Assessing the Ability of Gridded Datasets to Identify Local Extreme Weather Events

Nina Grant[1], Xinyue Luo[2], Delavane Diaz[3], and Erik T Smith[4]

[1]Rutgers University
[2]Rice University
[3]Electric Power Research Institute
[4]Kent State University

March 07, 2024

Abstract

Reanalysis products, or gridded datasets more broadly, are often used in place of surface observations. While they have been shown to capture long-term statistics on global or regional levels, it is still unclear how well they perform at the tails of the distribution, especially on daily timescales. Four widely used datasets, ERA5, ERA5-Land, MERRA-2, and PRISM, were assessed for their ability to capture extreme heat, extreme cold, and heavy precipitation events over the contiguous US (CONUS). While biases are evident in each dataset, particularly across the western US for temperature and along the Gulf Coast for heavy precipitation, all datasets do reasonably well in capturing extreme events and trends. Extreme heat is better represented than extreme cold or heavy precipitation. While no dataset emerges as a clear best for extreme heat, PRISM generally performs best for extreme cold and the bias-adjusted MERRA-2 dataset generally performs best for heavy precipitation days.

Hosted file

`985142_0_art_file_11794033_s7tpx3.docx` available at https://authorea.com/users/722939/articles/707826-assessing-the-ability-of-gridded-datasets-to-identify-local-extreme-weather-events

Hosted file

`985142_0_supp_11793766_s7tshf.docx` available at https://authorea.com/users/722939/articles/707826-assessing-the-ability-of-gridded-datasets-to-identify-local-extreme-weather-events

Assessing the Ability of Gridded Datasets to Identify Local Extreme Weather Events

N. Grant[1,3], X. Luo[2,3], D. Diaz[3], and E. Smith[3]

[1]Rutgers University. [2]Rice University. [3]Electric Power Research Institute.

Corresponding author: Erik Smith (esmith@epri.com)

Key Points:

- Extreme heat days are better captured than extreme cold days
- While all datasets reproduce extreme heat days comparably, PRISM is generally most capable of reproducing extreme cold days
- MERRA-2 bias-adjusted precipitation data generally reproduces heavy precipitation days best

Abstract

Reanalysis products, or gridded datasets more broadly, are often used in place of surface observations. While they have been shown to capture long-term statistics on global or regional levels, it is still unclear how well they perform at the tails of the distribution, especially on daily timescales. Four widely used datasets, ERA5, ERA5-Land, MERRA-2, and PRISM, were assessed for their ability to capture extreme heat, extreme cold, and heavy precipitation events over the contiguous US (CONUS). While biases are evident in each dataset, particularly across the western US for temperature and along the Gulf Coast for heavy precipitation, all datasets do reasonably well in capturing extreme events and trends. Extreme heat is better represented than extreme cold or heavy precipitation. While no dataset emerges as a clear best for extreme heat, PRISM generally performs best for extreme cold and the bias-adjusted MERRA-2 dataset generally performs best for heavy precipitation days.

Plain Language Summary

Ground-based observations are the most accurate records we have of current and past surface weather, but surface observations lack spatial and temporal completeness. To address this problem, gridded datasets were developed, however, these datasets are not observations and should thus be interpreted differently. This research finds that while these gridded datasets generally do well in capturing extreme events, some datasets may be a better option depending on the hazard of interest. For instance, PRISM data is generally a closer match to surface observations during extreme cold days than the other datasets while the bias-adjusted MERRA-2 precipitation data performs best for heavy precipitation days. Regardless of the dataset, locations with challenging terrain, like the Mountain West and Gulf Coast, tend to have higher discrepancies between the gridded dataset and the surface observations. With more companies conducting physical risk assessments, knowing which datasets are available and how well they represent extremes in the area of interest is critical. These results can help determine which dataset would be best to use if assessing extreme temperature or heavy precipitation events and trends in these events.

1 Introduction

Reanalysis and gridded datasets are used in numerous applications including detection and attribution studies, climate model validation, power system planning, renewable energy analysis, agricultural modeling, and more. As of Jan 2024, the NCEP-NCAR dataset, the first reanalysis dataset created by NOAA, had been citated more than 34,000 times (Kalnay et al., 1996; Parker, 2016). Despite their prolific usage, reanalysis datasets are not actual observations and the process by which gridded datasets are created makes errors inherent. However, gridded datasets are not meant to replace surface observations, rather they are meant to fill spatial and temporal gaps between observations. So, while differences between surface observations and gridded data can be misinterpreted as errors, they are simply different and should be interpreted as so. Nonetheless, it is an important exercise to quantify these differences so that end-users of gridded data products can better understand appropriate applications as well as the uncertainties that may arise from the use of gridded datasets.

Uncertainty in gridded datasets can arise from several sources including, quality of the observational data, the density or structure of the observational network, and the interpolation or assimilation method chosen (Dunn et al., 2014; Ge et al., 2023; Yin et al., 2015). Structural uncertainty (data selection and analysis method) has been found to have the greatest effect on extremes in gridded datasets (Dunn et al., 2014; Hofstra et al., 2010). Efforts to evaluate gridded datasets use techniques and metrics such as cross-validation, ensemble uncertainty estimates, probability distribution functions, and spatial correlation (Gross et al., 2018; Parker, 2016; Pitman & Perkins, 2009; Thorne & Vose, 2010). However, most studies focus on the skill at capturing the means, not the extremes, and focus on annual, seasonal, or monthly scales as opposed to daily or sub-daily time scales (Pitman & Perkins, 2009).

Past assessments of gridded dataset skill in capturing extremes have found that looking only at global spatial scales or annual to monthly timescales can obscure major differences between datasets (Ge et al., 2023; Gross et al., 2018; Pitman & Perkins, 2009). While most gridded datasets are consistent for mean temperatures, other moments like variance and skewness are not robust across datasets (Gross et al., 2018). These studies have also found that that method of calculation (interpolation or assimilation method, observational network used, observational record length, order of operation, etc.), especially the underlying observational network, has the largest effect on results, while changes to specific parameters have little effect (Dunn et al., 2014; Ge et al., 2023; Yin et al., 2015). Most studies concluded that it is best to use different datasets for different research questions (such as trend analysis, instantaneous field estimates, regional studies, particular variables of interest), and, where the network is sparce or datasets inconsistent, it is recommended that a range be determined rather than averaging the reanalyses (Coughlan de Perez et al., 2023; Dunn et al., 2014; Dunn et al., 2020; Ge et al., 2023; Lader et al., 2016; Pitman & Perkins, 2009; Thorne & Vose, 2010; Yin et al., 2015).

There is little peer-reviewed guidance for those using gridded datasets on how to evaluate the available datasets and which datasets may be best for specific regions or for specific climate hazards. With more companies conducting physical risk assessments, knowing which datasets are available and how well they represent extremes in the area of interest is critical. In this paper, we assess the ability of 4 gridded datasets, ERA5, ERA5-Land, MERRA-2, and PRISM, to capture the magnitude and timing of extreme heat, extreme cold, and heavy precipitation over the contiguous United States (CONUS). The magnitude of the differences are compared across datasets for each surface station to show dataset bias by location and hazard. This kind of analysis provides insight on the differences between gridded datasets and surface observations when it comes to representing extreme weather events at the local level. These results can help determine which dataset may be best to use if assessing extreme temperature or heavy precipitation events.

2 Methods

2.1 Weather Data

Observational weather station data were retrieved using the meteostat Python library (https://dev.meteostat.net/python/), which accesses publicly available surface station data from the National Oceanic and Atmospheric Administration. Only US stations with a temporal coverage of at least 80% from 1981 to 2021 were included. This resulted in 317 stations across the contiguous US (CONUS) for temperature and 267 stations for precipitation.

The gridded datasets chosen for this study include ERA5 (Hersbach et al., 2020), ERA5-Land (Muñoz-Sabater et al., 2021), PRISM (Daly et al., 2008), and MERRA-2 (Gelaro et al., 2017). ERA5 was developed by the ECMWF Copernicus Climate Change Service as the successor to ERA-Interim. ERA5 has the longest temporal coverage of the datasets, with hourly data extending back to 1940 and a spatial resolution of 31 km. ERA5-Land is a 9 km land-only model of ERA5, forced with the ERA5 atmospheric output. MERRA-2 is the second iteration of MERRA global reanalysis dataset. Unlike the other datasets, MERRA-2 includes bias-adjusted precipitation data, but has the lowest resolution of the selected datasets at approximately 50 km.

PRISM is an interpolated gridded dataset calculated with a climate-elevation regression based on location, elevation, topography, orography, proximity to coasts, the vertical atmospheric layer, and includes 10,000 spatially quality-controlled temperature surface stations and 13,000 precipitation stations (Daly et al., 2008). It has the highest resolution of our selected datasets (4 km) and is the only dataset based on statistical interpolation of surface station observations. Conversely, reanalysis data is created by assimilating various types of observational data into numerical weather models (Kalnay et al., 1996). In other words, reanalysis data relies on physical principals to create spatially and temporally continuous datasets, whereas interpolated data relies on statistical methods.

The grid points from the gridded datasets overlapping the surface stations were used to compare events. Land-only datasets (PRISM and ERA5-Land) did not always overlap certain coastal stations. Thus, the neighboring grid cells were assessed for the best match to the surface station. Island and buoy locations were removed from the station list.

2.2 Metrics

In this study, extreme heat days (EHDs) and extreme cold days (ECDs) are defined as temperatures above the 95th percentile of summer (June-August, JJA) temperatures or below the 5th percentile of winter (December-February, DJF) temperatures. Hot and cold days (HD and CD) are defined as days above or below the 80th and 20th summer and winter percentiles respectively. Heavy precipitation days (HPD) are defined as days that exceed the wet day (>1 mm) 95th percentile, and precipitation days (PD) are defined at wet days exceeding the 80th percentile. Percentiles are based on the climate normal period, 1991-2020.

Mean absolute error (MAE) is used as the primary metric to quantify the ability of each gridded product to capture the magnitude of the extreme event at the surface stations. The MAE is calculated between the station observations and gridded data based on the days in which the station observation exceeds the 5th or 95th percentile (just 95th for precipitation). For precipitation, we also include mean absolute percentage error (MAPE) to account and scale for the large range of daily precipitation values across the CONUS.

To assess the timing with which the gridded datasets can accurately represent extreme days, we calculate a match percentage. We define the match percentage of EHDs, ECDs, and HPDs as the number of days when both observations and the gridded datasets exceed the respective temperature or precipitation percentiles. To allow for near misses and locations with small annual temperature variation, we extend our percentile threshold from the 95th and 5th to the 80th and 20th and allow values within 0.5°C of the percentile threshold. We also use the 80th percentile as the threshold for precipitation.

Lastly, we assess the ability of the gridded datasets to capture trends in extremes by calculating the Theil-Sen slope estimation for the magnitude of all EHDs and ECDs and the frequency of HPDs. The frequency of HPDs is defined as number of days each year that exceed the climate normal 95th percentile. We utilize frequency rather than magnitude to account for the fact that precipitation is generally a small-scale process that is not well resolved in gridded datasets and does not capture magnitude as well as temperature.

3 Results

3.1 Extreme Temperature

Gridded datasets are generally better at reproducing the magnitude of extreme heat days (EHDs) than extreme cold days (ECDs; Figure 1). Apart from coastlines (Gulf Coast and Great Lakes), stations in the eastern half of the US tend to have lower MAEs for EHDs while stations in the western US tend to have higher MAEs for both EHDs and ECDs across all datasets. For ECDs, ERA5 and ERA5-Land generally have lower MAEs in the Midwest and Mid-Atlantic region, while MERRA-2 generally has lower MAEs in the southeast. While all four datasets reproduce the magnitude of EHDs comparably, PRISM is generally much better at reproducing the magnitude of ECDs at surface station locations across CONUS. This is likely, at least in part, due to PRISM having the highest spatial resolution of four datasets. PRISM is also the only non-reanalysis dataset, thus the statistical methods by which the gridded data is created may anchor

Figure 1. Mean absolute error of extreme heat days (>95th percentile) (left 4 panels) and extreme cold days (<5th percentile) (right 4 panels)

the data more closely to the surface observations and allow for a better match. These results averaged for all surface stations across the CONUS are shown in Figure S1.

While the MAE for ECDs was generally higher than EHDs across the gridded datasets, the timing, or match percentage of cold days (CDs) tends to be better captured than hot days (HDs; Figure S2). MERRA-2 reproduces the timing of HDs across the central US better than the other datasets but is generally comparable to or worse than other datasets in the western US. Outside of the central US where the MERRA-2 performs best, HDs are reproduced comparably across datasets. While all datasets reproduce CDs across the Midwest and Northeast well, PRISM generally reproduces CDs across the CONUS better than the other 3 datasets. The lowest match percentages for HDs tend to be in the western US and near the coast in Florida, the Gulf Coast, New England, and the Great Lakes. For CDs, the lowest match percentages tend to be in the western US and along the Gulf and Florida coastlines. It should be noted that bias in the gridded datasets may contribute to high match percentages. Figure S3 shows the direction of the errors and Figure S4 shows how often the gridded datasets produce false alarms (days exceeding the percentile threshold and the surface station by 1.5°C). PRISM and MERRA-2 tend to have more false alarms than ERA5 or ERA5-Land for HDs and CDs. In other words, high match percentages in PRISM and MERRA-2 may, in-part, be due to the datasets routinely overestimating the magnitude of HDs and CDs.

In contrast to Sheridan et al. (2020), which focused on apparent temperature in gridded datasets, we find the largest differences in temperature-only calculations for ECDs. The inclusion of humidity and 10-m wind speeds in the apparent temperature calculation may contribute to the differences in the findings. Other studies have also found larger discrepancies in the cold extremes (Gross et al., 2018; Pitman & Perkins, 2009; You et al., 2013). Like Gross et al. (2018), this research suggests additional studies are needed to determine why cold extremes generally exhibit larger differences from surface observations.

While the trends in EHDs and ECDs are similar to the trends observed at surface stations, there are notable differences. Trends in ECDs are found to be generally overestimated (positive) while trends in EHDs show more regional variation (Figure 2). Moreover, differences in the

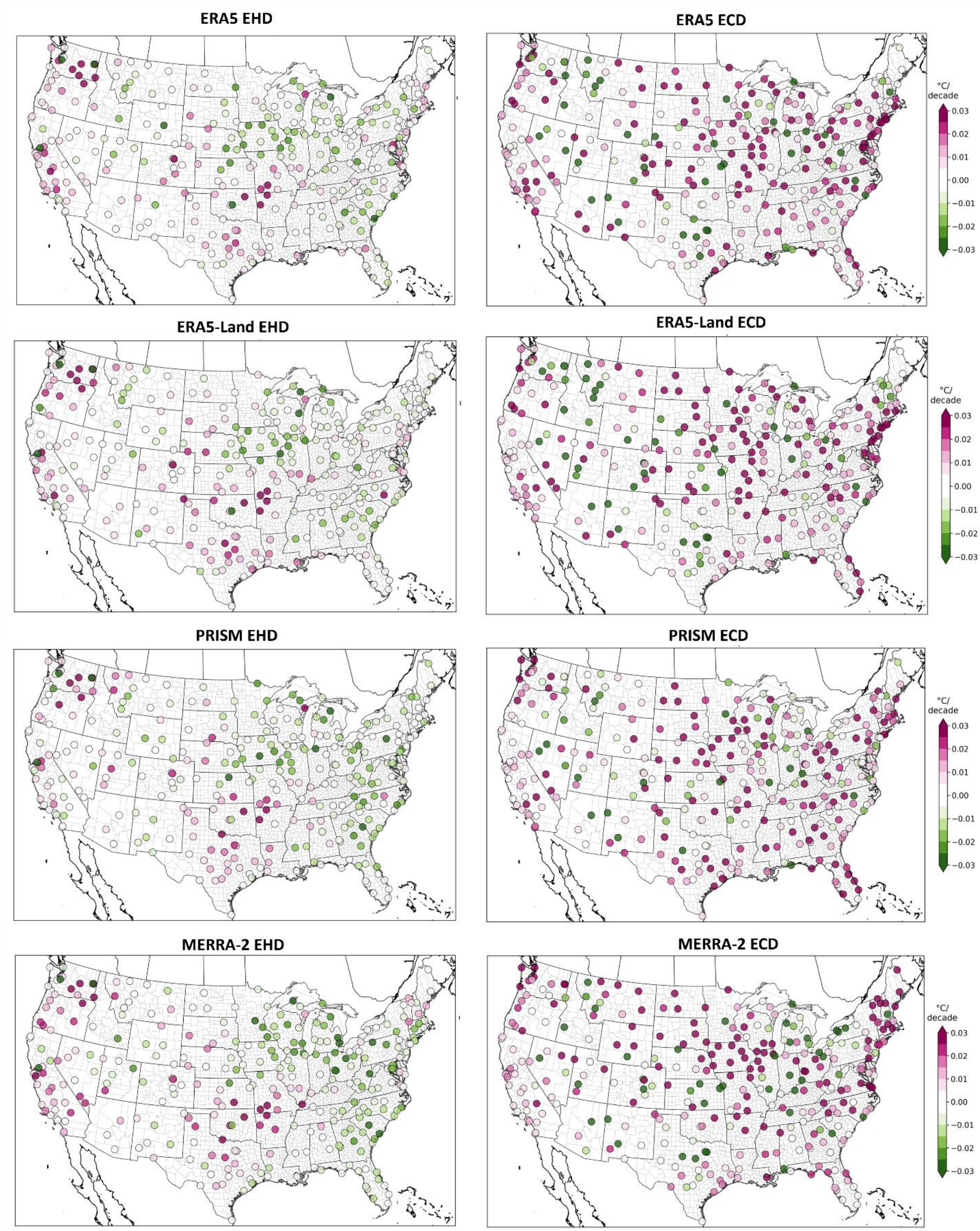

Figure 2. Difference in extreme heat day magnitude trends (left 4 panels) and extreme cold day magnitude trends (right 4 panels) between gridded datasets and surface observations.

ECD trends tend to be larger than for EHDs. For EHDs, the gridded datasets tend to underestimate the trends for eastern and midwestern US, while overestimating trends in the south-central and western US. All datasets reproduce trends comparably and have similar regional biases.

3.2 Extreme Precipitation

Figure 3 shows the MAE and MAPE for heavy precipitation days (HPDs). In contrast to EHDs and ECDs, the Mountain West region of the US consistently has the lowest errors in HPD magnitude, while the Gulf Coast and Southeast have the highest. This is likely due to the eastern US having higher precipitation than the western US. The MERRA-2 bias-adjusted product is better at reproducing HPDs and has a smaller MAE and MAPE compared to the other gridded datasets. To better account for the regionality of precipitation, MAPE is used. This effectively eliminates the spatial pattern observed in the MAE. Unlike EHDs and ECDs, PRISM tends to have the highest MAE and MAPE for HPDs across the CONUS. The statistical method PRISM uses to interpolate precipitation may contribute to larger differences in HPDs compared to stations. While bias-adjustment is shown to provide some benefit when representing HPDs with gridded data in MERRA-2, it should be considered whether this improvement holds for locations with no surface observations.

It would stand to reason that gridded datasets capture large-scale precipitation events better than small-scale, convective events, and, therefore, may do better representing heavy precipitation events in seasons where convective precipitation accounts for fewer of the heavy precipitation events. To help disentangle possible reasons for the large differences in HPDs, we separate summer (JJA) and winter (DJF) precipitation as a proxy for convective vs synoptic driven HPDs. For both MAE and match percentage, the differences were smallest for winter HPDs (defined as days exceeding the 1991-2020 DJF 95th percentile), suggesting convective precipitation is likely more challenging to represent in gridded datasets than large-scale, synoptic events (Figures S6 and S7).

Results for the PD match percentage skill are shown in Figure S6. Compared to HDs and CDs, the timing of PDs is not well captured by any gridded dataset. Match percentages are generally below 70% across CONUS in all four gridded datasets with PRISM match percentages generally below 40%. MERRA-2, ERA5 and ERA5-Land have higher match percentages along the West Coast and the lowest match percentages along the Gulf Coast. The spatial distribution

Figure 3. Mean absolute error (left 4 panels) and mean absolute percentage error (right 4 panels) of heavy precipitation days (>95th percentile) between gridded datasets and stations

of precipitation match percentage skill is much more uniform that those for extreme temperatures. As with extreme temperatures, false alarm rates are calculated for HPDs. ERA5, ERA5-Land, and PRISM have false alarm rates above 50% across much of the CONUS while MERRA-2 false alarm rates below 30% across most of the CONUS. Again, bias-adjustment proves to enhance the ability of gridded data to reproduce HPD timing relative to non-bias-adjusted datasets.

Observed trends in HPD frequency are shown in Figure 4. For much of the CONUS, trends in the magnitude of HPDs is nearly identical to surface observation trends. Larger differences are generally observed across the eastern half of the US. More specifically, differences in trends are largest around the Great Lakes in ERA5, ERA5-Land, and MERRA-2, whereas PRISM has larger differences across much of the eastern US. PRISM also tends to have a consistent positive bias in HPD trends relative to surface station trends. ERA5 and ERA5-Land have similar differences and spatial patterns in trends compared to MERRA-2, but MERRA-2 tends to overestimate trends (positive bias) compared to ERA5 and ERA5-Land which tends to underestimate trends in HPDs.

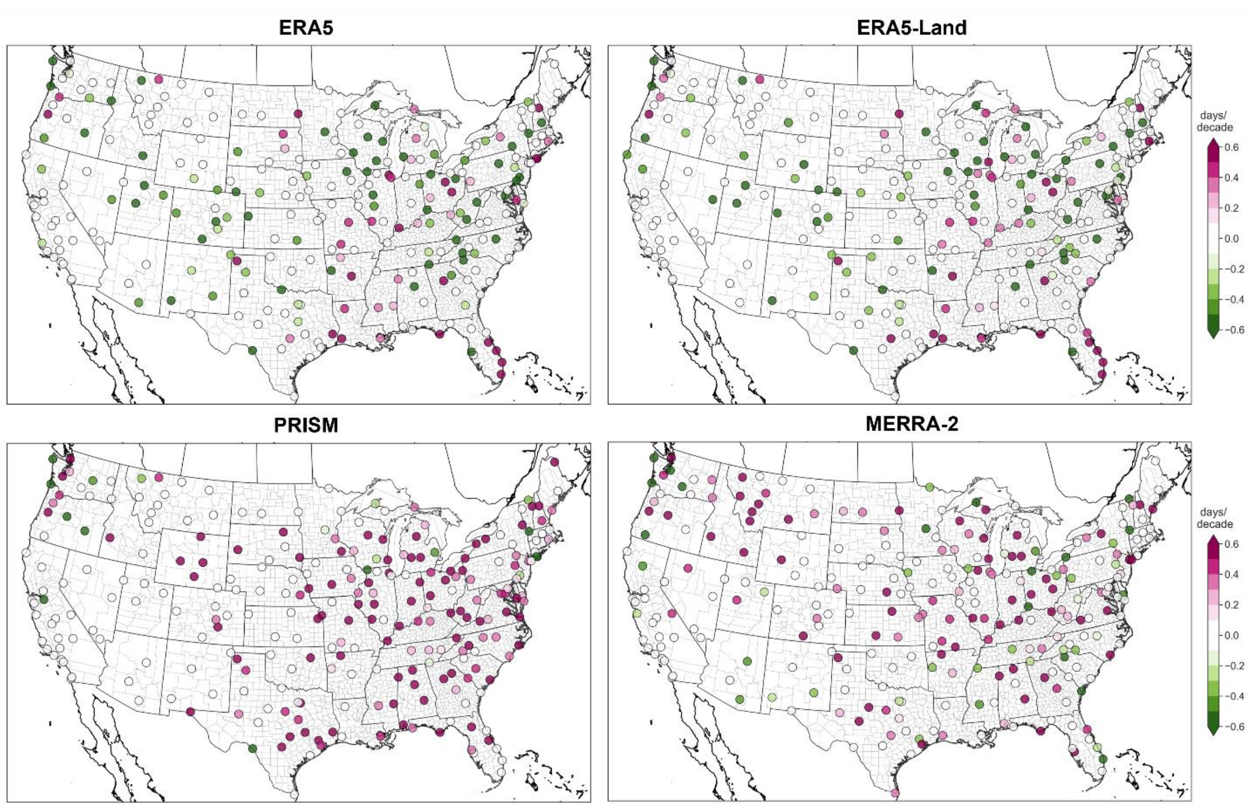

Figure 4. Difference in decadal extreme precipitation (>95th percentile) frequency trends (gridded dataset minus station observation).

4 Conclusion

Surface observations are the most accurate records we have of historical weather, but the spatial gaps, missing data, and inconsistent lengths of record create challenges in using the data. Gridded datasets were created to solve these problems, but they are inherently different from point observations. Most notably, gridded datasets are meant to represent a much larger area than point observations. In other words, a 55 km grid cell is tasked with representing a much larger and potentially more diverse area than a single measurement from a surface station. Thus, when gridded datasets differ from surface stations, these differences should not be viewed as errors. Conversely, surface observations are a key input into gridded datasets and these observations will intrinsically weight the value of the grid cell. For this reason, gridded data is often a very close match to surface station observations.

The results of this study suggest that while all four gridded datasets do reasonably well in reproducing extreme events, some datasets are better able to reproduce extremes in certain regions and for specific hazards. All four datasets reproduce the magnitude and timing (match percentage) of extreme heat days (EHDs) comparably, but PRISM most closely reproduces the magnitude and timing of extreme cold days (ECDs). While PRISM tends to represent ECDs better than EHDs, ERA5, ERA5-Land, and MERRA-2 all reproduce EHDs better than ECDs. For heavy precipitation days (HPDs), all four datasets have the larger MAEs from surface stations in wetter locations, like the Southeast, but the MAPE effectively eliminates the regional differences. Unlike extreme temperatures, PRISM is generally the least capable of reproducing HPDs across the CONUS. PRISM's larger differences may be attributed to the statistical interpolation being used as opposed to a physical approach (numerical weather prediction model). The bias-adjusted precipitation variable from MERRA-2, though the coarsest spatial resolution of all datasets, reproduces the magnitude of HPDs best. This suggests bias-adjustment can overcome at least some of the limitations with representing precipitation in coarse datasets.

Using gridded datasets for climate trends is generally not recommended, though there is little empirical evidence for this. This research shows that while the trends in EHDs are different, they are generally very close to the observed trends from stations. The trends in ECDs, however, have larger differences and a consistent positive bias. That is, the trends in the gridded datasets tend to show more warming than was otherwise observed at the surface station, thus one should use caution when estimating trends in extreme cold from gridded data. In general, all four datasets are comparable with respect to differences in EHD and ECD magnitude trends. For heavy precipitation, PRISM and MERRA-2 have a general positive bias and ERA5 and ERA5-Land, though generally a closer match to the station trend, have a general negative bias. In other words, ERA5 and ERA5-Land tend to overestimate the increase in HPDs while MERRA-2 and PRISM tend to show smaller increase or even decrease in HPDs.

All four gridded datasets tend to reproduce extreme temperature events in the eastern US better than the western US. Complex terrain and microclimate in the western US create challenging dynamics that are difficult to capture in coarse gridded datasets. For heavy precipitation, locations with more frequent heavy precipitation events and larger annual precipitation totals, such as the southeastern US, tend to have larger differences between surface observations and gridded datasets. While reanalysis products have been shown to represent averages well (Gross et al., 2018; Pitman & Perkins, 2009), these results suggest that caution should be used when using gridded datasets for extreme events. For a more robust analysis of extremes, it would be best to quantify the range of uncertainty across multiple gridded datasets as well as nearby surface observations.

Future work will add additional gridded datasets, such as the recently released CONUS404 data, which is a downscaled ERA5 dataset aimed at hydrological applications (Rasmussen et al., 2023).

Open Research

The data used in this study can be found online for download at the following locations. ERA5 and ERA5-Land daily surface temperature and precipitation data are available from the European Center for Medium-Range Weather Forecasts (ECMWF) via the Copernicus Climate Change Service (at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels and https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form). PRISM air temperature and precipitation data are available from the PRISM Climate Group at Oregon State (at https://prism.oregonstate.edu/downloads/). MERRA-2 data is made available on the National Aeronautics and Space Administration (NASA) Goddard Earth Sciences Data and Information Services Center (GES DISC). The code developed for this study can be made available upon request.

References

Coughlan de Perez, E., Arrighi, J., & Marunye, J. (2023). Challenging the universality of heatwave definitions: Gridded temperature discrepancies across climate regions. *Climatic Change*, *176*(12), 167. https://doi.org/10.1007/s10584-023-03641-x

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

Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A. A., Brunet, M., Caesar, J., Chagnaud, G., Cheng, V., Cinco, T., Durre, I., de Guzman, R., Htay, T. M., Wan Ibadullah, W. M., … Bin Hj Yussof, M. N. (2020). Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, *125*(16), e2019JD032263. https://doi.org/10.1029/2019JD032263

Dunn, R. J. H., Donat, M. G., & Alexander, L. V. (2014). Investigating uncertainties in global gridded datasets of climate extremes. *Climate of the Past*, *10*(6), 2171–2199. https://doi.org/10.5194/cp-10-2171-2014

Ge, S., Jiang, C., Wang, J., & Liu, S. (2023). Analyzing temperature and precipitation extremes in China using multiple gridded datasets: A comparative evaluation. *Weather and Climate Extremes*, *42*, 100614. https://doi.org/10.1016/j.wace.2023.100614

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. da, Gu, W., … Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gross, M. H., Donat, M. G., Alexander, L. V., & Sisson, S. A. (2018). The Sensitivity of Daily Temperature Variability and Extremes to Dataset Choice. *Journal of Climate*, *31*(4), 1337–1359. https://doi.org/10.1175/JCLI-D-17-0243.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, *35*(5), 841–858. https://doi.org/10.1007/s00382-009-0698-1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., … Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Lader, R., Bhatt, U. S., Walsh, J. E., Rupp, T. S., & Bieniek, P. A. (2016). Two-Meter Temperature and Precipitation from Atmospheric Reanalysis Evaluated for Alaska. *Journal of Applied Meteorology and Climatology*, *55*(4), 901–922. https://doi.org/10.1175/JAMC-D-15-0162.1

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Parker, W. S. (2016). Reanalyses and Observations: What's the Difference? *Bulletin of the American Meteorological Society*, *97*(9), 1565–1572. https://doi.org/10.1175/BAMS-D-14-00226.1

Pitman, A. J., & Perkins, S. E. (2009). Global and Regional Comparison of Daily 2-m and 1000-hPa Maximum and Minimum Temperatures in Three Global Reanalyses. *Journal of Climate*, *22*(17), 4667–4681. https://doi.org/10.1175/2009JCLI2799.1

Sheridan, S. C., Lee, C. C., & Smith, E. T. (2020). A Comparison Between Station Observations and Reanalysis Data in the Identification of Extreme Temperature Events. *Geophysical Research Letters*, *47*(15), e2020GL088120. https://doi.org/10.1029/2020GL088120

Thorne, P. W., & Vose, R. S. (2010). Reanalyses Suitable for Characterizing Long-Term Trends. *Bulletin of the American Meteorological Society*, *91*(3), 353–362. https://doi.org/10.1175/2009BAMS2858.1

Yin, H., Donat, M. G., Alexander, L. V., & Sun, Y. (2015). Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *International Journal of Climatology*, *35*(10), 2809–2827. https://doi.org/10.1002/joc.4174

You, Q., Fraedrich, K., Min, J., Kang, S., Zhu, X., Ren, G., & Meng, X. (2013). Can temperature extremes in China be calculated from reanalysis? *Global and Planetary Change*, *111*, 268–279. https://doi.org/10.1016/j.gloplacha.2013.10.003