ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Lack of clear standards and usable comparisons of downscaled climate projections pose a roadblock for US climate discovery and adaptation

To cite this article: Samantha H Hartke *et al* 2025 *Environ. Res. Lett.* **20** 054067

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
21 October 2024

REVISED
26 February 2025

ACCEPTED FOR PUBLICATION
31 March 2025

PUBLISHED
29 April 2025

LETTER

Lack of clear standards and usable comparisons of downscaled climate projections pose a roadblock for US climate discovery and adaptation

Samantha H Hartke[1,*], **Andrew J Newman**[1], **Ethan D Gutmann**[1], **Rachel R McCrary**[1], **Nicholas D Lybarger**[1], **Flavio Lehner**[2,3,4], **Andrew W Wood**[2,5], **Jeffrey R Arnold**[6] **and Keith W Dixon**[7]

[1] National Science Foundation (NSF) National Center for Atmospheric Research (NCAR) Research Applications Lab, Boulder, CO, United States of America
[2] National Science Foundation (NSF) National Center for Atmospheric Research (NCAR) Climate & Global Dynamics Lab, Boulder, CO, United States of America
[3] Department of Earth & Atmospheric Sciences, Cornell University, Ithaca, NY, United States of America
[4] Polar Bears International, Bozeman, MT, United States of America
[5] Civil and Environmental Engineering Department, Colorado School of Mines, Golden, CO, United States of America
[6] ERT, LLC, Greenbelt, MD, United States of America
[7] NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: anewman@ucar.edu

Keywords: downscaling, adaptation, projections, community needs, precipitation, resilience

Supplementary material for this article is available online

Abstract

The release of global climate projections coupled with the demand for local-resolution climate-forced meteorology has prompted many research groups to downscale these projections using various statistical, dynamical, and current machine learning techniques. Such downscaled datasets are being used to plan infrastructure and other community needs over the coming decades. Faced with roughly a dozen available US downscaled datasets, many practitioners ask, 'What are the relevant differences between datasets?' This work highlights the difficulty of comparing downscaled datasets and illustrates ways in which datasets differ even when using identical climate model input data. We show that substantial variability in precipitation projections arises from downscaling alone and that the downscaled dataset agreement varies depending on global climate projection. This analysis emphasizes the need for greater coordination and movement toward rigorous benchmarking of downscaling strategies within the downscaling research community, à la the land-modeling community, to better quantify downscaling dataset differences, strengths, and weaknesses for practitioners.

1. Introduction

Downscaled datasets translate coarse global climate model projections to higher spatial (and sometimes also temporal) resolutions that are more directly relevant for planners and decision-makers. There are roughly a dozen publicly available CMIP5 and CMIP6 downscaled datasets which provide daily precipitation and temperature data from 1950–2100 for the contiguous United States (CONUS), but comparisons between datasets are difficult to make because they often use differing methodologies and training datasets and downscale different GCM members (see [1–3] for a review of downscaling techniques and associated strengths and limitations). As a result, downscaled dataset users, defined here as those who utilize downscaled climate data for planning and adaptation purposes, including federal agencies, are often left wondering which downscaled dataset is most appropriate to use and how similar or dissimilar different downscaled projections are. This basic lack of understanding of whether, where and why downscaled datasets differ limits users' ability to plan for communities, infrastructure, and ecosystems under a changing climate [1].

While various studies have quantified the differences in future downscaled projections that arise due to specific methodological choices (e.g. [4–7]), few compare the currently available suite of downscaled datasets (e.g. LOCA2, MACA, NA-CORDEX runs, or NEX-GDDP) that CONUS adaptation practitioners, especially at the state and federal levels regard as potential data resources for future planning efforts. Analyzing the differences between available downscaled CONUS climate datasets (listed in figure 1(a)) is difficult because of the numerous subjective choices that characterize each dataset: not only are the downscaling methods and observation-based training datasets unique but each dataset downscales different GCM members and future emission scenarios (e.g. [8–11]). Such CONUS dataset comparisons are not coordinated by any overarching group and are complicated by the fact that new versions of datasets are released sporadically and hosted in various locations. Nonetheless, the current suite of CONUS datasets needs to have clearer comparable performance metrics for both producers and users to enable future improvements and appropriate usage.

Studies comparing publicly available downscaled datasets have shown that substantial uncertainty can arise due to methodology and training dataset choices, but these studies compare a limited number of datasets and usually average them across multiple GCMs [12–17]. Since downscaled datasets include only certain GCMs and their respective members, averaging across different GCM suites obscures the differences in downscales datasets that can be attributed to downscaling methodological choices alone. To isolate the downscaling process as the cause of variations, GCM input must be the same between the compared datasets. Kim *et al* [18] took this approach to quantify the disagreement between CMIP5 downscaled datasets; while their analysis showed that disagreement varied by region and downscaled variable, it was limited to five monthly statistical datasets. The current analysis extends the work of Kim *et al* [18] to more downscaled datasets and daily variables.

This work presents a simple but illuminating analysis of publicly available downscaled climate products which downscale the same GCM ensemble members. The results presented here explore how to answer an increasingly common and relevant question, 'Where do downscaled datasets agree? And on what? , in attempting to find suitable comparisons between downscaled datasets, this work also highlights the need for greater coordination among the groups that produce downscaled data.

2. Methods

2.1. Datasets

Overlapping GCM members among publicly available CMIP5 and CMIP6 downscaled datasets were identified to select GCM members and emission scenarios (e.g. RCPs and SSPs) to be included in this analysis (figure 1(a)). These datasets include statistical (BCSD, GARD, LOCA, MACA, NEX-GDDP, STAR-ESDM), dynamical (ICAR, NA-CORDEX, RegCM-BC (bias corrected [19]), WUS-D3), and machine-learning (DeepSD-BC) downscaling methods. See the references listed in figure 1(a) and the Open Research Statement for more information on each dataset. This resulted in 5 GCM members across eight downscaled CMIP5 datasets for the historical period and RCP4.5 and 3 GCM members across 8 downscaled CMIP6 datasets for the historical period and various SSP scenarios. Finding common ensemble members across CMIP6 datasets proved more difficult than for CMIP5 datasets, possibly due to the CMIP6 archive's more recent release and the availability of more GCM ensemble members in CMIP6. Additionally, since several of the CMIP6 downscaled datasets only downscale specific emission scenarios (e.g. SSP370 for GARD-LENS and WUS-D3, and SSP585 for RegCM), the signals of these datasets could not be fairly compared. Although all of the datasets listed in figure 1(a) are used in the analysis for this manuscript, datasets that are limited by the GCM simulations that they downscale or the period of record (such as Cordex or RegCM) may only appear in one or two figures.

We also note that, although the intent of this analysis is to compare downscaled datasets which ingest identical GCM members, some statistically downscaled datasets using multivariate approaches (e.g. GARD and MACA), ingesting wind and other atmospheric fields as well as surface temperature (tas) to predict daily temperature, for example and that dynamical downscaling is inherently multivariate; this means that these datasets are not all ingesting completely identical data. The statistical methods compared, however, all ingest at a minimum the GCM predictor variable with the highest predictive value, e.g. the target variable of the downscaling effort.

2.2. Comparing climate change signals

For all CMIP5 datasets and GCM ensemble members listed in figure 1(a), annual metrics were calculated across CONUS for daily precipitation: annual mean, maximum and standard deviation, and June, July, August (JJA) mean; for daily minimum temperature (tasmin): mean, December, February, January mean; for daily maximum temperature (tasmax): mean, JJA mean. Then, the 'change signal,' or trend, for each of these metrics were calculated as the difference between the 2070–2100 and 1970–2000 averages for each downscaled dataset and GCM ensemble member (see an example for the mean daily precipitation signal in figure 1(b), calculated for MIROC5 r1i1p1 for all CMIP5 downscaled datasets). We calculate the grid-point by grid-point minimum, mean and maximum signal for each metric across downscaled datasets, the standard deviation across downscaled datasets, the signal-to-noise ratio for each GCM simulation (as the mean signal divided by the standard

a. CMIP5 and CMIP6 downscaled datasets with overlapping GCM ensemble members

	Downscaled datasets	GCM ensemble members
CMIP5	BCCAv2 (Brekke et al., 2013) BCSD (Brekke et al., 2013) GARD (Gutmann et al., 2022) ICAR (Gutmann et al., 2016) LOCAv1 (Pierce et al., 2014) MACA (Abatzoglou & Brown, 2012) NASA NEX-GDDP (Thrasher et al., 2015) NA-CORDEX (**only CanESM2, Mearns et al. 2017)	CanESM2 r1i1p1 CNRM-CM5 r1i1p1 (excl. ICAR) GFDL-CM3 r1i1p1 (excl. BCCAv2) IPSL-CM5A-LR r1i1p1 (excl. ICAR) MIROC5 r1i1p1 MRI-CGCM3 r1i1p1 NorESM1-M r1i1p1
CMIP6	DeepSD (Chegwidden et al., 2022) GARD-LENS (Hartke et al., 2024) ICAR (Gutmann et al., 2016) LOCAv2 (Pierce et al., 2023) NASA NEX-GDDP (Thrasher et al., 2022) RegCM (Rastogi et al., 2022) STAR-ESDM (Hayhoe et al. 2023) WUS-D3 (Rahimi et al. 2024)	CanESM5 r1i1p1f1 (excl. WUS-D3 & RegCM) NorESM2-MM r1i1p1f1 (excl. DeepSD & GARD-LENS) MRI-ESM2-0 r1i1p1f1 (excl. GARD-LENS, ICAR & WUS-D3)

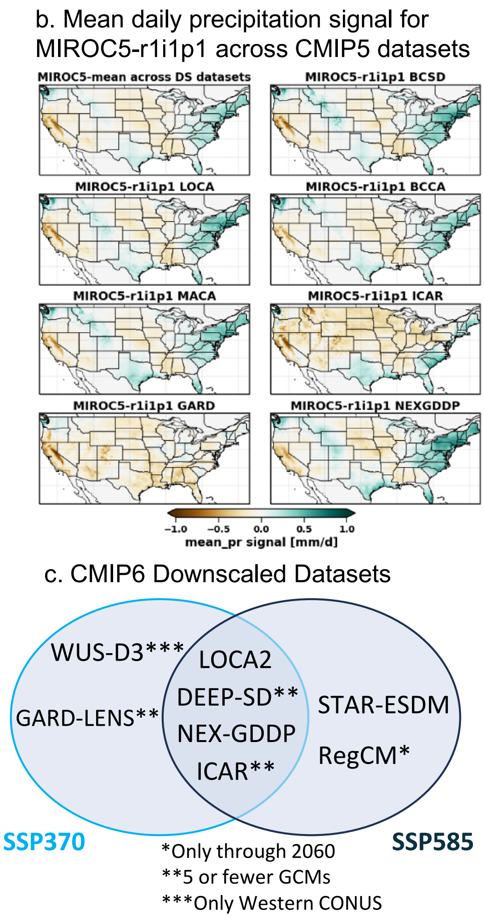

Figure 1. (a) Downscaled datasets and GCM members used in this analysis. Native resolutions of datasets are included in parentheses. (b) Example of climate signal being calculated for all CMIP5 downscaled (DS) datasets containing a specific GCM ensemble member. (c) Overlap in Shared Socioeconomic Pathways (SSPs) among CMIP6 downscaled datasets.

deviation across downscaled datasets) and the lowest and highest outlying downscaled datasets for each metric. Signals were calculated at the native resolution of each dataset, and the results were then regridded to a common 1/8° grid.

Although conservatively regridding the datasets to a common grid is necessary in order to calculate pixel-based statistics such as the mean and standard deviation of signals across datasets, this does not address downscaled dataset uncertainty that is due to resolution differences. While we may wish to separate the uncertainty due to resolution from that due to the downscaling method that would likely require additional downscaled datasets and is outside the scope of this analysis.

For all CMIP6 datasets and GCM members listed in figure 1(a), we also focused on an analysis of historical fields for CanESM5 r1i1p1f1 and NorESM2-MM r1i1p1f1 since there was little overlap in emission scenarios between datasets (figure 1(c)). The historical 1970–2000 daily values of precipitation, tasmax, and tasmin are plotted as an empirical probability density function for each dataset in several US cities. During this process, we note that DeepSD has a probability of precipitation equal to 1 at every location and, upon further inspection, does not actually contain any instances of zero precipitation. Instead, a minimum value (e.g. 0.625 in Seattle and 0.015 in New York City) seems to take the place of zero precipitation and is recorded for long stretches. We took the liberty of thresholding the DeepSD time series and setting each instance of the minimum value to zero for our analysis for each city; this reduced the DeepSD probability of precipitation to values more consistent with other datasets.

3. Results

Figure 2 shows a comparison of the change signal for mean JJA daily precipitation, calculated as the difference between 1970–2000 and 2070–2100 averages, among CMIP5 downscaled datasets in western CONUS. Each column shows the results for a single GCM simulation, and the far right column shows the results when the signal for each downscaled dataset is calculated from the average of all five GCM members. Rows b.–d. show that there is substantial spread in the signal among downscaled datasets even though they are all downscaling in the same GCM simulation. Row e. presents the signal to noise ratio of this trend;

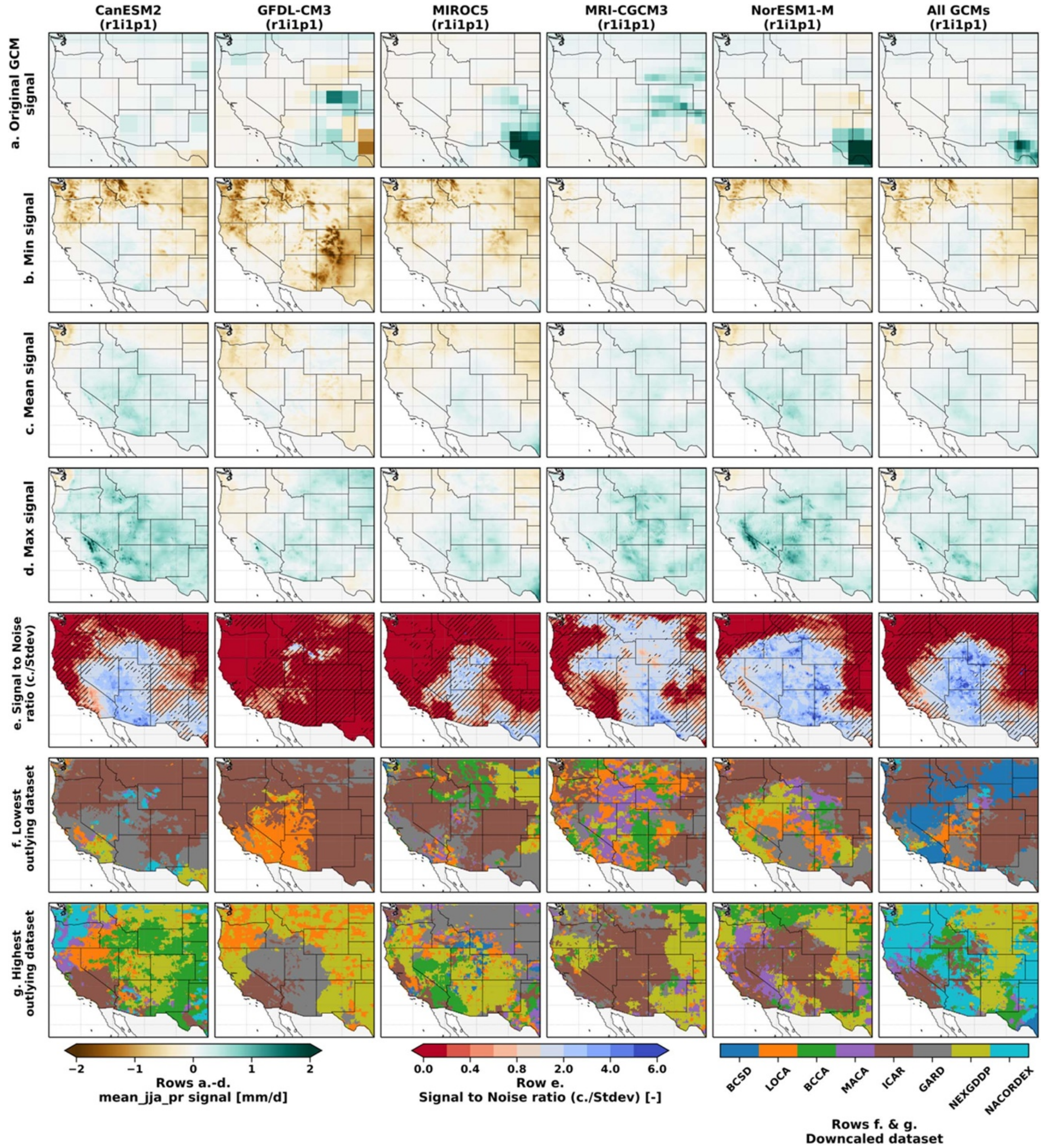

Figure 2. (a) Original GCM signal for mean June, July, August daily precipitation. (b) Minimum, (c) mean, and (d) maximum signal across all CMIP5 downscaled datasets for each GCM member, calculated on a grid point basis. (e) The signal-to-noise ratio calculated as (c) divided by the standard deviation across the estimated signals from all downscaled datasets. Hatched areas denote those where all datasets do not agree on the direction of the signal, and depending on the dataset, the predicted climate signal can be negative or positive. (f) The downscaled dataset with the lowest signal value. (g) The downscaled dataset with the highest signal value. Note that NA-CORDEX is only available for CanESM2.

values less than 1 indicate that the uncertainty (calculated as the standard deviation) in the trend among datasets is higher than the mean trend. Values greater than 1 show where the magnitude of the mean trend is greater than the uncertainty among downscaled datasets. Hatching in Row e. delineates areas where downscaled datasets do not agree on the direction of the climate signal and, therefore, where the trend can be estimated as positive or negative, depending on which dataset is consulted. The hatched areas where datasets do not agree on the direction of the climate signal are calculated using the following rules: The datasets agree on the signal if all downscaled datasets estimate either a positive or negative trend. If there are any differences between the signal directions (e.g. if one dataset estimates a positive trend and all others estimate a negative signal), then the datasets are considered to disagree on the direction of the climate signal. All the available downscaled datasets (between 6 and 8, depending on the GCM) are used to determine this agreement. Rows f. and g. show that there is no dataset which consistently estimates the lowest or highest signal across Western CONUS or across different GCM members, aligning with similar findings in Kim *et al* [18], although ICAR stands out as often estimating the lowest change signal, particularly for the CanESM2 and GFDL-CM3 members. Supplemental figure 1 shows additional

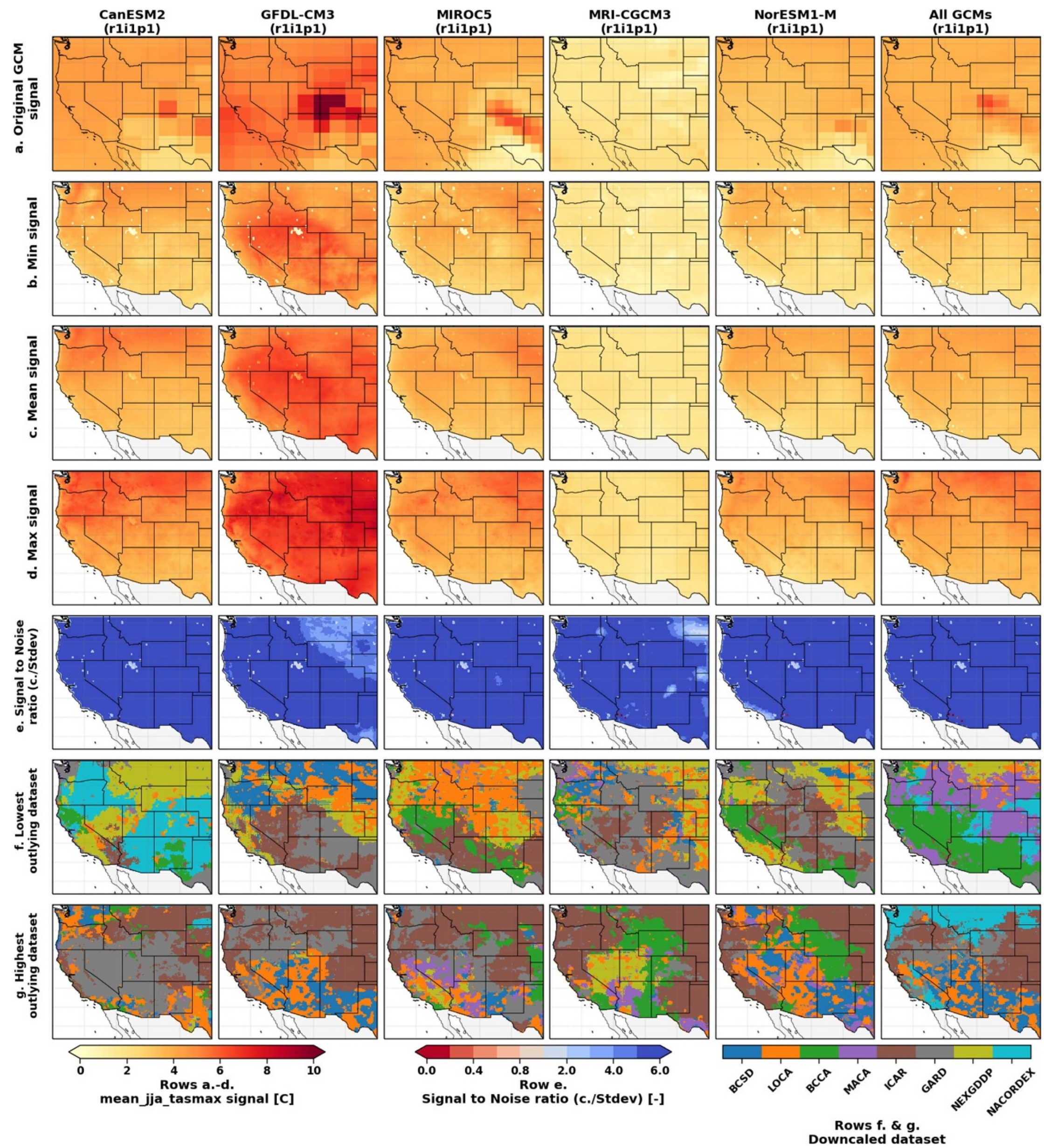

Figure 3. (a) Original GCM signal for mean June, July, August daily maximum temperature. (b) Minimum, (c) mean, and (d) maximum signal across all CMIP5 downscaled datasets for each GCM member, calculated on a grid point basis. (e) The signal-to-noise ratio calculated as (c) divided by the standard deviation across the estimated signals from all downscaled datasets. Hatched areas denote those where all datasets do not agree on the direction of the signal. (f) The downscaled dataset with the lowest signal value. (g) The downscaled dataset with the highest signal value. Note that NA-CORDEX is only available for CanESM2.

signal agreement and standard deviation analysis for figure 2, and supplemental figure 2 shows the same analysis as in figure 2 for the signal for 95th percentile daily precipitation.

Figure 3 shows a similar comparison in climate signal as in figure 2, but for the mean JJA daily maximum temperature. While all datasets agree on the direction of the signal for the mean JJA maximum daily temperature (figure 3(e)), there is disagreement on the magnitude of the signal among the datasets (figures 3(b)–(d)). The signal to noise ratio for this temperature trend is consistently high for all GCMs; from a user perspective, this indicates that the choice of downscaled dataset will not substantially impact the temperature trend used in adaptation plans. Similar to precipitation (figure 2), there is little consistency in which the dataset exhibits the lowest or highest signal (figures 3(f) and (g)). Supplemental figures 3 and 4 show the same analysis as in figure 3 for the mean annual daily minimum temperature and 95th percentile daily maximum temperature signals, respectively, with additional signal agreement and standard deviation plots.

In figure 4, the historical 1970–2010 daily precipitation, maximum temperature, and minimum temperature distributions at four US cities are compared for CanESM5 r1i1p1f1 and five CMIP6 downscaled datasets, as well as two observation-based datasets,

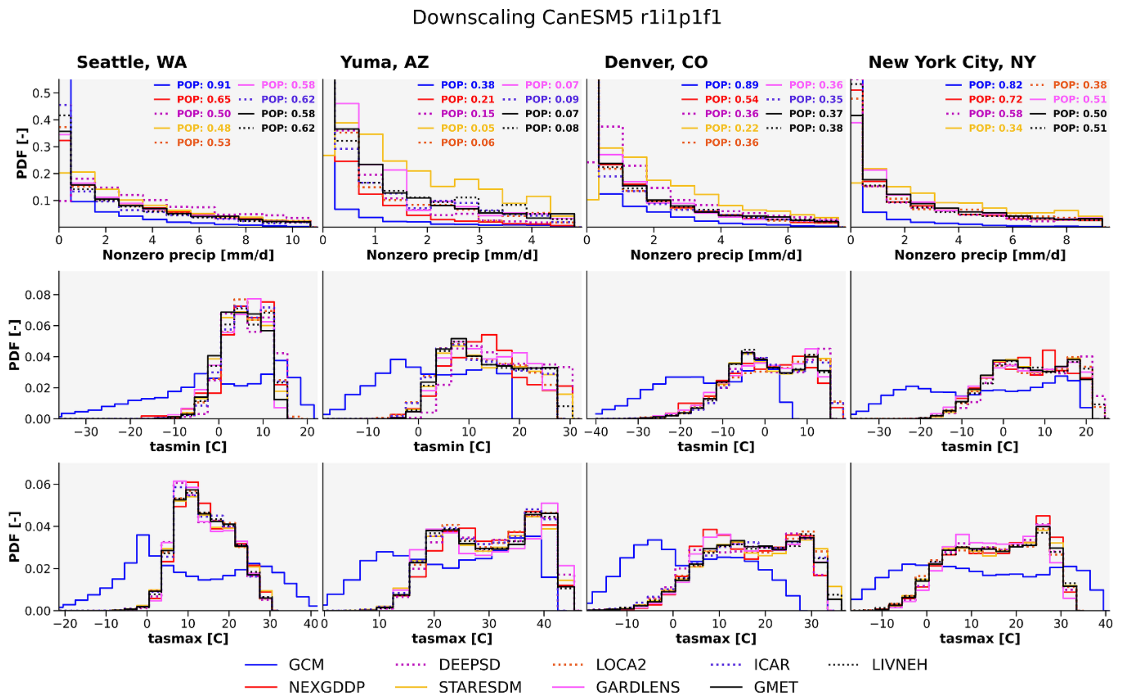

Figure 4. Historical 1970–2010 distributions of downscaled CanESM5 r1i1p1f1 data for (Top row) daily nonzero precipitation, (Middle row) minimum daily precipitation, and (Bottom row) maximum daily temperature at four U.S. cities. Note that some CMIP6 downscaled datasets do not contain this GCM member (see figure 1(a)) and are therefore not included in this plot.

GMET [20] and Livneh [21] (specifically, the unsplit version of Livneh described in Pierce *et al* [22]). In the top row, the nonzero precipitation distribution is plotted, and the probability of precipitation (POP) of datasets at each city is included in the upper right. The original CanESM5 member (blue) tends to overestimate the probability of precipitation, as could be expected due to its coarse spatial resolution. STAR-ESDM records fewer instances of near zero precipitation than other datasets and observations, i.e. not enough drizzle, most visibly in New York City and Denver. In the middle and bottom row, minimum and maximum daily temperatures are underestimated by the original CanESM5 member, which has a much wider temperature distribution than observations, but all downscaled datasets roughly match the observed temperature distributions (e.g. GMET and Livneh). Notably, differences in the downscaled datasets in figure 4 cannot be attributed to resolution differences alone, even though the observations datasets are at different resolutions (GMET at 1/12° and Livneh at 1/24°), they estimate nearly identical distributions at the cities in figure 4. The same analysis as in figure 4 is performed for NorESM2-MM r1i1p1f1 and presented in supplementary figure 5.

Figure 5 shows the comparisons in future climate signals that can be made across CMIP6 datasets; there are four downscaled datasets each that can be compared for the SSP370 and SSP585 scenarios for CanESM5 member r1i1p1f1. Signals for annual mean daily precipitation and annual mean daily maximum temperature are shown for datasets which downscale SSP370 and SSP585. There is general agreement across datasets on the precipitation signal, except for DeepSD; in Southern California where DeepSD estimates a positive signal while all other datasets project a negative signal. The maximum temperature signal across all datasets and both SSPs are positive, although GARD-LENS and DeepSD show a notably smaller signal than LOCA2, NEX-GDDP, and STAR-ESDM.

4. Summary

A growing suite of downscaled climate projection datasets, both CONUS-wide and global provides localized insight about our future world, but good recommendations for dataset producers and users on how to evaluate and choose between downscaled products are sorely lacking. This is due to a number of hurdles; the large size of these datasets makes them difficult to host for many research groups who seek to evaluate and compare downscaled data. Even after overcoming the not insubstantial challenge of gathering over a dozen downscaled datasets for analysis, comparing these datasets is still not straightforward; many datasets only downscale a few common GCM members and some datasets exclusively downscale different carbon emission scenarios (e.g. RCPs and SSPs). Noted by Barsugli *et al* [23] a decade ago, before most of the datasets in this analysis were even created, the 'practitioner's' dilemma is no longer the lack of down-scaled projections; it is how to choose

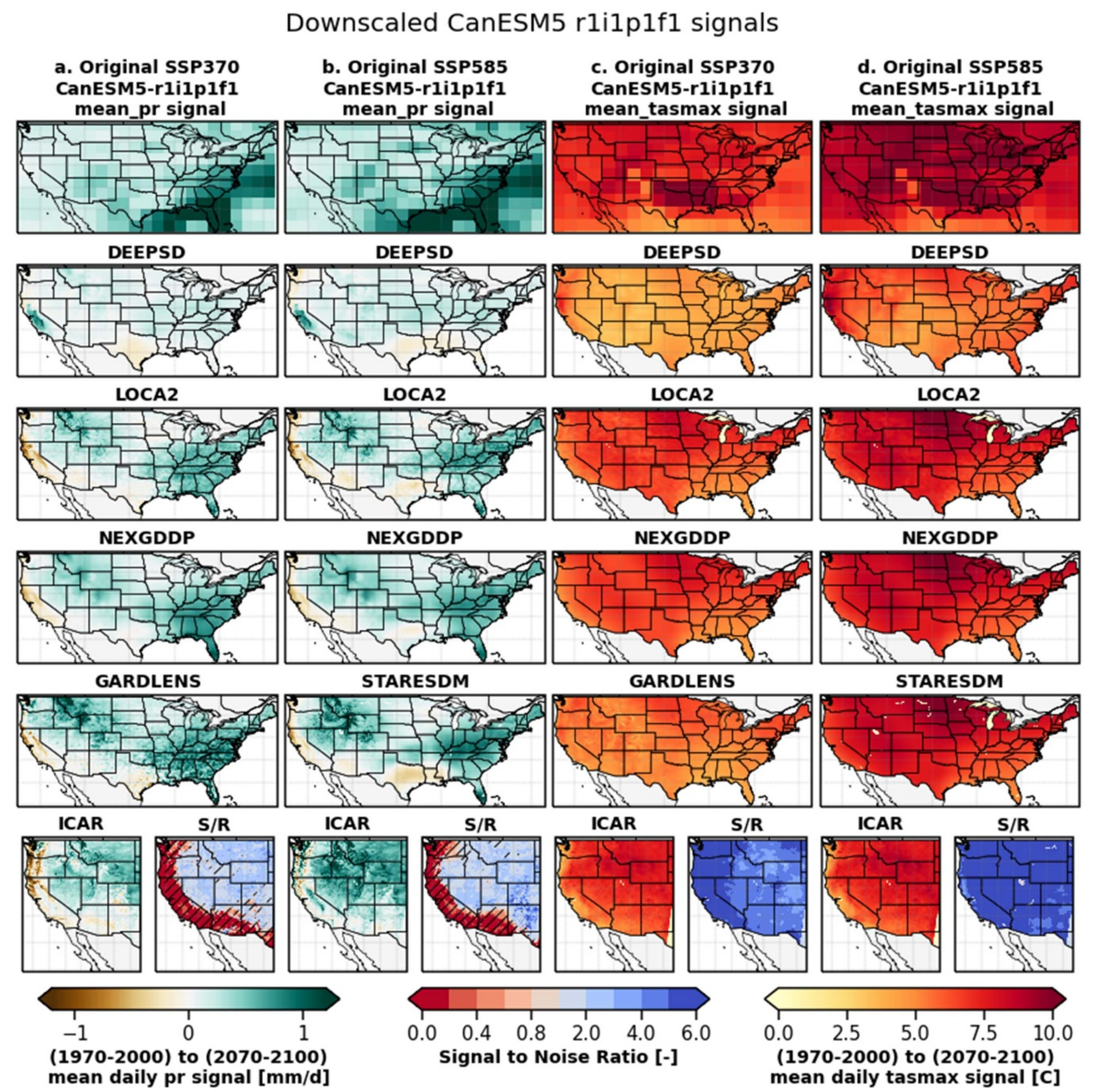

Figure 5. Signals for (a) annual mean daily precipitation based on SSP370, (b) annual mean daily precipitation based on SSP585, (c) annual mean daily maximum temperature based on SSP370, and (d) annual mean daily maximum temperature based on SSP585 for the original CanESM5 r1i1p1f1 simulation and available datasets which downscale CanESM5 r1i1p1f1. The signal to noise ratio, calculated as the mean signal across datasets divided by the standard deviation of the signal estimates from each dataset, is shown in the lower right of each column. Hatched areas denote those where all datasets do not agree on the direction of the signal.

an appropriate dataset, assess its credibility, and use it wisely'.

A decade later, we in the downscaling community have still not sufficiently addressed this dilemma, although funding is often available for single 'offshoot' downscaling datasets, there is less investment in deciding which gaps in the catalog of CONUS downscaled datasets need to be filled and who is responsible for maintaining and updating datasets (e.g. generating a dataset for the next CMIP). Other communities such as the land-modeling and the global climate modeling communities have made great strides in the past 20 years to formalize benchmarking activities via publications, standardized testing protocols and datasets, and subsequent websites and public software packages [24–28]. These general concepts can be leveraged here as the downscaling community still relies on ensembles of opportunities and does not coordinate upstream on what downscaled datasets and associated testing protocols users and the evaluation community really need. Previous studies that compare statistical downscaling techniques (e.g. bias correction spatial disaggregation (BCSD) vs. quantile mapping vs. delta scaling, e.g. [7, 29–31]) are informative, but do not compare the currently available daily datasets that practitioners must choose between. Kim *et al* [18] was the first to set out to compare publicly available, downscaled datasets with identical GCM inputs but were limited to CMIP5 statistically downscaled products and monthly data.

Even so, Kim *et al* [18] found that the choice of downscaled datasets introduced substantial uncertainty in future precipitation estimates.

In this analysis, we show that publicly available daily datasets for the contiguous US (CONUS) can differ substantially even when downscaling the same GCM data, particularly for precipitation projections. Aligning with previous studies downscaled temperature estimates show greater consistency than precipitation across downscaling methods, and the agreement between the datasets is metric- and region-dependent [18, 29, 32]. However, this work goes further to show that the downscaling uncertainty around precipitation trends is frequently greater than the trends that practitioners are trying to plan for (figures 2 and 5); the selection of downscaled dataset can be the difference between planning for a −90 mm or 45 mm end-of-century summer precip trend in the Pacific Northwest. Interestingly, the downscaled dataset agreement is not always consistent across GCMs, pointing to the sensitivity of the methods to the input data, supporting Kim *et al*'s [18] finding that 'for some parts of CONUS, the choice of [downscaled] dataset may be as significant as the choice of GCM'. Unsurprisingly, methods with similar underlying techniques, such as those based on historical analogs like BCSD, LOCA, BCCA, MACA, and NEX-GDDP often show very similar results (figures 1(b) and 5). Crucially, this analysis shows that, for adaptation practitioners, the choice of downscaled dataset can be the difference between preparing for a positive or negative future change in precipitation. With differences this large between downscaled datasets, even when GCM input is held constant, there needs to be more guidance for producers and users on how to evaluate and choose between datasets, as a prerequisite, downscaled datasets need to be more intercomparable.

The downscaling community can improve dataset comparability with several actions:

- If downscaling a single member of a GCM ensemble, selecting the r1i1p1f1 member increases the chances that it will be comparable to other datasets.
- Downscaled dataset producers should coordinate to select a core subset of GCM simulations to prioritize when generating downscaled datasets, perhaps including those which already have a substantial amount of overlap in downscaled datasets, such as CanESM5 and NorESM2-MM (figure 1(a)).
- A practice that could be quite beneficial when introducing downscaled datasets is that of the 'perfect model experiment,' as used in Dixon *et al* [13] and Hayhoe *et al* [33], in which a reanalysis or other high resolution climate dataset are coarsened and downscaled. This experiment isolates downscaling methods' ability to reproduce the high resolution properties of a coarse climate dataset, which is difficult to assess in downscaled GCM simulations due to the imperfect, biased relationship between GCM simulations and historical observations. Selecting one or several 'perfect model' datasets, such as a coarsened ERA5 or high resolution GCM, to include when generating downscaled datasets would be very beneficial for downscaling evaluation work.
- The European downscaling community utilized a common repository from 2012–2015 for uploading downscaled CMIP5 datasets for validation and comparison (The EU Cooperation in Science and Technology (COST) Action ES1102 VALUE, www.value-cost.eu [34, 35]); a similar repository for CONUS downscaled datasets can be envisioned where developers upload a 'perfect model' run which downscales ERA5 and a predefined set of downscaled GCM members (e.g. CanESM5 r1i1p1f1 and NorESM2-MM r1i1p1f1) and include dataset descriptions and links to downscaled datasets in their entirety.

  - This website would streamline comparisons of publicly available CONUS downscaled projections without requiring users to navigate access to the entirety of a downscaled dataset, a substantial hurdle for many research groups and users.
  - By streamlining the comparability of CONUS downscaled projections, this website would allow researchers to provide more comprehensive guidance on dataset selection.
  - Although dozens of metrics may be of interest to practitioners, this website, and other coordinate evaluation efforts, can focus on a select subset of metrics, including several for trends and extremes (e.g. historical annual precipitation trend or the 99th percentile of daily mean temperature). These metrics should present a way to effectively verify and compare the 'fitness for purpose' of downscaled datasets, after which interested users can calculate niche metrics of interest (e.g. JJA max daily precipitation) for the dataset(s) that are best suited to their specific needs.
  - Although such a website would broadly benefit the CONUS downscaling community, without one entity or designated group responsible for funding, standing up, and, perhaps more crucially, maintaining such a website, this need will not be met.

Downscaled datasets are already an integral part of adaptation and resilience planning at community, state, federal, and global levels [36], it is paramount that, in addition to providing high resolution climate projections, we also provide sufficient resources and guidance for users to effectively compare and weigh the trade-offs of downscaled datasets. If we want to foster a comprehensive impact assessment community, we need to discuss the standardization and coordination of downscaling runs to enable more informative comparisons of downscaled climate change signals. A coordinated website dedicated to downscaled dataset comparison for specific GCM simulations and a perfect model benchmark (e.g. ERA5) would help both downscaled dataset developers and users answer that frequent question, 'Where do datasets agree, on what, and why?'. Lessons learned from other communities' implementation of these ideas should be used to speed our progress, yet such activities will require more frequent and direct communication among practitioners and resources, including funding support, for such activities.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.13891667.

Acknowledgment

This work was supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977. This work was funded by DoD ESTCP award #W912HQ23C0002 (Project CR22-7259) and US Bureau of Reclamation Dam Safety Office Grant #R22AC00313. Rachel McCrary was partially supported by the US Department of Energy Regional and Global Climate Modeling program award DOE DE-SC0016605. We thank the downscaling groups who have made their data accessible for this work and the many practitioners who have taken part in conversations around downscaling comparison and evaluation.

Open research

All data analysis and plotting were conducted in Python. The Jupyter notebooks used for this manuscript are available at https://https://doi.org/10.5281/zenodo.13891668. Downscaled datasets were obtained from a variety of original archives (e.g. NASA NEX-GDDP on Amazon Web Services, a local server for LOCA2, etc). The references and documentation for each downscaled dataset are as follows: BCCAv2 [12], BCSD [12], GARD [8], ICAR [37], LOCAv1 [10], MACA [38], CMIP5 NASA NEX-GDDP [11], NA-CORDEX [39], DeepSD [40], GARD-LENS [41], LOCAv2 [22], CMIP6 NASA NEX-GDDP [42], RegCM [43], STAR-ESDM [44], WUS-D3 [45].

ORCID iDs

Samantha H Hartke https://orcid.org/0000-0002-0239-4723
Andrew J Newman https://orcid.org/0000-0001-8796-0861
Nicholas D Lybarger https://orcid.org/0000-0001-6835-3003

References

[1] Ekström M *et al An Appraisal of Downscaling Methods Used in Climate Change Research* 2015 (Wiley-Blackwell) (https://doi.org/10.1002/wcc.339)
[2] Kotamarthi D *et al* 2016 *Use of Climate Information for Decision-Making and Impacts Research : State of Our Understanding* (Strategic Environmental Research and Development Program (DOD)) pp 1–55
[3] Nielsen-Gammon J W *et al* 2023 Climate model downscaling for Texas (available at: www.gyclimate.org/ch4)
[4] Mandal S, Breach P A and Simonovic S P 2016 Uncertainty in precipitation projection under changing climate conditions: a regional case study *Am. J. Clim. Change* **5** 116–32
[5] Michalek A T, Villarini G and Kim T 2024 Understanding the impact of precipitation bias-correction and statistical downscaling methods on projected changes in flood extremes *Earths Future* **12** e2023EF004179
[6] Wootten A, Terando A, Reich B J, Boyles R P and Semazzi F 2017 Characterizing sources of uncertainty from global climate models and downscaling techniques *J. Appl. Meteorol. Climatol.* **56** 3245–62
[7] Wootten A M, Dixon K W, Adams-Smith D J and McPherson R A 2021 Statistically downscaled precipitation sensitivity to gridded observation data and downscaling technique *Int. J. Climatol.* **41** 980–1001
[8] Gutmann E D, Hamman J J, Clark M P, Eidhammer T, Wood A W and Arnold J R 2022 En-GARD: a statistical downscaling framework to produce and test large ensembles of climate projections *J. Hydrometeorol.* **23** 1545–61
[9] Bukovsky M S and Mearns L O 2020 Regional climate change projections from NA-CORDEX and their relation to climate sensitivity *Clim. Change* **162** 645–65
[10] Pierce D W, Cayan D R and Thrasher B L 2014 Statistical downscaling using localized constructed analogs (LOCA) *J. Hydrometeorol.* **15** 2558–85
[11] Thrasher B, Maurer E P, McKellar C and Duffy P B 2012 Technical Note: bias correcting climate model simulated daily temperature extremes with quantile mapping *Hydrol. Earth Syst. Sci.* **16** 3309–14
[12] Brekke L, Thrasher B P, Maurer E P and Pruitt T 2013 Downscaled CMIP3 and CMIP5 climate projections: release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs bureau of reclamation climate analytics group climate central (available at: http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/)
[13] 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
[14] Lanzante J R, Nath M J, Whitlock C E, Dixon K W and Adams-Smith D 2019 Evaluation and improvement of tail behaviour in the cumulative distribution function transform downscaling method *Int. J. Climatol.* **39** 2449–60
[15] Lafferty D C, Sriver R L, Haqiqi I, Hertel T W, Keller K and Nicholas R E 2021 Statistically bias-corrected and

downscaled climate models underestimate the adverse effects of extreme heat on U.S. maize yields *Commun. Earth Environ.* **2** 196

[16] Lafferty D C and Sriver R L 2023 Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6 *npj Clim. Atmos. Sci.* **6** 158

[17] Ullrich P 2023 Validation of LOCA2 and STAR-ESDM statistically downscaled products

[18] Kim J B, Jiang Y, Hawkins L R and Still C J 2022 A comparison of multiple statistically downscaled climate change datasets for the conterminous USA *Environ. Res. Commun.* **4** 125005

[19] Kao S-C, Ashfaq M, Rastogi D and Gangrade S 2022 CMIP6-based multi-model hydroclimate projection over the conterminous US (https://doi.org/10.21951/SWA9505V3/1887469)

[20] Newman A J, Clark M P, Craig J, Nijssen B, Wood A, Gutmann E, Mizukami N, Brekke L and Arnold J R 2015 Gridded ensemble precipitation and temperature estimates for the contiguous United States *J. Hydrometeorol.* **16** 2481–500

[21] Livneh B, Bohn T J, Pierce D W, Munoz-Arriola F, Nijssen B, Vose R, Cayan D R and Brekke L 2015 A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013 *Sci. Data* **2** 150042

[22] Pierce D W, Cayan D R, Feldman D R and Risser M D 2023 Future increases in North American extreme precipitation in CMIP6 downscaled with LOCA *J. Hydrometeorol.* **24** 951–75

[23] Barsugli J J *et al* 2013 The practitioner's dilemma: how to assess the credibility of downscaled climate projections *Am. Geophys. Union* **94** 424–5

[24] Abramowitz G 2005 Towards a benchmark for land surface models *Geophys. Res. Lett.* **32** 1–4

[25] Abramowitz G 2012 Towards a public, standardized, diagnostic benchmarking system for land surface models *Geosci. Model Dev. Discuss.* **5** 549–70

[26] Abramowitz G *et al 2024* On the predictability of turbulent fluxes from land: PLUMBER2 MIP experimental description and preliminary results *Biogeosciences* **21** 5517–38

[27] Eyring V *et al* 2020 Earth System Model Evaluation Tool (ESMValTool) v2.0—an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP *Geosci. Model Dev.* **13** 3383–438

[28] Best M J *et al* 2015 The plumbing of land surface models: benchmarking model performance *J. Hydrometeorol.* **16** 1425–42

[29] Kim T and Villarini G 2024 Projected changes in daily precipitation, temperature and wet-bulb temperature across Arizona using statistically downscaled CMIP6 climate models *Int. J. Climatol.* **44** 1994–2010

[30] Tabari H, Mendoza Paz S, Buekenhout D and Willems P 2021 Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought *Hydrol. Earth Syst. Sci.* **25** 3493–517

[31] Werner A T and Cannon A J 2016 Hydrologic extremes—an intercomparison of multiple gridded statistical downscaling methods *Hydrol. Earth Syst. Sci.* **20** 1483–508

[32] Wang F, Tian D, Lowe L, Kalin L and Lehrter J 2021 Deep learning for daily precipitation and temperature downscaling *Water Resour. Res.* **57** e2020WR029308

[33] Hayhoe K *et al* 2021 STAR-ESDM: a new bias correction method for high-resolution station- and grid-based climate projections *EGUGA* pp EGU21–13931

[34] Maraun D, Widmann M, Gutiérrez J M, Kotlarski S, Chandler R E, Hertig E, Wibig J, Huth R and Wilcke R A I 2015 VALUE: a framework to validate downscaling approaches for climate change studies *Earths Future* **3** 1–14

[35] Gutiérrez J M *et al* 2019 An intercomparison of a large ensemble of statistical downscaling methods over Europe: results from the VALUE perfect predictor cross-validation experiment *Int. J. Climatol.* **39** 3750–85

[36] Marvel K *et al* 2023 Climate trends *Fifth National Climate Assessment* (U.S. Global Change Research Program (USGCRP)) ch 2

[37] Gutmann E, Barstad I, Clark M, Arnold J and Rasmussen R 2016 The intermediate complexity atmospheric research model (ICAR) *J. Hydrometeorol.* **17** 957–73

[38] Abatzoglou J T and Brown T J 2012 A comparison of statistical downscaling methods suited for wildfire applications *Int. J. Climatol.* **32** 772–80

[39] Mearns L O 2017 The NA-CORDEX dataset *version 1.0* (https://doi.org/10.5065/D6SJ1JCH)

[40] Chegwidden O *et al* 2022 Open data and tools for multiple methods of global climate downscaling (available at: https://carbonplan.org/research/cmip6-downscaling-explainer) (Accessed 08 November 2023)

[41] Hartke S H, Newman A J, Gutmann E D, McCrary R, Lybarger N D and Lehner F 2024 GARD-LENS: a downscaled large ensemble dataset for understanding the future climate and its uncertainties *Sci. Data* **11** 1374

[42] thrasher B *et al* 2022 NASA global daily downscaled projections, CMIP6 *Sci. Data* **9** 1–6

[43] Rastogi D, Kao S-C and Ashfaq M 2022 How may the choice of downscaling techniques and meteorological reference observations affect future hydroclimate projections? *Earths Future* **10** e2022EF002734

[44] Hayhoe K, Stoner A, Wuebbles D J and Scott-Fleming I 2024 STAR-ESDM: A generalizable approach to generating high-resolution climate projections through signal decomposition *Earth's Futur.* **12** 7

[45] Rahimi S *et al* 2024 An overview of the Western United States dynamically downscaled dataset (WUS-D3) *Geosci. Model Dev.* **17** 2265–86