grid (Spero et al., 2014). This approach is designed to reduce

185   internal model drift away from the GCM state. One-way nesting is then used to dynamically downscale the 45-km result to the 9-km grid, on which spectral nudging is not implemented. The 9-km grid encompasses the entirety of the WECC's U.S. coverage area. For all nests, a sponge layer of 5 grid points is used.

The lateral boundary conditions are updated at 6-hourly intervals, and adaptive time stepping is

190   used. Convective precipitation is parameterized following Tiedtke (1989) and Zhang et al. (2011). P3 microphysics is used (Morrison and Milbrandt, 2015), shortwave and longwave radiation schemes of Iacono et al. (2008) are implemented, and the Noah land surface model with multi-parameterizations (Noah-MP) is used (Niu et al., 2011).

195   **2.2 GCM Selection**

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

200 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
205 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
210 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
215 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.

250

We only dynamically downscale GCMs with the following outputs archived on the Earth System Grid Federation system: 3-D atmospheric temperature (ta), horizontal winds (ua and va), and specific humidity (hus), surface pressure (ps), soil layer-specific temperature and water content (tsl and

255 mrsol, respectively), and sea surface temperature (SST; tos). Furthermore, we only dynamically downscale GCMs with 6-hourly instantaneous atmospheric outputs defined on native model levels ("6hrLev") rather than on isobaric surfaces ("6hrPlevPt"). Generally, 6hrPlevPt GCM outputs are only defined on 3-10 pressure surfaces which may be problematic for atmospheric phenomena characterized by more granular vertical structures. In testing, we found that this vertical resolution can have a large

260 impact on the downscaled solution in cases where 6 versus 23 vertical levels were used.

Additionally, we require that the full time series of SSTs be available in GCM outputs. These SSTs are then prescribed in WRF to update daily, which may be problematic for atmospheric processes subject to a strong atmospheric-ocean coupling evolving on sub-daily time scales. To bypass this issue, we tested using a slab ocean model in WRF. With time, strange artifacts in the SST and outgoing

265 longwave radiation fields gradually developed, so slab ocean physics were not enabled and its use is discouraged for simulations on regional climate time scales (Ming Chen, personal communication; https://forum.mmm.ucar.edu/threads/weird-pixilated-skin-temperatures-when-using-sf_ocean_physics-1.12693/). Daily SSTs are available for most GCMs, except for FGOALS-g3 and GISS-E2-1-G, which

only made monthly SST outputs available. Thus, in the cases of FGOALS-g3 and GISS-E2-1-G, linear

interpolation is used to upsample monthly mean SSTs (assumed to be valid at the midpoint of each

275   month) to daily values.

2.3. Sea surface temperatures in the Gulf of California

SSTs in the Gulf of California (GoC) are known to modulate the North American Monsoon,

which provides roughly a third of Arizona and New Mexico's annual precipitation (Mitchell et al.,

280   2002). However, the GoC is poorly resolved in CMIP6 GCMs; in the best case, the GoC is expressed as

a subtle bay that barely intrudes into the North American continent. As a result, there is generally no

SST information from GCMs across the GoC that can be used to directly prescribe SSTs in the WRF-

resolved GoC. An additional problem is that the adjacent open Pacific SSTs are on average about 10 K

lower and undergo less seasonal variability than in the GoC (Figure 2). Hence, linearly extrapolating

285   from the adjacent open Pacific to the GoC would produce a representation of GoC SSTs that is clearly

unphysical. Fortunately, there are predictable relationships in ERA5 between the climatological GoC

entrance region temperature (Fig. 2c), which can be taken directly from GCMs, and the along-axis GoC

SST gradient (Fig. 2d), which can be used to produce reasonable SSTs within the GoC. Thus, in most

GCMs, we apply the following linear extrapolation to estimate GoC SSTs based on the entrance-region

290   SSTs:

$$T_{GoC} = -\frac{\partial T}{\partial n}\bigg]_{ERA5} (n) + T_{entry,GCM} \qquad (1)$$

where $\frac{\partial T}{\partial n}\big]_{ERA5}$ is the monthly varying climatological GoC temperature gradient from ERA5 and is always positive, n is the along-GoC axis coordinate (pointing towards the northwest), and $T_{entry,GCM}$ is the GoC entrance temperature, which is resolved in GCMs. The relevant regions are outlined in Figure

295 2a. To our knowledge, the difficulty in dealing with SSTs in coastal estuaries and gulfs has not been generally addressed in regional climate modeling efforts, and this is the first time that a physically based mathematical relationship has been used to address this issue across this region.

We apply the above linear extrapolation to all GCMs, except CESM2, CNRM-ESM2-1, and MPI-ESM1-2-LR, which were all downscaled prior to implementation of this improvement.

300 Consequently, for CESM2 and MPI-ESM1-2-LR, there is a spurious SST discontinuity (Figure S1). This is due to the default extrapolation routine used in WRF, which uses a nearest weighted gridpoint averaging approach to prescribe GoC SSTs. Thus, in the southern GoC, the default extrapolation uses the nearest GCM grid points from the warm GoC entrance region, whereas further north the closest GCM ocean grid cells are (inappropriately) taken from the Open Pacific. We list the GCMs with the

305 SST modification in the rightmost column of Table 1. The discontinuity and unrealistically low SSTs in the northern GoC in these simulations may affect the simulation of the North American Monsoon but are unlikely to affect other WUS phenomena documented in this paper. For CNRM-ESM2-1, we masked out the southern GoC to remove this discontinuity in extrapolation, leading to its SSTs being homogeneously populated by Open Pacific SSTs. Despite the absenc of a SST discontinuity (Fig. S1),

310 this approximation is less physical than the improvement described above (Fig. 2).

2.4 Interpolation strategy

Deleted: While smooth

WRF requires all atmospheric, land, and ocean GCM inputs to be defined on a single rectilinear

315 grid with atmospheric variables defined on isobaric surfaces. However, some GCMs' outputs are given

on irregular atmospheric grids, whose latitude coordinates are not equally spaced from pole to pole.

FGOALS-g3 for instance is characterized by ~5° latitudinal grid spacing near the poles and ~2° grid

spacing near the equator. Thus, for the GCMs without a native rectilinear grid, we interpolate the output

to rectilinear grids with grid lengths defined by their respective absolute minimum latitude or longitude

320 grid spacing. This technique preserves the smallest-scale features resolved on the native GCM grid.

Since GCMs use different land surface models (LSMs) containing differently defined vertical

coordinates, we generally interpolate volumetric soil moisture and soil temperature from the native

LSM levels to 3.5, 14, 64, and 195 cm. In instances where vertical interpolation was not used, we used

the GCM's native grid soil information. Volumetric soil moisture was computed using the CMIP6

325 variable, mrsol, the layer total water content, and dividing it by the layer thickness and the density of

water. GCM soil fields were generally available daily.

2.5 Other Technical Challenges

In this section, we present additional technical challenges and known issues in the downscaled

330 data. First, WRF is not designed to ingest GCM inputs that are, depending on the modeling center,

defined on different vertical coordinates. For instance, CESM2 uses a hybrid-pressure, FGOALS-g3 a

sigma, and UKESM1-0-LL a hybrid-height vertical coordinate system. As a result, unique routines had

to be developed for each GCM to convert their model level output to WRF-usable inputs on isobaric

pressure surfaces. Further, MPI-ESM1-2 simulations had to be converted from NetCDF4 to NetCDF4-

Classic in order for input-output processing times to be tractable in binary processing. These issues alone prevented the development of a one-size-fits all routine to preprocess GCM outputs for ingestion by WRF. These issues were compounded by the fact that some GCMs, such as UKESM1-0-LL and ACCESS-CM2, contain staggered outputs on their native Arakawa C-grids.

340   Second, 6hrLev GCM atmospheric fields are generally provided at 0000, 0600, 1200, and 1800 UTC. However, for the entire FGOALS-g3 and historical (1980-2014) component of the NorESM2-MM experiments, data were provided at 0300, 0900, 1500, and 2100 UTC. For FGOALS-g3, we simply integrated all experiments from 1 August 1980 0300 UTC through 1 September 2100 0300 UTC. For NorESM2-MM however, we linearly interpolated the historical GCM data to 0000, 0600, 1200, and

345   1800 UTC before downscaling. As a further aside, since UKESM1-0-LL uses a 360-day calendar, we had to modify WRF's source code accordingly. WRF is designed by default to function with Proleptic Gregorian calendars (e.g., ERA5, MPI-ESM1-2-LR, EC-Earth3-Veg), but we compiled the model with no-leap calendars for other GCM experiments (e.g., CESM2, GISS-E2-1-G, TaiESM1).

350   **2.6. Spin-up Strategy Consequences**

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

355   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

360 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

365 inadequate across regions with a strong land-atmosphere coupling.

3. Simulation of the Historical Climate

Next, we present a review of simulated historical (1981-2010) precipitation and surface air

temperature across the WUS. Figure 3 shows the added value introduced by dynamical downscaling in

370 simulating these patterns, as well as the relative fidelity of the GCMs when downscaled with WRF. We

compare the downscaled ensemble mean against the native-resolution GCM ensemble mean, in addition

to 9-km WRF-ERA5 and observational estimates from the 4-km Parameter-elevation Regressions on

Independent Slopes Model (PRISM; Daly et al., 1994). The inability of the raw GCMs to capture the

complex terrain of the WUS is illustrated by major warm and cold biases over mountains and valleys,

375 respectively. In particular, California's Central Valley is 5-7 K too cool, while the Sierra Nevada is

warm-biased by the same magnitude. By contrast, the dynamically downscaled simulations, whether

GCM- or ERA5-driven, better resemble the regional temperature and precipitation patterns shown by

PRISM. Despite this improvement, the downscaled GCM experiments are generally colder than

PRISM, by as much as 5 K during part of the year in some states (Figure 4). The annual-mean spatial

380 patterns in Fig. 3 reveal the cold biases to be predominantly over mountains. The cold bias is generally

most prominent in the winter months (shown by spatial patterns in Figure S2) but persists year-round.

385   Additionally, dynamical downscaling generally reduces the simulated temperature spread from that of the parent GCMs, as indicated by the red circles (Fig. 4). Exceptions are noted across western states, especially in winter; we speculate that dynamical downscaling is increasing the spread proportional to the magnitude of GCM biases in temperature, winds, and SSTs which, when inherited by WRF, leads to varying magnitudes of downscaled precipitation and temperature bias. GCM bias impacts on the

390   dynamically downscaled solution are a current core focus by our research team.

       The dynamically downscaled ensemble mean is generally too wet across the states of Washington, Oregon, and California (Figure 3; Figure 4). A preexisting wet bias in the parent GCMs is increased by downscaling, an impact seen primarily over mountains during winter (Figure S3). These biases vary substantially within the ensemble, with individual downscaled GCMs exhibiting meaningful

395   state-wide biases of hundreds of percent (e.g., California in May for CNRM-ESM2-1; not shown). However, the downscaled results greatly improve on large wet biases across Nevada, Colorado, Wyoming, and Montana in the parent GCMs, which are as much as 50% in the ensemble mean across Wyoming, and hundreds of percent in some GCMs. Also, across Arizona, the summertime precipitation maximum is completely missed in all GCMs. Meanwhile, the downscaled results capture it well, albeit

400   with some simulations far too wet (~100% bias) compared to PRISM. Difficulties in simulating summertime precipitation across the southwestern U.S. have been noted in previous studies (Liu et al., 2017; Rahimi et al., 2022). Dynamical downscaling generally increases the simulated precipitation spread across Arizona, California, Oregon, and Washington whilst decreasing the spread across interior states.

In general, overly wet and cold dynamically downscaled GCMs have previously been noted across the region with a different RCM (Rastogi et al., 2022), indicating that biases in the GCM forcing data may be to blame. The effects of GCM bias propagation are being explored in Rahimi et al., (2023; in revisions) and Risser et al. (2023; in revisions). The absence of such large biases in WRF-ERA5 (Figs. 3, 4, and 5), which is equivalent to the downscaled GCMs, except driven by ERA5, lends further evidence in support of this hypothesis.

Finally, we evaluate historical extreme precipitation (rx1day) in WUS-D3. Dynamical downscaling markedly improves the spatial distribution of rx1day across the region compared to the parent GCMs (Figure 5, top), as with mean precipitation (Fig. 3). Across individual states, dynamical downscaling produces rx1day magnitudes that are in many cases about double their parent GCM values, particularly across Arizona, California, Oregon, and Washington. While generally too wet compared to PRISM, downscaled simulations are much closer to the downscaled reanalysis (WRF-ERA5). We attribute the greater rx1day values in the downscaled simulations to the much better representation of topography and orographic precipitation in WRF compared to the parent GCMs. As such, the wetter behavior of WRF solutions is generally localized to the highest elevations across each state. These locations are precisely where observational uncertainties are also maximized (Lundquist et al., 2019). Thus, we characterize downscaled rx1day simulations as being wetter than PRISM, rather than clearly being wet biased. Because of the rareness of rx1day events, the computation of rx1day is also sensitive to the phasing of internal climate variability, which is different in GCMs relative to PRISM and WRF-ERA5. Hence, differences between downscaled GCMs and PRISM rx1day precipitation may be partially explainable by these phasing differences.

As indicated by the shaded bars in Figure 5, downscaling may alter the original GCM spread in simulated rx1day magnitudes. Specifically, WRF significantly increases the GCM spread in Oregon, Arizona, New Mexico, and Colorado, but significantly decreases the spread in California, Nevada, Idaho, and Montana. The states with large increases in spread are generally where rx1day is more likely to occur during summer, indicating disagreements in monsoon-related extreme precipitation across downscaled results. The amplification of model uncertainty in precipitation extremes by dynamical downscaling is yet to be addressed by the regional modeling community and is a current focus of our research efforts.

4. Climate Response Across the Western U.S.

Next, we provide an overview of the WUS-D3's climate response to anthropogenic forcing (following SSP3-7.0). Figure 6 shows the mid-century (2030–2060; MC) and end-century (2070–2100; EC) projected changes in annual-mean precipitation scattered against warming, averaged across 11 WUS states in each GCM. The native GCM projections (indicated by letters) are connected to their downscaled counterparts (indicated by circles) by thick arrows. The purpose of Fig. 6 is to illustrate the degree to which downscaling can modify the original GCM projections on regional scales.

According to the downscaled ensemble, the WUS will experience 2.25 ±0.58 K of warming by MC, and 4.65 ± 1.14 K by EC (relative to 1980-2010). A considerably more uncertain but generally wetter future is also predicted, with an ensemble mean precipitation change of 0.039 ± 0.93 mm d^{-1} by MC and 0.083 ± 0.13 mm d^{-1} by EC. Despite a positive mean change, a handful of simulations suggest drying across the region (Fig. 6, right). Downscaling generally preserves the inter-model variation

among the parent GCMs in the 11-state mean. For temperature change, there are correlation coefficients of 0.96 and 0.98 for MC and EC, respectively, between the raw GCM and downscaled ensembles. Correlation coefficients are lower for precipitation change but remain high: 0.88 and 0.78 for MC and EC, respectively. Regional-mean GCM warming is typically modified by no more than 0.5 K.

465 Interestingly, downscaling generally reduces warming (leftward pointing arrows). This effect is most prominent during winter and spring (Figure S4), indicating that much better resolution of topography and hence climatological snowpack improvements (e.g., Walton et al., 2017) in the downscaling may be reducing the overall snow albedo feedback intensity and hence the surface's temperature sensitivity to anthropogenic forcing. Summer will see the largest mean temperature increases across the WUS by EC,

470 5.2 ± 1.2 K in the WRF simulations compared to 5.3 ± 1.2 K in the GCMs. In contrast to temperature, downscaling does alter the regional precipitation signals significantly, but not in any systematic or obviously predictable way; downscaling can either wetten or dry the GCM precipitation projection. These modifications are generally no more than 0.05 mm d^{-1}, but notably CanESM5 and FGOALS-g3's projections are altered by -0.2 mm d^{-1} and $+0.15$ mm d^{-1}, respectively by EC. In the case of CanESM5,

475 this transforms strong wetting to weak drying.

4.1 Spatial Patterns of Temperature and Precipitation Change in WRF versus GCMs

Although domain-mean changes are minimally unaffected by downscaling, the spatial patterns of temperature and precipitation change in the downscaled solutions are significantly different from

480 those of the raw GCM projections (Figure 7, individual downscaled GCM annual changes are shown in Figs. S5, S6). To account for large intermodel spread in climate sensitivity, the local warming is

485 normalized by EC changes in global warming. A value of 2 K K^{-1} indicates that a gridcell warms at

twice the rate of the global average. Examining the upper panels of Figure 7, large-scale spatial patterns

of warming are preserved in the downscaling, but there are seasonal and local differences. Notably, we

see enhanced (and likely more realistic) warming adjacent to mountainous areas of the Rockies during

winter and spring and at the highest elevations of the Sierra Nevada during summer. This is primarily

490 tied to the improved representation of topography, and thus more expansive historical snow cover. We

expect simulations with greater snow cover to exhibit more warming across these areas because they

would be able to lose more snow under warming (Figure S7) and therefore have a stronger snow albedo

feedback (SAF: Hall 2004; Qu and Hall 2006; Thackeray and Fletcher 2016). The addition of value is

clear in terms of the granularity of future snow loss and subsequent impact on warming in winter and

495 spring and is comparable to previous studies (e.g., Walton et al. 2017). However, enhanced summertime

high-elevation warming may be somewhat overestimated due to large cold-season wet biases (Figures

4, S3); excessive snow survives into the warm season and creates an unrealistically large snow albedo

feedback effect under climate change. We also hypothesize that the lapse rate feedback (Hansen et al.,

1984, Colman and Soden, 2021) may be contributing to the enhanced warming at high elevations during

500 summertime. For example, in the GCMs, 850 hPa temperatures warm by 3.75 K across the WUS, while

300 hPa temperatures warm by 4.75 K (not shown). This enhanced warming at high altitudes likely

contributes to enhanced surface warming at high elevations as well. Lastly, the downscaled ensemble

exhibits enhanced warming across the interior during fall, perhaps associated with drying and a

reduction in evaporative damping of surface temperature (Zhou et al., 2019).

505   Spatial patterns of precipitation change illustrate greater contrasts between the downscaling and

GCMs across seasons (Figure 7, lower panels). EC changes are normalized by global-mean warming

here to take into account the large spread in climate sensitivity among CMIP6 models, and to focus

attention on the components of the hydrologic response that do not simply scale with temperature. The

large-scale precipitation response is generally preserved in downscaling, with statistically significant

510   wetting (drying) in the northwestern (southwestern) U.S. during winter and spring. The lack of

statistical significance along a transition region, extending from southern California through northern

Arizona and New Mexico, is symptomatic of GCM disagreement on the location of the transition of

subtropical drying to mid-latitude wetting (Meehl et al., 2007; Neelin et al., 2013). Consistent with

other studies (e.g., Mahoney et al. 2021; Rupp et al. 2022), the downscaled ensemble appears to

515   produce greater wetting across major WUS mountain ranges during spring and winter. The locally more

intense change signal is tied to increased water vapor within atmospheric rivers and other synoptic

disturbances, which interacts with WRF's more realistically simulated terrain to produce more realistic

orographic uplift relative to native-resolution GCMs (Huang et al., 2020; see mean changes in vertical

velocities; Fig. S8). There are also instances where WRF simulates a locally more intense drying signal

520   compared to the native GCMs, which is also clearly linked to topography, e.g., the Sierra Nevada in

autumn and spring, the upslope of the Cascades in summer, and northwestern Mexico in winter and

summer.

        We also examine ensemble-mean fractional precipitation changes (again normalized by

warming), to focus attention on where the largest changes are relative to the climatology (Figure

525   8). Here, a value of -20% K^{-1} indicates that EC-era precipitation has decreased by 20% relative to the

historical-era while the global temperature has warmed by 1 K. One of the most notable and robust

signals, seen during all seasons and almost entirely missed in the parent GCMs (Fig. S9), is significant

wetting in the lee of major WUS mountain ranges. This effect was explored in an idealized context in

530 Siler and Roe (2014). They concluded that higher cloud bases associated with decreased surface relative

humidity values in a warmer world will lead to enhanced hydrometeor fallout further upslope and

downwind of mountain ranges. In our simulations, these lee-side changes are large in magnitude. For

example, in winter, precipitation increases by 7-10% K^{-1} in the lee of the Cascades, 10-20% K^{-1} in the

lee of the Sierra Nevada, 6-20 % K^{-1} over California's Central Valley (i.e., the lee of the coastal ranges),

535 and otherwise 5-20% K^{-1} in lee-side watersheds of the intermountain West, including the entire western

Great Plains. In spring, this lee-side wetting response is limited to northern mountain ranges such as the

Cascades, Wyoming ranges, and the Northern Great Plains. During summer, the downscaling also

shows a dipole of drying (wetting) over the windward (leeward) side of the Sierra Nevada. This could

be related to the mechanism identified by Siler and Roe (2014), although given the importance of

540 mountain-top convection to summertime precipitation here, it may also result from changes in other

mechanisms. In general, because the lee side of WUS mountain ranges are typically arid, these large

and robust fractional increases in lee-side precipitation will likely have a significant impact on local

water resources and ecology.

545 **4.2 Changes in Extremes**

The future fractional change in extreme (rx1day) precipitation is much more consistent across

the WUS than for the mean, with intensified extremes occurring over most of the domain in both the

parent and dynamically downscaled GCMs (Figure 9, left column, Fig S10 for individual GCMs). These

changes in rx1day vary from roughly 0-12 %/K across the domain in the downscaled ensemble mean. In

550   both the GCM and downscaling cases, the spatial variations in the changes in rx1day can be traced in

part to vertical velocity (defined by the pressure velocity) changes: Spatial correlations of -0.7 (-0.3) are

found between rx1day precipitation and vertical-velocity changes in GCM (downscaling) experiments.

Here, a negative correlation coefficient implies a positive relationship between upward vertical

velocities and positive rx1day changes. However, the patterns of vertical velocity change are very

555   different in the two cases. In the downscaling experiments, the largest intensification of rx1day occurs

via upward vertical-velocity increases on the lee side of latitudinally-oriented mountain chains that are

not resolved in the parent GCMs (Figure 9, right column). Over large parts of these areas, the increases

are super-Clausius-Clapeyron (>7%/K). This indicates that extreme precipitation intensifies at a greater

rate than saturation specific humidity, commonly termed the thermodynamic component of extreme-

560   precipitation scaling. Thus, WRF simulates greater dynamical intensification of extreme precipitation

(e.g., Norris et al., 2019) than the GCMs, and in a distinct topographically-modulated pattern.

   Next, we examine future changes in extreme heat, defined by the number of days exceeding the

99th percentile of the historical daily-maximum surface air temperature (Tmax99). Consistent with

extreme precipitation, these changes are normalized by global warming to account for the large

565   intermodel spread in climate sensitivity (Figure 10, left). Averaged across the WUS, Tmax99

exceedances increase by 11.9 ± 2.1 days per year K^{-1}. California, Oregon, and Washington see increases

of 4-7 days per year K^{-1}, with coastal areas seeing increases of less than 5 days per year K^{-1}. The power

of dynamical downscaling is particularly evident, as the GCMs (top row, Fig. 10) cannot simulate (i) the

correct coastline geometry, leading to an unphysical intrusion of maximized ocean-influenced Tmax99

exceedances, and (ii) the complex terrain of the WUS, which strongly modulates the snow coverage and

subsequently the land surface sensitivity to warming. Additional examination reveals that the GCMs

575   with the greatest regional mean warming are not necessarily the GCMs with the largest increase in

exceedances (Figs. S5, S12). This discrepancy may be due to GCM differences in the simulation of

synoptic-scale events that produce heat waves. Anthropogenic changes in such events may occur

independently of mean temperature shifts (Fig. 7).

Next, we explore whether changes in Tmax99 exceedances are explainable by mean shifts in the

580   temperature distribution. As shown in Fig. 10 (right column), the number of actual Tmax99

exceedances from parent and dynamically downscaled GCMs can be quite different compared to the

case where all quantiles in the temperature distribution are shifted equally based on the amount of local

mean warming in Tmax99 (Fig. 10; middle column). Red (blue) pixels indicate regions where the tails

of the temperature distribution are warming more (less) than can be explained by mean warming in

585   Tmax99. Assuming a mean shift in Tmax99 significantly underpredicts the increase in exceedances by

3-4 days per year K^{-1} across portions of California, Oregon, and Washington. Still greater

underpredictions of future exceedances assuming a mean shift are seen across western Montana, Idaho,

and portions of western Wyoming, particularly at higher elevations. Further south however, the number

of exceedances in Tmax99 can be explained mostly by mean shifts in Tmax99. Assuming a mean shift,

590   exceedances are slightly overpredicted across portions of New Mexico and western Texas relative to

GCM and WRF simulations. This analysis highlights that the intensification of extreme temperature

events may not be entirely explainable by mean shifts in the temperature distribution alone, and parent

and downscaled GCMs are broadly similar in this respect. However, there is also significant spatial structure in the downscaling patterns not seen in the GCMs, indicating that local atmospheric dynamics and local land-atmosphere feedbacks play a role in shaping change in the right tail of the temperature distribution.

7. Summary and Conclusions

Future regional climate change remains difficult to project, given the low resolution of GCMs, particularly over a region of complex terrain such as the western U.S. In this study, we present a dataset containing sixteen CMIP6 models dynamically downscaled with WRF over the region from 1980 to 2100 at 9-km grid spacing: the Western U.S. Dynamically Downscaled Dataset (WUS-D3). The future projections are primarily based on the SSP3-7.0 high-emissions scenario, but we include two additional downscaled experiments with CESM2 of the SSP2-4.5 and SSP5-8.5 scenarios. An extensive evaluation of CMIP6 models' historical simulations over the western U.S. has been conducted (Krantz et al., 2021; Goldenson et al., 2023) to identify the most suitable candidates for downscaling over this region. However, GCM selection was also based on the availability of data required to provide initial and boundary conditions to WRF. The optimal configuration of WRF over the western U.S. was established via an extensive evaluation of an ERA5-driven WRF run (Rahimi et al., 2022). Numerous other challenges of using the CMIP6 data to force WRF are outlined in the methods.

Aside from the obvious improved representation of spatial patterns of meteorological variables, there are many notable improvements of the downscaling over raw GCMs when compared to observations over the historical period. For example, the WRF simulations largely correct for major wet

Deleted: , in revisions

630   biases (~100%) in the raw GCMs over Nevada, Wyoming, Montana, and Colorado. These bias

reductions apply to both winter and summer, depending on the state. Moreover, the GCMs completely

fail to represent the summertime precipitation maximum over Arizona and New Mexico, which is

corrected in the downscaled experiments, albeit with some large wet biases therein. The WRF

simulations also correct large summertime warm biases over much of the domain, particularly the

635   interior states, due to the improved representation of terrain and resulting snowpack improvements.

Finally, extreme precipitation (measured by rx1day) is greatly increased (generally about doubled) from

the GCM values. In some cases, this amounts to wet biases from WRF, according to PRISM, but these

apparent biases are mostly at high elevations where observational uncertainties are maximized

(Lundquist et al., 2019).

640   There are, however, extensive and systematic biases that remain from the parent GCMs and in

some cases are exacerbated. For example, the GCMs generally overestimate winter precipitation along

the west coast, which likely results from unrealistically high moisture contained within atmospheric

rivers (Norris et al., 2021) and other GCM biases transmitted to WRF. And in the WRF simulations,

these wet biases are amplified, likely as excessive moisture is forced up steeper orographic gradients

645   than in the GCMs. Also, unlike the GCMs, the downscaled experiments are generally too cold

compared to PRISM, particularly in winter, with some states exhibiting as much as a 5 K bias. These

results are comparable to Rastogi et al. (2022), who used a different regional climate model, implying

that inherited GCM biases may be to blame. The dynamical downscaling community should be frank

about such biases, particularly in lieu of the fact that these biases are often artificially removed post-

650   downscaling using bias correction. This practice is ubiquitous in hydrology and demand forecast

modeling, as well as in statistical downscaling. End-users of WUS-D3 should be open-eyed and wary about the possibility that these large historical biases may compromise the trustworthiness of its climate change signals.

The future downscaled climate change signals are shaped In physically credible ways by the regional model's more realistic coastlines and topography. Large-scale warming patterns are generally preserved from the parent GCMs, but with enhanced warming adjacent to high terrain during winter and spring and over high elevations during summer. This locally enhanced warming occurs where relative snow losses are maximized in the future, a feature that cannot be captured at the GCMs' coarse resolution. Meanwhile, precipitation patterns undergo much greater transformation with downscaling. Although WRF preserves the broad pattern of subtropical drying and midlatitude wetting, WRF simulates additional local precipitation changes. In particular, mean precipitation changes are often consistent with wetting on the windward side of mountain complexes, as warmer, moister air masses are uplifted orographically during precipitation events, similar to Huang et al. (2020). There are large fractional precipitation increases on the lee side of mountain complexes, consistent with the theoretical work of Siler and Roe (2014). This could lead to significant changes in water resources and ecology across these arid landscapes. The intensification of precipitation and temperature extremes is also modified in significant ways by dynamical downscaling. Over complex terrain, precipitation extremes scale at much greater rates, on the order of 12%/K. This greater scaling in WRF is likely due to greater dynamical enhancement of extreme precipitation over mountain ranges, as evidenced by the intensification of upward vertical velocity changes conditioned on extremes. Temperature extremes also intensify, as measured by future exceedances of historical 99th-percentile surface air temperature, per

degree global warming. These are on the order of +5 days year^{-1} K^{-1} along the west coast and

675 approaching 15 days year^{-1} K^{-1} in the interior west. The simulated changes are mostly greater than that

predicted by a simple mean shift of the temperature distribution, indicating the effect of an extension of

the right tail. The imprint of topography is evident in this change in the temperature distribution's

shape, indicating the importance of local atmospheric dynamics and land surface feedbacks.

Despite the care taken in creating WUS-D3, this manuscript provides a forum to scrutinize

680 dynamical downscaling technique. For instance, here we assume that the ocean-atmosphere coupling is

adequately preserved in downscaling since SSTs are prescribed to update regularly, and large-scale

winds and temperatures are preserved in downscaling via spectral nudging. But, is this a good

assumption given that half of our 45-km grid covers the open Pacific, so should a version of WRF with

coupled ocean capabilities be used in future dfownscaling across the region? Also, as discussed

685 previously, unrealistically large surface air temperature and precipitation biases in the parent GCMs

were in some cases replaced by equally egregious biases in the downscaled solution. Despite a careful

GCM selection process employed in this study, does this result motivate the consideration of a pre-

downscaling bias correction procedure of GCM fields in future studies?

WUS-D3 constitutes the first comprehensive dataset of landscape-resolving climate projections

690 over the western U.S. Although only temperature and precipitation projections have been evaluated

here, the dataset includes all 2-D and 3-D meteorological and land-surface variables at 6-hourly

intervals with a auxiliary datastream of more than 20 land-surface variables needed to drive downstream

models (e.g., hydrology) models offline. Thus, it represents a unique opportunity to explore potential

future changes to a wide diversity of weather/climate phenomena over the region. These include but are

Deleted:

Deleted: 's

Deleted: resolution

not limited to atmospheric rivers, the North American monsoon, summer convective storms, intense

heat waves, wildfire-related downslope winds, and ventilation by sea breezes. Moreover, these data may

700   be used to drive offline and calibrated hydrology and fire-weather models to obtain more detailed

projections of water resources/flooding and wildfire. Nevertheless, there are biases in the downscaled

simulations, briefly documented here, which should be understood and appreciated when using the data

for future projections. We strongly encourage the community to use these results with other

dynamically and statistically downscaled products to develop risk assessments and bound uncertainty.

705   Such intercomparisons of different downscaled products are critical to assessing a product's usefulness

and applicability. Finally, for the express purpose of dynamical downscaling, we implore the CMIP7

protocol development team to mandate that new GCM outputs for ta, ua, va, and hus be reported on 20-

30 isobaric pressure levels and at 6-hourly intervals (along with ps), as increasingly powerful computing

platforms are beginning to enable the community to consider dynamically downscaling large ensembles

710   of GCMs.

Data Availability

The versions of WRF used in this study, a Jupyter noetebook reproducing the figures in the main text,

their attendant files, and the geography files are arhived with zenodo in an open DOI subject to a

715   Creative Commons License version 4 (DOI:10.5281/zenodo.10286544;

https://zenodo.org/records/10286545. All downscaled data for WUS-D3, including the full 6-hourly

WRF datastream (Tier 1), hourly data for select land-surface variables (Tier 2), and a daily post-

processed datastream (Tier 3) are located in the following open-data bucket on Amazon S3: s3://wrf-

cmip6-noversioning/ at https://registry.opendata.aws/wrf-cmip6/. These data are completely open and

720   free to the public. We have also developed a technical access and usage document that details these

three data tiers which can be found at

https://dept.atmos.ucla.edu/sites/default/files/alexhall/files/aws_tiers_dirstructure_nov22.pdf and on

ResearchGate at

(https://www.researchgate.net/publication/374504614_Data_tier_descriptions_directory_structure_and_

725   data_access_of_the_Western_US_Dynamically_Downscaled_Dataset_WUS-D3_version_1;

DOI:10.13140/RG.2.2.11385.85609). As recommended in the document, these data are most easily

downloaded when using Amazon Web Service's (AWS') Command Line Interface (CLI) or with wget.

An example is presented in the technical access and usage document.

730   **Code Availability**

Individualized preprocessing codes were developed to create the intermediate binary files for each

GCM before ingestion into WRF. As such, we have archived these codes for various GCMs on zenodo

(DOI:10.5281/zenodo.10286544).

> **Deleted:** we have created uploaded these files into the Amazon S3 bucket in a subdirectory labeled 'downscaling_codes'.

735   **Team List**

Stefan Rahimi, Lei Huang, Alex Hall, Naomi Goldenson, Will Krantz, Benjamin Bass, Chad

Thackeray, Henry Lin, Di Chen, Eli Dennis, Emily Slinskey, Sara Graves, Surabhi Biyani, Stephen

Copper, Bowen Wang, and Elease Liu Stemp from the UCLA Center for Climate Science contributed to

the evaluation of this product, along with Zachary Lebo from the University of Oklahoma nad Ethan Collins from the University of Wyoming.

Author Contributions

745 SR and LH executed the simulations; SR and AH wrote the manuscript draft; JN, AH, and ZL edited the manuscript, NG, WK, and LH developed workflow to evaluate GCM performance and download GCM outputs, while BB, CT, HL, DC, ED, EC, ZL, ES, SG, SB, SC, and BW led analyses to explore the downscaled results in more detail.

Competing Interests

750 The authors declare that they have no conflict of interest.

Funding and Acknowledgements

We acknowledge the supporting funding agencies supporting this project: the Department of Energy's HyperFACETSS (DE-SC0016605) project, the Strategic Environmental Research and Development

755 Program (SERDP) under Project (RC19-1391), the California Energy Commission (EPC-20-006), and the University of California's Climate Ecosystems Future (LRF-18-542511) project. We also thank the computational support through the NCAR-Wyoming Supercomputing Center (NWSC) and the Computational and Information Systems Laboratory (CISL). We also acknowledge the TORNERDO consortium, as well as Cora for writing assistance.

760

References

765 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.

Bruyère, C. L., Done, J. M., Holland, G. J., and Fredrick, S.: Bias corrections of global models for

770 regional climate simulations of high-impact weather, Clim Dyn, 43, 1847–1856,

https://doi.org/10.1007/s00382-013-2011-6, 2014.

Bukovsky, M. S. and Karoly, D. J.: A Regional Modeling Study of Climate Change Impacts on Warm-

Season Precipitation in the Central United States, Journal of Climate, 24, 1985–2002,

https://doi.org/10.1175/2010JCLI3447.1, 2011.

775 Bukovsky, M. S., Gao, J., Mearns, L. O., and O'Neill, B. C.: SSP-Based Land-Use Change Scenarios:

A Critical Uncertainty in Future Regional Climate Change Projections, Earth's Future, 9,

e2020EF001782, https://doi.org/10.1029/2020EF001782, 2021.

Chen, D., Norris, J., Thackeray, C., and Hall, A.: Increasing precipitation whiplash in climate change

hotspots, Environ. Res. Lett., 17, 124011, https://doi.org/10.1088/1748-9326/aca3b9, 2022.

780 Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., Ban, N., Bastin, S., Belda,

M., Belusic, D., Caldas-Alvarez, A., Cardoso, R. M., Davolio, S., Dobler, A., Fernandez, J., Fita, L.,

Fumiere, Q., Giorgi, F., Goergen, K., Güttler, I., Halenka, T., Heinzeller, D., Hodnebrog, Ø., Jacob, D.,

Kartsios, S., Katragkou, E., Kendon, E., Khodayar, S., Kunstmann, H., Knist, S., Lavín-Gullón, A.,

Lind, P., Lorenz, T., Maraun, D., Marelle, L., van Meijgaard, E., Milovac, J., Myhre, G., Panitz, H.-J.,

785   Piazza, M., Raffa, M., Raub, T., Rockel, B., Schär, C., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L.,

Stocchi, P., Tölle, M. H., Truhetz, H., Vautard, R., de Vries, H., and Warrach-Sagi, K.: A first-of-its-

kind multi-model convection permitting ensemble for investigating convective phenomena over Europe

and the Mediterranean, Clim Dyn, 55, 3–34, https://doi.org/10.1007/s00382-018-4521-8, 2020.

Coppola, E., Nogherotto, R., Ciarlo', J. M., Giorgi, F., van Meijgaard, E., Kadygrov, N., Iles, C., Corre,

790   L., Sandstad, M., Somot, S., Nabat, P., Vautard, R., Levavasseur, G., Schwingshackl, C., Sillmann, J.,

Kjellström, E., Nikulin, G., Aalbers, E., Lenderink, G., Christensen, O. B., Boberg, F., Sørland, S. L.,

Demory, M.-E., Bülow, K., Teichmann, C., Warrach-Sagi, K., and Wulfmeyer, V.: Assessment of the

European Climate Projections as Simulated by the Large EURO-CORDEX Regional and Global

Climate Model Ensemble, Journal of Geophysical Research: Atmospheres, 126, e2019JD032356,

795   https://doi.org/10.1029/2019JD032356, 2021.

Colman, R. and Soden, B. J.: Water vapor and lapse rate feedbacks in the climate system, Rev. Mod.

Phys., 93, 045002, https://doi.org/10.1103/RevModPhys.93.045002, 2021.

Dennis, E. J. and Berbery, E. H.: The Role of Soil Texture in Local Land Surface–Atmosphere

Coupling and Regional Climate, Journal of Hydrometeorology, 22, 313–330,

800   https://doi.org/10.1175/JHM-D-20-0047.1, 2021.

Daly, C., Neilson, R. P., and Phillips, D. L.: A Statistical-Topographic Model for Mapping

Climatological Precipitation over Mountainous Terrain, Journal of Applied Meteorology and

Climatology, 33, 140–158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

805   Durack, Paul J., Taylor, Karl E., Eyring, Veronika, Ames, Sasha K., Doutriaux, Charles, Hoang, Tony, Nadeau, Denis, Stockhause, Martina, and Gleckler, Peter J.. *input4MIPs: Making [CMIP] model forcing more transparent*. United States: N. p., 2017. Web. doi:10.2172/1463030.

Goldenson, Naomi, L. Ruby Leung, Linda O. Mearns, David W. Pierce, Kevin A. Reed, Isla R.

810   Simpson, Paul Ullrich, et al. "Use-Inspired, Process-Oriented GCM Selection: Prioritizing Models for Regional Dynamical Downscaling." *Bulletin of the American Meteorological Society* 104, no. 9 (September 22, 2023): E1619–29. https://doi.org/10.1175/BAMS-D-23-0100.1.

Hall, A.: The Role of Surface Albedo Feedback in Climate, Journal of Climate, 17, 1550–1568,

815   https://doi.org/10.1175/1520-0442(2004)017<1550:TROSAF>2.0.CO;2, 2004.

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., and Lerner, J.: Climate Sensitivity: Analysis of Feedback Mechanisms, in: Climate Processes and Climate Sensitivity, American Geophysical Union (AGU), 130–163, https://doi.org/10.1029/GM029p0130, 1984.

820

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R.,

Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E.,

830 Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P. de, Rozum, I.,

Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the

Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Huang, X., Swain, D. L., Walton, D. B., Stevenson, S., and Hall, A. D.: Simulating and Evaluating

Atmospheric River-Induced Precipitation Extremes Along the U.S. Pacific Coast: Case Studies From

835 1980–2017, Journal of Geophysical Research: Atmospheres, 125, e2019JD031554,

https://doi.org/10.1029/2019JD031554, 2020.

Huang, X., Swain, D. L., and Hall, A. D.: Future precipitation increase from very high resolution

ensemble downscaling of extreme atmospheric river storms in California, Science Advances, 6,

eaba1323, https://doi.org/10.1126/sciadv.aba1323, 2021.

840 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.:

Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models,

Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD009944, 2008.

Komurcu, M., Emanuel, K. A., Huber, M., and Acosta, R. P.: High-Resolution Climate Projections for

the Northeastern United States Using Dynamical Downscaling at Convection-Permitting Scales, Earth

845 and Space Science, 5, 801–826, https://doi.org/10.1029/2018EA000426, 2018.

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.

850 Lanzante, J. R., Dixon, K. W., Nath, M. J., Whitlock, C. E., and Adams-Smith, D.: Some Pitfalls in Statistical Downscaling of Future Climate, Bulletin of the American Meteorological Society, 99, 791–803, https://doi.org/10.1175/BAMS-D-17-0046.1, 2018.

Liu, C., Ikeda, K., Thompson, G., Rasmussen, R., and Dudhia, J.: High-Resolution Simulations of Wintertime Precipitation in the Colorado Headwaters Region: Sensitivity to Physics Parameterizations,
855 Monthly Weather Review, 139, 3533–3553, https://doi.org/10.1175/MWR-D-11-00009.1, 2011.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen, L., Clark, M., Dai, A., Dudhia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute, S., Li, Y., Thompson, G., and Yates, D.: Continental-scale convection-permitting modeling of the current and future climate of North America, Clim Dyn, 49, 71–95, https://doi.org/10.1007/s00382-016-3327-9, 2017.

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

Mahoney, K., Scott, J. D., Alexander, M., McCrary, R., Hughes, M., Swales, D., and Bukovsky, M.: Cool season precipitation projections for California and the Western United States in NA-CORDEX
865 models, Clim Dyn, 56, 3081–3102, https://doi.org/10.1007/s00382-021-05632-z, 2021.

Mankin JS, Simpson I, Hoell A, Fu R, Lisonbee J, Sheffield A, Barrie D.: NOAA Drought Task Force Report on the 2020–2021 Southwestern U.S. Drought. NOAA Drought Task Force, MAPP, and NIDIS, 2022

870 Meehl, G.A., et al. (2007) Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge.

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., Gettelman, A.,
875 Lawrence, D. M., Sanderson, B. M., and Strand, W. G.: Climate Change Projections in CESM1(CAM5) Compared to CCSM4, Journal of Climate, 26, 6287–6308, https://doi.org/10.1175/JCLI-D-12-00572.1, 2013.

Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., Caya, D., Correia, J.,
880 Flory, D., Gutowski, W., Takle, E. S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A. M. B., Qian, Y., Roads, J., Sloan, L., and Snyder, M.: The North American Regional Climate Change Assessment Program: Overview of Phase I Results, Bulletin of the American Meteorological Society, 93, 1337–1362, https://doi.org/10.1175/BAMS-D-11-00223.1, 2012.

Mitchell, D. L., Ivanova, D., Rabin, R., Brown, T. J., and Redmond, K.: Gulf of California Sea Surface
885 Temperatures and the North American Monsoon: Mechanistic Implications from Observations, Journal

of Climate, 15, 2261–2281, https://doi.org/10.1175/1520-0442(2002)015<2261:GOCSST>2.0.CO;2, 2002.

Morrison, H. and Milbrandt, J.: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests, Journal of the Atmospheric Sciences, 72, 287–311, https://doi.org/10.1175/JAS-D-14-0065.1, 2015.

Neelin, J. D., Langenbrunner, B., Meyerson, J. E., Hall, A., and Berg, N.: California Winter Precipitation Change under Global Warming in the Coupled Model Intercomparison Project Phase 5 Ensemble, Journal of Climate, 26, 6238–6256, https://doi.org/10.1175/JCLI-D-12-00514.1, 2013.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2010JD015139, 2011.

Norris, J., Carvalho, L. M. V., Jones, C., and Cannon, F.: Deciphering the contrasting climatic trends between the central Himalaya and Karakoram with 36 years of WRF simulations, Clim Dyn, 52, 159–180, https://doi.org/10.1007/s00382-018-4133-3, 2019.

Norris, J., Hall, A., Chen, D., Thackeray, C. W., and Madakumbura, G. D.: Assessing the Representation of Synoptic Variability Associated With California Extreme Precipitation in CMIP6

Models, Journal of Geophysical Research: Atmospheres, 126, e2020JD033938,

905 https://doi.org/10.1029/2020JD033938, 2021.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R.,

Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The

Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development,

9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

910 Qu, X. and Hall, A.: Assessing Snow Albedo Feedback in Simulated Climate Change, Journal of

Climate, 19, 2617–2630, https://doi.org/10.1175/JCLI3750.1, 2006.

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,

915 https://doi.org/10.1029/2021JD035699, 2022.

Rahimi, S., Huang, L., Norris, J., Hall, A., Thackeray, C., Goldenson, N., Risser, M., Feldman, D., and

Lebo, Z. J.: Dynamical Downscaling Should Use Physically Explainable Bias Correction, *Geophys Res.

Lett.*, 2023. (in revisions)

Risser, M., Rahimi, S., Goldenson, G., Lebo, Z. J., Hall, A., and Feldman, D.: Is bias correction in

920 dynamical downscaling defensible? *Geophys. Res. Lett.*, 2023, (in revisions)

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., Tewari, M., Barlage, M., Dudhia, J.,

Yu, W., Miller, K., Arsenault, K., Grubišić, V., Thompson, G., and Gutmann, E.: High-Resolution

Coupled Climate Runoff Simulations of Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate, Journal of Climate, 24, 3015–3048, https://doi.org/10.1175/2010JCLI3985.1, 2011.

Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., Gutmann, E., Dudhia, J., Chen, F., Barlage, M., Yates, D., and Zhang, G.: Climate Change Impacts on the Water Balance of the Colorado Headwaters: High-Resolution Regional Climate Model Simulations, Journal of Hydrometeorology, 15, 1091–1116, https://doi.org/10.1175/JHM-D-13-0118.1, 2014.

Rastogi, D., Kao, S.-C., and Ashfaq, M.: How May the Choice of Downscaling Techniques and Meteorological Reference Observations Affect Future Hydroclimate Projections?, Earth's Future, 10, e2022EF002734, https://doi.org/10.1029/2022EF002734, 2022.

Rupp, D. E., Hawkins, L. R., Li, S., Koszuta, M., and Siler, N.: Spatial patterns of extreme precipitation and their changes under ~ 2 °C global warming: a large-ensemble study of the western USA, Clim Dyn, 59, 2363–2379, https://doi.org/10.1007/s00382-022-06214-3, 2022.

Scalzitti, J., Strong, C., and Kochanski, A.: Climate change impact on the roles of temperature and precipitation in western U.S. snowpack variability, Geophysical Research Letters, 43, 5361–5369, https://doi.org/10.1002/2016GL068798, 2016.

Siler, N. and Roe, G.: How will orographic precipitation respond to surface warming? An idealized thermodynamic perspective, Geophysical Research Letters, 41, 2606–2613, https://doi.org/10.1002/2013GL059095, 2014.

945   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.

950   Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., Wang, W., Powers, G., Duda, G., Barker, D., and Huang, X.: A Description of the Advanced Research WRF Model Version 4, https://doi.org/10.5065/1dfh-6p97, 2019.

Spero, T. L., Otte, M. J., Bowden, J. H., and Nolte, C. G.: Improving the representation of clouds, radiation, and precipitation using spectral nudging in the Weather Research and Forecasting model,
955   Journal of Geophysical Research: Atmospheres, 119, 11,682-11,694, https://doi.org/10.1002/2014JD022173, 2014.

Swain, D. L., Langenbrunner, B., Neelin, J. D., and Hall, A.: Increasing precipitation volatility in twenty-first-century California, Nature Climate Change, 8, 427–433, https://doi.org/10.1038/s41558-018-0140-y, 2018.

960   Thackeray, C. W. and Fletcher, C. G.: Snow albedo feedback: Current knowledge, importance, outstanding issues and future directions, Progress in Physical Geography: Earth and Environment, 40, 392–408, https://doi.org/10.1177/0309133315620999, 2016.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale

Models, Mon. Wea. Rev., 117, 1779–1800, https://doi.org/10.1175/1520-

965   0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Walton, D. B., Hall, A., Berg, N., Schwartz, M., and Sun, F.: Incorporating Snow Albedo Feedback into

Downscaled Temperature and Snow Cover Projections for California's Sierra Nevada, J. Climate, 30,

1417–1438, https://doi.org/10.1175/JCLI-D-16-0168.1, 2017.

Wang, J. and Kotamarthi, V. R.: Assessment of Dynamical Downscaling in Near-Surface Fields with

970   Different Spectral Nudging Approaches Using the Nested Regional Climate Model (NRCM), Journal of

Applied Meteorology and Climatology, 52, 1576–1591, https://doi.org/10.1175/JAMC-D-12-0302.1,

2013.

Wang, J. and Kotamarthi, V. R.: High-resolution dynamically downscaled projections of precipitation in

the mid and late 21st century over North America, Earth's Future, 3, 268–288,

975   https://doi.org/10.1002/2015EF000304, 2015.

White, A. B., Moore, B. J., Gottas, D. J., and Neiman, P. J.: Winter Storm Conditions Leading to

Excessive Runoff above California's Oroville Dam during January and February 2017, Bulletin of the

American Meteorological Society, 100, 55–70, https://doi.org/10.1175/BAMS-D-18-0091.1, 2019.

White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., Harley, C. D. G., Henderson,

980   S. B., Jakob, M., Lau, C.-A., Mareshet Admasu, L., Narinesingh, V., Rodell, C., Roocroft, E.,

Weinberger, K. R., and West, G.: The unprecedented Pacific Northwest heatwave of June 2021, Nat Commun, 14, 727, https://doi.org/10.1038/s41467-023-36289-3, 2023.

Zhang, C., Wang, Y., and Hamilton, K.: Improved Representation of Boundary Layer Clouds over the Southeast Pacific in ARW-WRF Using a Modified Tiedtke Cumulus Parameterization Scheme, Mon.
985   Wea. Rev., 139, 3489–3513, https://doi.org/10.1175/MWR-D-10-05091.1, 2011.

Zhou, S., Williams, A. P., Berg, A. M., Cook, B. I., Zhang, Y., Hagemann, S., Lorenz, R., Seneviratne, S. I., and Gentine, P.: Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity, Proceedings of the National Academy of Sciences, 116, 18848–18853, https://doi.org/10.1073/pnas.1904955116, 2019.

990   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.

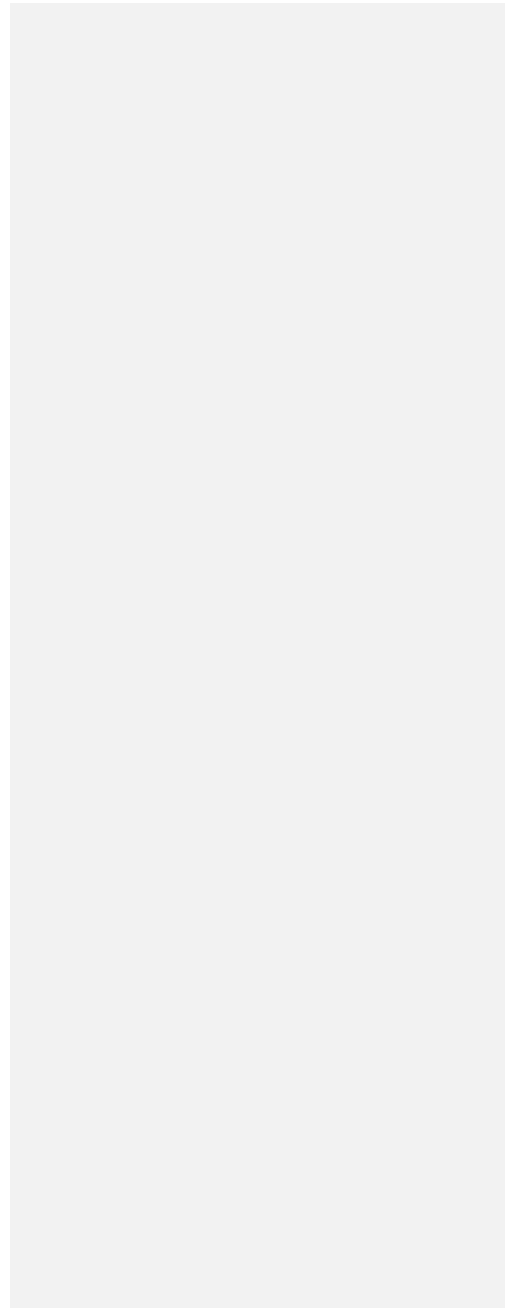
Zobel, Z., Wang, J., Wuebbles, D. J., and Kotamarthi, V. R.: Evaluations of high-resolution dynamically downscaled ensembles over the contiguous United States, Clim Dyn, 50, 863–884,
995   https://doi.org/10.1007/s00382-017-3645-6, 2018.

1000

1005

1010

1015

1020

1025 **Table 1. List of GCMs dynamically downscaled in this study. Approximate near-equatorial latitude-longitude resolutions are given. The rightmost column titled 'SST mod' indicates whether or not our CoC entrance region-based SST extrapolation is utilized.**

GCM	Variant	Center	Resolution	Source	SSP	SST mod?
ACCESS-CM2	r5i1p1f1	Commonwealth Scientific and Industrial Research Organization	1.25°x1.25°	(Bi et al., 2020)	3-7.0	yes
CanESM5	r1i1p2f1	Canadian Climate Center	2.8°x2.8°	Swart et al., (2019)	3-7.0	yes
CESM2*	r11i1p1f1	National Center for Atmospheric Research	0.94°x1.25°	Danabasoglu et al., (2020)	2-4.5, 3-7.0, 5-8.5	no
CNRM-ESM2-1	r1i1p1f2	Centre Europeen de Recherche et de Formation	1.4°x1.4°	Séférian et al., (2019)	3-7.0	no

GCM	Variant	Center	Resolution	Source	SSP	SST mod?
		Avancee en Calcul Scientifique				
EC-Earth3	r1i1p1f1	EC-Earth Consortium	0.7°x0.7°	Döscher et al., (2022)	3-7.0	yes
EC-Earth3-Veg	r1i1p1f1	EC-Earth Consortium	0.7°x0.7°	Döscher et al., (2022)	3-7.0	yes
FGOALS-g3	r1i1p1f1	Chinese Academy of Sciences	2°x2°	Li et al., (2020)	3-7.0	yes
GISS-E2-1-G	r1i1p1f2	National Aeronautic and Space Administration	2°x2.5°	Kelley et al. (2020)	3-7.0	yes
MIROC6	r1i1p1f1	Japan Agency for Marine-Earth Science and Technology	1.4°x1.4°	Tatebe et al. (2019)	3-7.0	yes
MPI-ESM1-2-HR	r7i1p1f1	Max Planck Institute	0.94°x0.94°	Gutjahr et al. (2019)	3-7.0	yes

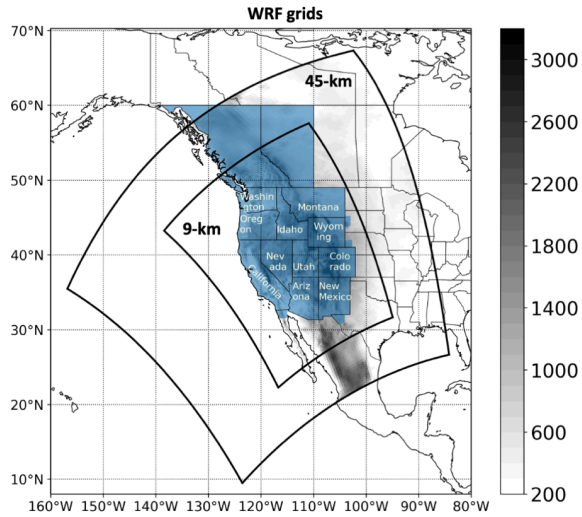
GCM	Variant	Center	Resolution	Source	SSP	SST mod?
MPI-ESM1-2-LR	r7i1p1f1	Max Planck Institute	1.9°x1.9°	Mauritsen et al., (2019)	3-7.0	no
NorESM2-MM	r1i1p1f1	NorESM Climate modeling Consortium	0.94°x1.25°	Seland et al. (2020)	3-7.0	yes
TaiESM1	r1i1p1f1	Research Center for Environmental Changes	0.94°x1.25°	Wang et al. (2021)	3-7.0	yes
UKESM1-0-LL	r2i1p1f2	Met Office Hadley Centre	1.25°x1.25°	Sellar et al., (2020)	3-7.0	yes

1030

Figure 1: WRF grids used in this study. Topography [m] is shaded to its highest resolution, and the blue shading indicates the Western Electricity Coordinating Council (WECC) coverage area.

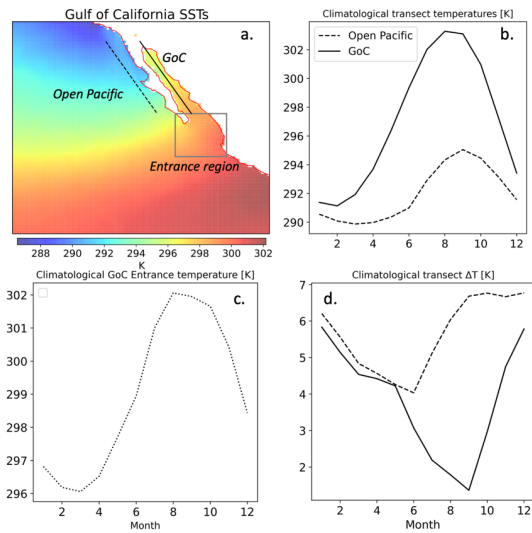



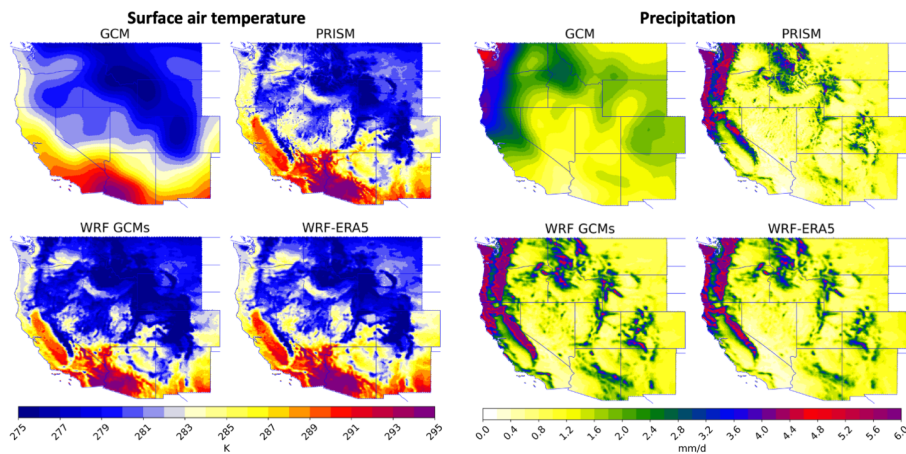
Figure 2. Panel (a) Climatological (1980-2014) mean SSTs from ERA5 along with transects across the Gulf of California (GoC; solid black curve) and Open Pacific (dashed black curve). The gray bounded zone is our GoC entrance region. Panel (b) shows the latitudinally weighted transect-mean temperatures from the Open Pacific and GoC. Panel (c) shows the area-weighted GoC entrance region temperature, while panel (d) depicts the temperature gradient along the Open Pacific and GoC transects from northwest to southeast.

1040

Figure 3. 1981-2010 annual mean (left) surface air temperature [K] and (right) precipitation rate [mm d⁻¹] from the native GCMs (GCM; 14-GCM ensemble mean), dynamically downscaled ERA5 (WRF-ERA5), dynamically downscaled GCMs (WRF GCMs; 14-member ensemble mean), and PRISM. All GCM and PRISM data are interpolated from their native grids to the 9-km WRF

1045 **grid.**

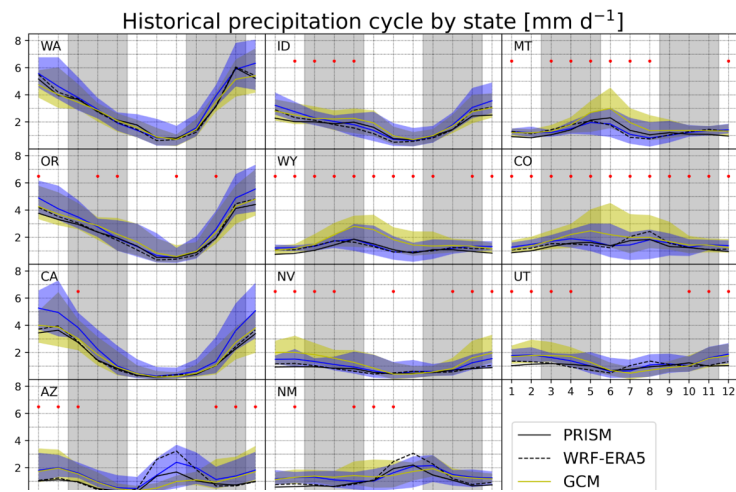
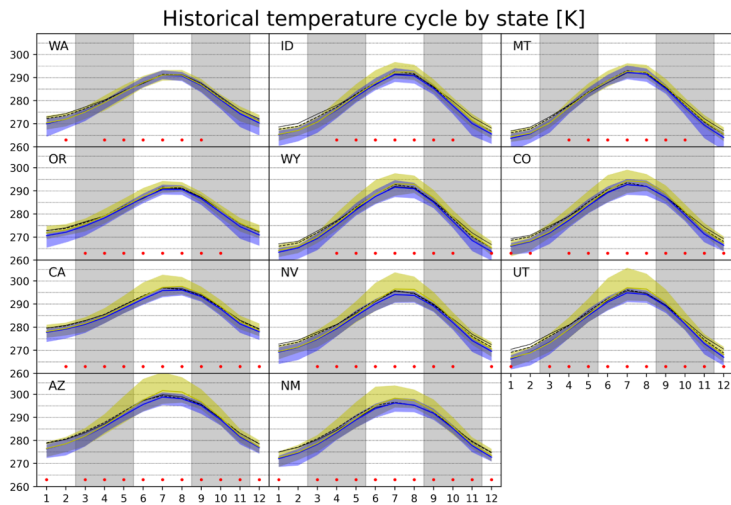



Figure 4. 1981-2010 seasonal cycles of state-mean surface air temperature [K] and precipitation [mm d⁻¹] from native GCMs (GCM), dynamically downscaled ERA5 (WRF-ERA5), dynamically

1050 downscaled GCMs (WRF GCMs; 14-member ensemble mean), and PRISM. The parent and

downscaled GCM ensemble spreads are presented in yellow and blue shading, respectively. Red

circles indicate months where the dynamically downscaled spread is smaller than the parental

GCM spread.

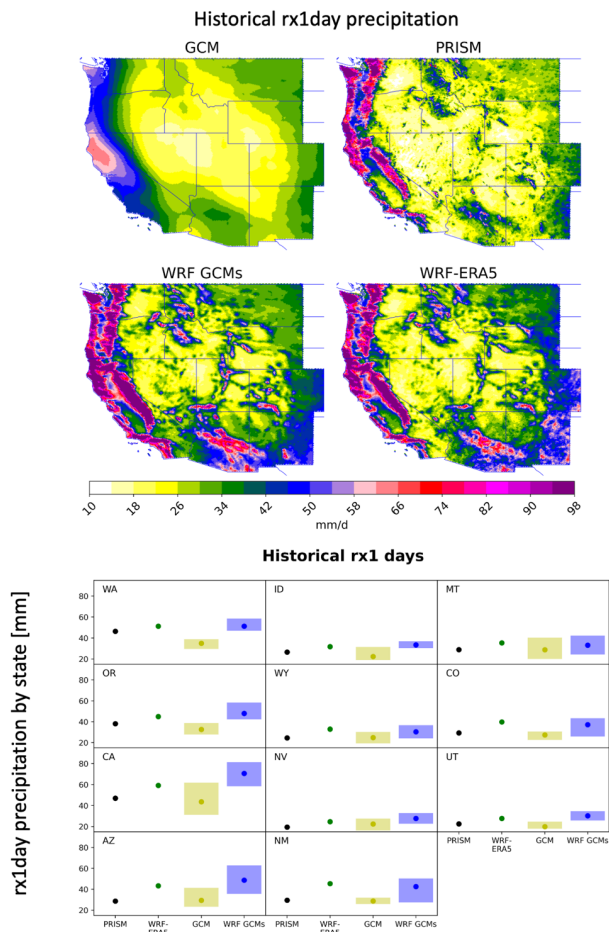



Figure 5. Historical (1981-2010) mean rx1day (annual-maximum daily precipitation) precipitation

amounts [mm] from native GCMs (14-GCM ensemble mean), dynamically downscaled ERA5

(WRF-ERA5), dynamically downscaled GCMs (WRF GCMs; 14-member ensemble mean), and

PRISM. The top figure presents the spatial distribution of rx1day precipitation, while the bottom figure presents rx1day precipitation amounts averaged across each western U.S. state. Ensemble mean values are presented as colored circles, while the GCM spread in rx1day values is shaded. GCM data were interpolated to a 1° rectilinear grid before computations.

1065

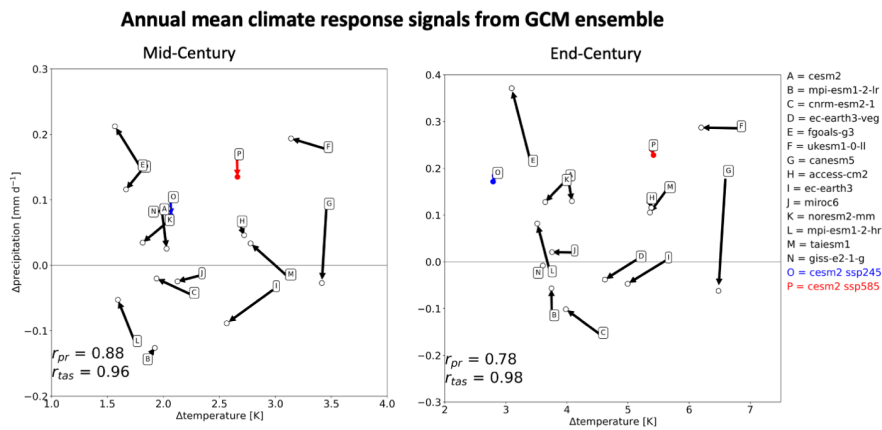
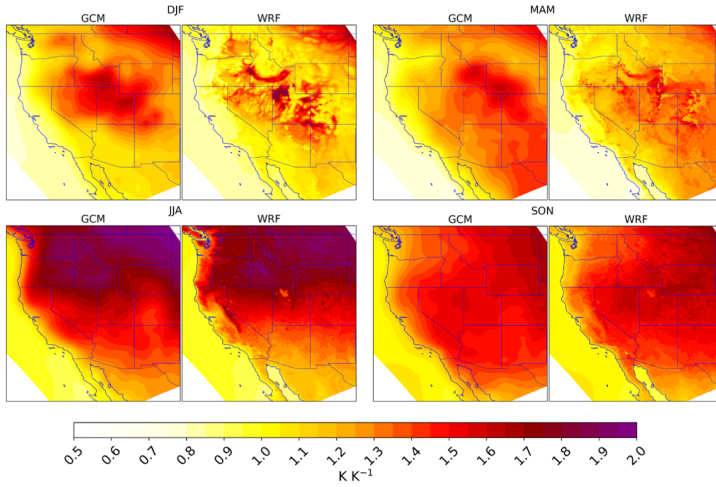



Figure 6. Future climate response for parent GCMs (indicated by lettering) and their downscaled counterparts (indicated by open circles) on the 9-km WRF grid averaged across 11 western U.S. states. Non-colored circles are for SSP3-7.0 projections only, while blue (red) circles represent the SSP2-4.5 (SSP5-8.5) projections. Arrows point away from parent GCMs towards downscaled counterparts.

1070

Surface air temperature

Precipitation

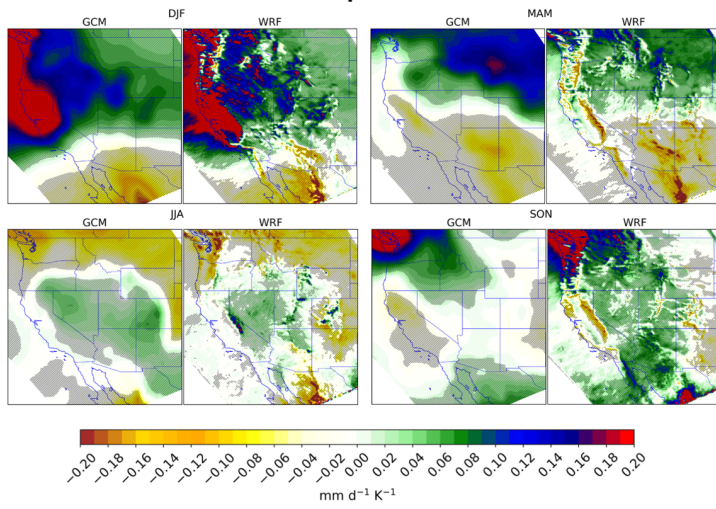



Figure 7. Ensemble-mean future changes in (top) seasonal surface air temperature [K K^{-1}] and (bottom) precipitation per degree of global warming [mm d^{-1} K^{-1}] from 16 downscaled GCMs. Hatching indicates statistical significance to the 95% confidence interval when grid point distributions are subjected to a two-sided Student's t-test. Stippling is not included for temperature because every grid point returns a p value smaller than 0.05.

Fractional precipitation change from WRF

1080

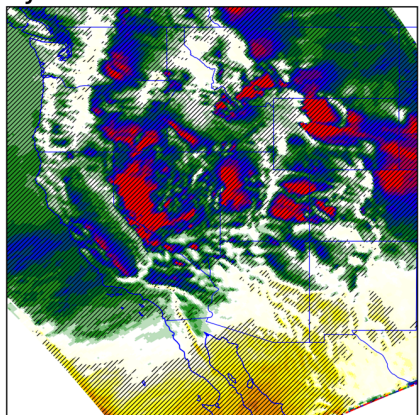
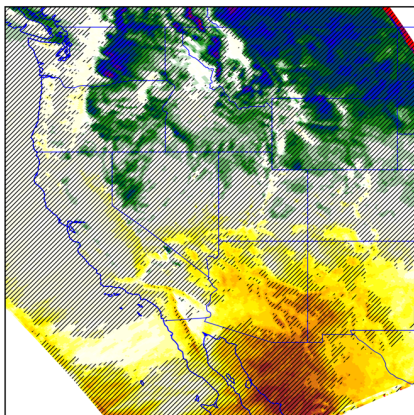
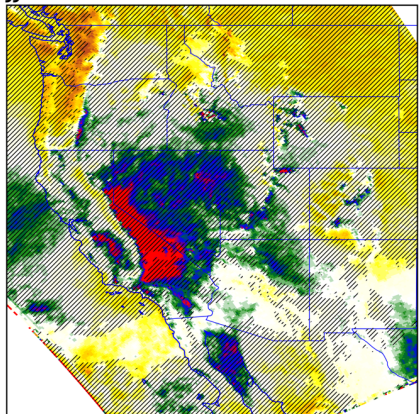
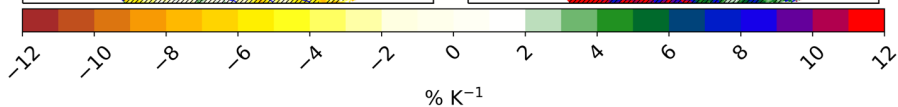
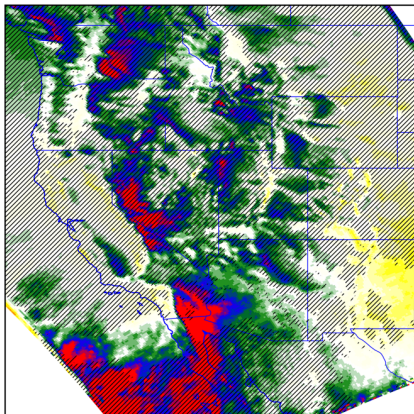



Figure 8. 16-GCM-mean future changes in dynamically downscaled fractional precipitation normalized by the amount of global warming [% K^{-1}]. Hatching indicates statistical significance to the 95% confidence interval when grid point distributions are subjected to a two-sided Student's t-test. Stippling is not included for temperature because every grid point returns a p value smaller than 0.05.

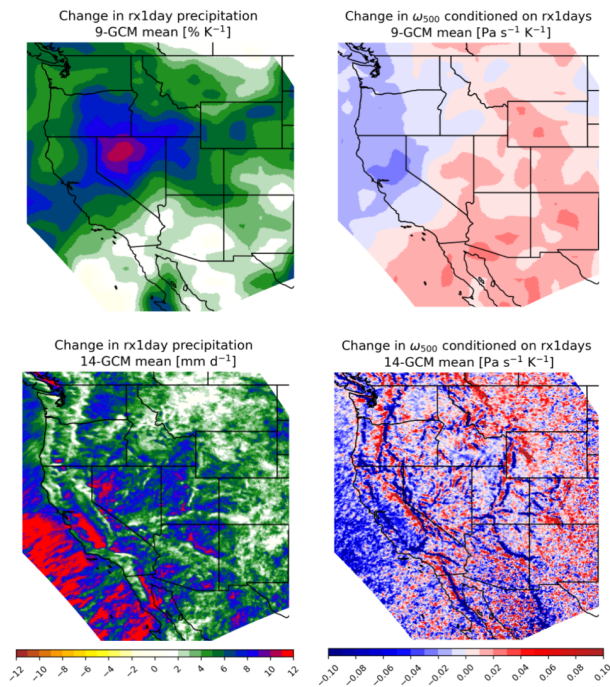



Figure 9. Future response in (left) rx1day precipitation [% K^{-1}] and (right) 500 hPa pressure velocity [Pa s^{-1} K^{-1}] conditioned on rx1day occurrence for the (top) GCM and (bottom) WRF ensembles per degree of global warming [% K^{-1}]. Due to the lack of GCM data with daily vertical velocity outputs, we use a 14-GCM (9-GCM) mean for rx1day (pressure velocity); the WRF patterns of rx1day generally are similar for a 9-GCM mean (Figure S11).

1090

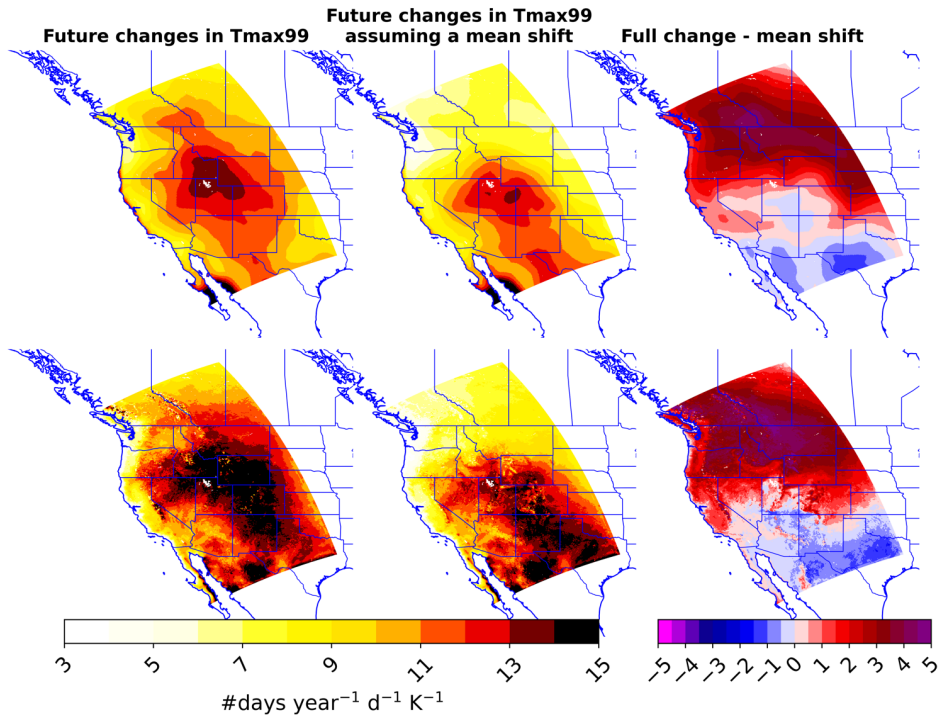



Figure 10. Future changes in the 99th percentile of surface daily maximum air temperature (Tmax99) per degree of global warming [days year^{-1} K^{-1}] considering the full change (left) and assuming a mean shift in the temperature distribution (middle) for (top) GCMs and (bottom) WRF GCMs. The right panel presents the difference between the left and center panels. Parent GCM (WRF GCM) calculations utilize a 11-member (16-member) ensemble. When using the same 11-member ensemble, the WRF panels look similar (Figure S13).

1100