Validation of LOCA2 and STAR-ESDM Statistically Downscaled Products

Paul Ullrich
Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory

October 19, 2023

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Abstract

This report represents an independent validation of 30 years of historical temperature and precipitation data from two statistically downscaled climate data products included in the fifth National Climate Assessment. The two products examined are LOCalized Analogs version 2 (LOCA2), which is trained on the Livneh-unsplit gridded climate data product, and Seasonal Trends and Analysis of Residuals Empirical-Statistical Downscaling Model (STAR-ESDM), which is trained on the NClimGrid-Daily gridded climate data product. Both datasets are compared alongside the widely accepted and scientifically validated Parameter-elevation Relationships on Independent Slopes Model (PRISM). The goal of this report is to assess consistency among these products using a suite of metrics and diagnostics, and use the results of this analysis to develop recommendations on the careful use of these data products. With that said, care should be taken when using any climate data product, and users are encouraged to be aware of caveats related to their use. Overall, this report presents the following observations and recommendations:

- In general, LOCA2 and STAR-ESDM are largely consistent with each other and with PRISM under most metrics. There are minor differences between products, related to the downscaling methodology and observational uncertainty.
- In some coastal regions (particularly Northern California and New England) there are differences in temperature metrics related to how these data products treat oceanic modulation of extreme temperatures.
- Days with temperature above a threshold can be very sensitive to the choice of data product in regions where these events are rare.
- The most extreme minimum daily temperatures vary substantially between STAR-ESDM and LOCA2, with differences of up to 6 degrees Celsius in the eastern U.S. and up to 8 degrees Celsius in the western U.S. These differences are attributed to observational uncertainty in mountainous regions and methodological choices.
- In several regions and seasons, precipitation in the statistically downscaled products can deviate from the suite of gridded observations (e.g., a drier summer and autumn in the Northeast and drier autumn in the Southeast).
- Extreme precipitation intensity appears to be slightly underestimated (5-10%) in both statistically downscaled products throughout the Midwest and Northeast.
- Spatial imprinting from the observational network is apparent in the precipitation fields of both products throughout the eastern United States. Spatial imprinting from the driving model is apparent in individual ensemble members of STAR-ESDM. This suggests that particular care should be taken when using precipitation data at the gridpoint scale.

This report concludes that both products are scientifically sound and are of high-quality. Both products represent the state of the art in high-resolution regional climate data, and so are viable for use in scientific research and for use in relevant applications. Notably, the validation effort undertaken in this report applies only to the version of the data available at time of writing.

1 Overview of this Report

The National Climate Assessment (NCA) is the preeminent national report examining current and future risks posed by climate change. Countless agencies, policymakers, stakeholders and other end-users rely upon guidance from the NCA to plan for an uncertain future. These groups all depend on modern curated data, provided alongside the NCA, to quantify the impact of climate change on metrics of relevance for their decision processes. In its fifth iteration (NCA5), two statistically downscaled ensemble products, each providing daily temperature and precipitation data at grid spacing of approximately 5km over the contiguous United States, were selected to accompany the report. These include LOCalized Analogs version 2 (LOCA2) and Seasonal Trends and Analysis of Residuals Empirical-Statistical Downscaling Model (STAR-ESDM). Both data products are produced through a process known as statistical downscaling, where relatively coarse Global Climate Model (GCM) data is refined to locally relevant scales through the application of scientifically-supported empirical and algorithmic relationships. In support of the NCA effort, this report provides an independent validation of these two products against historical observations, with a focus on precipitation and near-surface temperature variables. Based on the results of this validation, several recommendations are provided that relate to the use of these data products.

The structure of this report is as follows: In section 2, we review the three gridded observational products that are part of our intercomparison. In section 3, we describe the two statistical downscaling techniques and their corresponding datasets. In section 4, the methodology we employ for validation is described. Section 5 provides results of the validation, which in turn motivate our recommendations on the use of these data products. A brief summary is provided in section 6.

2 Gridded Observational Products

The uneven spatial and temporal distribution of in-situ and satellite-based meteorological observations has motivated the development of gridded observational products, which provide a spatially and temporally contiguous representation of meteorological fields. Over the past several decades, a number of such products have been developed, featuring varied regional coverage, grid spacing, temporal resolution, and levels of quality. We focus our investigation here on three such products, which are widely viewed in the scientific community to be of high-quality and produced through scientifically sound means: the Parameter-elevation Relationships on Independent Slopes Model (PRISM), NClimGrid-Daily and Livneh-unsplit.

PRISM was developed at Oregon State University and consists of a spatially continuous high-resolution (800m and 4km) dataset spanning 1970 to present day for the United States (Daly et al., 2008). Using a digital elevation model (DEM), 13,000 precipitation and 10,000 temperature observational stations, and several key regional climate indicators (i.e., location, elevation, coastal proximity, topographic orientation and position, vertical atmospheric layer, and terrain slope) the PRISM precipitation and surface temperature dataset was constructed. PRISM has demonstrated strong characterization of coastal effects, cold air drainage, elevational gradients, inversion layers, and rain shadows. PRISM is widely regarded throughout the climate community to be a high-quality and well-supported gridded observational product. In this evaluation, only the PRISM data at 4km is used as reference due to the similar grid spacing to the LOCA2 and STAR-ESDM products.

NClimGrid-Daily (Durre et al., 2022) is a modern daily gridded precipitation product covering the period 1951-present that is produced and served by the National Atmospheric and Oceanic Administration (NOAA). Gridded fields are generated by interpolating morning and midnight observations from the Global Historical Climatology Network Daily (GHCNd) dataset using thin-plate smoothing splines. Further processing steps are then applied to limit inhomogeneous artifacts from station density, observation time, and other factors. NClimGrid-Daily is relatively new, and so is not featured as widely in the scientific literature as PRISM.

Livneh-Unsplit (Pierce et al. 2021) is an update of the Livneh et al. (2015) precipitation product covering the period 1915-2018. The update to the Livneh dataset was motivated by an observed underestimation in extreme daily precipitation in the original product which was shown to emerge because of a time adjustment applied to the gauge observations. Gridded fields are generated by interpolating once-daily precipitation data from the National Climatic Data Center (NCDC) Cooperative Observer Network (COOP) stations using the SYMAP algorithm introduced in Shepard (1984) and following Maurer et al. (2002). To adjust for topographic effects at unobserved locations, data are scaled to PRISM gridded climatology (Daly et al., 2008).

3 Downscaled Products

The Coupled Model Intercomparison Project (CMIP) is a widely-recognized international effort organized by the World Climate Research Programme (WCRP) Working Group on Coupled Modeling (WGCM). CMIP provides a framework for Global Climate Model (GCM) developers to perform global simulations under a coordinated experimental protocol, so as to ensure consistency among contributed model datasets. CMIP6 includes over 100 models from more than 50 modeling centers worldwide. Publicly accessible GCM data from the most recent sixth phase of CMIP (CMIP6) is generally available with grid spacing between 100 and 300 kilometers in the mid-latitudes. GCM data is far too coarse to be useful to most end-users, since its coarse spatial scale provides inadequate representation of finer scale features, including topography and extreme storms. Consequently, a number of methods for downscaling data have been developed to extract physically-consistent and high-spatial scale information from the GCM output. Data provided in conjunction with the NCA5 come from two widely-used and scientifically vetted techniques for downscaling – LOCalized Analogs version 2 (LOCA2) and Seasonal Trends and Analysis of Residuals Empirical-Statistical Downscaling Model (STAR-ESDM). These methods and their corresponding CMIP6-derived datasets are briefly reviewed here.

LOCA2 is a technique that employs statistical downscaling with analogs – i.e., finding days in the historical record that exhibit regional meteorology similar to the regional patterns of a particular GCM day. The LOCA algorithm first bias-corrects historical GCM outputs to observations using a form of quantile mapping that preserves future model-projected changes by quantile, and additionally by adjusting the amount of variability seen in different frequency bands to match observations using a digital filter applied in frequency space (Pierce et al., 2015). To downscale data, the 30 observed days that best match the model day in the wider region around the point being downscaled are first found, then the single one of those 30 days that best matches the model day in the local neighborhood is used as the analog day. Precipitation is downscaled as a value, while temperature is downscaled as an anomaly. To get the final downscaled temperature the downscaled anomaly is added to the downscaled projected climatological change, which is obtained by differencing 30-year normals from future and historical periods and downscaling using similar methodology. The LOCA2 North American product uses Livneh-unsplit data with 6km grid spacing as the precipitation training data set (Pierce et al., 2021) and an updated version of Livneh et al. 2015 as the temperature training data set. In this report, 27 LOCA2 downscaled datasets derived from CMIP6 GCMs are examined (Table 1a).

STAR-ESDM is a statistical downscaling technique based on signal decomposition (Hayhoe et al., 2023). The STAR-ESDM algorithm first disaggregates observations and the GCM output into four separate components: the long-term trend, the static climatology (mean annual cycle over historical), the dynamic climatology (mean annual cycle accounting for annual variations) and high-frequency (daily) anomalies. For each of these components, mappings are constructed between observations and historical GCM output. Future projections are bias-corrected using these mappings, then components are recombined to produce a consistent estimate of future time series. The STAR-ESDM contiguous United States product uses NClimGrid-Daily with 5km grid spacing for training (Durre et al., 2022). In this report, 23 downscaled datasets derived from CMIP6 GCMs are examined (Table 1b).

(a) LOCA2 GCMs		**(b) STAR-ESDM GCMs**	
ACCESS-CM2	HadGEM3-GC31-LL	ACCESS-CM2	INM-CM5-0
ACCESS-ESM1-5	HadGEM3-GC31-MM	ACCESS-ESM1-5	IPSL-CM6A-LR
AWI-CM-1-1-MR	INM-CM4-8	BCC-CSM2-MR	KIOST-ESM
BCC-CSM2-MR	INM-CM5-0	CanESM5	MIROC6
CanESM5	IPSL-CM6A-LR	CMCC-ESM2	MPI-ESM1-2-HR
CESM2-LENS	KACE-1-0-G	EC-Earth3	MPI-ESM1-2-LR
CNRM-CM6-1	MIROC6	EC-Earth3-Veg-LR	MRI-ESM2-0
CNRM-CM6-1-HR	MPI-ESM1-2-HR	EC-Earth3-Veg	NESM3
CNRM-ESM2-1	MPI-ESM1-2-LR	FGOALS-g3	NorESM2-LM
EC-Earth3	MRI-ESM2-0	GFDL-CM4	NorESM2-MM
EC-Earth3-Veg	NorESM2-LM	GFDL-ESM4	TaiESM1
FGOALS-g3	NorESM2-MM	INM-CM4-8	
GFDL-CM4	TaiESM1		
GFDL-ESM4			

Table 1: GCM datasets used in this report from each statistically downscaled ensemble. Shaded GCMs in the LOCA2 column are not included in the STAR-ESDM ensemble, and vice versa. Although the KACE-1-0-G model was also present in the STAR-ESDM ensemble, it is excluded from our analysis because several daily timeslices were found to be corrupted.

Both LOCA2 and STAR-ESDM are scientifically sound techniques, although at the time of writing this report the paper describing the STAR-ESDM methodology and dataset remains under review. Our validation focuses on three impacts-relevant meteorological variables included as part of both statistically downscaled ensemble data products: daily total precipitation (pr), daily maximum temperature (tasmax), and daily minimum temperature (tasmin).

4 Validation Methodology

4.1 Time Period

For the purposes of this validation, the 30-year period between January 1, 1985 and December 31, 2014 is used. This period coincides with the last 30 years of the "historical" period conducted under the CMIP6 protocol. Periods of 30 years are also commonly employed as "climate normals," which are sufficiently long periods of time for convergence of common statistical measures of the climate system.

4.2 Regridding

Regridding refers to the technique of translating data defined at one set of locations (referred to as the source grid) to another set of locations (referred to as the target grid). Regridding is necessary when differencing two sets of data, since differences must be performed at the same location in space. However, depending on how it is performed, regridding of meteorological fields has the potential to mute extrema or lead to inconsistency with the underlying topography. Consequently, instead of regridding precipitation and temperature fields prior to computing metrics, we instead compute evaluation metrics first on the native grid of the downscaled or observational product. Nearest neighbor regridding is then employed when regridding is needed. This technique draws the target data value from the nearest grid point in the source data, does not mute or exaggerate extrema and can be used when the source and target grid spacings are similar. However, nearest neighbor regridding will still produce spurious behavior near rough topography because climatological conditions can vary rapidly over scales of even a few kilometers.

4.3 Metrics

Overall performance of the downscaled datasets is assessed using a suite of commonly employed metrics. For each of these metrics, it was independently observed that individual ensemble members from LOCA2 and STAR-ESDM do not vary substantially enough from one another to affect our conclusions (not shown). With this in mind, only the ensemble mean is evaluated. Metrics are first computed using TempestExtremes (Ullrich et al., 2021) on the native grid for LOCA2, STAR-ESDM and PRISM products. The PRISM results for each metric are regridded to the LOCA2 and STAR-ESDM grids using nearest neighbor regridding; when a difference is needed between LOCA2 and STAR-ESDM data, computations are performed on the LOCA2 and STAR-ESDM grids. For precipitation metrics and frequency metrics, relative differences are computed by dividing by the PRISM product values. Absolute differences and relative differences at each grid point are then sorted within each NCA5 region to obtain 25th percentile, median, and 75th percentile values. This final step allows us to convey the spread in metric values without being sensitive to outliers. A metric is flagged as being significantly different between the source product and PRISM when the interquartile range (25th percentile - 75th percentile) does not bracket zero.

The suite of metrics employed in this report are as follows. These metrics are commonly employed in understanding the impacts of climate change, and describe both mean and extreme characteristics of the precipitation and temperature. These metrics are also selected as they are included among the standard set of impacts-relevant metrics provided as part of the NCA5:

- annualmean_pr: Annual mean precipitation
- seasonalmean_pr: Seasonal mean daily precipitation
- pr_q50: Median daily precipitation on days with more than 1mm of precipitation
- pr_q99p9: 99.9th percentile daily precipitation on days with more than 1mm of precipitation
- annual_pxx: Average annual maximum daily precipitation
- annualmean_tasmax: Annual mean daily maximum temperature
- seasonalmean_tasmax: Seasonal mean daily maximum temperature
- annual_txx: Average annual maximum daily maximum temperature
- annual_tasmax_ge_95F: Fraction of days with daily maximum temperature greater than 95°F
- annual_tasmax_ge_100F: Fraction of days with daily maximum temperature greater than 100°F
- annual_tasmax_ge_105F: Fraction of days with daily maximum temperature greater than 105°F
- tasmax_q50: Median daily maximum temperature
- tasmax_q99p9: 99.9th percentile daily maximum temperature
- annualmean_tasmin: Annual mean daily minimum temperature
- annual_tasmin_le_32F: Fraction of days with daily minimum temperature less than 32°F
- annual_tnn: Average annual minimum daily minimum temperature

Because uncertainties related to observational biases and climate variability are unavoidable in any data product, there is no expectation that the metrics above should be exactly zero, even for a "perfect" dataset. As observed later in this report, significant differences can even arise between the three gridded observational products described in section 2. These differences reflect observational uncertainties, which are particularly persistent in regions where the density of the observational network is sparse (e.g., the Intermountain West). In fact, many of the differences between LOCA2, which is trained using Livneh-Unsplit, and STAR-ESDM, which is trained using NClimGrid-Daily, can be attributed to differences among these two observational products, rather than differences in the downscaling methodology itself. Our results and recommendations must be considered in light of this uncertainty. In this report we avoid using the term "bias" to characterize differences between the statistically downscaled datasets and our chosen reference dataset (PRISM, in this case).

5 Results and Recommendations

5.1 Overall Performance

The results obtained from the metrics in section 4.3 are tabulated in Table 2 for LOCA2 and Table 3 for STAR-ESDM. Overall, both statistically downscaled datasets exhibit generally agree with PRISM (again using the inclusion of zero in the interquartile range as an indicator of a significant difference). These results suggest that for many regions, observational uncertainties are small, and reflect that both products agree well with their training products. However, differences tend to have a larger spread in the western United States, as a result of observational uncertainty, along with the regridding issues described in section 4.2.

There are some notable exceptions which indicate substantial divergence between these products. In the next several sections, we focus only on those metrics where significant differences are present and explore their cause.

5.2 Temperature

Except for annual_tnn, essentially all absolute metrics for temperature are within 1 degree Celsius of reference. Differences tend to be slightly larger in LOCA2, which is attributed to larger temperature differences between PRISM and Livneh than between PRISM and NClimGrid.

Looking to annual_txx (Table 2, Table 3 and Figure 1), two areas are immediately evident as exhibiting significant differences from PRISM. First, in the eastern United States (particularly annual, JJA, and SON) both LOCA2 and STAR-ESDM are warmer than PRISM by 0.2 to 1.0 degrees Celsius. These slightly warmer temperatures are not apparent in Livneh and NClimGrid, suggesting that they are an artifact of the downscaling process. Second, inland from the Pacific seaboard, annual_txx in both LOCA2 and STAR-ESDM are lower than in PRISM, by around 4°C in LOCA2 and 2°C in STAR-ESDM (Figure 1). Unlike in the eastern U.S., this pattern is apparent in NClimGrid and Livneh, suggesting that this difference is attributable to observational uncertainty. Notably, the related metric, 99.9th percentile daily maximum temperatures (not plotted), shows greater agreement across products, and so suggests this to be related to a slight overestimation of intraannual variability.

Along the close coastal periphery (i.e., within 10 kilometers) of both the Pacific Ocean and Atlantic Ocean, substantial differences are apparent in annual_txx in LOCA2 relative to PRISM (Figure 2). Here it is found that annual_txx can be up to 17°C larger along the California coast and up to 11°C larger along the Maine coast. This difference appears to be inherited from the Livneh product, which does not modulate temperatures in coastal regions (in contrast with PRISM and NClimGrid).

Metrics measuring the fraction of days above a given temperature threshold (95°F, 100°F and 105°F) show substantial variation across the contiguous United States. Since this analysis uses the relative difference in the fraction, grid points with small values of the denominator (particularly rare high temperatures) can inflate these results. Nonetheless, it is apparent that the warmer temperatures in both LOCA2 and STAR-ESDM, compared to PRISM, yield far more days with temperatures above these thresholds.

Annual mean daily minimum temperatures and days below 32°F largely agree throughout the analysis domain. However, annual_tnn (average annual minimum daily temperature) is consistently lower in LOCA2 than PRISM across all regions, whereas STAR-ESDM is largely in agreement with PRISM. In the Western U.S., this difference appears to be inherited from far lower values of annual_tnn in Livneh than in PRISM or NClimGrid (Figure 3). However, the Eastern U.S. is also substantially cooler in LOCA2 than PRISM, even though both Livneh and NClimGrid largely agree on this metric. The exact cause of this difference is currently unknown, though may be related to the fact that LOCA2 does not downscale daily minimum temperature directly, but instead diagnoses it from daily maximum temperature and daily temperature range. Notably, STAR-ESDM is also conspicuously colder than NClimGrid, again possibly related to more intraannual variability than in observations.

Recommendations to Users:

- Care should be taken when using temperatures from downscaled products very near their Pacific or Atlantic coastal periphery. Notably, Livneh (and hence, LOCA2) does not account for oceanic modulation of temperatures in these regions.
- When using the metric of "days with temperature above a threshold" care should be taken in areas where these events are rare, as small biases in temperature can lead to exaggerated differences in event frequency.
- Differences against PRISM in annual average daily minimum temperature are substantially larger in LOCA2 than in STAR-ESDM throughout the U.S.

5.3 Precipitation

Relative precipitation metric differences, even among extreme metrics, are generally within 10% throughout the contiguous United States (Table 2 and Table 3). In LOCA2, the Midwest, Northeast and N. Great Plains are drier in the annual mean than in PRISM, largely due to drier summer (JJA) and autumn (SON) seasons; whereas the Southeast agrees with PRISM in the annual mean, largely due to compensating differences related to a wetter spring (MAM) and drier autumn (SON). In STAR-ESDM, the Northeast and Southeast exhibit similar differences to PRISM, while agreement is greater in the Midwest and N. Great Plains. Notably, the downscaled products are slightly drier than their training datasets, as Livneh and NClimGrid are in better agreement with PRISM for these metrics.

For both LOCA2 and STAR-ESDM, drier extreme precipitation (pr_q99p9 and annual_pxx) is apparent throughout much of the eastern U.S., with median differences around 10-20% less than PRISM (Figure 4). Although some of this difference is attributable to observational uncertainty, it is still clear that the downscaled products are drier than their training datasets. This suggests that the downscaling process for both LOCA2 and STAR-ESDM contribute to drying of extremes in this region. In the western U.S., particularly the Intermountain West, both products show more intense extreme precipitation than

PRISM. However, examining NClimGrid and Livneh pr_q99p9 fields, it's clear that both products lie within the observational spread in this region.

Recommendations to Users:

- Be aware of precipitation biases that may exist for each downscaled product, in a given analysis region and during a given season.
- Take care when using extreme precipitation data from gridded data products, as observational spread is large and downscaled products tend to be drier than observations in the eastern U.S.

5.4 Spatial Imprinting of Precipitation

Throughout the eastern United States (and particularly in the southeastern United States), all observational and statistically downscaled products examined here exhibit some degree of spatial imprinting on the precipitation field related to proximity to the observing network. PRISM, Livneh and LOCA2 products all exhibit larger values in most quantile fields near observing stations (note the dimpling in Figures 5 and 6, particularly in the southeastern U.S.). NClimGrid and STAR-ESDM have relatively smooth extreme precipitation fields, but lower median precipitation near observing stations (Figure 5).

Spatial imprinting of the observational network generally originates in the observational product, and is attributed to the data gridding process. As precipitation intensity can vary substantially over short distances, gridding methods that rely on linear or polynomial interpolation between observing stations often do not characterize the spatial pattern of the precipitation. For example, consider four observing stations laid out in a 2x2 grid. If precipitation is only present at one station, a method such as bilinear interpolation will also flag all grid points interior to that grid as precipitating. Consequently, precipitation events in these interior points will tend to occur more frequently and last longer.

A related artifact of the downscaling process is visually present in individual members of the STAR-ESDM data (but not in LOCA2). Likely because of the way precipitation is downscaled to the fine resolution grid, the precipitation field exhibits small unphysical anomalies from the source grid. For instance, precipitating features in the downscaled data tend to align zonally or meridionally, following the GCM grid. These artifacts are clearly apparent in the precipitation autocorrelation field (Figure 7), where autocorrelation increases on the downscaled grid between cell centers of the GCM.

Recommendations to Users:

- Care should be taken when using precipitation data at the gridpoint scale, noting that proximity to an observing station can affect precipitation statistics. Although data is provided at scales of (4-6km), the credible resolution of the statistically downscaled products is likely closer to 25km.

6 Summary

In this report, we have calculated a number of metrics from the LOCA2 and STAR-ESDM statistically downscaled products which will accompany the NCA5. In general, under the validation procedure described in section 4, both products are largely in agreement with our best observations of the contiguous United States over the 1985-2014 validation period. A few caveats where clear differences are present among these data products have been highlighted, and subsequent recommendations on the use of these data products provided in section 5. For these few cases, end-users should be aware of these issues and understand their impact on the decision process. We note that the work in this report is only relevant to the versions of the data products considered here. These techniques are under constant development, and so the issues raised in this report are likely to be addressed in future versions of these downscaled products.

References

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. and Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: a Journal of the Royal Meteorological Society*, *28*(15), pp.2031-2064, doi: 10.1002/joc.1688.

Durre, I., M. F. Squires, R. S. Vose, A. Arguez, W. S. Gross, J. R. Rennie, and C. J. Schreck, 2022: NOAA's nClimGrid-Daily Version 1 – Daily gridded temperature and precipitation for the Contiguous United States since 1951. NOAA National Centers for Environmental Information, since 6 May 2022, doi: 10.25921/c4gt-r169.

Hayhoe, K., Stoner, A., Wuebbles, D.J. and Scott-Fleming, I., 2023. STAR-ESDM: A Generalizable Approach to Generating High-Resolution Climate 2 Projections through Signal Decomposition 3. *Authorea Preprints*, doi: 10.22541/essoar.169462036.65393270/v1.

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 US, and Southern Canada 1950–2013. *Scientific Data*, *2*(1), pp.1-12, doi: 10.1038/sdata.2015.42.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P. and Nijssen, B., 2002. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, *15*(22), pp.3237-3251, doi: 10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

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. *Journal of Hydrometeorology*, *16*(6), pp.2421-2442, doi: 10.1175/JHM-D-14-0236.1.

Pierce, D.W., Su, L., Cayan, D.R., Risser, M.D., Livneh, B. and Lettenmaier, D.P., 2021. An extreme-preserving long-term gridded daily precipitation dataset for the conterminous United States. *Journal of Hydrometeorology*, *22*(7), pp.1883-1895, doi: 10.1175/JHM-D-20-0212.1.

Shepard, D.S., 1984. Computer mapping: The SYMAP interpolation algorithm. In *Spatial statistics and models* (pp. 133-145). Dordrecht: Springer Netherlands.

Ullrich, P.A., Zarzycki, C.M., McClenny, E.E., Pinheiro, M.C., Stansfield, A.M. and Reed, K.A., 2021. TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets. *Geoscientific Model Development*, *14*(8), pp.5023-5048, doi: 10.5194/gmd-14-5023-2021.

	Midwest			Northeast			N. Great Plains			Northwest		
Metric	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th
annualmean_pr (%)	-5.2	-3.2	-1.0	-5.0	-2.8	-0.7	-7.6	-4.6	-1.3	-5.0	0.3	5.4
seasonalmean_pr (DJF) (%)	-5.9	-2.2	2.1	-1.7	1.3	4.4	-6.0	1.8	10.7	-1.3	5.2	11.7
seasonalmean_pr (MAM) (%)	-5.1	-1.8	1.5	-3.4	-0.7	2.0	-8.1	-3.6	0.8	-12.6	-7.6	-2.0
seasonalmean_pr (JJA) (%)	-7.0	-3.9	-0.7	-8.6	-5.2	-1.5	-9.0	-5.3	-0.5	-5.5	1.5	10.1
seasonalmean_pr (SON) (%)	-7.1	-4.0	-0.8	-8.3	-5.7	-3.2	-12.4	-7.6	-1.8	-5.2	1.0	7.2
pr_q50 (%)	-11.8	-7.2	-2.6	-11.5	-7.6	-3.7	-8.0	-2.3	4.2	-1.9	4.0	10.7
pr_q99p9 (%)	-15.7	-8.4	-0.9	-20.0	-12.6	-5.2	-11.6	-1.4	11.4	-5.2	5.4	18.2
annual_pxx (%)	-12.5	-7.5	-2.6	-14.6	-9.3	-4.0	-10.6	-3.8	4.9	-5.6	2.8	12.4
annualmean_tasmax (°C)	0.5	0.6	0.8	0.2	0.4	0.6	0.3	0.6	0.8	-0.4	0.2	0.6
seasonalmean_tasmax (DJF) (°C)	0.2	0.4	0.5	0.0	0.2	0.4	-0.3	0.1	0.4	-1.2	-0.3	0.3
seasonalmean_tasmax (MAM) (°C)	0.2	0.4	0.6	0.0	0.3	0.6	0.1	0.3	0.6	-0.2	0.3	0.8
seasonalmean_tasmax (JJA) (°C)	0.7	0.8	1.0	0.3	0.5	0.8	0.8	1.1	1.3	-0.1	0.5	1.1
seasonalmean_tasmax (SON) (°C)	0.7	0.8	1.0	0.3	0.5	0.8	0.7	0.9	1.1	-0.4	0.2	0.6
tasmax_q50 (°C)	0.5	0.7	0.9	0.1	0.4	0.6	0.4	0.7	0.9	-0.1	0.3	0.7
tasmax_q99p9 (°C)	-0.2	0.2	0.6	-0.1	0.2	0.6	-0.3	0.1	0.5	-1.0	-0.3	0.4
annual_tasmax_ge_95F (%)	20.1	47.2	72.8	-8.4	28.3	94.4	10.9	32.6	59.6	-54.0	-8.9	30.0
annual_tasmax_ge_100F (%)	-32.4	22.2	75.5	-49.4	-6.2	69.4	-1.4	33.3	71.2	-72.2	-25.9	37.0
annual_tasmax_ge_105F (%)	16.0	172.2	423.7	-92.6	-85.2	-75.9	-29.1	9.3	58.0	-90.1	-48.1	35.2
annual_txx (°C)	0.4	0.7	1.1	0.2	0.6	1.0	-0.1	0.2	0.7	-0.9	-0.2	0.4
annualmean_tasmin (°C)	0.0	0.2	0.4	-0.6	-0.2	0.1	-0.7	-0.1	0.2	-2.5	-1.0	-0.1
annual_tasmin_le_32F (%)	-4.9	-3.0	-1.1	-1.7	1.3	4.9	-4.4	-2.0	1.4	-2.1	10.4	25.3
annual_tnn (°C)	-2.4	-1.6	-0.9	-2.4	-1.5	-0.7	-5.1	-3.8	-2.8	-5.4	-3.5	-2.3

	Southeast			S. Great Plains			Southwest		
Metric	25th	50th	75th	25th	50th	75th	25th	50th	75th
annualmean_pr (%)	-2.7	-0.6	1.7	-7.9	-5.6	-3.2	-9.4	-4.3	1.6
seasonalmean_pr (DJF) (%)	-2.3	0.8	4.0	-15.1	-10.9	-6.2	-14.3	-7.6	-0.3
seasonalmean_pr (MAM) (%)	1.8	5.7	10.4	-7.3	-3.0	1.7	-11.4	-4.3	3.3
seasonalmean_pr (JJA) (%)	-6.8	-3.0	0.7	-14.1	-10.1	-5.8	-10.4	-1.8	9.5
seasonalmean_pr (SON) (%)	-9.0	-5.5	-2.4	-4.6	-0.6	3.8	-9.1	-2.5	6.8
pr_q50 (%)	-11.4	-6.5	-1.3	-10.8	-4.2	2.5	-6.6	1.0	9.6
pr_q99p9 (%)	-14.9	-7.1	0.8	-13.2	-4.1	5.6	-8.1	4.7	20.4
annual_pxx (%)	-10.2	-4.9	0.2	-11.4	-5.8	-0.1	-9.2	0.3	11.7
annualmean_tasmax (°C)	0.3	0.5	0.6	0.4	0.6	0.8	0.0	0.4	0.8
seasonalmean_tasmax (DJF) (°C)	0.3	0.5	0.7	0.3	0.6	0.8	-0.3	0.3	0.7
seasonalmean_tasmax (MAM) (°C)	0.3	0.5	0.7	0.3	0.5	0.7	-0.1	0.3	0.7
seasonalmean_tasmax (JJA) (°C)	0.3	0.4	0.6	0.4	0.7	0.9	0.1	0.5	1.0
seasonalmean_tasmax (SON) (°C)	0.4	0.5	0.7	0.5	0.7	0.9	0.1	0.5	0.9
tasmax_q50 (°C)	0.2	0.4	0.5	0.2	0.4	0.7	-0.1	0.3	0.7
tasmax_q99p9 (°C)	0.0	0.3	0.7	-0.1	0.2	0.6	-0.5	0.0	0.5
annual_tasmax_ge_95F (%)	15.4	32.3	54.8	11.7	20.1	31.0	-10.2	9.6	39.1
annual_tasmax_ge_100F (%)	9.1	41.5	87.7	14.5	30.5	50.7	-28.4	7.4	50.6
annual_tasmax_ge_105F (%)	-35.9	18.5	111.1	5.5	41.1	88.8	-45.3	3.5	64.0
annual_txx (°C)	0.5	0.8	1.1	0.3	0.6	1.1	-0.4	0.0	0.5
annualmean_tasmin (°C)	-0.2	0.0	0.3	0.0	0.2	0.5	-2.0	-0.7	0.2
annual_tasmin_le_32F (%)	-6.4	-0.5	6.8	-7.6	-4.3	2.4	-3.0	7.7	27.1
annual_tnn (°C)	-3.0	-2.3	-1.5	-4.8	-4.0	-3.4	-5.3	-3.7	-2.1

Table 2: Grid point evaluation metrics from the LOCA2 ensemble mean.

Metric	Midwest			Northeast			N. Great Plains			Northwest		
	25th	50th	75th	25th	50th	75th	25th	50th	75th	25th	50th	75th
annualmean_pr (%)	-4.8	-2.4	0.1	-6.6	-4.0	-1.6	-5.6	-2.0	2.5	-6.7	2.9	13.2
seasonalmean_pr (DJF) (%)	-8.0	-3.3	5.0	-2.2	1.5	5.3	0.5	8.9	20.2	-2.0	9.3	23.3
seasonalmean_pr (MAM) (%)	-6.6	-3.4	0.4	-5.1	-1.9	1.2	-7.8	-2.3	3.3	-16.2	-7.4	2.2
seasonalmean_pr (JJA) (%)	-5.4	-1.5	2.8	-11.0	-6.1	-0.8	-6.9	-2.1	4.0	-8.9	1.4	13.9
seasonalmean_pr (SON) (%)	-6.6	-2.3	2.7	-11.8	-8.8	-5.9	-11.3	-5.0	2.4	-5.8	5.2	17.3
pr_q50 (%)	-6.2	-2.6	1.0	-8.6	-5.2	-2.0	-5.2	0.3	6.9	-3.9	3.5	13.0
pr_q99p9 (%)	-20.2	-13.0	-5.7	-25.2	-18.1	-10.7	-12.3	-2.3	9.2	-8.5	4.1	16.9
annual_pxx (%)	-13.9	-9.2	-4.1	-17.0	-11.9	-7.1	-9.9	-4.0	3.7	-5.9	4.6	16.2
annualmean_tasmax (°C)	0.2	0.4	0.5	0.1	0.3	0.6	0.0	0.3	0.5	-0.3	0.2	0.6
seasonalmean_tasmax (DJF) (°C)	-0.3	0.0	0.3	-0.2	0.1	0.4	-0.7	-0.4	-0.1	-0.3	0.1	0.5
seasonalmean_tasmax (MAM) (°C)	0.0	0.2	0.4	0.0	0.3	0.6	-0.2	0.1	0.5	-0.4	0.2	0.7
seasonalmean_tasmax (JJA) (°C)	0.5	0.7	0.9	0.3	0.6	0.8	0.5	0.9	1.2	-0.3	0.3	0.9
seasonalmean_tasmax (SON) (°C)	0.3	0.5	0.7	0.1	0.4	0.7	0.2	0.5	0.8	-0.4	0.0	0.5
tasmax_q50 (°C)	0.3	0.5	0.7	0.1	0.4	0.6	0.0	0.3	0.7	-0.4	0.1	0.6
tasmax_q99p9 (°C)	0.0	0.4	0.9	0.2	0.6	1.0	0.0	0.5	1.1	-0.6	0.2	0.9
annual_tasmax_ge_95F (%)	18.5	48.9	86.1	-0.2	48.1	152.2	8.2	33.5	74.0	-34.9	1.6	47.8
annual_tasmax_ge_100F (%)	3.5	56.5	108.7	4.3	56.5	147.8	25.9	73.1	144.7	-33.9	30.4	124.6
annual_tasmax_ge_105F (%)	65.2	217.4	471.7	21.7	52.2	73.9	33.2	104.3	234.8	-60.3	30.4	193.3
annual_txx (°C)	0.3	0.6	0.8	0.3	0.7	1.1	-0.4	0.0	0.5	-1.0	-0.3	0.4
annualmean_tasmin (°C)	-0.1	0.2	0.4	-0.4	0.0	0.4	-0.1	0.2	0.6	0.0	0.4	0.9
annual_tasmin_le_32F (%)	-3.9	-2.0	-0.3	-3.4	-0.3	3.4	-4.2	-1.6	0.6	-13.6	-7.0	-1.2
annual_tnn (°C)	-1.1	-0.5	0.0	-1.0	-0.4	0.1	-1.4	-0.8	-0.2	-1.8	-0.8	0.1

Metric	Southeast			S. Great Plains			Southwest		
	25th	50th	75th	25th	50th	75th	25th	50th	75th
annualmean_pr (%)	-1.2	1.2	4.1	-3.9	-1.1	1.8	-6.3	1.4	10.7
seasonalmean_pr (DJF) (%)	-0.3	4.0	9.2	-13.5	-9.1	-4.4	-11.9	-1.1	11.7
seasonalmean_pr (MAM) (%)	1.6	6.8	11.7	-0.5	4.5	10.2	-9.9	0.9	13.4
seasonalmean_pr (JJA) (%)	-2.2	2.3	6.5	-10.3	-5.6	-0.4	-9.0	1.8	13.9
seasonalmean_pr (SON) (%)	-10.9	-7.7	-4.0	-1.8	3.0	8.3	-7.2	3.1	16.8
pr_q50 (%)	-0.7	5.0	10.6	0.5	9.4	21.4	1.5	10.9	21.4
pr_q99p9 (%)	-10.1	-1.5	7.5	-4.3	6.5	19.1	-2.1	12.5	29.2
annual_pxx (%)	-7.6	-2.4	2.7	-5.4	0.8	7.7	-4.6	5.7	18.6
annualmean_tasmax (°C)	0.2	0.4	0.6	0.1	0.3	0.5	-0.3	0.1	0.6
seasonalmean_tasmax (DJF) (°C)	0.1	0.3	0.5	0.0	0.2	0.4	-0.2	0.2	0.6
seasonalmean_tasmax (MAM) (°C)	0.1	0.3	0.5	-0.1	0.1	0.3	-0.6	-0.2	0.3
seasonalmean_tasmax (JJA) (°C)	0.2	0.4	0.6	0.3	0.6	0.9	-0.3	0.2	0.8
seasonalmean_tasmax (SON) (°C)	0.2	0.4	0.6	0.1	0.4	0.7	-0.3	0.2	0.6
tasmax_q50 (°C)	0.2	0.4	0.5	-0.1	0.1	0.4	-0.5	-0.1	0.4
tasmax_q99p9 (°C)	0.1	0.4	0.7	0.0	0.4	0.8	-0.5	0.1	0.8
annual_tasmax_ge_95F (%)	1.6	22.0	47.8	2.4	13.3	26.7	-20.4	0.5	29.1
annual_tasmax_ge_100F (%)	23.0	63.2	122.4	9.4	29.7	54.8	-20.3	12.6	74.8
annual_tasmax_ge_105F (%)	4.3	78.3	188.4	9.1	54.3	123.9	-31.3	13.4	123.4
annual_txx (°C)	0.2	0.5	0.8	-0.1	0.2	0.5	-0.9	-0.3	0.3
annualmean_tasmin (°C)	-0.2	0.0	0.3	0.2	0.5	0.7	-0.2	0.4	1.0
annual_tasmin_le_32F (%)	-10.7	-4.2	1.4	-13.6	-7.4	-3.1	-13.0	-4.0	2.6
annual_tnn (°C)	-0.9	-0.5	-0.1	-0.9	-0.5	0.0	-0.9	0.0	0.9

Table 3: Grid point evaluation metrics from the STAR-ESDM ensemble mean.

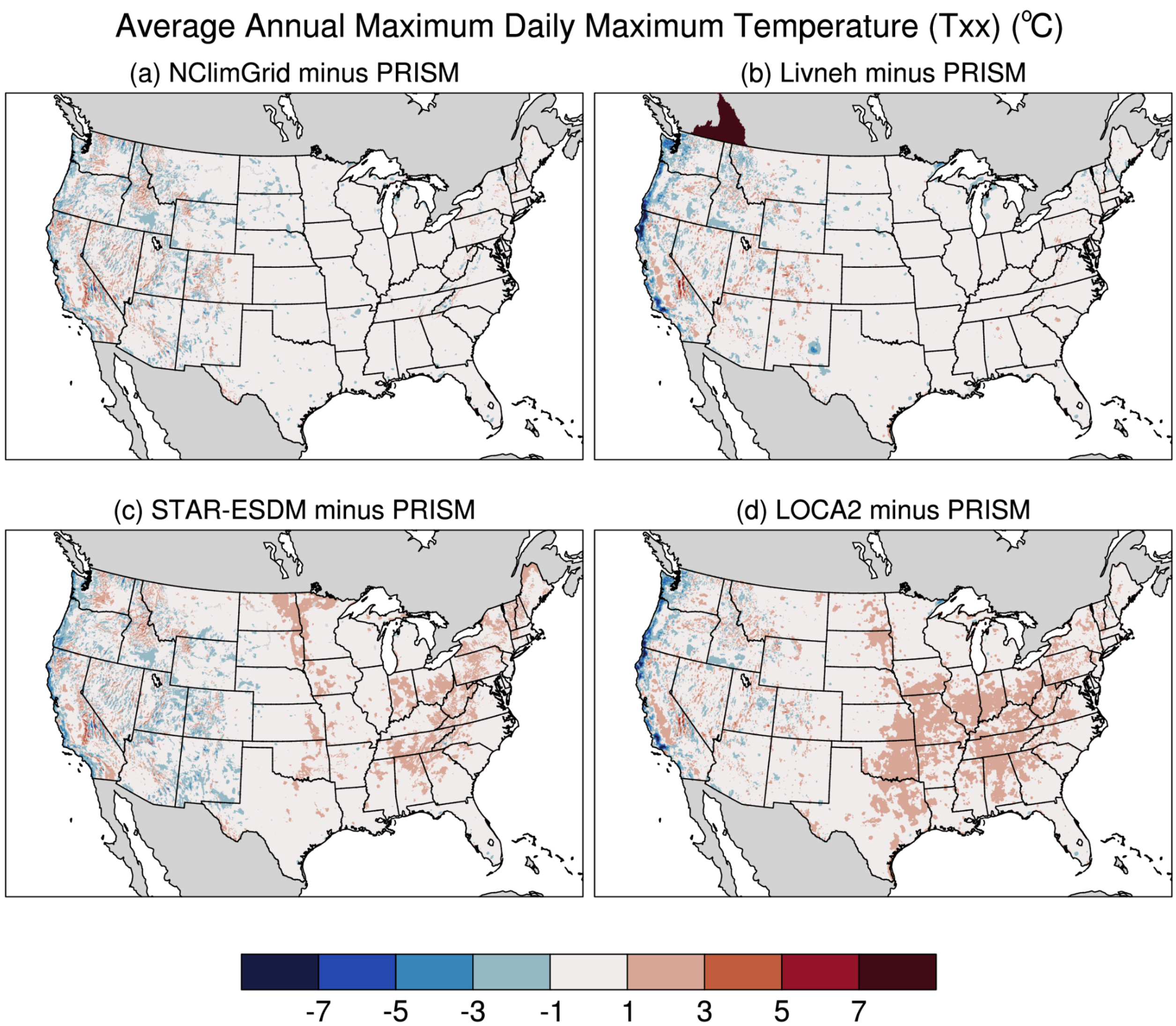

Figure 1: Average annual maximum temperature (Txx) difference from gridded observations and statistically downscaled products. Note the substantially cooler temperatures along the Pacific coast in Livneh and LOCA2, and warmer temperatures throughout the eastern United States in STAR-ESDM and LOCA2.

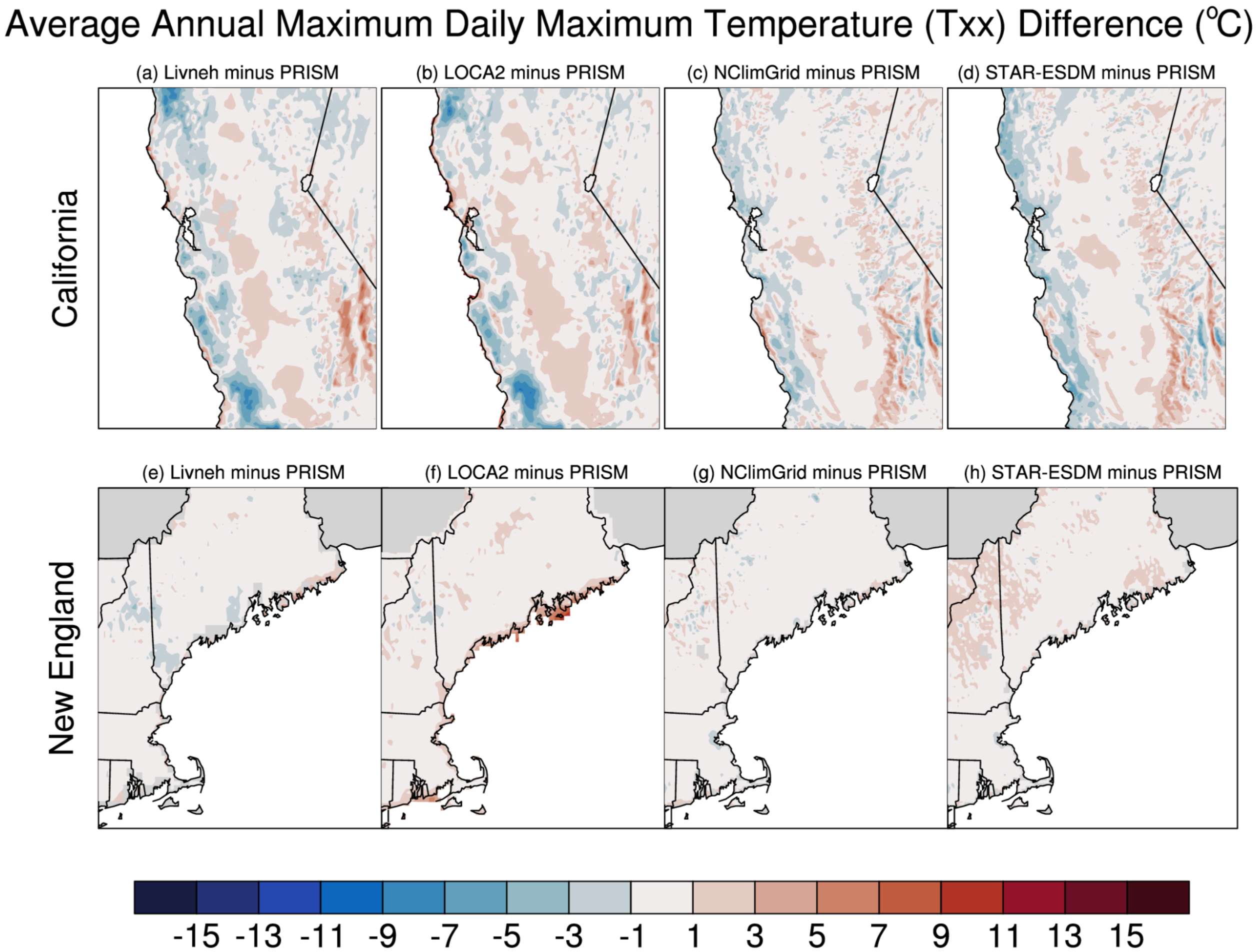

Figure 2: Average annual maximum temperature (Txx) difference between the gridded observations and statistically downscaled products and PRISM. Note the enhancement in this metric along the California coast and New England coast in the LOCA2 product that is partially inherited from the Livneh dataset.

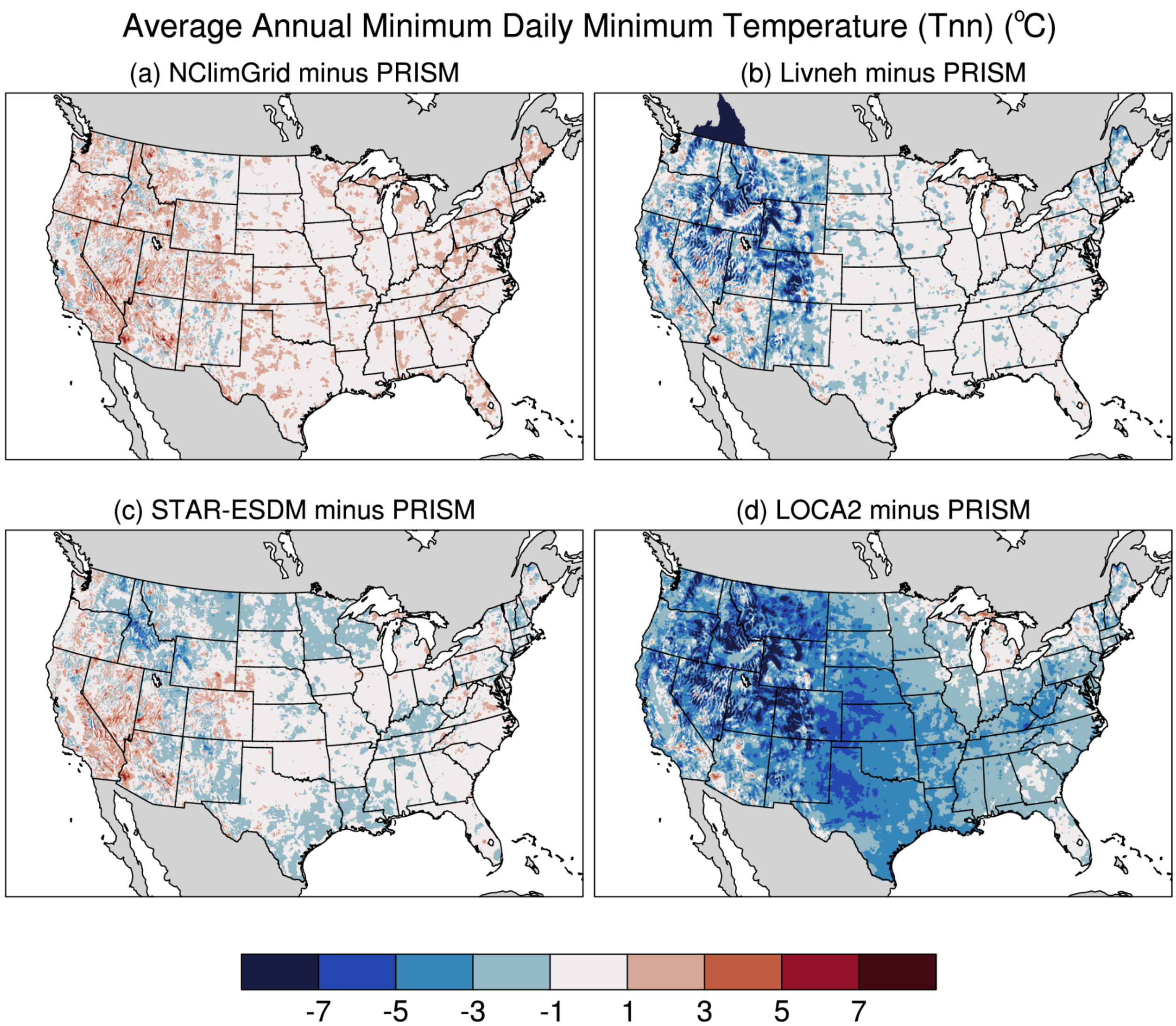

Figure 3: Average annual minimum daily minimum temperature difference from gridded observations and statistically downscaled products. Note the substantially cooler temperatures across the western United in Livneh, and throughout the United States in LOCA2.

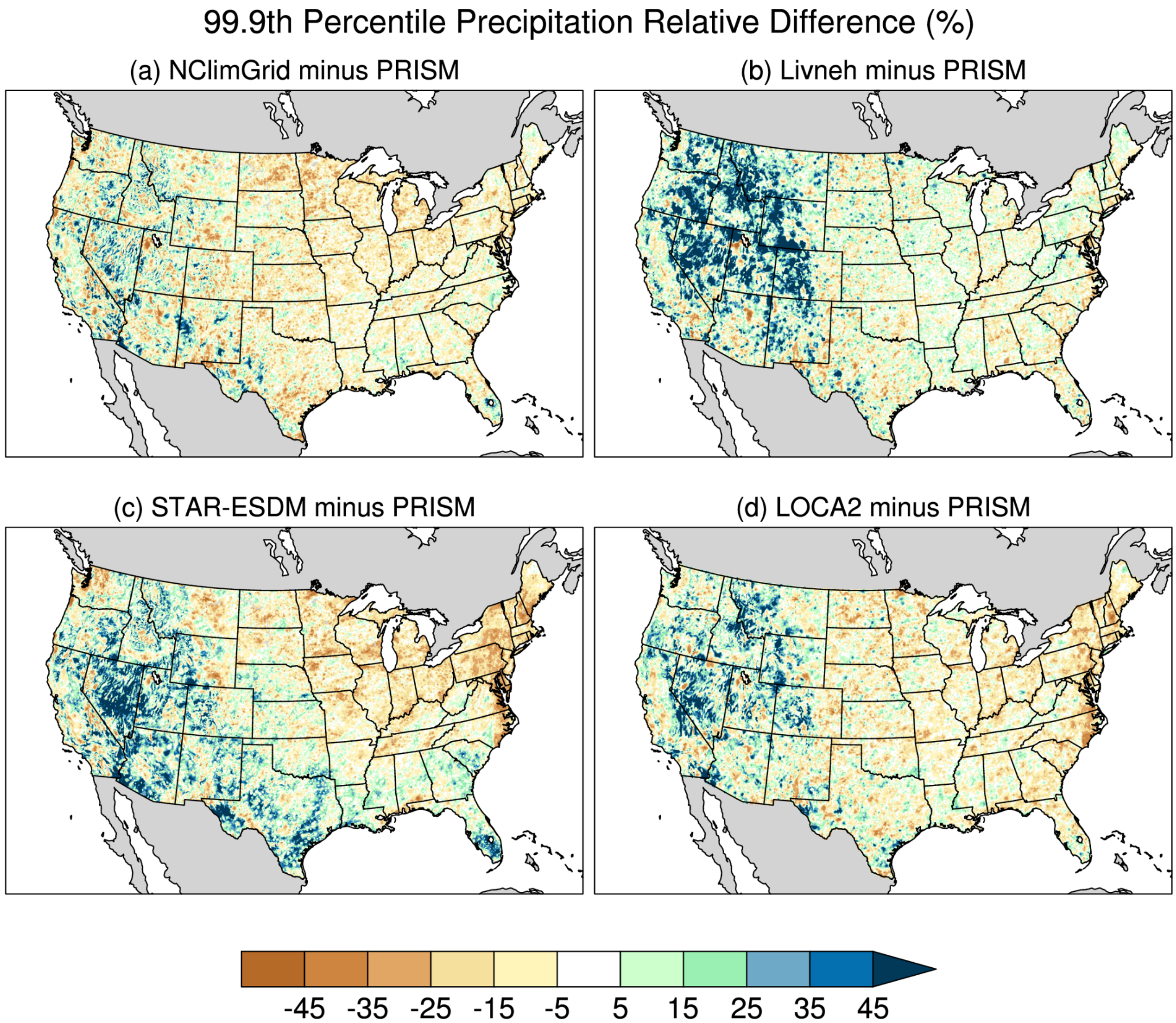

Figure 4: Relative difference in the 99.9th percentile of precipitation from observational and the statistically downscaled ensemble mean. Throughout the Northeast and Midwest, both LOCA2 and STAR-ESDM are drier than their driver product. LOCA2 is also drier than its driver product elsewhere in the U.S., while STAR-ESDM is wetter elsewhere.

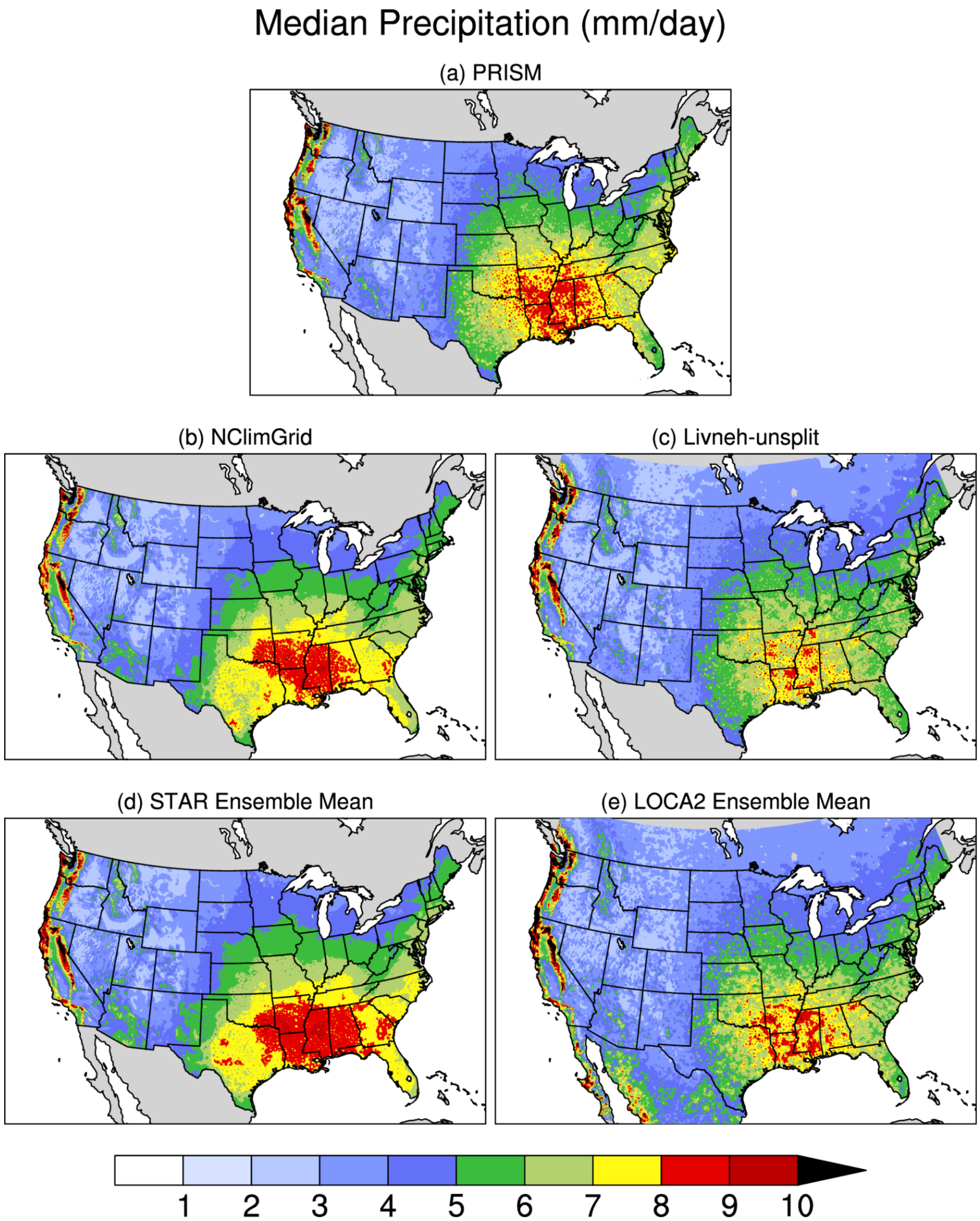

Figure 5: Median precipitation from gridded observations, STAR-ESDM and LOCA2. Dimpling around observing stations is apparent throughout the eastern United States in all products. PRISM, Livneh-unsplit and LOCA2 show enhancement in median precipitation near observing stations, while NClimGrid and STAR-ESDM exhibit less intense precipitation.

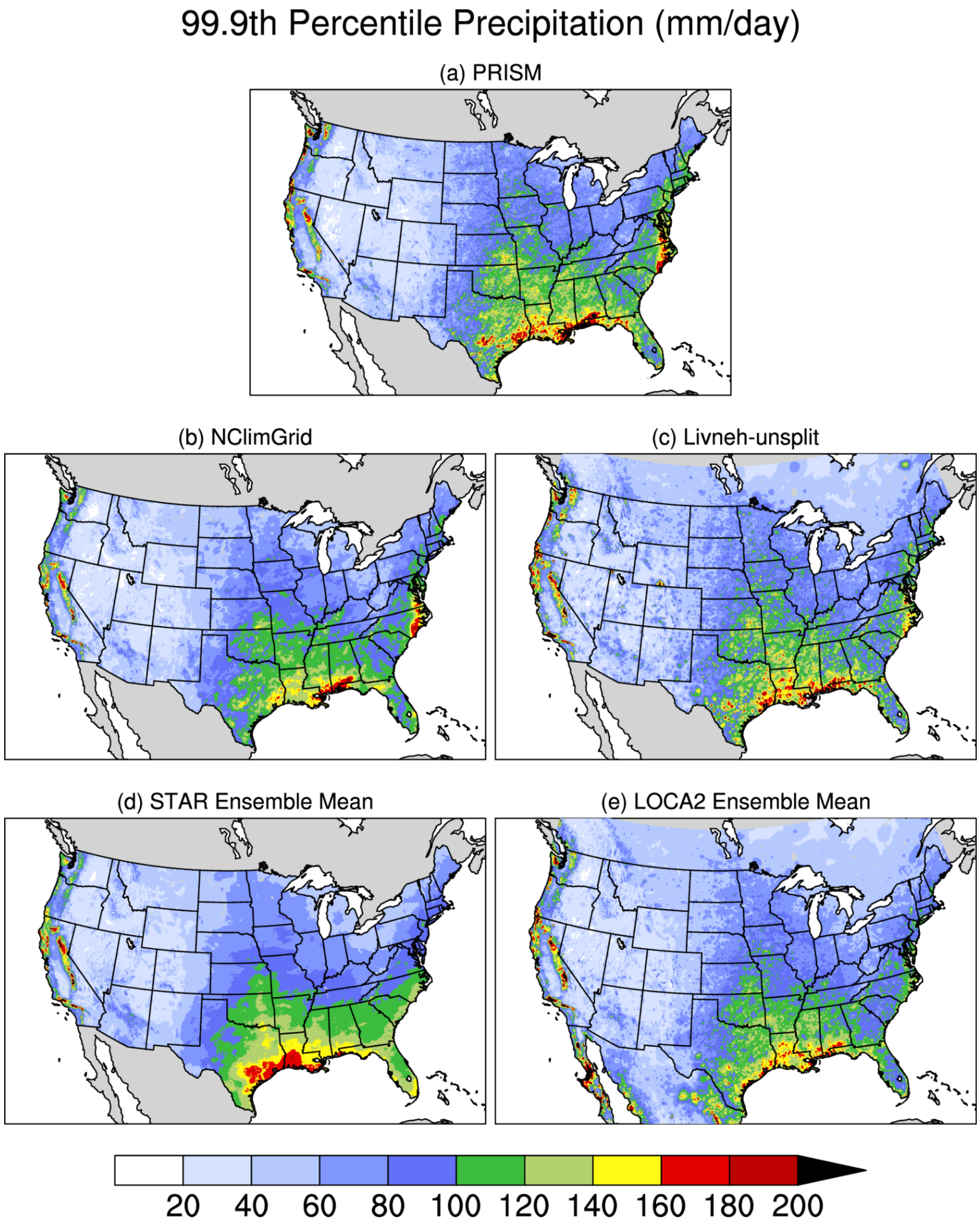

Figure 6: 99.9th percentile precipitation from gridded observations, STAR-ESDM and LOCA2. Dimpling around observing stations is apparent throughout the eastern United States in PRISM, Livneh-unsplit and LOCA2.

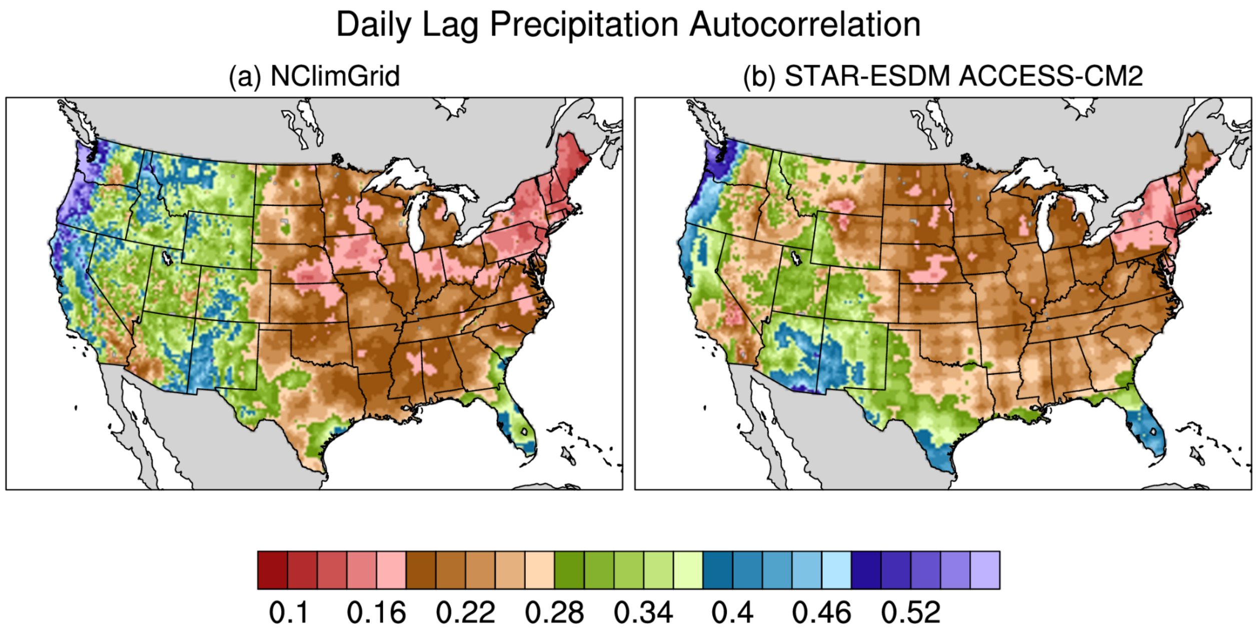

Figure 7: Daily lag precipitation autocorrelation from NClimGrid and one ensemble member of STAR-ESDM derived from ACCESS-CM2 r1i1p1f1. Note the presence of grid imprinting from the GCM over much of the United States.