Effect of Baseline Period on Quantification of Climate Extremes over the United States

Natalie P. Thomas[1,2], Allison B. Marquardt Collow[1,2], Michael G. Bosilovich[2], and Amin Dezfuli[1,2]

[1]University of Maryland Baltimore County, Baltimore, Maryland

[2]Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, Maryland

Corresponding author: Natalie Thomas (natalie.p.thomas@nasa.gov)

Key Points:

- Updating the baseline period from 1981-2010 to 1991-2020 leads to significant changes in percentile-based extreme climate indices in the US
- Temperature indices show generally increased cold extremes and decreased warm extremes across the US when the baseline period is updated
- For precipitation indices, the later baseline period indicates fewer but more intense extreme events in the south and central US

Abstract

Extreme climate events are societally harmful and have increased in frequency and intensity in recent decades. Indices based on temperature and precipitation are a valuable way to quantify climate extremes. Certain indices are defined relative to percentiles, which are dependent on a climatological baseline period. In this study, indices computed using temperature and precipitation from the Modern Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) are calculated using percentiles from three baseline periods: 1981-2010, 1991-2020 and 1981-2020. Updating the baseline period from 1981-2010 to 1991-2020 leads to significant changes in the quantification of temperature and precipitation extremes over the United States over 1980-2021. Using the later baseline period indicates more cold extremes, fewer warm extremes, and fewer but more intense precipitation extremes throughout the US, with regional variation. Changing the baseline period can mislead the public and decision makers, potentially undermining the appropriate response to climate-related health risks.

Plain Language Summary

Indices computed using 2-meter air temperature and precipitation are used to represent extreme climate events such as heat waves, cold waves, heavy precipitation, and drought. Some indices are defined relative to percentile-based thresholds, which are computed using a baseline climatology period. The baseline climatology is typically a thirty-year period and is updated every ten years. This study examines how updating the baseline climatology period from 1981-2010 to 1991-2020 affects the quantification of climate extremes in the United States over 1980-2021. In general, since the 1991-2020 period is warmer than 1981-2010 throughout the United States, there are fewer warm extremes detected and more cold extremes detected when it is used as the baseline. The differences are most notable in the southwest and northeast United States. The changes in the precipitation indices vary throughout the country, but in certain parts of the southern and central United States, updating the baseline period leads to the detection of fewer but heavier extreme precipitation events. It is important to communicate the choice of baseline climatology period to prevent misinterpretation of the extreme climate indices and the comparison of different studies.

1 Introduction

Extreme climate events, including heatwaves, heavy precipitation, and drought, have a large impact on society through human health, destruction of infrastructure, ecological change, and economic losses. Indices where daily temperature or precipitation is compared to a threshold are a valuable tool for the monitoring and quantification of extremes across different regions (Zhang et al., 2011; Alexander et al., 2019; Dunn et al., 2022). Some indices use a percentile-based threshold, and thus are dependent on the choice of baseline period used to define the percentiles (Zhang et al., 2005; Dunn & Morice, 2022). As global and regional climate continues to change, the interpretation of extreme events is increasingly reliant on this baseline period, and this can be a source of confusion and ambiguity for the policy making community.

To have the best representation of the current climate, operational centers typically use a 30-year climate baseline period that shifts in time every ten years (Arguez et al., 2012), also known as a normal. However, due to the non-stationarity of climate, alternatives to the 30-year climate normal have been suggested (Livezey et al., 2007; Wilks, 2013; Wilks & Livezey, 2013). The World Meteorological Organization (WMO) suggests that the maximum amount of data should be included for the detection of extreme events due to their rare occurrence (Trewin, 2007). The appropriate baseline period may differ based on the application (i.e., Schreck et al., 2021).

The climate changes between long-term means (e.g., Kendon et al., 2020), so the shifting of the baseline period can affect the magnitude and interpretation of climate anomalies (Scherrer et al., 2006; Arguez & Vose, 2011). This issue has the potential to be exacerbated in the situation of climate extremes. Previous studies have shown linear trends in percentile-based extreme temperature indices to vary significantly with different baseline periods (Yosef et al., 2020; Dunn & Morice, 2022). Conversely, the transition to a new baseline was found to affect a drought index only marginally (Cammalleri et al., 2021).

In this study, we examine how updating the baseline period from 1981-2010 to either 1981-2020 or 1991-2020 affects the quantification and classification of climate extremes across the continental United States. We employ indices defined using 2-m temperature and precipitation from NASA's Modern Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al., 2017). Changing the baseline period can affect the perception of the public and decision makers, so it is crucial to understand and communicate how to interpret this change. This evaluation expands on the above-mentioned studies by focusing on distinct regions within the United States, including heatwave and precipitation indices, and examining seasonal variability of changes in the indices. This manuscript also serves to document differences between Version 1 (GMAO, 2020) and Version 2 (GMAO, 2022) of the MERRA-2 Monthly Extremes Detection Indices dataset. Data and methods are described in section 2, while section 3 outlines the changes in temperature and precipitation extreme climate indices with the updated baseline. Conclusions follow in section 4.

2 Data and Methods

2.1 MERRA-2

Data used in this study is from the MERRA-2 reanalysis (Gelaro et al., 2017) and is akin to the extremes detection indices file collection (Collow et al., 2022; GMAO 2020; GMAO 2022). Daily 2-m temperature and precipitation data from MERRA-2 are available at a spatial resolution of 0.625° longitude by 0.5° latitude from January 1980 to present (GMAO 2015a, b), though the current analysis is for 1980-2021. Precipitation used to generate the climate statistics is the model generated output, and not the observation corrected land-forcing precipitation (Reichle et al., 2017). An evaluation of the climate of MERRA-2 can be found in Bosilovich et al. (2015).

2.2 Percentile Calculation

Percentiles and extreme indices were derived using daily mean fields of precipitation and 2-m temperature, as well as daily minimum and maximum 2-m temperature from MERRA-2. Percentiles for each calendar day of the year were computed with the multi-year daily running percentile values (*ydrunpctl)* function from the Climate Data Operators toolbox (Schulzweida, 2022) with a 15-day running window. This differs from the 5-day window recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI), because this shorter window resulted in too much day-to-day variability in the percentiles across the United States. The 15-day window has been utilized in the past by Collow et al. (2016) and Thomas et al. (2020). Depending on the location, this may result in additional exceedances of warm extreme thresholds during the summer and fewer exceedances during the winter. There is minimal influence during the shoulder seasons. Zhang et al. (2005) evaluated differences between a 5-day and 25-day window and demonstrated that the 25-day window results in a smaller bias within the baseline period but could complicate the interpretation of more intense extreme events. Only days with at least 1 mm of precipitation were included in the percentile calculation for precipitation. Three baseline periods, 1981-2010, 1981-2020, and 1991-2020, were used for the percentile calculations to determine the dependency on the climatology period used.

2.3 Indices Calculation

Daily exceedances of the percentiles were detected in the MERRA-2 dataset for the years of 1980 through 2021 using the three sets of percentiles and aggregated into monthly indices representing extreme temperature and precipitation events, as well as heatwaves. These indices are analogous to those included in the MERRA-2 monthly extremes detection indices file collection (GMAO 2020; GMAO 2022), and most have been recommended for use by the ETCCDI (Alexander et al., 2006). More specific details pertaining to the extreme indices are given in Table 1. The selected indices are included in the MERRA-2 extremes detection indices data product and are also available for visualization on the Global Modeling and Assimilation Office's Framework for Live User-Invoked Data (FLUID) webpage, https://fluid.nccs.nasa.gov/reanalysis/extreme_merra2/. The heatwave related indices (HWD, HWF, and HWM) are based on Perkins and Alexander (2013) in which a heatwave occurs if the mean 2-meter temperature exceeds the calendar day 90th percentile for at least three consecutive days. The frequency of extreme precipitation events, R90d, R95d, and R99d, are given as a count of the number of events as opposed to the percentage of the total precipitation that is considered extreme as included in the Climpact list of indices (https://climpact-sci.org/indices/). The frequency of 99th percentile precipitation events was previously used to evaluate the underlying general circulation in MERRA-2 with respect to teleconnection patterns (Collow et al., 2017). The dependence on baseline period is assessed using the difference between a given index computed using two baseline periods over the entire MERRA-2 period (1980-2021). Significance of differences is assessed using a two-sample student's t-test at the 90% confidence level.

Table 1. Percentile-based indices included in this study. *Recommended by the Expert Team on Sector-Specific Climate Indices (ET-SCI; https://climpact-sci.org/indices/).

Index	Name	Calculation	Units
HWD*	Heat wave duration	Maximum length of consecutive days that satisfy heat wave criteria (daily mean 2 m temperature exceeds the 90th percentile for at least three consecutive days)	days
HWF*	Heat wave frequency	Count of days that satisfy heat wave criteria (see HWD)	count
HWM*	Heat wave magnitude	Mean temperature anomaly on days that satisfy heat wave criteria (see HWD)	K
R90p	Wet day precipitation	Mean precipitation on days that exceed the 90th percentile of precipitation	mm day^{-1}
R90d	Wet days	Count of days that exceed the 90th percentile of precipitation	count
R95p*	Very wet precipitation	Mean precipitation on days that exceed the 95th percentile of precipitation	mm day^{-1}
R95d	Very wet days	Count of days that exceed the 95th percentile of precipitation	count
R99p*	Extremely wet precipitation	Mean precipitation on days that exceed the 99th percentile of precipitation	mm day^{-1}
R99d	Extremely wet days	Count of days that exceed the 99th percentile of precipitation	count
TN10p*	Cold Nights	Percent of days with a minimum temperature below the 10th percentile	%
TX10p*	Cold Days	Percent of days with a maximum temperature below the 10th percentile	%
TN90p*	Warm Nights	Percent of days with a minimum temperature above the 90th percentile	%
TX90p*	Warm Days	Percent of days with a maximum temperature above the 90th percentile	%

3 Results

3.1 Percentile changes with changing baseline period

The average over all calendar days of the 90th percentile of daily mean 2-meter temperature and precipitation is shown in Fig. 1. During the 1981-2010 period, the 90th percentile of 2-meter temperature is greatest in the Southern Great Plains and smallest in the high-elevation areas in the Rocky Mountain range (Fig. 1a). When the baseline period is updated to 1991-2020 (Fig. 1c), the 90th percentile of 2-meter temperature increases throughout the US, with largest differences in the Southwest. Differences are significant everywhere except for a small region in the Northern Great Plains. The spatial pattern is similar to the differences in 30-year normals produced by the National Centers for Environmental Information (NCEI 2021). The differences when the baseline period is 1981-2020 are smaller, but still statistically significant throughout the continental US (Fig. 1e).

For precipitation, the 90th percentile averaged over all calendar days over the 1981-2010 period (Fig. 1b) shows higher values in the south-central US, eastern US, and Pacific Northwest, and lower values in the intermountain west. With the updated baseline period of 1991-2020, changes in the percentiles for precipitation are less spatially consistent than for temperature, but still significant in many regions. Parts of the Southern Great Plains through the Midwest and Southwest US show significantly larger precipitation percentiles with the updated climatology (Fig. 1d). This differs from the change in the NCEI precipitation normals, which shows a decrease over the Southwest US with the later period (NCEI 2021). With the 1981-2020 baseline period, the changes are more muted, but still significantly positive in these regions (Fig. 1f).

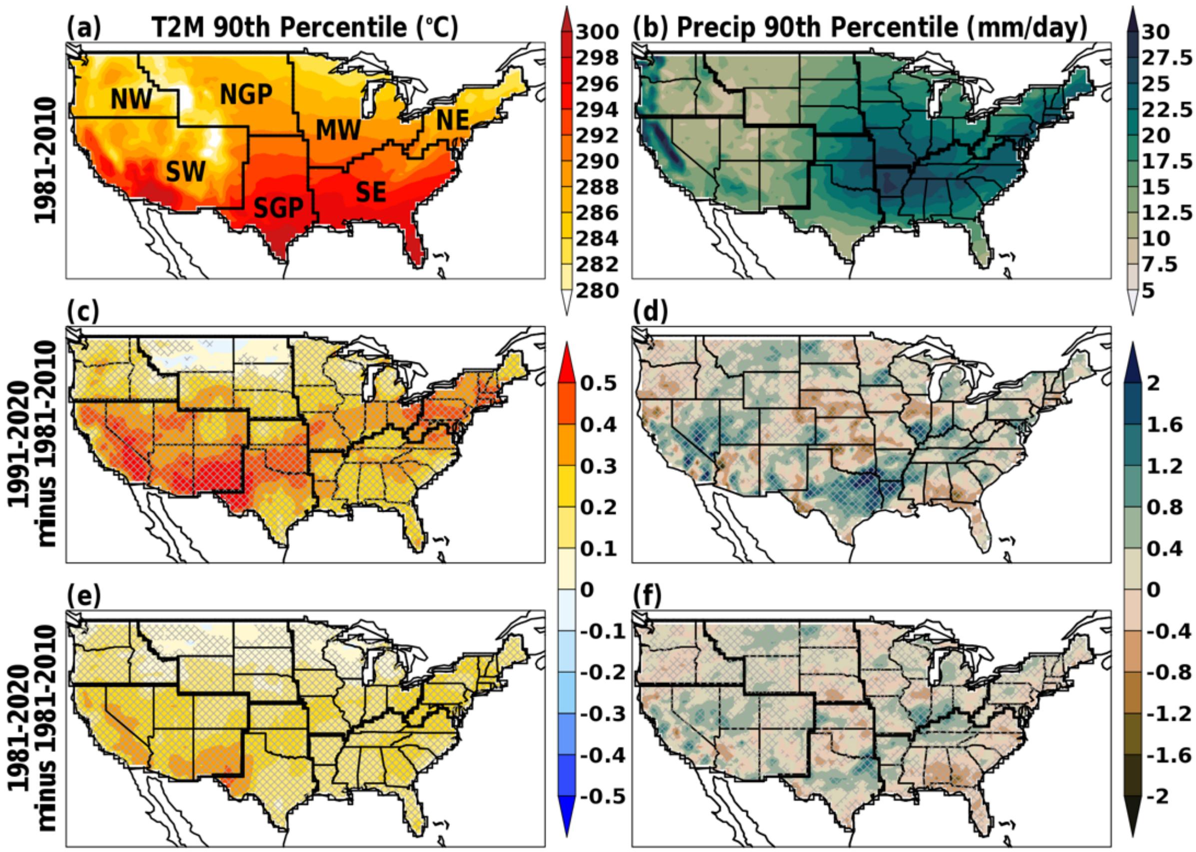

Figure 1. (a) Average of all calendar-day 90th percentiles of 2-m temperature computed using 1981-2010, (c) average difference over all calendar days between percentiles computed using 1991-2020 and 1981-2010, (e) average difference over all calendar days between percentiles computed using 1981-2020 and 1981-2010. (b,d,f) as in (a,c,e) but for the 90th percentile of precipitation. Grey hatching indicates where differences are significant at the 90% confidence level. Labels in (a) denote the regions used in Figure 2.

3.2 Changes in extreme climate indices

The identified changes in the temperature and precipitation percentiles will lead to changes in the extreme climate indices that are defined relative to them. Fig. 2 shows a summary of each of the percentile-dependent indices and how they change, on average, when the baseline period for percentiles is changed. Fig. 2 separates the percent differences relative to the 1981-2010 climatology period for each index into the regions of the US used in the National Climate Assessment (NCA; Wuebbles et al., 2017) and denoted in Fig. 1. In general, changes with baseline period are largest for the temperature indices using the 1991-2020 climatology (Fig. 2a). Temperature extremes defined using the 10th percentile (TN10p and TX10p) are more frequent with the updated climatology – with the later climatology period, there are more days and nights identified below the 10th percentile. The opposite is true for indices defined using the 90th

percentile (TN90p and TX90p) - the 1991-2020 climatology results in fewer identified extreme warm days and nights. Changes in heat wave frequency (HWF) indicate that the 1991-2020 climatology leads to fewer heat wave days on average in most regions, except in the Northern Great Plains, where changes in the 90th percentile of temperature were weak and insignificant when updating the baseline period (Fig. 1c). In general, the changes when using the 1981-2020 climatology (Fig. 2b) result in the same sign, but weaker in magnitude and significance, as is expected under a warming climate when moving to a longer reference period that includes more recent years (i.e., Fig. 1c, e).

For the precipitation indices, in general the updated climatology periods result in fewer days with extreme precipitation (R99d, R95d, R90d) but more precipitation on extreme days (R99p, R95p, R90p), with differences most significant for the 99th percentile indices (Fig. 2c-d). Unlike the temperature indices, the differences for R99d and R99p are larger when using the 1981-2020 climatology rather than the 1991-2020, likely due to the very rare nature of these events.

While the focus here is on the annual changes, variations across the seasonal cycle are also an important consideration and are included in the supplementary material (Figs. S1-S4). In general, changes in the minimum temperature indices (TN10p and TN90p) are stronger and more significant in summer (Fig. S3a) while the maximum temperature indices (TX10p and TX90p) are stronger and more significant in winter (Fig. S1a). The increase in detected heat wave frequency (HWF) in the Northern Great Plains is primarily a spring phenomenon (Fig. S2a). For precipitation indices, there are no significant changes during the winter or spring seasons when updating to the climatology period of 1991-2020 (Figs. S1c, S2c). Changes are most significant during summer, when the Southeast and Southwest regions show a decrease in the frequency of extreme precipitation events and increase in the amount of precipitation from an event (Figure S3c).

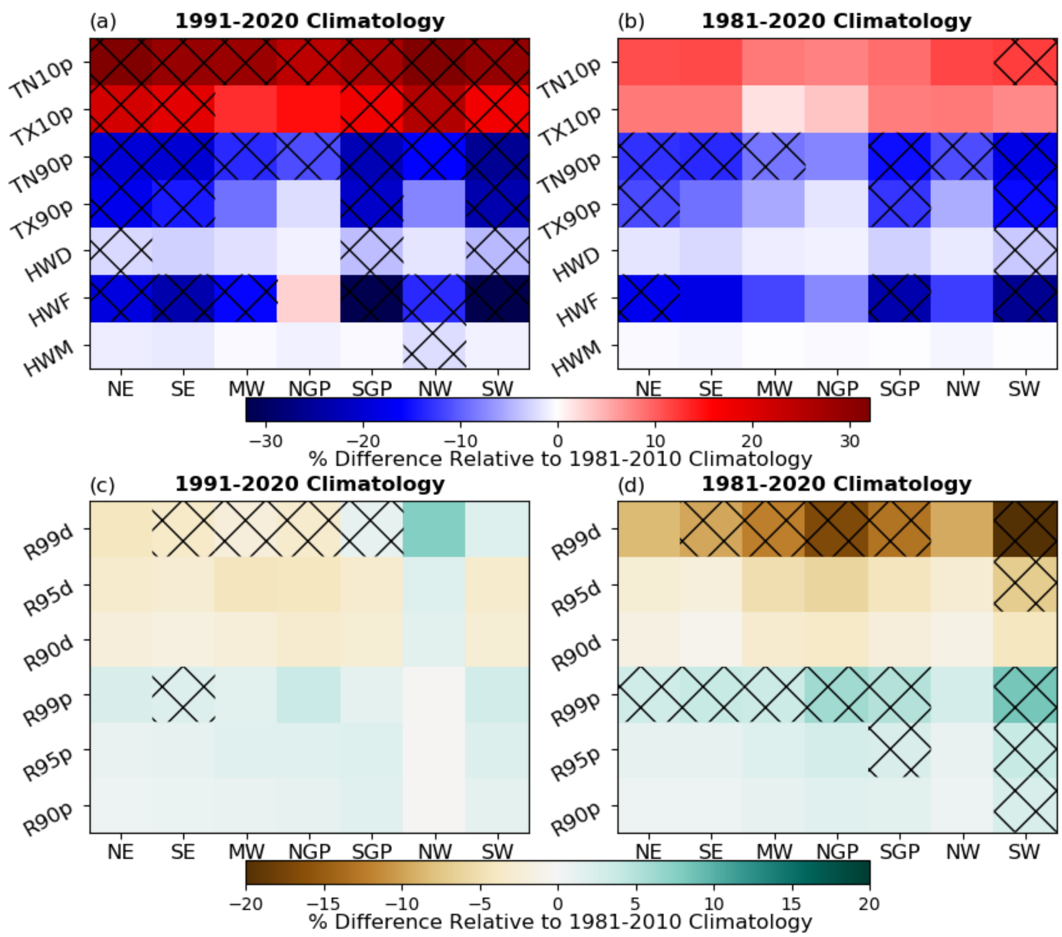

Figure 2. Average percent difference relative to the baseline climatology of 1981-2010 in area averaged over regions of the United States for (a) temperature indices using a baseline climatology of 1991-2020, (b) temperature indices using a baseline climatology of 1981-2020, (c) precipitation indices using a baseline climatology of 1991-2020, and (d) precipitation indices using a baseline climatology of 1981-2020. Hatching denotes the two climatologies result in statistically significant differences at 90% confidence.

Based on the regionally area-averaged changes shown in Fig. 2, spatial variability of changes in selected indices are shown in Fig. 3 (the other indices are shown in Figs. S5-S8). Here, differences between indices defined with the two baseline periods are averaged over all months in 1980-2021. Figure 3a shows the spatial variability of warm nights (TN90p) averaged over all months in 1980-2021 using the 1981-2010 climatology period. On average, there are relatively more warm nights detected in the Southwest US and fewer in the Northern Great Plains. When the baseline climatology is updated to 1991-2020, TN90p is reduced everywhere throughout the United States– strongest in the Southwest, and weakest in the Northern Great Plains (Fig. 3b). This spatial pattern is a result of the change in the 90th percentile of temperature (Fig. 1c) with the 1991-2020 baseline. Differences are similar when the baseline period is 1981-2020, but with smaller magnitude throughout the United States (Fig. 3c). For cool days (TX10p), there are climatologically more in the Northern Great Plains and Midwest, and fewer in the Southwest (Fig. 3d). TX10p increases on average throughout the United States when the baseline period is 1991-

2020 (Fig. 3e), except for the Northern Great Plains where differences are small and insignificant. When the baseline period is 1981-2020, the changes are smaller and not significant in most regions of the United States, likely due to the thirty-year overlap of the two baseline periods (Fig. 3f).

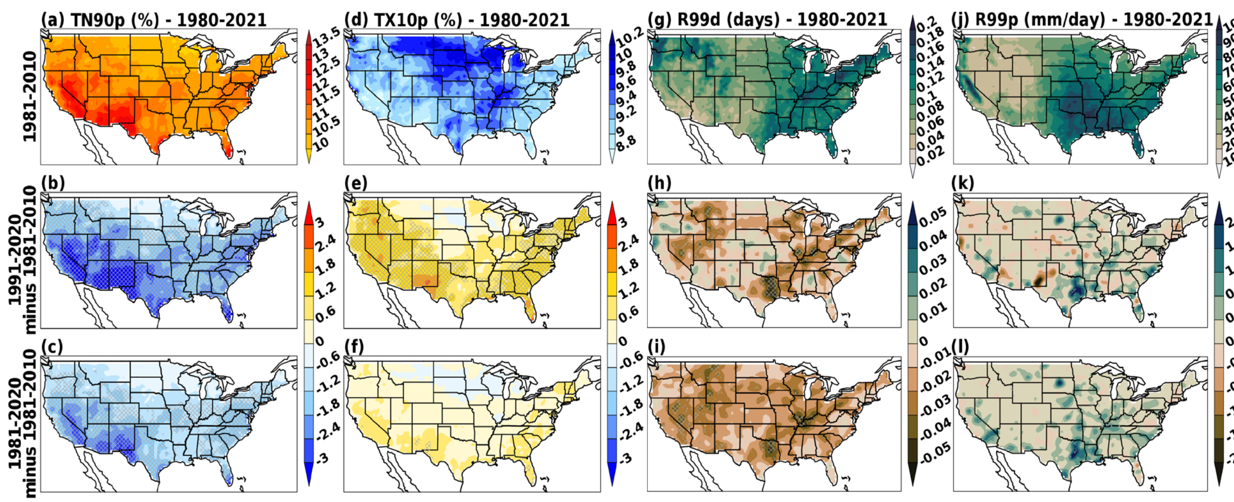

Figure 3. (a) TN90p defined using percentiles from the 1981-2010 baseline period, averaged over all months in 1980-2021, (b) the difference between TN90p defined using percentiles from the 1991-2020 baseline period and TN90p defined using percentiles from the 1981-2010 baseline period, averaged over all months 1980-2021; grey hatching indicates where difference is significant at the 90% confidence level, (c) the difference between TN90p defined using percentiles from the 1981-2020 baseline period and TN90p defined using percentiles from the 1981-2010 baseline period, averaged over all months 1980-2021, (d,e,f), (g,h,i), and (j,k,l) as in (a,b,c) but for TX10p, R99d, and R99p. For readability, panels h-i, k-l are plotted with a 9-point smoother, i.e., a weighted average of the values of the grid point and the 8 surrounding ones.

R99d (days with precipitation above the 99th percentile; Fig. 3g) and R99p (precipitation on these days; Fig. 3j) are both largest, on average, in the eastern US and pacific northwest. When the climatology period is updated, R99d is decreased over much of the US, i.e., fewer days with precipitation above the 99th percentile. Differences are largest and most significant over eastern Texas, parts of the west and the Midwest US. The changes are larger when the climatology period is 1981-2020 (Fig. 3i) than 1991-2020 (Fig. 3h). The differences in R99p are less consistent across the country, and mostly consist of increases in eastern Texas and parts of the Southeast and Midwest (Fig. 3k-l).

Finally, Fig. 4 shows the monthly time series of select indices averaged over the Southwest region of the US (as shown in Fig. 1a). The Southwest is chosen due to the relatively large changes observed in this region when updating the baseline period (Fig. 2). Monthly indices are shown as computed from the three baseline climate periods: 1981-2020 (red line), 1981-2010 (black line) and 1991-2020 (blue line). For the index representing warm days (TX90p; Fig. 4a), values are consistently lower when the index is defined using the 1991-2020 percentiles than the 1981-2010 percentiles. The difference between them increases later in the time series, indicating implications for trends in the indices; Dunn and Morice (2022) showed that positive trends in warm indices such as TX90p were reduced when a later baseline period was used. The opposite is true for cold nights (TN10p; Fig. 4b), where the newer climatology period results in higher values for the index throughout the period. For the precipitation indices shown (Fig. 4c-d), differences become most apparent after 2010, when the 1991-2020 baseline period results in fewer very wet days (Fig. 4d), but more precipitation on very wet days (Fig. 4c).

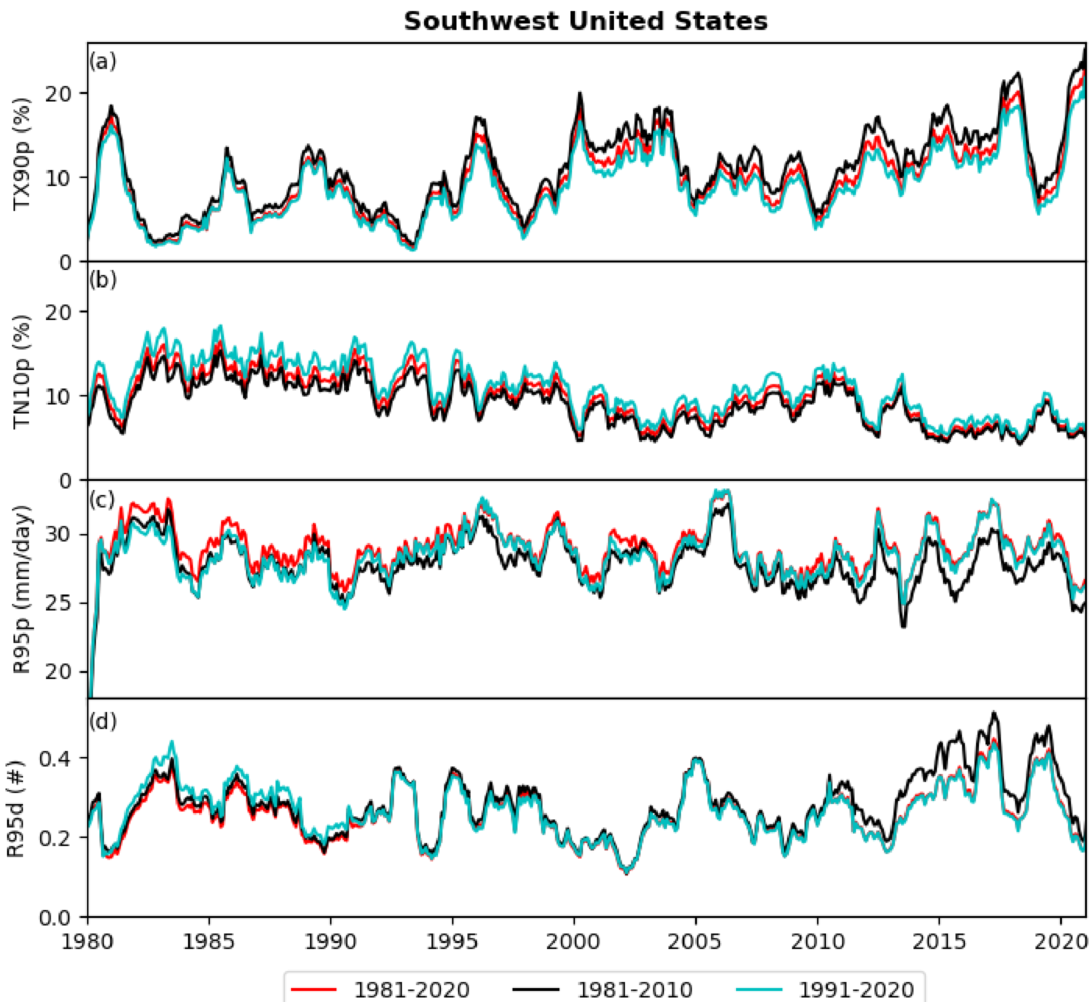

Figure 4. Time series of 12-month running means for (a) TX90p, (b) TN10p, (c) R95p, and (d) R95d area averaged over the Southwest region using a base climatology of 1981-2020 (red lines), 1981-2010 (black lines), and 1991-2020 (cyan lines).

4 Conclusions

Defining a climatological baseline period is necessary for the computation of percentile-based extreme climate indices. However, in a non-stationary climate, updating this baseline period leads to significant changes in the quantification of these climate extremes. In summary, over the United

States, updating the baseline period from 1981-2010 to 1991-2020 (or 1981-2020) generally leads to more days identified as a cold temperature extreme, fewer days identified as a warm temperature extreme, and extreme precipitation events that are classified as being less frequent but more intense. There is regional variability in these changes: temperature indices are most affected by the baseline period in the Southwest and Northeast, and least affected in the Northern Great Plains; precipitation changes are localized but typically greatest in the Southeast, Midwest, and inter-mountain west. This work has focused on the United States. However, the effect of the baseline period on the definition of extremes could be even more pronounced in other regions more sensitive to climate change.

The goal of this study has been to quantify the changes in detected temperature and precipitation extremes with an updated baseline. The cause of the differences in temperature and precipitation percentiles between the baseline periods is potentially related to several factors. The relative roles of human-induced climate change and multidecadal variability need to be assessed, especially to better quantify how extreme climate indices will change in the future (Sillmann et al. 2013a). The changing observing system of the reanalysis (e.g., McCarty et al., 2016) will also be explored in future work. While the focus of this work has been on MERRA-2, future work should involve analysis of extreme indices in other data sets, as Sillmann et al. (2013b) showed these can vary among reanalysis datasets.

While it is standard practice to use a 30-year climate baseline period that shifts in time every ten years (Arguez et al., 2012), the results here suggest that it may be useful to consider alternatives for defining climate extremes. The baseline period could be updated more frequently than every ten years, though it is not an easy task for operational centers to update their climatology period every year. To minimize the effect of multidecadal climate variability on extremes, the climatology could be extended to consist of the longest-record possible (Trewin, 2007). However, if one considers that society may adapt to the impacts of extremes over time, a shorter, more recent climatology may be a more logical comparison point. Furthermore, if using an observational record to define a baseline period, it is important to note whether in situ observational sites reported data within the reference period used.

The most appropriate baseline likely depends on the application, so data centers could create versions of indices using multiple baseline periods (e.g., Dunn et al., 2020), or provide users with the option to develop their own baseline climatology best suited to their purpose. Some users may need a more frequently updated baseline, while other users may need older baselines retained. It should also be noted that extreme climate events can be defined without percentile-based thresholds, such as using indices with a fixed threshold, though these have limited regional relevance. Methods based on return periods or time of emergence (Lewis et al., 2017) could also be used. Regardless of the approach, it is important to clearly communicate how extremes are defined and interpreted as this choice and the unique statistics produced can influence public perception.

Acknowledgments

This work was made possible by NASA's Center for Climate Simulation and was supported by the NASA Earth Science Research Program for Modeling and Analysis (MAP). MERRA-2 data and the extreme indices discussed here are disseminated by the Goddard Earth Sciences Data Information and Services Center (GES DISC). We thank Dr. Robert Dunn and Dr. Colin Morice for constructive feedback on the paper. We also thank Randy Koster, Anthony DeAngelis, Siegfried Schubert, Young-Kwon Lim and Yehui Chang for helpful discussions.

Open Research

CDO is available for download at https://code.mpimet.mpg.de/projects/cdo/#:~:text=Climate%20Data%20Operators,more%20than%20600%20operators%20available.
MERRA-2 data is publicly available through the GES DISC at

https://disc.gsfc.nasa.gov/information/glossary?title=MERRA-2.

References

Alexander, L.V., Fowler, H.J., Bador, M., Behrangi, A., Donat, M.G., Dunn, R.D., et al. (2019) On the use of indices to study extreme precipitation on sub-daily and daily timescales. *Environmental Research Letters, 14,* 125008. https://doi.org/10.1088/1748-9326/ab51b6

Alexander, L. V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A.M.G., et al. (2006) Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res., 111*, D05109. https://doi.org/10.1029/2005JD006290

Arguez, A., Durre, I., Applequist, S., Vose, R.S., Squires, M.F., Yin, X., et al. (2012) NOAA's 1981-2010 U.S. Climate Normals: An Overview. *BAMS, 93*, 1687-1697. https://doi.org/10.1175/BAMS-D-11-00197.1

Arguez A. & Vose, R.S. (2011) The Definition of the Standard WMO Climate Normal. *BAMS 92*, 699-704. https://doi.org/10.1175/2010BAMS2955.1

Bosilovich, M.G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., et al. (2015) MERRA-2: Initial Evaluation of the Climate. NASA Tech. Memo. NASA/TM-2015-104606/Vol. 43, 145 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf]

Cammalleri, C., Spinoni, J., Barbosa, P., Toreti, A. & Vogt, J.V. (2021) The effects of non-stationarity on SPI for operational drought monitoring in Europe. *International Journal of Climatology, 42*, 1-13. https://doi.org/10.1002/joc.7424

Collow, A., Bosilovich, M. G., & Koster, R. D. (2016) Large Scale Influences on Summertime Extreme Precipitation in the Northeastern United States. *J. Hydrometeor*, *17*, 3045-3061. https://doi.org/10.1175/JHM-D-16-0091.1

Collow, A., Thomas, N., Bosilovich, M., Dezfuli, A., & Lucchesi, R. (2022) File Specification for MERRA-2 Climate Statistics Products. GMAO Office Note No. 19 (Version 1.2), 15 pp, available from http://gmao.gsfc.nasa.gov/pubs/office_notes.

Dunn, R.J.H. & Morice, C.P. (2022) On the effect of reference periods on trends in percentile-based extreme temperature indices. *Environmental Research Letters*, *17*, 034026. https://doi.org/10.1088/1748-9326/ac52c8

Dunn, R.J.H., Alexander, L.V., Donat, M.G., Zhang, X., Bador, M., Herold, N., et al. (2020) Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. *J. Geophys. Res. Atmos., 125.* https://doi.org/10.1029/2019JD032263

Dunn, R.J.H., Donat, M.G., & Alexander, L.V. (2022) Comparing extremes indices in recent observational and reanalysis products. *Front. Clim*., *4*, 989505. doi: 10.3389/fclim.2022.989505

Gelaro, R., McCarty, W., Suarez, M.J., Todling, R., Molod, A., Takacs, L., et al. (2017) The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim*., *30*, 5419-5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Global Modeling and Assimilation Office (GMAO), 2015a: MERRA-2 statD_2d_slv_Nx: 2d, Daily, Aggregated Statistics, Single-Level, Assimilation, Single-Level Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), doi: 10.5067/9SC1VNTWGWV3.

Global Modeling and Assimilation Office (GMAO), 2015b: MERRA-2 tavg1_2d_flx_Nx: 2d, 1-Hourly, Time-Averaged, Single-Level, Assimilation, Surface Flux Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), doi: 10.5067/7MCPBJ41Y0K6.

Global Modeling and Assimilation Office (GMAO) (2020) MERRA-2 statM_2d_edi_Nx: 2d, Single-Level, Monthly Extremes Detection Indices V1, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), doi: 10.5067/QFJ13GEGDI99.

Global Modeling and Assimilation Office (GMAO) (2022) MERRA-2 statM_2d_edi_Nx: 2d, Single-Level, Monthly Extremes Detection Indices based on 1991-2020 V2, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), doi: 10.5067/O8AX56DO60MI.

Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., Sparks, T. & Garforth, J. (2020) State of the UK Climate 2020. *International Journal of Climatology, 41*, 1-76. https://doi.org/10.1002/joc.7285

Lewis, S.C., King, A.D., & Perkins-Kirkpatrick, S.E. (2017) Defining a New Normal for Extremes in a Warming World. *BAMS, 98*, 1139-1151. https://doi.org/10.1175/BAMS-D-16-0183.1

Livezey, R.E., Vinnikov, K.Y., Timofeyeva, M.M., Tinker, R. & van den Dool, H.M. (2007) Estimation and Extrapolation of Climate Normals and Climatic Trends. *Journal of Applied Meteorology and Climatology, 46*, 1759-1776. https://doi.org/10.1175/2007JAMC1666.1

McCarty, W., Coy, L., Gelaro, R., Huang, A., Merkova, D., Smith, E.B., et al. (2016) MERRA-2 input observations: Summary and initial assessment. Technical Report Series on Global Modeling and Data Assimilation, Vol. 46, NASA Tech. Rep. NASA/TM–2016–104606, 61 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/McCarty885.pdf.] https://www.ncei.noaa.gov/products/land-based-station/us-climate-normals

National Centers for Environmental Information (NCEI) (2021, May 4) NOAA Delivers New U.S. Climate Normals, https://www.ncei.noaa.gov/news/noaa-delivers-new-us-climate-normals, last accessed 29 June 2023.

Perkins, S. E., & Alexander, L. V. (2013). On the Measurement of Heat Waves, Journal of Climate, 26(13), 4500-4517. https://doi.org/10.1175/JCLI-D-12-00383.1.

Reichle., R.H., Liu, Q., Koster, R.D., Draper, C.S., Mahanama, S.P.P. & Partyka, G.S. (2017) Land Surface Precipitation in MERRA-2. *Journal of Climate*, *30*, 1643-1664. https://doi.org/10.1175/JCLI-D-16-0570.1

Scherrer, S.C., Appenzeller, C., & Liniger, M.A. (2006) Temperature trends in Switzerland and Europe: Implications for Climate Normals. *International Journal of Climatology, 26*, 565-580. https://doi.org/ 10.1002/joc.1270

Schreck III, C.J., Klotzbach, P.J., & Bell, M.M. (2021) Optimal climate normal for North Atlantic Hurricane activity. *Geophysical Research Letters, 48*, 1-9. https://doi.org/10.1029/2021GL092864

Schulzweida, U. (2022) CDO User Guide (2.1.0). Zenodo. https://doi.org/10.5281/zenodo.7112925

Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X., and Bronaugh, D. (2013a), Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections, *J. Geophys. Res. Atmos*., 118, 2473- 2493, doi:10.1002/jgrd.50188

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D. (2013b), Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, J. Geophys. Res. Atmos., 118, 1716- 1733, doi:10.1002/jgrd.50203

Trewin, B. (2007) The role of climatological normals in a changing climate. WCDMP No. 61 – WMO/TD-No. 1377. https://library.wmo.int/doc_num.php?explnum_id=4546

Thomas, N. P., Bosilovich, M. G., Collow, A. B. M., Koster, R. D., Schubert, S. D., Dezfuli, A., & Mahanama, S. P. (2020) Mechanisms associated with Daytime and Nighttime Heat Waves over the Contiguous United States. *J. Appl. Meteorol. Clim.*, *59 (11),* 1865-1882. https://doi.org/10.1175/JAMC-D-20-0053.1

Wilks, D.S. (2013) Projecting "Normals" in a Nonstationary Climate. *Journal of Applied Meteorology and Climatology, 52*, 289-302. https://doi.org/10.1175/JAMC-D-11-0267.1

Wilks, D.S. & Livezey, R.E. (2013) Performance of Alternative "Normals" for Tracking Climate Changes, Using Homogenized and Nonhemogenized Seasonal U.S. Surface Temperatures. *Journal of Applied Meteorology and Climatology, 52*, 1677-1687. https://doi.org/10.1175/JAMC-D-13-026.1

Wuebbles, D.J., Fahey, D.W., Hibbard, K.A., Dokken, D.J., Stewart, B.C., & Maycock, T.K. Eds. (2017) Climate Science Special Report: Fourth National Climate Assessment. U.S. Global Change Research Program. Vol. I. 470 pp., https://doi.org/10.7930/J0J964J6

Yosef, Y., Agilar, E., & Alpert, P. (2021) Is it possible to fit extreme climate change indices together seamlessly in the era of accelerated warming? *Int. J. Climatol., 41*. https://doi.org/10.1002/joc.6740

Zhang, X., Hegerl, G., Zwiers, F. W., & Kenyon, J. (2005) Avoiding inhomogeneity in percentile-based indices of temperature extremes. *J. Climate*, *18*, 1641–1651. https://doi.org/10.1175/JCLI3366.1

Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., & Zwiers, F.W. (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Clim Change*, *2*, 851-70. https://doi.org/10.1002/wcc.147