

Environmental Research Communications



PAPER

OPEN ACCESS

RECEIVED
27 May 2022

REVISED
17 October 2022

ACCEPTED FOR PUBLICATION
17 November 2022

PUBLISHED
25 November 2022

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



A comparison of multiple statistically downscaled climate change datasets for the conterminous USA

John B Kim¹ , Yueyang Jiang² , Linnia R Hawkins³ and Christopher J Still³

¹ Western Wildland Environmental Threat Assessment Center, USDA Forest Service, Corvallis, OR, United States of America

² Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, United States of America

³ Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR, United States of America

E-mail: john.kim@usda.gov

Keywords: global climate models, statistical downscaling, climate change, cmip5, climate change impacts, representative concentration pathways

Supplementary material for this article is available [online](#)

Abstract

Climate change projections provided by global climate models (GCM) are generally too coarse for local and regional applications. Local and regional climate change impact studies therefore use downscaled datasets. While there are studies that evaluate downscaling methodologies, there is no study comparing the downscaled datasets that are actually distributed and used in climate change impact studies, and there is no guidance for selecting a published downscaled dataset. We compare five widely used statistically downscaled climate change projection datasets that cover the conterminous USA (CONUS): ClimateNA, LOCA, MACAv2-LIVNEH, MACAv2-METDATA, and NEX-DCP30. All of the datasets are derived from CMIP5 GCMs and are publicly distributed. The five datasets generally have good agreement across CONUS for Representative Concentration Pathways (RCP) 4.5 and 8.5, although the agreement among the datasets vary greatly depending on the GCM, and there are many localized areas of sharp disagreements. Areas of higher dataset disagreement emerge over time, and their importance relative to differences among GCMs is comparable between RCP4.5 and RCP8.5. Dataset disagreement displays distinct regional patterns, with greater disagreement in ΔT_{max} and ΔT_{min} in the interior West and in the North, and disagreement in ΔP in California and the Southeast. LOCA and ClimateNA are often the outlier dataset, while the seasonal timing of ClimateNA is somewhat shifted from the others. To easily identify regional study areas with high disagreement, we generated maps of dataset disagreement aggregated to states, ecoregions, watersheds, and forests. Climate change assessment studies can use the maps to evaluate and select one or more downscaled datasets for their study area.

1. Introduction

Climate change assessment projects need guidance for selecting climate datasets (Vano *et al* 2015). Coupled Model Intercomparison Project (CMIP) publishes climate projections, generated by an ensemble of global climate models (GCM) (Meehl *et al* 2000). The native spatial resolutions of the GCMs, ranging from 1 to 3°, are generally too coarse for local and regional applications, and do not capture finer scale weather patterns and processes. Therefore, climate scientists have downscaled GCM outputs to finer spatial scales, ranging from 30' to 0.44°. Statistical downscaling methods establish a statistical model between coarse scale GCM outputs and fine scale observation-based weather grids and apply the relationship to GCMs future projections to estimate fine scale projections (Jakob Themeßl *et al* 2011, Takayabu *et al* 2016). The downscaled datasets better reflect fine scale variations in observed surface meteorology. Currently various efforts to downscale CMIP Phase 6 (Eyring *et al* 2016) data are underway, and the most widely available downscaled datasets are based on CMIP Phase 5 (CMIP5).

Many studies have compared and evaluated different downscaling methodologies to improve our understanding of downscaling methods and their consequences. These studies use experiments designed to focus on differences in downscaling methods while minimizing differences in other factors. For example, they hold the GCM, spatio-temporal resolution, and training data constant (Dixon *et al* 2016, Yang *et al* 2019, Lanzante *et al* 2020). While these types of studies are critical for climate scientists to advance downscaling science, they are less relevant for climate change impact modelers and other end-users of the data, who need pragmatic information on how published downscaled datasets differ. Statistically downscaled datasets currently available to the scientific community have not produced in a coordinated and controlled manner. They may differ from one another not only due to the downscaling method used, but also due to spatio-temporal resolution, training data, computer coding and data storage (table 1). A comparison of downscaled datasets focusing on Northwest United States demonstrates that the largest source of variability is the choice of training data (Jiang *et al* 2018).

Heretofore no study has characterized the differences among the published downscaled climate datasets for the conterminous USA (CONUS). Such a comparison has the potential to inform the choice of datasets for use in climate change impact modeling and assessments. Perhaps more importantly, where studies already occurred a comparison study has the potential to provide context for interpreting the use of particular datasets.

In this study, we focus on five statistically downscaled climate change datasets that cover CONUS that have been published and have been widely used in climate change impact studies: ClimateNA (Wang *et al* 2006, Wang *et al* 2012, Hamann *et al* 2013, Wang *et al* 2016), LOCA (Pierce *et al* 2014), MACAv2- LIVNEH and MACAv2-METDATA (Abatzoglou and Brown 2012, Abatzoglou 2013), and NEX-DCP30 (Thrasher *et al* 2013). All of these datasets are derived from CMIP5 GCM output and have been applied to a wide array of climate change impact studies spanning multiple disciplines. For example, ClimateNA was used to create grassland bird habitat projections in the Northeast (McCauley *et al* 2017) and to develop water quality projections for regional watersheds in New York state (Gelda *et al* 2019). LOCA has the distinction of having been selected for the Fourth National Climate Assessment (NCA4) (USGCRP 2018). MACAv2-LIVNEH was used to model growth of loblolly pine (*Pinus taeda*) in the Southeast (Gonzalez-Benecke *et al* 2017), and MACAv2-METDATA has been selected by the US Forest Service to assess climate change impacts on forests and rangelands across the U.S. (Joyce and Coulson 2020). NEX-DCP30 was used to project habitat for whitebark pine (*Pinus albicaulis*) in the Greater Yellowstone Area, (Chang *et al* 2014, Buotte *et al* 2016) and to project American pika (*Ochotona princeps*) habitat in eight national parks in the West (Schwalm *et al* 2016). Many studies—if not most—use a single downscaled climate dataset as a basis for future projections. It is unclear how using only one downscaled dataset affects a study's conclusions on climate change effects on the ecosystem being studied.

In this paper, we characterize the similarities and differences among the five selected datasets, to serve as practical guidance for dataset end-users. This work expands the analysis performed by Jiang *et al* (2018) for the Pacific Northwest to CONUS. Our objectives are to identify where, when and by how much the datasets vary. We aggregate dataset variability by potential study areas—states, watersheds, ecoregions, and national forests—to serve as a visual aid and identify outlier datasets for multiple GCMs. The resulting maps can be used to determine whether more than one dataset may be needed and to help select datasets for a given study area in a climate change impacts study.

2. Methods

2.1. Study area

We compare the downscaled datasets for the conterminous USA (CONUS), comprising 8,080,464 km² of land surface from 25.15° to 49.4625° latitude and −124.59° to −67.9° longitude. Broadly speaking, the northern two-thirds of CONUS has temperate climate and the southern third has subtropical climate. The western third of CONUS contains many high elevation mountain ranges and vast, arid basins. The Northwest and California have Mediterranean climate, where precipitation falls predominantly in the winter, and summers are warm and dry. The Northwest receives copious rainfall, with many coastal areas receiving more than 2 m annually. The Southwest also has warm summers but receives rain via summer monsoon. The Great Plains, which are predominantly grasslands and croplands, occupy the center of CONUS, stretching from the northern border to Canada down to the southern border to Mexico. Winters are cold and snowy in the northern half of the Great Plains, and more moderate in the Southern Plains. Precipitation is highest in the summertime, when collision of air masses from the north and the south over low topographic relief results in frequent thunderstorms. Most of the eastern CONUS has humid, continental climate, where precipitation is spread throughout the seasons. The Northeast has cold, snowy winters and the coastal southern states have a humid, subtropical climate, where winters are mild and summers are hot and humid.

Table 1. Statistically downscaled climate datasets selected for this study. Details of each dataset are given in the publication listed for each dataset.

Dataset	Spatial extent & resolution	Time periods & resolution	Training data	Downscaling method
ClimateNA (Wang <i>et al</i> 2016)	North America @ 30'	1961-1990, 1981-2010, 2011-2040, 2041-2070, 2071-2100 monthly normals	PRISM, ANUSLIN	Delta
LOCA (Pierce <i>et al</i> 2014)	CONUS, S. Canada & N. Mexico @ 1/16°	1950-2100 @ daily	LIVNEH	Localized Constructed Analogs
MACAv2-LIVNEH (Abatzoglou & Brown 2012)	CONUS @ 1/16°	1950-2099 @ daily & monthly	LIVNEH	Multivariate Adaptive Constructed Analogs
MACAv2-METDATA (Abatzoglou & Brown 2012)	CONUS @ 1/24°	1950-2099 @ daily & monthly	METDATA	Multivariate Adaptive Constructed Analogs
NEX-DCP30 (Thrasher <i>et al</i> 2013)	CONUS @ 30'	1950-2100 @ monthly	PRISM	Bias Corrected Spatial Disaggregation

2.2. Downscaled climate datasets

We focused on statistically downscaled climate datasets for CONUS that have been widely cited and are freely available to download through public portals without requiring further computations or processing. We selected the climate datasets that are derived from a common set of GCMs published by Coupled Model Intercomparison Project Phase 5 (CMIP5), simulating Representative Concentration Pathway (RCP) 4.5 and 8.5 climate change scenarios (Van Vuuren *et al* 2011). The selected datasets are ClimateNA, LOCA, MACAv2-LIVNEH, MACAv2-METDATA and NEX-DCP30 (table 1). Except for LOCA, the datasets were previously compared for Oregon and Washington states (Jiang *et al* 2018). The selected datasets represent climate at the land surface at relatively fine resolution generally applicable for local and regional scale studies, ranging from 30'' to 1/16° grid resolution. The datasets, the associated methodology, their characteristics, and their evaluations for the historical period are described in detail by the dataset authors (Abatzoglou & Brown 2012, Thrasher *et al* 2013, Pierce *et al* 2014, Wang *et al* 2016). Here, we briefly summarize the key features of the datasets.

ClimateNA (Wang *et al* 2016) uses the Delta method (Mitchell and Jones 2005), the simplest downscaling technique, to downscale GCM output to 30'' resolution at a monthly time step, using PRISM gridded climate data (Daly *et al* 2002) and ANUSLIN (Hutchinson 1989) as reference data. It has several appealing features for modeling studies: a large spatial extent (North America); a large suite of derived bioclimatic variables; and it is available as software, to downscale data on demand. No bias correction is applied to the coarse resolution GCM output before spatial disaggregation.

NEX-DCP30 (Thrasher *et al* 2013) uses the bias correction-spatial disaggregation (BCSD) downscaling method (Maurer *et al* 2010). GCM output is first bias corrected at the coarse GCM resolution to match the reference data using stationary quantile mapping before it is spatially disaggregated to 30'' resolution using PRISM as reference data.

MACAv2-METDATA and MACAv2-LIVNEH are both produced with the multivariate adaptive constructed analogs (MACA) downscaling method. GCM output is first bias corrected using non-stationary quantile mapping. GCM output is spatially disaggregated by identifying 100 best coarse scale analogs from the reference dataset, then constructing a single analog from the corresponding fine scale patterns using a weighted mean. They are created using the METDATA (Abatzoglou 2013) and the LIVNEH gridded climate dataset (Livneh *et al* 2013), respectively. The datasets span CONUS at 1/24° and 1/16° resolution, respectively, at daily time steps.

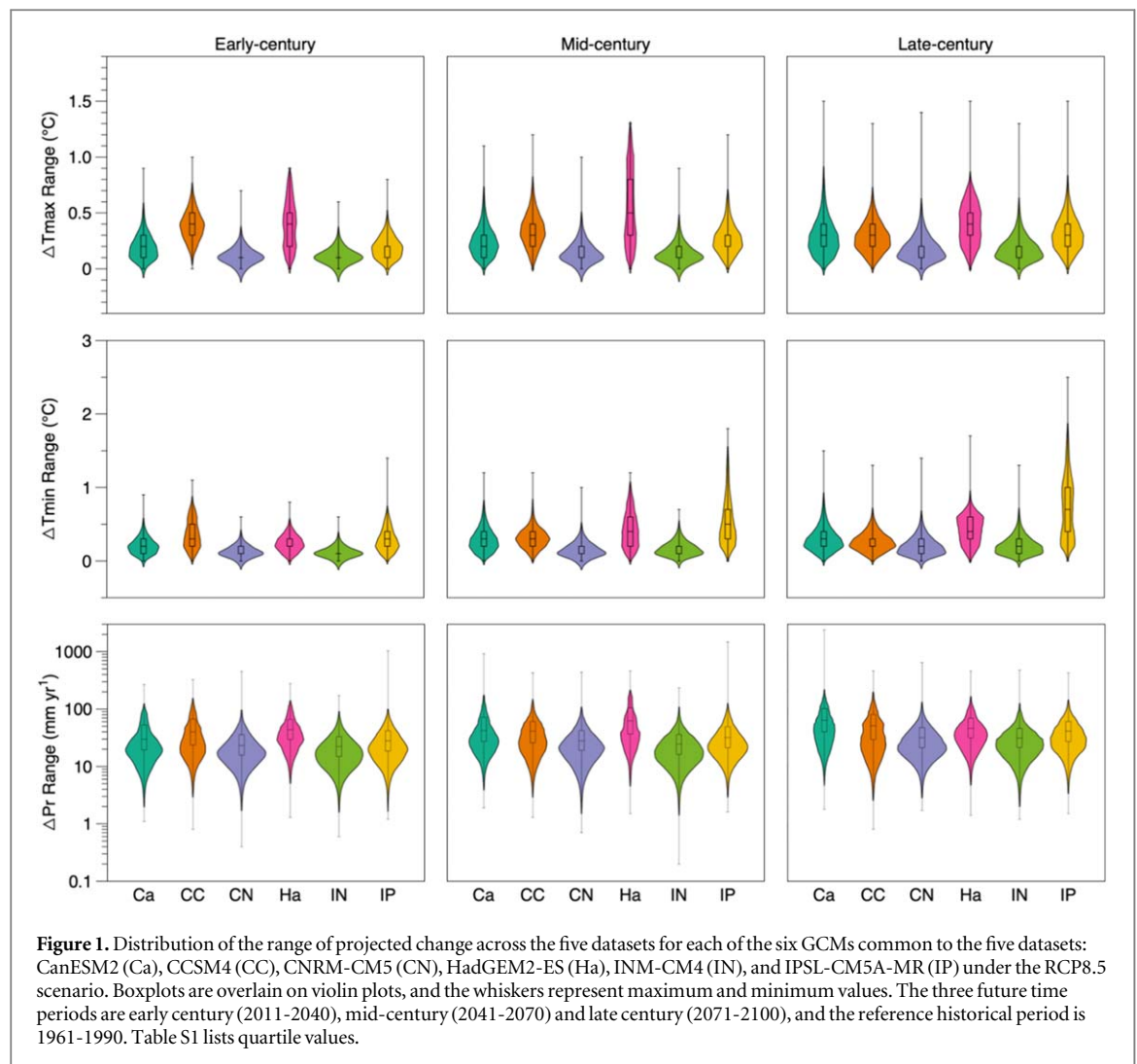
LOCA (Pierce *et al* 2014) is produced with a downscaling method similar to MACA, where GCM output is first bias corrected using frequency-dependent method, then spatially disaggregated by identifying analogs in the reference dataset. While MACA considers the entire CONUS when search for analogs and combines the 100 best analogs, LOCA breaks up the domain into smaller areas, and selects only a single analog. LOCA downscales to 1/16° resolution on a daily time step using the LIVNEH gridded climate dataset (Livneh *et al* 2013) as reference data, and extends from southern Canada to northern Mexico at 1/16° resolution.

2.3. Comparisons

RCP4.5 and RCP8.5 climate change scenarios are common to the five selected datasets. While each dataset includes between 7 and 33 GCMs, only six GCMs are common to the selected datasets: CanESM2, CCSM4, CNRM-CM5, HadGEM2-ES, INMCM4, and IPSL-CM5A-Mr. For the two scenarios and six GCMs common to the selected datasets, we compared three climate variables that are available in all five datasets: monthly mean of daily maximum temperature (Tmax), monthly mean of daily minimum temperature (Tmin), and monthly precipitation (Pr). We calculated the projected change (delta) from a reference period (1961-1990) to three future periods: 2011-2040 (early century), 2041-2070 (mid-century), and 2071-2100 (late century). To compare the datasets for spatial differences, we resampled the deltas of all datasets to 1/16° resolution, the lowest resolution among the five datasets. In the resampling procedure, for each 1/16° pixel we computed the weighted average of all original pixels covered by the 1/16° pixel, accounting for partial overlap.

To capture the disagreement among the five datasets, we calculated the range (max—min) of the deltas across the five datasets, for each GCM, scenario and time period. The standard deviation of delta was also considered as a measure of dataset variability, but it is nearly perfectly correlated with range for the three variables ($r^2 = 98.4$ to 99.6). For each climate variable and time period, we averaged the range across the six common GCMs to obtain the 'inter-dataset range'. For comparison, we obtained the 'inter-GCM range' by calculating the range across GCMs within each dataset, then averaging over the datasets.

To examine temporal patterns, we divided CONUS into the seven regions used by NCA4 (USGCRP 2018), and aggregated the original, un-interpolated delta values by month and region, resulting in thirty time series (six GCMs x five datasets) for each time period. To average values to each region, we weighted each pixel by its area. As with above, we calculated the inter-dataset range by first calculating the range over datasets per GCM, then



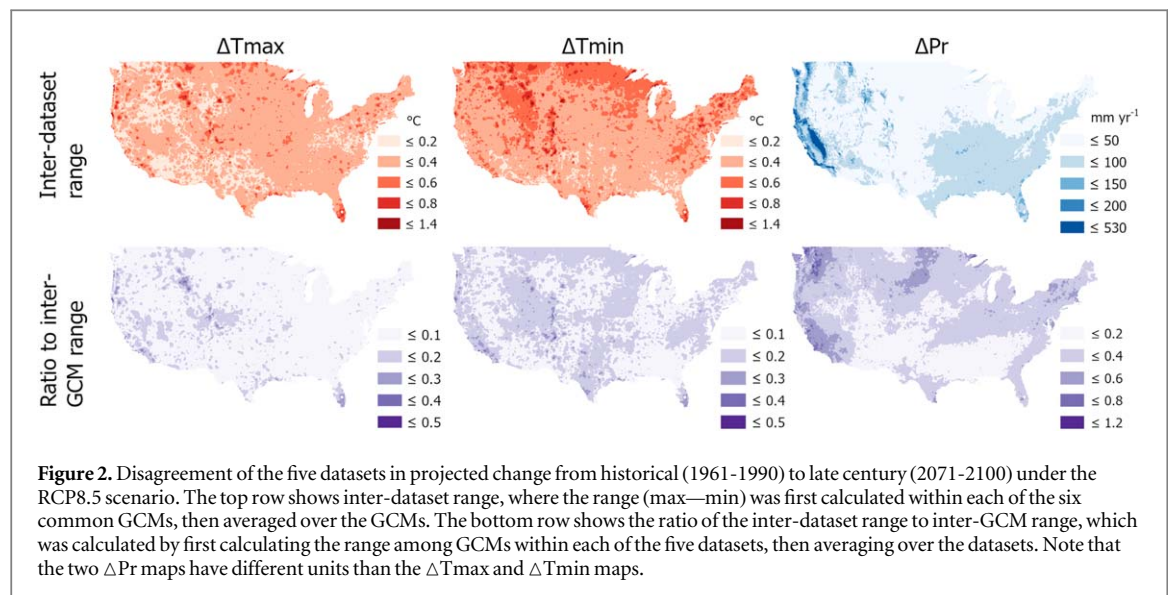
averaging over the GCMs. We calculated inter-GCM range by first calculating the range over GCMs within each dataset, then averaging over the datasets.

Many regional climate change studies begin with a predetermined study area, before downscaled climate datasets are sought. Also, there are many studies that have been completed for a study area, that may benefit from having dataset variability as context. To render our analysis easily comparable and visually accessible to existing study areas, we aggregate variability by several polygon schemes: states, EPA Level III Ecoregions (Omernik and Griffith 2014), USGS 4-digit subbasins (Seaber *et al* 1987), and USDA national forest administrative units (Anon 2022). We calculated the range of deltas across the datasets per GCM, before averaging across the six common GCMs and then finally aggregating for each polygon, where we weighted each pixel by its area contribution to the polygon. In addition, for each GCM we created a map of outlier datasets, where if a dataset's delta value in a pixel is 1.65 standard deviations or more away from the mean of the five datasets for that pixel it is considered an outlier.

3. Results

3.1. Continental & centennial scale patterns

The five datasets exhibit general agreement across CONUS for RCP8.5, with the distributions of inter-dataset range for $\Delta T_{\max}$ and $\Delta T_{\min}$ falling primarily between 0 and 0.5 °C for all but one GCM, and the distributions for ΔPr range falling primarily below 100 mm yr⁻¹ during all three time periods (figure 1). However, there are clear differences in the distributions among the GCMs, where some GCMs have sizable distributions of inter-dataset range values above 0.5 °C. For example, in mid-century the inter-dataset range values for HadGEM2-ES are distributed nearly evenly up to 1.0 °C. The inter-dataset range values for $\Delta T_{\min}$ for IPSL-CM5A-MR in the late century exhibits a similar pattern. While there is little increase in the medians of the distributions for $\Delta T_{\max}$ and $\Delta T_{\min}$ as time progresses from early to mid- to late century, all distributions extend upward, as evidenced



by the increases in both the third quartile values and the maximum in most GCMs (Table S1). With $\Delta Tmax$ and $\Delta Tmin$, all GCMs have higher maximum values by the late century than the early century, indicating that many parts of CONUS have far larger dataset disagreement by the late century than early.

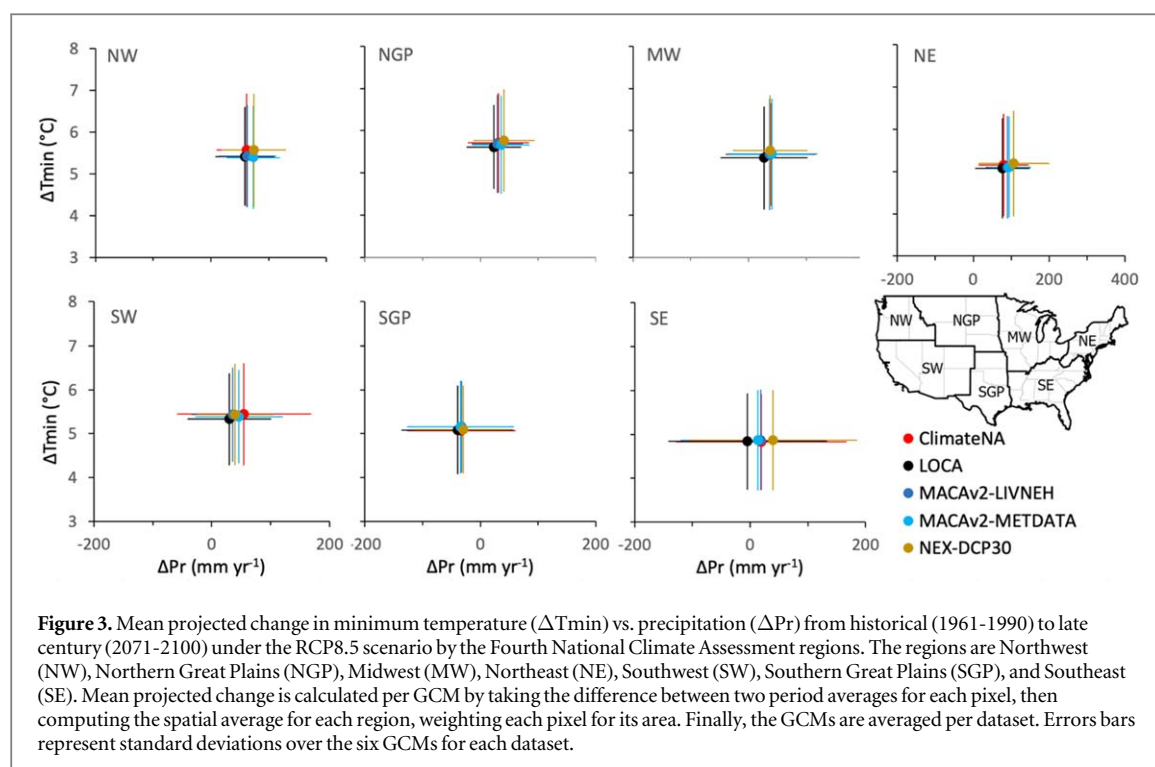
For ΔPr , the distributions of the inter-dataset range values skew heavily toward 10–50 mm yr^{−1} range (note the logarithm scale for ΔPr range in figure 1), and there is less contrast in the shape of the distributions among the GCMs. As with $\Delta Tmax$ and $\Delta Tmin$, the medians of the distributions do not shift sharply as time progresses. The first three quartiles of range values in all six GCMs remain below 103 mm yr^{−1} in all three time periods. Unlike $\Delta Tmax$ and $\Delta Tmin$, however, the maximums of the distributions do not all increase uniformly from early to late century. IPSL-CM5A-MR has the highest inter-dataset differences in ΔPr during the early and mid-century, of 1,031 and 1,480 mm yr^{−1} respectively, but by the late century its maximum is reduced to 426 mm yr^{−1}. CanESM2 has moderate maximum range of 265 mm yr^{−1} in early century but in late century its maximum range is extremely high, exceeding 2,385 mm yr^{−1} (Table S1).

The areas of dataset disagreement in temperature and precipitation appear spatially disjunct (figure 2). For $\Delta Tmax$ and $\Delta Tmin$ the areas of high dataset differences are somewhat scattered across CONUS, with localized high values the interior west and the northern latitudes, and with $\Delta Tmin$ exhibiting generally higher disagreement. For ΔPr the areas of high disagreement are located along the Westcoast, mountains in the interior West, and the Southeast. Pixels with the highest dataset disagreement in ΔPr are located in the Sierra Nevada Mountains in California, where the inter-dataset differences are between 200 and 530 mm yr^{−1}. One GCM is a dominant contributor to these high values: the inter-dataset range for CanESM2 exceeds 2,500 mm yr^{−1} in the Sierra Nevada Mountains. At the continental scale, linear regression reveals no significant relationship between the inter-dataset range and elevation.

While the inter-dataset range expressed as mm yr^{−1} may appear modest across much of CONUS, they can be significant when compared to the differences among the GCMs (figure 2). The inter-dataset range for $\Delta Tmax$ is generally less than 10% of the inter-GCM range for most of CONUS by late century under the RCP8.5 scenario. For $\Delta Tmin$, the inter-dataset range is 10%–20% of inter-GCM range for approximately half of CONUS. This ratio is the highest for ΔPr , where parts of the Northwest, northern Midwest, and the Southwest have inter-dataset range that is as much as 60%–120% of inter-GCM differences. Under the RCP4.5 scenario, the inter-dataset range is generally lower, but ratio to inter-GCM differences are generally comparable, with some locations having higher ratios (figure. S1). Relative to projected amount of warming, the dataset differences for $\Delta Tmax$ and $\Delta Tmin$ are small across most of CONUS, except limited areas associated with mountain ranges and the southern tip of Florida (figure S2). For precipitation, the disagreements are large as a percent of projected change across much of CONUS, but it is due to only small changes being projected for most of CONUS except for the Southwest (figure S2).

3.2. Variability by region

Dataset disagreements exhibit regional differences when averaged for the seven NCA4 regions (figure 3). The range of $\Delta Tmin$ is higher in the northern regions (0.15 °C–0.19 °C) relative to the Southeast and the Southern Great Plains (0.04 and 0.09 °C, respectively). The range of $\Delta Tmax$ is small, less than 0.09 °C for all regions. For ΔPr , the smallest range occurs in the Southern Great Plains and the Midwest (10 mm yr^{−1}), and the highest in



the Southeast (44 mm yr^{-1}). Relative to total precipitation, the Southwest has the highest range (25 mm yr^{-1}). This disagreement is geographically located primarily in California, which has a strong Mediterranean climate with dry summers (figure 2). In the interior Southwest, where there is significant summer monsoon precipitation from convective storms, the disagreement is limited to the high elevation subregion in northwestern Arizona (figure 2). LOCA's ΔPr is the lowest in all the regions, and in Southern Great Plains and the Southeast LOCA projects the greatest reduction in precipitation of all the datasets. LOCA's ΔT_{min} is the lowest in six of the seven regions. NEX-DCP30 has the largest or the second largest average ΔT_{min} in six regions. ClimateNA, MACAv2-LIVNEH and MACAv2-METDATA occupy the middle positions. MACAv2-LIVNEH and MACAv2-METDATA are similar, while MACAv2-LIVNEH and LOCA are dissimilar despite having a common reference dataset.

3.3. Variability by months

Monthly comparisons show ΔT_{min} of datasets tracking each other closely (solid lines, figure 4). ClimateNA is an exception, which projects greater warming in the spring and less warming in the winter than the others. ClimateNA also exhibits a relative phase shift, with more warming in early summer and less warming in late summer. ΔPr of datasets also track each other closely for most of the year (dotted lines, figure 4). Some divergence is observed in the winter in the Northwest when precipitation is the heaviest. In the Southwest, portions of which include significant monsoon precipitation, datasets diverge moderately in the summer. In the remaining regions, where convective systems produce rain, divergence occurs in summer and fall. ClimateNA's ΔPr diverges from the others only in the Southwest and the Southern Great Plains. The seasonal differences in datasets, when quantified as a ratio to the differences among the GCMs, range from 0.09 to 0.35 for ΔT_{min} and 0.14 to 0.43 for ΔPr (figure S3). The highest ratios for ΔT_{min} occur in the winter in the Northwest and in the spring in the Midwest when the inter-dataset differences can be as much as 35% of the differences among the GCMs. The highest ratios for ΔPr occur in the Northwest, the Northern Great Plains, and the Southwest, where multiple seasons have inter-dataset range exceeding 30% of inter-GCM range.

3.4. Guidance summaries and outliers

Area averaging smooths away high disagreement pixels but aggregates results into familiar polygons (figure 5). Dataset disagreements in some polygons are as high as 0.5°C for ΔT_{max} or ΔT_{min} , or over 100 mm yr^{-1} in ΔPr for late century under the RCP8.5 scenario. No state has ΔT_{max} or ΔT_{min} range greater than 0.3°C , except for Idaho and Wyoming. Only California and southeastern states have high ΔPr ranges. The subregions of high variability seen in figure 2 are visible in ecoregion and watershed averages, such as those containing the Rocky Mountains. Some areas with high disagreement are evident in one set of polygons but in not another. For example, the Willamette Valley ecoregion, Oregon, has moderately high ΔT_{max} range, but the enclosing

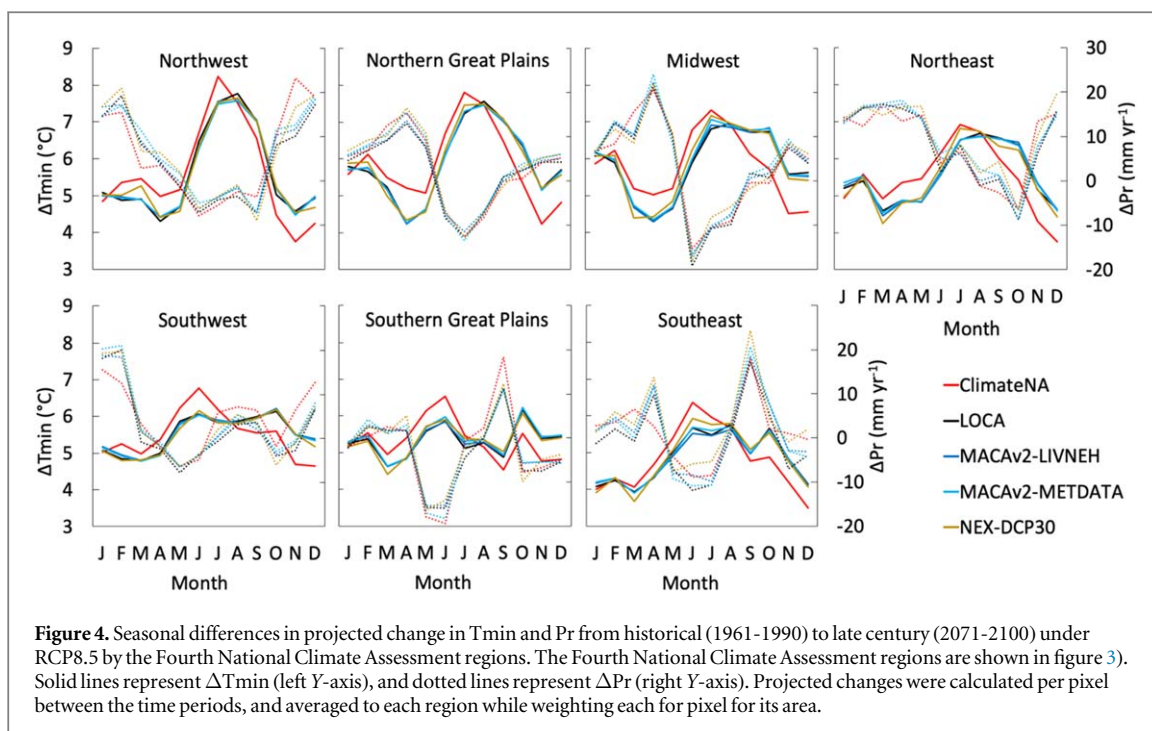


Figure 4. Seasonal differences in projected change in T_{min} and Pr from historical (1961-1990) to late century (2071-2100) under RCP8.5 by the Fourth National Climate Assessment regions. The Fourth National Climate Assessment regions are shown in figure 3). Solid lines represent ΔT_{min} (left Y-axis), and dotted lines represent ΔPr (right Y-axis). Projected changes were calculated per pixel between the time periods, and averaged to each region while weighting each for pixel for its area.

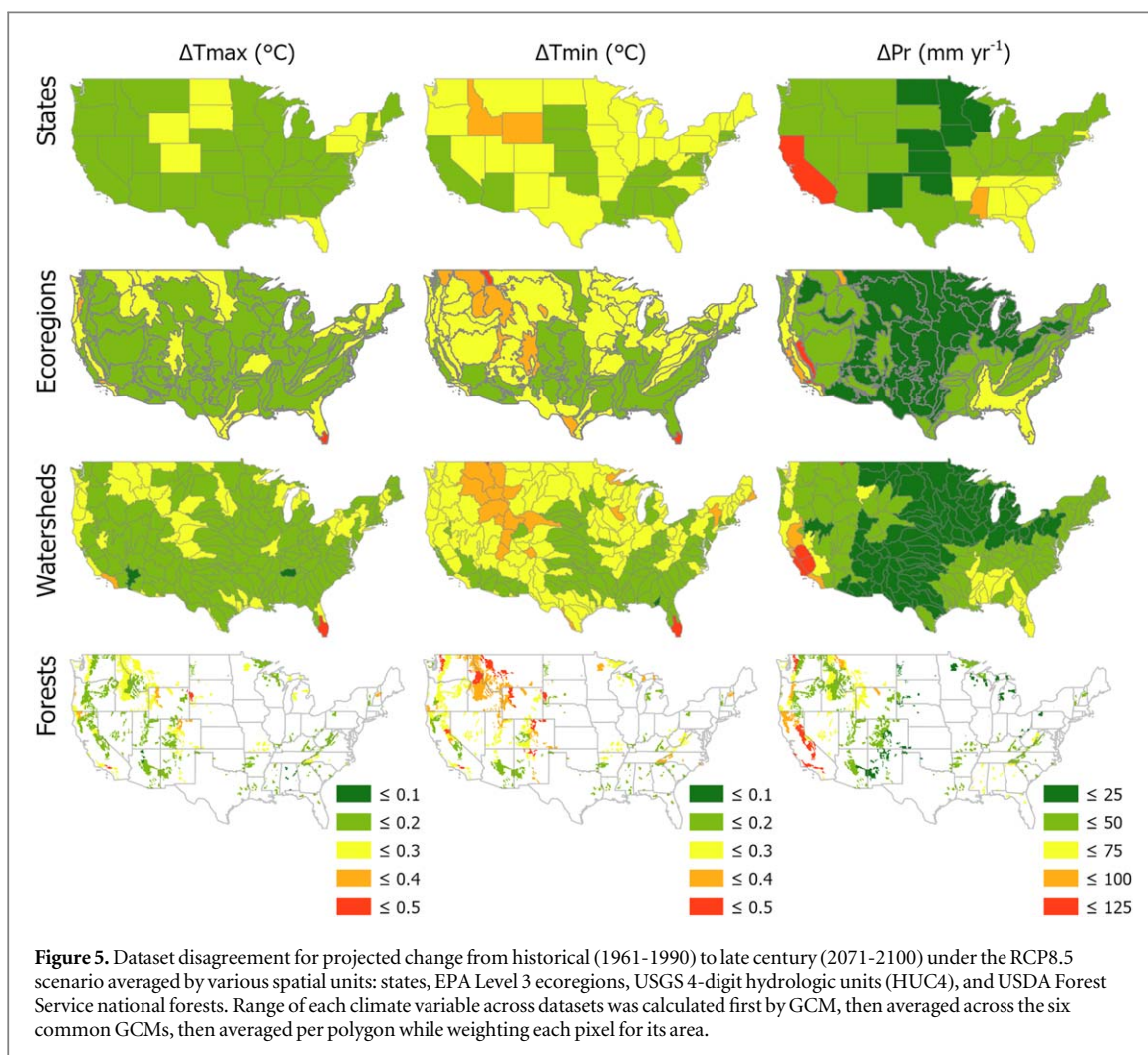
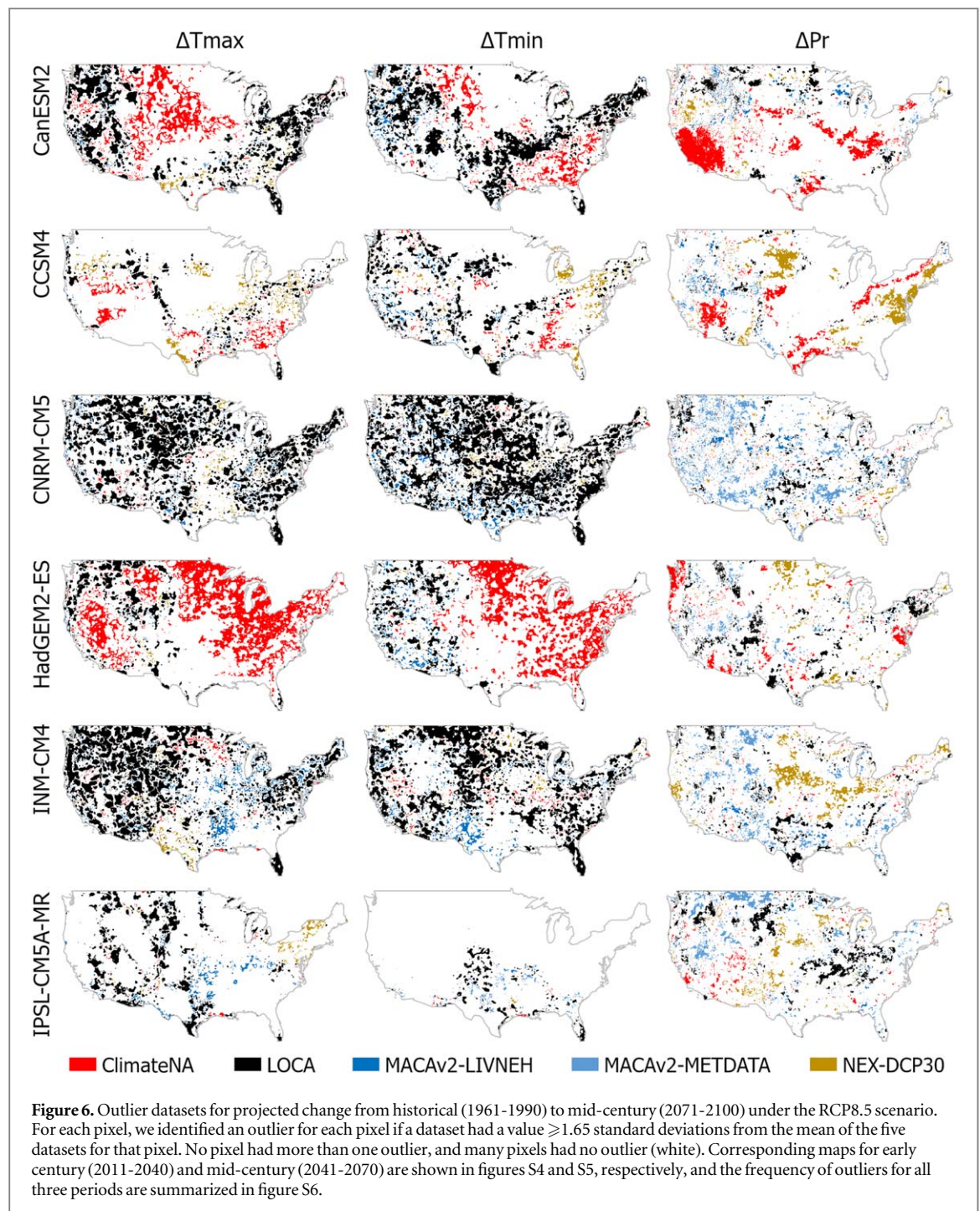


Figure 5. Dataset disagreement for projected change from historical (1961-1990) to late century (2071-2100) under the RCP8.5 scenario averaged by various spatial units: states, EPA Level 3 ecoregions, USGS 4-digit hydrologic units (HUC4), and USDA Forest Service national forests. Range of each climate variable across datasets was calculated first by GCM, then averaged across the six common GCMs, then averaged per polygon while weighting each pixel for its area.

watershed has lower range. The Upper Hudson watershed, New York, has moderately high ΔT_{min} range, but the enclosing ecoregion has lower range. National forest polygons do not fully cover CONUS and therefore exclude some high disagreement areas, such as southern Florida.



The location, frequency and the identity of outliers varied markedly and primarily by GCM and climate variable (figure 6). For example, with CNRM-CM5, LOCA was an outlier across much of CONUS for $\Delta T_{\max}$ and $\Delta T_{\min}$ (34% and 40% of all pixels, respectively), but only a small portion (5%) of the pixels for ΔPr . Instead, MACAv2-METDATA was the most frequent outlier (7%) for ΔPr . For HadGEM2-ES, ClimateNA was the most frequent outlier for much of eastern CONUS for $\Delta T_{\max}$ and $\Delta T_{\min}$ (24% and 18% of CONUS, respectively), but for ΔPr it was an infrequent outlier (4%). No pixel had more than one outlier, and some GCMs had relatively few outliers, indicating good dataset agreement.

The importance of GCM and climate variable held throughout time: the maps of outliers for early and mid-century resemble the late century maps in terms of overall location and identities of the outliers, although the frequencies are different (figures S4, S5). In other words, for a given GCM, a dominant outlier dataset remained dominant throughout all three time periods (figure S6). While no single dataset was the dominant outlier for all GCMs, LOCA was the most frequent outlier among the six GCMs, on average covering 18% of CONUS as outliers for $\Delta T_{\max}$ and $\Delta T_{\min}$. That was followed by ClimateNA, which on average covered 6% and 4% of CONUS for $\Delta T_{\max}$ and $\Delta T_{\min}$, respectively. The two MACA datasets and NEX-DCP30 never exceeded 3%

coverage of CONUS as $\Delta T_{\max}$ and $\Delta T_{\min}$ outliers for a single GCM, and on average cover only 0 to 2% of CONUS. The most frequent outlier for ΔPr in a single GCM is ClimateNA in CanESM2 (11%), followed by MACAv2-METDATA in CNRM-CM5 (7%) and LOCA in IPSL-CM5A-MR (7%).

4. Discussion

For those studying climate change impacts the five downscaled datasets compared herein present a wealth of options. Our comparison results suggest that careful consideration is warranted in choosing a combination of downscaled climate dataset and GCMs. Even by the mid-century, in many parts of CONUS one dataset may differ from others by $\geq 0.5^\circ\text{C}$ in $\Delta T_{\max}$ and $\Delta T_{\min}$, an amount comparable to the difference between RCP4.5 and RCP6.0 in projected global average surface temperature for late century (IPCC 2013). Likewise, in some parts of CONUS ΔPr may differ by over 100 mm yr^{-1} . This is a significant amount in arid regions like Southern California, where mean annual precipitation ranges only $263\text{--}494\text{ mm yr}^{-1}$. For some parts of CONUS, the choice of dataset may be as significant as the choice of GCM (figure 2), and the upward extension of the distribution of differences by late century (figure 1) suggests this is a more important issue for studies modeling late century climate impacts. Dataset differences are accentuated in some seasons (figure 4), which can affect models sensitive to seasonality of climate. The intensity and the seasonality of climate change can affect plant-water dynamics (Knighton *et al* 2019), stream flow (Grantham *et al* 2018), plant and wildlife biogeography (McCauley *et al* 2017, Parker and Abatzoglou 2018, Case *et al* 2020), human health (Diem *et al* 2017), and civil infrastructure (Auffhammer *et al* 2017).

Climate change impact studies typically use a single downscaled dataset. Our comparison suggests that could result in a systematic bias, especially in the Northwest, the Southwest, and the Southeast, where the differences are relatively large (figure 3). This may be particularly true for temperature-dependent impacts in western CONUS, where LOCA is an outlier in many regions, with many GCMs (figure 6). Dataset disagreement is greater for $\Delta T_{\min}$ than $\Delta T_{\max}$ across much of CONUS (figures 2, 5). This suggests caution for studies where $T_{\min}$ is an important driver, such as ecosystem models where $T_{\min}$ governs respiration and nighttime transpiration, or heat wave projections where $T_{\min}$ is important for nighttime recovery (Guirguis *et al* 2018). Where snowpack and phenology are important, choosing ClimateNA over the others may have a significant impact, given ClimateNA's relative bias in the magnitude and timing of seasonal warming and cooling (figure 4). Selecting a large ensemble of GCMs within a downscaled dataset may help capture a wide range of possible future climate and help mitigate any bias arising from the use of a single downscaled dataset. Dynamical downscaling approaches, where regional climate models are run under future climate scenarios is another option. For North America, NA-CORDEX provides dynamically downscaled dataset at 0.22° ($\sim 25\text{ km}$) or 0.44° ($\sim 50\text{ km}$) resolution (Mearns *et al* 2017).

There are many localized clusters of high range pixels that one might call 'hotspots' of dataset disagreement (figure 2). For $\Delta T_{\min}$, many of those hotspots are concentrated in the Rocky Mountains, an area with high topographic relief, complex weather dynamics that present challenges for downscaling climate data. While some hotspots, such as those covering Yellowstone National Park and the Colorado Front Range, appear to be correlated to high elevation and snowpack dominance, there are many high elevation areas that are not hotspots. For example, the Grand Mesa, Uncompahgre, and Gunnison National Forests in Colorado have an average elevation of $2,887\text{ m}$, yet the range of $\Delta T_{\min}$ is less than 0.2°C by the late century (figure 5). For ΔPr , many hotspots also occur in the Rocky Mountains, at Flathead, Bighorn, and Medicine Bow-Routt National Forests. Yellowstone National Park is also a hotspot, with 104 mm yr^{-1} ΔPr range, which is high relative to its 516 mm yr^{-1} precipitation. The largest concentrations of ΔPr hotspots occur on the West Coast, in parts of the Cascade Range in the Northwest, and in the Sierra Nevada mountains and the Coast Ranges of California. High dataset disagreement may have important modeling implications in California, a biodiversity hotspot (Myers *et al* 2000). While many of these pixels occur at high elevations with high average precipitation, several occur at the arid lower elevations. Finally, in the Southeast, curious hotspots occur in a tight band along the East Coast, and across Mississippi and Alabama, areas with little topographic relief.

Outlier datasets identify the source high inter-dataset variability where they occur (figure 6). LOCA is the most frequent outlier (figure S6), which likely arises from its use of relatively small regions for identifying analogs; and its use of a single observed record to downscale, instead of a weighted combination of multiple observations. ClimateNA is the second most frequent outlier, and which may stem from its lack of bias correction. Outliers generally do not spatially coincide with disagreement hotspots, indicating that a single outlier dataset does not often drive high disagreement; instead, hotspots generally occur where multiple datasets disagree. The outlier designation does not imply inaccuracy, only uncertainty. Assessing accuracy of the datasets for the historical period is difficult to measure (Takayabu *et al* 2016), since there is no single pure observation gridded historical weather dataset. The training gridded data used by each dataset (e.g., PRISM, gridMet, and

LIVNEH) are created from point-based weather station data using various complex algorithms and models, and each dataset is bias-corrected to fit the training data (except for ClimateNA). Therefore, assessing accuracy using the training datasets would be circular logic. Innovative and controlled simulation experiments are needed to assess accuracy and diagnose the sources (e.g., Gutiérrez *et al* 2012, Dixon *et al* 2016, Lanzante *et al* 2018, Wang *et al* 2018).

Although diagnosing the causes of the differences of the datasets is not the focus of the current study, the comparison naturally raises questions about why the datasets differ. Jiang *et al* (2018) identified reference datasets as an important driver of differences among downscaled datasets in the Northwest. We see no evidence of this for CONUS, although in theory some are expected (Brands *et al* 2012). MAVAv2-LIVNEH and LOCA are based on the LIVNEH reference dataset and the other three are based on PRISM (or METDATA, which is based on PRISM). We observe no clear differences between these two sets. Instead, LOCA's most frequent outlier status suggests the importance of LOCA's unique algorithm in identifying local analogs of weather patterns. The two MACA datasets are rare outliers, which may reflect MACA's unique multivariate weighting schemes to ensure fields are physically compatible, whereas the other methods downscale each field independently. Yet another important source of differences may be bias correction, which can affect trends, variability, and extremes in downscaled data (Pierce *et al* 2015). ClimateNA does not bias correct GCM data. NEX-DCP30 bias corrects by mapping quantiles of GCM output to the respective quantiles in training data, which assumes variance remains stationary into the future. MACA uses equidistant quantile mapping (Li *et al* 2010) to preserve changes in variability represented by GCM output, but only at one frequency. LOCA uses a frequency dependent bias correction method designed to preserve the projected change in both the mean and extremes. In many regions GCMs project larger changes in extremes relative to the mean (Janssen *et al* 2014, Guirguis *et al* 2018). Therefore, it may be important to preserve changes in variability simulated by GCMs when downscaling. Finally, small amounts of differences may arise from irregularity in choice of GCM ensemble members to downscale. For CMIP5, GCMs were run with variations in parameters, where each run is called a realization and noted with a triad of numbers, e.g., r1i1p1 (Taylor *et al* 2009, Taylor *et al* 2012). NEX-DCP30, MACAv2-METDATA, MACAv2-LIVNEH and LOCA all downscaled the r1i1p1 realization for all GCMs, except the two MACA datasets and LOCA downscaled the r6i1p1 realization for CCSM4. ClimateNA used the average of all published realizations, up to five maximum realizations.

Regional and sub-regional climate change impact studies are often limited by resources, and the study area may be defined *a priori*. When the question of choice of dataset arises, the dataset disagreement described above suggests caution is needed in many parts of CONUS, and the use of multiple datasets may generate more robust results. The inter-dataset range maps (figure 2) and the spatially aggregated maps (figure 5) serve as an entry point for exploring dataset disagreement for a given study area. For completed studies, the maps of variability hotspots (figure 2) and outliers (figure 6), and the relative biases of each dataset by region (figures 3, 4), provide some context for uncertainty arising from the choice of datasets. It is important to note that the results presented herein are based on the six GCMs common to the five datasets, while the published datasets include a larger set of GCMs. ClimateNA has 7, NEX-DCP30 has 33, MACAv2-LIVNEH and MACAv2-METDATA have 20, and LOCA has 32 GCMs. Relationships among the datasets may differ if a larger set of GCMs are compared.

Acknowledgments

Yueyang Jiang was supported by a Joint Venture Agreement between the USDA Forest Service and Oregon State University 19-JV-11261952-100. Linnia Hawkins was supported by the National Science Foundation, Division of Environmental Biology, through the macrosystems biology and NEON-enable science program grant number DEB-1802885. We thank incisive comments and suggestions provided by anonymous reviewers, which significantly improved this work.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

John B Kim  <https://orcid.org/0000-0002-3720-7916>

Yueyang Jiang  <https://orcid.org/0000-0002-5313-1210>

Linnia R Hawkins  <https://orcid.org/0000-0001-6678-4248>

Christopher J Still  <https://orcid.org/0000-0002-8295-4494>

References

- Anon 2022 Administrative Forest Boundaries USDA Forest Service FSGeodata Clearinghouse(<http://data.fs.usda.gov/geodata/edw/datasets.php>)
- Abatzoglou J T 2013 Development of gridded surface meteorological data for ecological applications and modelling *Int. J. Climatol.* **33** 121–31
- Abatzoglou J T and Brown T J 2012 A comparison of statistical downscaling methods suited for wildfire applications *Int. J. Climatol.* **32** 772–80
- Auffhammer M, Baylis P and Hausman C H 2017 Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the united states *Proc. Natl Acad. Sci.* **114** 1886–91
- Brands S, Gutiérrez J M, Herrera S and Cofiño A S 2012 On the use of reanalysis data for downscaling *J. Climate* **25** 2517–26
- Buotte P C, Hicke J A, Preisler H K, Abatzoglou J T, Raffa K F and Logan J A 2016 Climate influences on whitebark pine mortality from mountain pine beetle in the Greater Yellowstone Ecosystem *Ecological Applications* **26** 2507–24
- Case M J, Kim J B and Kerns B K 2020 Using a vegetation model and stakeholder input to assess the climate change vulnerability of tribally important ecosystem services *Forests* **11** 618
- Chang T, Hansen A J and Piekielek N 2014 Patterns and variability of projected bioclimatic habitat for *Pinus albicaulis* in the greater yellowstone area *PLoS One* **9** e111669
- Daly C, Gibson W P, Taylor G H, Johnson G L and Pasteris P 2002 A knowledge-based approach to the statistical mapping of climate *Clim. Res.* **22** 99–113
- Diem J E, Stauber C E and Rothenberg R 2017 Heat in the southeastern United States: characteristics, trends, and potential health impact *PLoS One* **12** e0177937
- Dixon K W, Lanzante J R, Nath M J, Hayhoe K, Stoner A, Radhakrishnan A, Balaji V and Gaitán C F 2016 Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Clim. Change* **135** 395–408
- Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geoscientific Model Development* **9** 1937–58
- Gelda R K, Mukundan R, Owens E M and Abatzoglou J T 2019 A practical approach to developing climate change scenarios for water quality models *J. Hydrometeorology* **20** 1197–211
- Gonzalez-Benecke C A, Teskey R O, Dinon-Aldridge H and Martin T A 2017 *Pinus taeda* forest growth predictions in the 21st century vary with site mean annual temperature and site quality *Global Change Biol.* **23** 4689–705
- Grantham T E W, Carlisle D M, McCabe G J and Howard J K 2018 Sensitivity of streamflow to climate change in California *Clim. Change* **149** 427–41
- Guirguis K, Gershunov A, Cayan D R and Pierce D W 2018 Heat wave probability in the changing climate of the Southwest US *Clim. Dyn.* **50** 3853–64
- Gutiérrez J M, San-Martín D, Brands S, Manzanar R and Herrera S 2013 Reassessing statistical downscaling techniques for their robust application under climate change conditions *J. Climate* **26** 171–88
- Hamann A, Wang T, Spittlehouse D L and Murdock T Q 2013 A comprehensive, high-resolution database of historical and projected climate surfaces for western North America *Bull. Am. Meteorol. Soc.* **94** 1307–9
- Hutchinson M F 1989 A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun. *Technical Memorandum, Division of Water Resources* (89/5) CSIRO Australia 95–104 (<http://hdl.handle.net/102.100.100/261944?index=1>)
- IPCC 2013 Summary for Policymakers. In: *Climate Change 2013 The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed T F Stocker et al (Cambridge: Cambridge University Press)
- Jakob Themeßl M, Gobiet A and Leuprecht A 2011 Empirical-statistical downscaling and error correction of daily precipitation from regional climate models *Int. J. Climatol.* **31** 1530–44
- Janssen E, Wuebbles D J, Kunkel K E, Olsen S C and Goodman A 2014 Observational- and model-based trends and projections of extreme precipitation over the contiguous United States *Earth's Future* **2** 99–113
- Jiang Y, Kim J B, Still C J, Kerns B K, Kline J D and Cunningham P G 2018 Inter-comparison of multiple statistically downscaled climate datasets for the Pacific Northwest, USA *Sci. data* **5** 180016
- Joyce L A and Coulson D 2020 Climate scenarios and projections: a technical document supporting the USDA forest service 2020 RPA assessment *Gen. Tech. Rep. RMRS-GTR-413. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station* 85
- Knighton J, Conneely J and Walter M T 2019 Possible increases in flood frequency due to the loss of Eastern Hemlock in the Northeastern United States: Observational insights and predicted impacts *Water Resour. Res.* **55** 5342–59
- Lanzante J R, Adams-Smith D, Dixon K W, Nath M and Whitlock C E 2020 Evaluation of some distributional downscaling methods as applied to daily maximum temperature with emphasis on extremes *Int. J. Climatol.* **40** 1571–85
- Lanzante J R, Dixon K W, Nath M J, Whitlock C E and Adams-Smith D 2018 Some pitfalls in statistical downscaling of future climate *Bull. Am. Meteorol. Soc.* **99** 791–803
- Li H, Sheffield J and Wood E F 2010 Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change AR4 models using equidistant quantile matching *J. Geophys. Res.* **115** D10101
- Livneh B, Rosenberg E A, Lin C, Nijssen B, Mishra V, Andreadis K M, Maurer E P and Lettenmaier D P 2013 A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions *J. Clim.* **26** 9384–92
- Maurer E P, Hidalgo H G, Das T, Dettinger M D and Cayan D R 2010 The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California *Hydrol. Earth Syst. Sci.* **14** 1125–38
- McCauley L A, Ribic C A, Pomara L Y and Zuckerberg B 2017 The future demographic niche of a declining grassland bird fails to shift poleward in response to climate change *Landscape ecology* **32** 807–21
- Mearns L O et al 2017 The NA-CORDEX dataset, version 1.0. Accessed Oct 1, 2022(<https://doi.org/10.5065/D6SJ1JCH>)
- Meehl G A, Boer G J, Covey C, Latif M and Stouffer R J 2000 The coupled model intercomparison project (CMIP) *Bull. Am. Meteorol. Soc.* **81** 313–8
- Mitchell T D and Jones P D 2005 An improved method of constructing a database of monthly climate observations and associated high-resolution grids *Int. J. Climatol.: A J. Royal Meteorol. Soc.* **25** 693–712
- Myers N, Mittermeier R A, Mittermeier C G, Da Fonseca G A B and Kent J 2000 Biodiversity hotspots for conservation priorities *Nature* **403** 853–8

- Omernik J M and Griffith G E 2014 Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework *Environ Manage* **54** 1249–66
- Parker L E and Abatzoglou J T 2018 Shifts in the thermal niche of almond under climate change *Clim. Change* **147** 211–24
- Pierce D W, Cayan D R, Maurer E P, Abatzoglou J T and Hegewisch K C 2015 Improved bias correction techniques for hydrological simulations of climate change *J. Hydrometeorology* **16** 2421–42
- Pierce D W, Cayan D R and Thrasher B L 2014 Statistical downscaling using localized constructed analogs (LOCA) *J. Hydrometeorology* **15** 2558–85
- Schwalm D, Epps C W, Rodhouse T J, Monahan W B, Castillo J A, Ray C and Jeffress M R 2016 Habitat availability and gene flow influence diverging local population trajectories under scenarios of climate change: a place-based approach *Global Change Biol.* **22** 1572–84
- Seaber P R, Kapinos F P and Knapp G L 1987 Hydrologic Unit Maps, USGS Publications Warehouse (<https://doi.org/10.3133/wsp2294>)
- Takayabu I, Kanamaru H, Dairaku K, Benestad R, von Storch H and Christensen J H 2016 Reconsidering the quality and utility of downscaling *J. Meteorol. Soc. Japan. Ser. II* **94** 31–45
- Taylor K E, Balaji V, Hankin S, Juckes M and Lawrence B 2010 CMIP5 and AR5 data reference syntax (DRS) Version 0.25 (https://pcmdi.llnl.gov/mips/cmip5/docs/cmip5_data_reference_syntax_v0-25_clean.pdf?id=53)
- Taylor K E, Stouffer R J and Meehl G A 2012 An overview of CMIP5 and the experiment design *Bull. Am. Meteorol. Soc.* **93** 485–98
- Thrasher B, Xiong J, Wang W, Melton F, Michaelis A and Nemani R 2013 Downscaled climate projections suitable for resource management *Eos, Trans. Am. Geophys. Union* **94** 321–3
- 2018 *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment* D R Reidmiller et al (ed) (Washington, DC: U.S. Global Change Research Program) II, 1515 USGCRP
- Vano J A, Kim J B, Rupp D E and Mote P W 2015 Selecting climate change scenarios using impact-relevant sensitivities *Geophys. Res. Lett.* **42** 5516–25
- Van Vuuren D P et al 2011 The representative concentration pathways: an overview *Clim. Change* **109** 5
- Wang T, Hamann A, Spittlehouse D and Carroll C 2016 Locally downscaled and spatially customizable climate data for historical and future periods for North America *PLoS One* **11** e0156720
- Wang T, Hamann A, Spittlehouse D L and Aitken S N 2006 Development of scale-free climate data for western Canada for use in resource management *Int. Climatol.: A J. Royal Meteorol. Soc.* **26** 383–97
- Wang T, Hamann A, Spittlehouse D L and Murdock T Q 2012 ClimateWNA—high-resolution spatial climate data for western North America *J. Appl. Meteorol. Climatol.* **51** 16–29
- Wang Y, Sivandran G and Bielicki J M 2018 The stationarity of two statistical downscaling methods for precipitation under different choices of cross-validation periods *Int. J. Climatol.* **38** e330–48
- Yang Y, Tang J, Xiong Z, Wang S and Yuan J 2019 An intercomparison of multiple statistical downscaling methods for daily precipitation and temperature over China: present climate evaluations *Clim. Dyn.* **53** 4629–49