Title
Quantitative Precipitation Estimation of Extremes in CONUS With Radar Data

Journal
Geophysical Research Letters, 48(16)

ISSN
0094-8276

Authors
Molter, Edward M
Collins, William D
Risser, Mark D

Publication Date
2021-08-28

DOI
10.1029/2021gl094697

Peer reviewed

Quantitative Precipitation Estimation of Extremes in CONUS with Radar Data

Edward M. Molter[1,2], **William D. Collins**[2,3], **and Mark D. Risser**[2]

[1]Astronomy Department, University of California, Berkeley; Berkeley CA, 94720, USA
[2]Climate and Ecological Sciences Division, Lawrence Berkeley National Laboratory, Berkeley CA, 94720
[3]Earth and Planetary Science Department, University of California, Berkeley; Berkeley CA, 94720, USA

Key Points:

- The 5-year return values of extreme rainfall vary only minimally on spatial scales smaller than 100 kilometers.
- The gauge-only Mountain Mapper algorithm, used to spatially interpolate rain-gauge data in the operational Next-Generation Radar (NEXRAD) data products where radar data is unavailable, underestimates 5-year return values far from the locations of rain gauges.
- The Risser et al 2019 algorithm for spatially interpolating extreme value statistics between rain gauges represents these statistics more accurately than Mountain Mapper over the majority of CONUS.

Corresponding author: Edward M. Molter, `emolter@berkeley.edu`

Abstract

Constructing an accurate, continental, in-situ-based, kilometer-scale, long-term record of the precipitation field and its spatiotemporal changes remains a significant challenge. Here we determine the extreme-value behavior of the NEXRAD Stage IV radar-based quantitative precipitation estimate (QPE). We find that the climatology of 5-year daily return values in CONUS East of the Rocky Mountains shows only slight variability on spatial scales smaller than ∼100 km. In light of this finding, we test whether rain-gauge-only daily precipitation datasets can produce accurate extreme-value behavior at spatial scales finer than the spacing between gauges. We find that the 5-year daily return values are accurate at locations far from rain gauges only if the interpolation between gauges is carried out appropriately for extremes. Precipitation statistics derived from in-situ rain gauge data are therefore of sufficient spatial resolution to faithfully capture daily extremes over much of the eastern United States.

Plain Language Summary

Accurate measurement of the amount of precipitation that falls within a given region and time period is crucial for environmental modeling, climate change research, and resource and risk management. For all of those applications, it is desirable to understand not only how much precipitation falls on average, but also how much precipitation falls during an extreme event, such as a severe storm. Using data from weather radar, we show that certain statistical properties of extreme rainfall are highly correlated on spatial scales up to 100 kilometers over the eastern United States. This means that rain gauge networks, which have typical inter-gauge spacings of roughly 30 kilometers over the eastern United States, are dense enough to accurately measure these statistical properties. However, it's imperative to interpolate between the rain gauge measurements in a way that explicitly captures extremes if the application of interest requires capturing extremes accurately. Our research represents a step toward constructing an accurate, continental-scale, long-term, high-resolution precipitation dataset.

1 Introduction

Accurate measurement of the amount of precipitation that falls within a given region and time period is crucial for environmental modeling (e.g. Jones et al., 2001; Parra et al., 2004; Abatzoglou, 2013), climate change research (e.g. Groisman et al., 1999; Alexander et al., 2006; Schär et al., 2016), and resource and risk management (e.g. Rosenzweig et al., 2002; Schumann, 2011; Vogel et al., 2019). All of these applications require understanding not only the seasonal-mean or annual-mean precipitation but also the extreme tail of the daily or sub-daily precipitation distribution. Precise measurements of the total rainfall over a specific area at scales of tens of kilometers or less, such as a city or watershed, are often also needed.

Estimating the true spatiotemporal distribution of precipitation from observational data is known as quantitative precipitation estimation (QPE), and is currently obtained from three main data sources: satellite, rain gauges, and ground-based radar. Each source provides unique advantages subject to specific limitations. Satellite observations provide spatially continuous measurements, but are subject to severe uncertainty because precipitation must be inferred from cloud top height or temperature as derived from microwave and/or infrared spectra (Iguchi et al., 2009; Tapiador et al., 2012). This uncertainty generally leads to overestimation of extreme precipitation events relative to gauge- or radar-based estimates (AghaKouchak et al., 2011; Mehran et al., 2014). Over the contiguous United States (CONUS), where the density of both rain gauges and radar stations is high, satellite-based QPE products tend to perform compared unfavorably to other estimates. For example, Timmermans et al. (2019) found significant biases in the representation of daily precipitation extremes from satellite-based gridded QPEs compared with rain gauge estimates. Satellite products are therefore not considered further here.

Rain gauges provide the most accurate and temporally continuous point measurements of precipitation despite errors from undercatch, variance in management quality, and changes in location or equipment (see Tapiador et al., 2012, for a recent review). They also provide the longest time record of any precipitation measurement by far. However, gauges yield point measurements only, and one must interpolate spatially between them to estimate precipitation over an area. Ground-based radar observations provide very high native spatial and temporal resolution. Each Weather Surveillance Radar 88 Doppler Radar (WSR-88D) radar stations in CONUS (NOAA, 2006) completes a full scan of the sky every $\sim$10 minutes, and the station's preprocessing algorithm bins the scan into 1 km range by 1° azimuth sections, amounting to sub-hourly precipitation estimates on a $\lesssim$4 km grid (Fulton et al., 1998). However, the relationship between radar reflectivity and precipitation rate is degenerate and differs for different types of storms, and therefore it must thus be determined empirically via comparison with rain gauge data (Fulton et al., 1998; Young et al., 2000). Multisensor estimates combine the high spatial and temporal coverage of the radar data with the high fidelity of the gauge data and hence represent the state-of-the-art in operational QPE. However, experiments to validate these QPEs often take place over small regions and on timescales shorter than a few months (e.g., Willie et al., 2017; Spies et al., 2018; B.-C. Seo et al., 2018).

A variety of studies make use of gauge-based QPE and solve the aforementioned spatial interpolation problem in various ways, sometimes making use of elevation corrections or models of climatology (e.g. Daly et al., 1994, 2015; Schaake et al., 2004; Sheridan et al., 2010; Livneh et al., 2013). We refer to this class of gauge-based analysis of extremes as "grid-then-fit" techniques because they interpolate at the native temporal scale (e.g., daily) and then calculate statistical properties of the interpolated data. These approaches tend to underestimate extreme precipitation, especially at small (0.25°) scales (Sun & Barros, 2010; Gervais et al., 2014; Behnke et al., 2016). To rectify this issue, Risser et al. (2019) have developed a statistical "fit-then-grid" technique in which Generalized Extreme-Value (GEV) statistics (see Coles et al., 2001) are calculated at individual rain gauges, the GEV parameters are spatially interpolated, and then the gridded GEV distributions are reconstructed from these interpolated parameters. This method implicitly assumes that the parameters of the GEV distribution vary smoothly in space such that high-quality inference about extremes can be made in between stations. The optimal gauge interpolation technique depends on both the grid resolution and the application of interest (Chen & Knutson, 2008; Gervais et al., 2014), and best practices for interpolating to smaller scales than the inter-gauge spacing have not been established.

This study seeks to determine whether the extreme statistics of daily precipitation vary smoothly between rain gauges over the CONUS, testing the assumption of Risser et al. (2019) over that domain, and to evaluate the accuracy of this novel fit-then-grid technique as compared with standard grid-then-fit algorithms. To these ends, we consider the GEV statistics of a dataset at very high (4 km) spatial resolution, namely the NEXRAD Stage IV daily dataset (Fulton et al., 1998; D.-J. Seo & Breidenbach, 2002; Lin & Mitchell, 2005; Lin, 2011), a radar-based multisensor QPE, from 2002-2019. Stage IV is available at hourly, six-hourly, and daily frequencies. We focus on daily maxima here to facilitate comparison with the GHCN-D network, the most extensive network of rain gauges in the CONUS, and to test the results of Risser et al. (2019) directly. Stage IV has been evaluated extensively in the literature, including via the use of percentile-based metrics to capture extreme value behavior (Prat & Nelson, 2015; Nelson et al., 2016). However, percentile-based metrics have been shown to produce different results depending on the specific metric used (Schär et al., 2016). McGraw et al. (2019) previously employed GEV statistics in a comparison between rain gauges and Stage IV data at hourly, 3-hourly, 6-hourly, and daily frequencies. However, they only reported those statistics at the locations of ∼500 rain gauges and did not consider spatial variability in their GEV fits. Our paper is the first (to our knowledge) to publish GEV statistics at every grid cell in Stage IV.

We describe our data processing and GEV fitting in Section 2. In Section 3, we use this new data product to explore whether the high spatial resolution of this QPE provides new information on the climatology of extremes at finer spatial scales than accessible using gauge-only estimates. In Section 4, our product is compared with the Risser et al. (2019, hereafter R19) gauge-only interpolation technique, as well as with the Mountain Mapper algorithm (Schaake et al., 2004), a more conventional gridded QPE that incorporates the Parameter-elevation Relationships on Independent Slopes Model (PRISM) climatological model (Daly et al., 2015) and is widely used for operational weather analyses. Specifically, Mountain Mapper is the official rainfall product distributed by the California-Nevada, Colorado Basin, and Northwest River Forecast Centers of the National Oceanographic and Atmospheric Administration (NOAA). We contextualize our findings within existing literature on the spatial scales of extremes in Section 5, then summarize our work in Section 6.

2 Data Processing

We computed and made use of three distinct extreme-value datasets in this paper; these are summarized below.

Stage IV GEV: We downloaded the 4-km-resolution daily NEXRAD Stage IV gridded multisensor QPE for every day between 1 January 2002 and 31 December 2019, totaling 6573 days (18 years) on a 881 × 1121 grid. We compared the Stage IV daily measurements to Global Historical Climatology Network Daily (GHCN-D) rain gauge data (Menne et al., 2012) in grid cells that contained a GHCN station. These validation steps are outlined in Supporting Information S1. Stage IV was found to agree very well with

GHCN-D in both means and extremes in grid cells that contained a gauge, meaning that the normalization of the radar data to nearby rain gauges in the Stage IV processing pipeline preserves information about extremes.

As noted by other authors (Prat & Nelson, 2015; Nelson et al., 2016), Stage IV is a fundamentally heterogeneous dataset. The product from the three western NOAA River Forecast Centers (California-Nevada, Northwest, and Colorado Basin, hereafter "Western RFCs") differs substantially from that of the nine other RFCs comprising the CONUS (hereafter "Eastern RFCs"). Specifically, the Western RFCs produce their QPE using the gauge-only Mountain Mapper technique discussed below and do not incorporate radar data at all, while the Eastern RFCs use the radar-inclusive procedures outlined in Fulton et al. (1998) and Lin and Mitchell (2005). Data from the Western RFCs are therefore not actually made using a multisensor technique, and so are for the most part not considered further in this paper.

In each grid cell of Stage IV, we extracted seasonal maximum precipitation amounts for each season (DJF, MAM, JJA, and SON), and then fit the GEV distribution to the 18 seasonal maxima over our period of record. This is a fairly short period of record over which to apply GEV statistics; however, in this work we draw our conclusions from the 5-year return values only, which are well sampled by 18 years of data. To assess whether our GEV fits provided an adequate representation of the data, we performed a 2-sided Kolmogorov-Smirnov (K-S) test to quantify the likelihood that the observed seasonal maxima were drawn from the GEV distribution. We found a p-value of < 0.05 in at least 94% of grid cells in the Eastern RFCs in all seasons, meaning that the data were plausibly drawn from the GEV distribution. We used the GEV fit to generate 5-year return value estimates in each grid cell for each season. Following R19, we assessed the errors in our fit parameters using a bootstrap resampling technique: the seasonal maxima at each grid cell were resampled with replacement and then re-fit 250 times, and the standard deviation of the fit parameters in those 250 fits were used to define the 1-σ error on the parameters.

R19 GEV: We used the same extreme-value dataset as R19, which is based on GHCN-D rain gauge measurements, but extracted 5-year return values instead of 20-year return values as in that paper. To create a mean climatology from the R19 analysis, the exact same procedures described in R19 were applied to GHCN measurements of seasonal average daily precipitation (instead of seasonal maximum daily precipitation), and an ordinary least-squares fit was applied at each station such that the spatially-interpolated parameters were the mean and variance instead of the GEV parameters. We refer to this mean climatology hereafter as the "R19 mean".

Mountain Mapper: The Mountain Mapper dataset (Schaake et al., 2004) is a gauge-only gridded precipitation product at 4-km resolution that is widely used in operational weather analysis. It interpolates from rain gauges to the 4-km grid using an inverse-square weighting scheme, incorporating also the PRISM climatological model (Daly et al., 2015). However, the official Mountain Mapper product is not archived at NOAA. Instead, we have created several versions of the dataset using an identical procedure to Schaake et al. (2004). The creation of our own versions of the dataset is beneficial for direct comparison to R19 and Stage IV for three reasons. First, we have used the same rain gauge network as Risser, namely stations from the GHCN-D network in CONUS with at least 66.7% nonmissing values over our time period (8097 stations). Second, we can interpolate the gauge network onto any grid we choose. Finally, we can force the long-term mean of Mountain Mapper to equal any chosen climatology; versions of Mountain Mapper constrained to the R19 mean and to the mean of our 18-year slice of Stage IV are employed in this paper. We compute GEV statistics and their errors at each grid cell in the same way as for Stage IV. Our computations using the Mountain Mapper procedure are discussed further in Supporting Information S2.

3 Spatial Scales of Extremes

If the hypothesis that the extreme statistics of daily precipitation vary smoothly between rain gauges over the CONUS is true, then a spatial power spectrum of the 5-year return value map should show little power at $\lesssim$50-km scales. We thus used a wavelet decomposition to compute a spatial power spectrum of the 5-year return values in the Stage IV dataset using a 2-D continuous wavelet transform. Following the procedure outlined in Torrence and Compo (1998), we used Morlet wavelets with non-dimensional frequency $\omega_0 = 6$, and 40 widths equally spaced on a log scale from 4 km to $\sim$2000 km.[1] Maps of the power on various representative spatial scales are shown for JJA and DJF in Fig-

[1] Using a Morlet wavelet in this way is mathematically identical to a "short-time" Fourier transform using a Gaussian window function

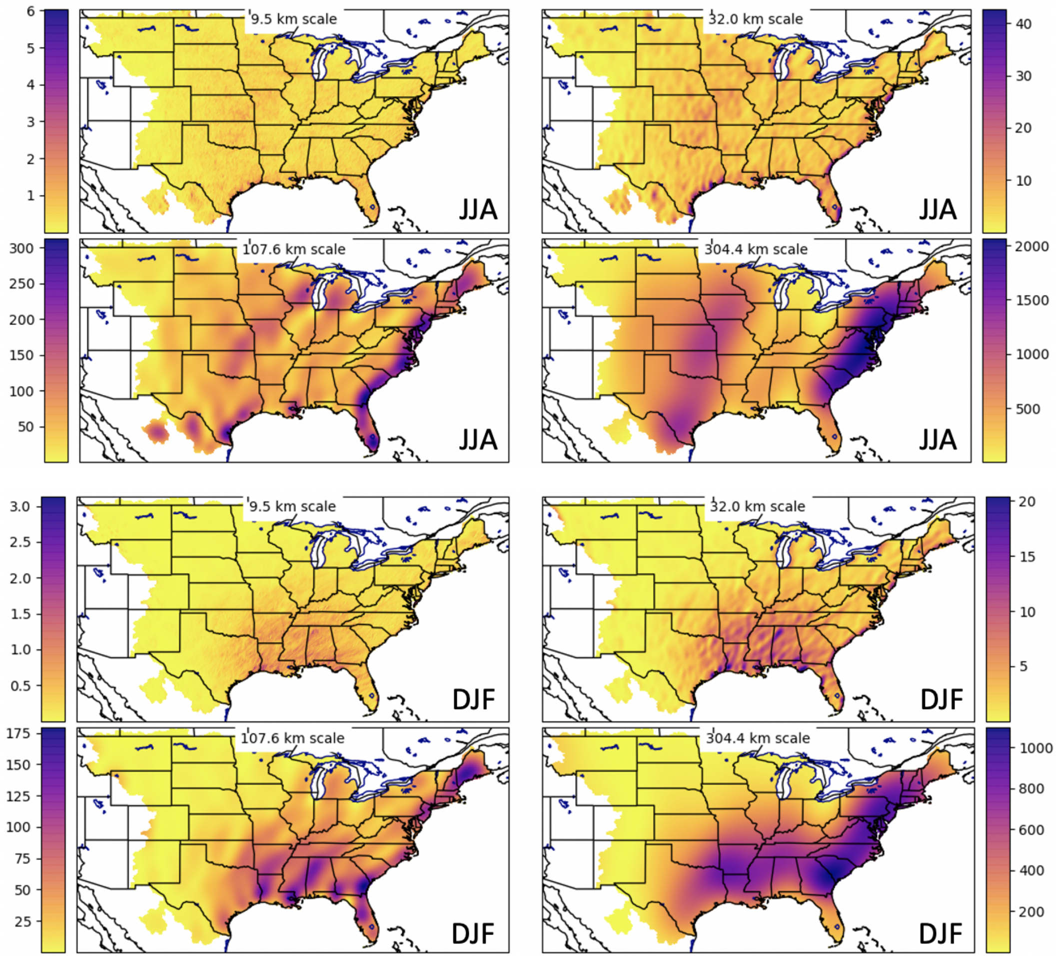

Figure 1. Wavelet decomposition of Stage IV 5-year return value map for **(Top:)** DJF and **(Bottom:)** JJA at four representative spatial scales. The colormap denotes power spectral density in arbitrary units.

ure 1 to help visualize the wavelet decomposition. The power spectrum of our 5-year return value map is presented in Figure 2. (The same maps and power spectra are shown for the MAM and SON seasons in the Supporting Information.) The spatial scales s plotted on the x-axis are nearly equal to the Fourier wavelength λ for this choice of wavelet (formally $\lambda = 1.02s$ following Torrence & Compo, 1998), and should be interpreted in the same manner as a Fourier wavelength, namely as the combined length scale of a positive and negative fluctuation about the mean. Note that substantial edge effects obscure any useful information at scales larger than $s = 1000$ km, so these are not plotted.

To aid in understanding the implications of Figure 2, we have overplotted the power spectrum of a test dataset that contains pure white noise within the entire domain (the

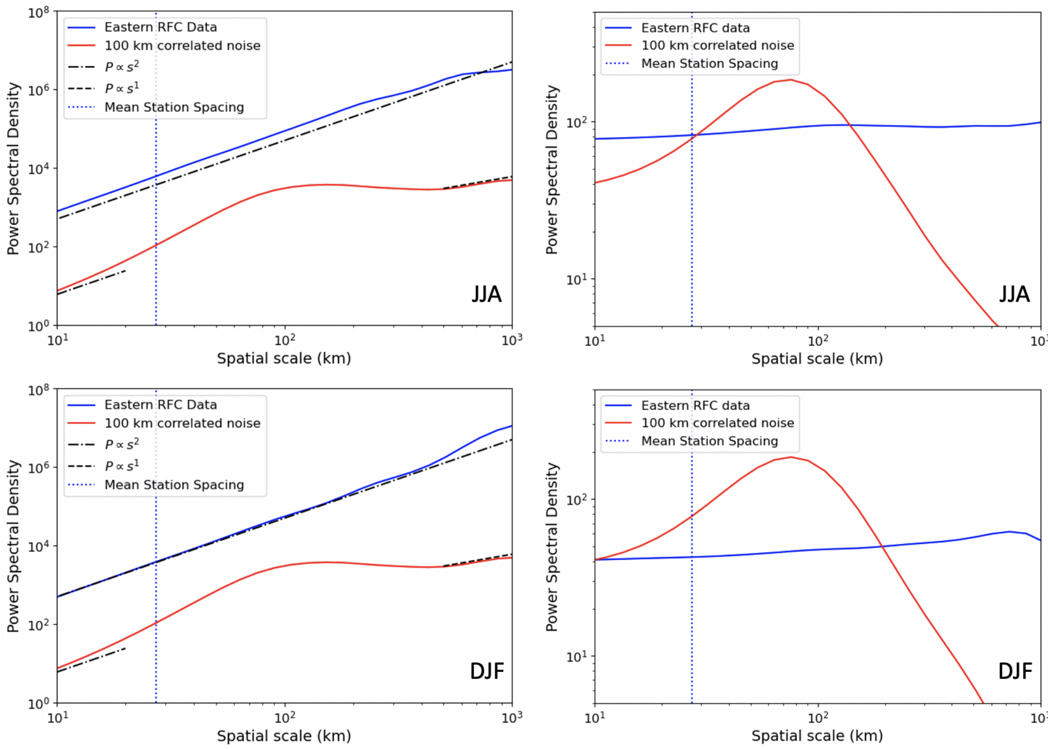

Figure 2. Seasonal wavelet power spectrum of the Stage IV 5-year return value map for **Top:** JJA and **Bottom:** DJF over the eastern RFCs only (blue lines). Vertical lines show the mean spacing of 27 km between GHCN stations in the eastern RFC domain. The power spectrum of white noise correlated at the 100 km scale is also shown (red line). The left panels show the raw power spectra; the right panels show the spectra after being divided by the $P = s^2$ line.

eastern RFCs), correlated at the 25-pixel (100-km) scale; that is, we made a map of pure Gaussian noise then oversampled it by a factor of 25. A log-linear power spectrum takes the form

$$S_\nu(f) = cf^{-\beta} \tag{1}$$

where f is the spatial frequency and β is the spectral scaling. The correlated noise test spectrum can be interpreted as transitioning between $\beta = 2$ at the smallest length scales, where the map is highly autocorrelated, and $\beta = 1$ at the largest length scales, where the map is completely uncorrelated and looks like pure white noise. Note that white noise is not spectrally flat on a log-linear scale, but instead follows a $\beta = 1$ scaling. In between these two regimes is stored all the information content in the map, and as such, the power spectrum is strongly peaked at 100 km length scales. This can be seen most clearly after the spectra have been divided by the $P = s^2$ line in the right panel of Figure 2. The 5-year return value maps in both JJA and DJF show similar behavior to the correlated noise map, with strong autocorrelation ($P \propto s^2$) at small spatial scales but with a broader, less prominent spike in power that begins near 200-km scales and continues out to 800-km scales. This means that 5-year return values are strongly autocorrelated at $s < 200$ km, confirming the hypothesis that extreme statistics of daily precipitation vary smoothly between rain gauges over the CONUS. The power spectral density is maximized at very large length scales of ∼800 km. The strong autocorrelation at small scales is present in each of the four major Köppen-Geiger climate classes within the eastern CONUS, as shown in the Supplementary Material.

4 QPE Product Comparison

Section 3 validated the implicit assumption of the R19 fit-then-grid technique that extreme statistics of daily rainfall vary smoothly between rain gauges. It is *a priori* unclear, though, whether this fit-then-grid algorithm is actually more accurate than a standard grid-then-fit algorithm when applied to an identical set of rain gauges and given an identical mean climatology. We set up this test by comparing the extreme-value behavior between the Mountain Mapper and R19 datasets. To ensure a direct comparison, we used a version of Mountain Mapper constructed such that its long-term seasonal mean ($\bar{Y}$ in Supporting Information S2) was equal to the R19 mean. As Figure 3 shows, the return values are substantially different between the two datasets, with Mountain Mapper underestimating R19 by greater than 10% over much of the CONUS in both DJF

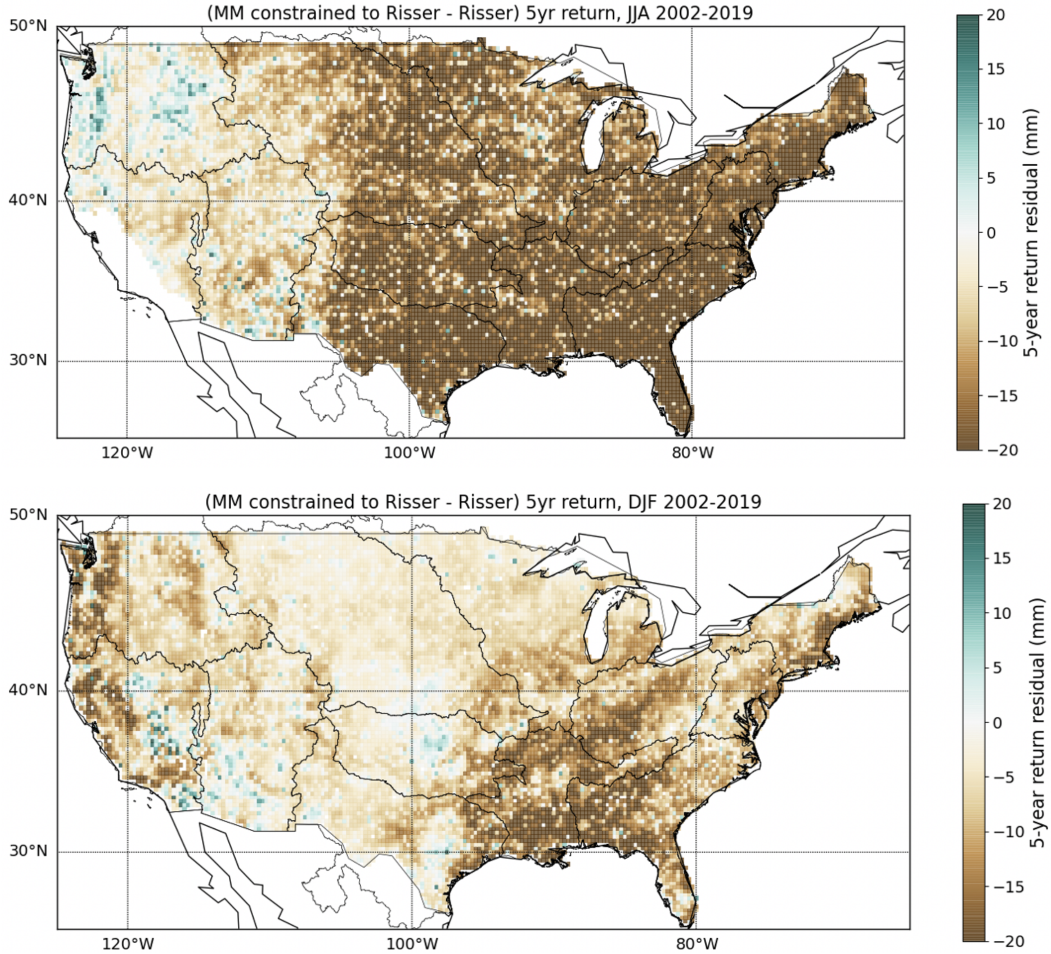

Figure 3. Difference between 5-year return value from R19 and our Mountain Mapper implementation constrained to the R19 mean climatology for **Top:** JJA and
Bottom: DJF. The extremes in Mountain Mapper are lower in magnitude than in R19 over the majority of CONUS in both seasons.

and JJA. An assessment of the statistical significance of this difference is given in Supporting Information S3.

We next evaluated the R19 and Mountain Mapper 5-year return values against Stage IV. This comparison is somewhat difficult to probe directly because the long-term means of R19 and Stage IV are not strictly equal, so differences in extremes may be partially caused by differences in the long-term means of those datasets. To get around this, we computed return values from a version of Mountain Mapper that is forced to equal the long-term means of Stage IV, and compared both this Mountain Mapper version and Stage IV itself with R19 (see Figure 4). In so doing, any differences are isolated to the treatment

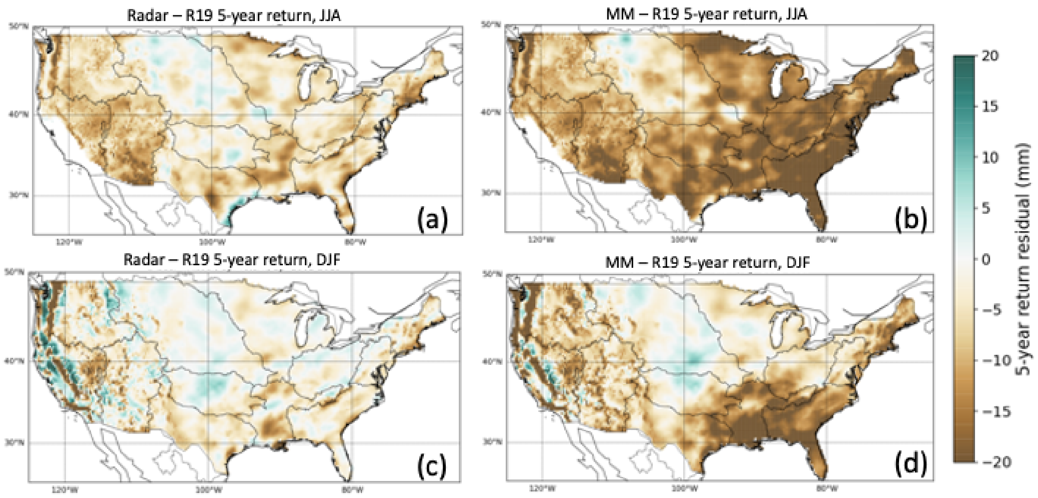

Figure 4. Difference between 5-year return value from **(a)** R19 and Stage IV in JJA, **(b)** R19 and Mountain Mapper in JJA, **(c)** R19 and Stage IV in DJF, and **(d)** R19 and Mountain Mapper in DJF. Here the Mountain Mapper datasets have been constrained to the Stage IV mean climatology. The R19 extremes agree more closely with Stage IV than Mountain Mapper in both JJA and DJF (i.e., the discrepancies are smaller in panels **a** and **c** than in panels **b** and **d**), validating the ability of the R19 technique to interpolate extremes to smaller spatial scales.

of extremes. This Mountain Mapper version is found to underestimate extremes relative to Stage IV over large portions of the Eastern RFCs, whereas R19 agrees more closely.

The difference between Mountain Mapper and R19 is attributable to the grid-then-fit approach taken by Mountain Mapper: using an inverse-square weighting scheme to interpolate between grid points makes it unlikely for extremes to occur at grid points far from any one rain gauge. This hypothesis is confirmed by considering the difference between Mountain Mapper and Stage IV as a function of distance from the nearest rain gauge over the eastern RFCs. The 5-year return values from Mountain Mapper agree well with Stage IV at distances $\lesssim$10 km from the nearest gauge, but begin to underestimate Stage IV at larger distances in both DJF and JJA (Figure 5). It is important to note that the spatial averaging inherent in the Mountain Mapper technique is not a deficiency *per se*, and is in fact the appropriate way to measure the spatial average of extremes over a large grid box (Gervais et al., 2014) for comparison to climate models at $\sim$100 km resolution. However, we have shown that the Risser technique provides a more

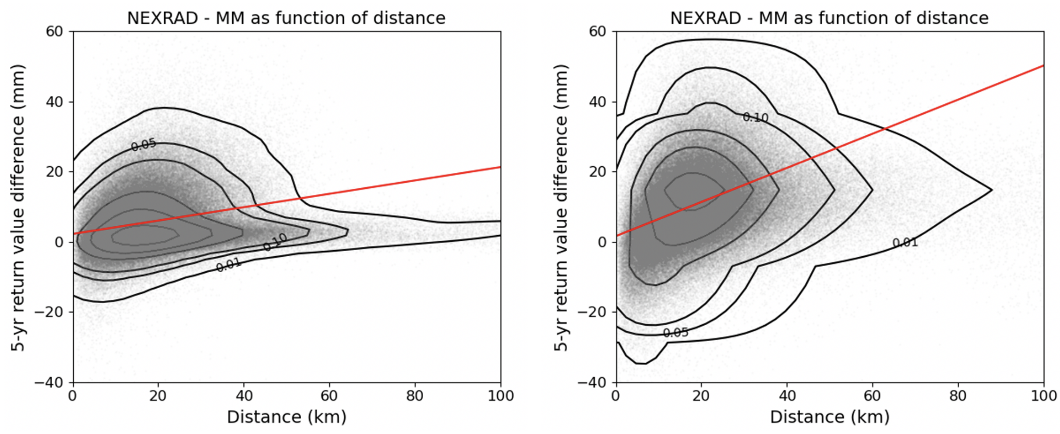

Figure 5. Five-year return value difference between Stage IV and Mountain Mapper as a function of distance from the nearest rain gauge used in Mountain Mapper for **Left:** DJF and **Right:** JJA over the eastern CONUS. The small grey points denote individual grid cells; the contours describe the cumulative density of points. The red line shows a least-squares linear fit to the grey points. The return values agree in grid cells near rain gauges, but Mountain Mapper begins to underestimate Stage IV as the distance from a gauge is increased.

accurate estimate of rainfall extremes at ∼25 km scales, assuming the radar-aided Stage IV dataset to be a “ground truth”.

5 Discussion

The long correlation lengths of 5-year return values in the eastern CONUS derived in Section 3 are perhaps unsurprising in the context of the dynamical systems that produce extreme precipitation in that region. In the central and eastern United States, extreme precipitation is most often associated with one of three categories of storm: mesoscale convective systems (MCSs), landfalling tropical cyclones (TCs), and synoptic forcing events (i.e., extratropical cyclones). MCSs are organized groups of thunderstorms that produce distinct circulations at scales longer than 100 km and persist over timescales of 3 hours to 1 day (Parker & Johnson, 2000; Houze, 2004; Feng et al., 2019). These systems account for over half of extreme rainfall events at 24-h duration in the warm season in these regions (Schumacher & Johnson, 2006; Stevenson & Schumacher, 2014). Landfalling tropical cyclones (TCs) also contribute substantially in the summer and fall in the eastern and southeastern United States, especially in coastal regions (Shepherd et al., 2007; Knight & Davis, 2009; Miniussi et al., 2020). In the cool season, extreme precipitation results primarily from strong synoptic forcing events (Maddox et al., 1979; Schumacher & Johnson, 2006; Stevenson & Schumacher, 2014). Although synoptic forcing events occur with nearly unchanging frequency throughout the year, MCSs and TCs are much more seldom present in DJF (Stevenson & Schumacher, 2014), and extremes in DJF tend to be of lower magnitude than in JJA (Maddox et al., 1979; Stevenson & Schumacher, 2014). Individual storms of these types all tend to produce heavy precipitation over length scales of 100 km or more. The long correlation lengths in the statistics of extreme precipitation presented here can thus be partially attributed to the long correlation lengths of individual events. This interpretation is in good agreement with Touma et al. (2018), who used indicator semivariograms to assess the correlation scales of 90th percentile rainfall days over CONUS. Although that analysis was split into more climatological regions, their North, Northeast, South, and Southeast regions all display DJF length scales within 1σ of 300 km.

Previous studies (e.g. Kursinski & Mullen, 2008) have shown, perhaps in apparent tension with the above, that individual extreme storms can be highly localized in both space and time, with heavy precipitation falling over spatial scales of $\sim$50 km or less. Importantly, though, the spatial statistics of rainfall depend strongly on the time cadence considered. In a case study of the Cévennes-Vivarais region of France, Lebel et al. (1987) and Kirstetter et al. (2010) showed that the decorrelation distance of rainfall amounts increases with lengthening temporal scale from hourly to daily cadence. We wish to stress that our results are only valid at the daily cadence we considered; the assumption of smoothly-varying GEV statistics between rain gauges, and therefore the R19 technique, may not be justified at sub-daily cadences.

It is interesting to consider Figure 2 in terms of the fractal properties of rainfall explored by Lovejoy (1982) and Lovejoy and Mandelbrot (1985). Those authors describe the spatial structure cloud and rain areas according to $N \sim L^{-D}$, where N is the extent to which a fractal fills space as measured at scale L. Bies et al. (2016) explored the relationship between the fractal and power spectrum interpretations of scaling fields, finding that the fractal dimension D and β in Equation 1 are related according to $D = 1 + (4 - \beta)/2$ for a 2-dimensional field. (In the terminology of Bies et al. (2016), we measured here a "surface β" and the cited papers use a "coastal edge D".) We find $\beta \approx 2$ (which leads to $D \approx 2$) for this process up to scales of a few hundred km; that is, the extreme precipitation field is 2-dimensional. This is another way of interpreting the high level of autocorrelation on small scales. $D = 2$ is larger than the fractal dimension found by Lovejoy and Mandelbrot (1985) for cloud and rain areas, meaning that the spatial scales over which extreme statistics vary in Stage IV are larger than the spatial scales of individual precipitation events.

6 Conclusions

We have tested the assumption that the climatology of extremes varies only minimally at length scales smaller than the average inter-rain-gauge spacing of $\sim$30 km in the eastern CONUS. We find that this assumption is valid: 5-year daily return values are strongly autocorrelated at scales up to at least 100 km in both DJF and JJA. We also find that the fit-then-grid algorithm of R19 substantially improves the fidelity of daily extreme statistics compared with the grid-then-fit Mountain Mapper technique. On both 4-km and 25-km scales, the grid-then-fit Mountain Mapper technique underestimates extremes relative to the more spatially complete multi-sensor Stage IV QPE in the eastern United States, whereas the Risser et al. (2019) technique measures extremes more accurately than Mountain Mapper at 25 km scales. Taken together, these findings show that rain gauge observations are sufficient to capture the large majority of the extreme-value information in the climatology of the true rain field, but only if interpolated appropriately for the application of interest. This paper improves confidence that appropriately-constructed gauge-only gridded products provide an accurate historical record of daily extreme statistics beyond the years in which radar data are available, an important step toward creating an accurate, continental-scale, in-situ-based, long-term precipitation record for use in hydrological modeling, resource management, and climate change studies. As the resolution of global circulation models continues to increase into the future, QPEs will be required at finer and finer scales, and standard gauge-interpolation techniques will fail to accurately represent precipitation within these grid boxes. The human impacts of extreme events are felt at human scales, e.g. homes (10m), farms (1 km), and watersheds (10 km). Our work moves toward casting measurements of extremes into a risk framework at those scales.

Acknowledgments

This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No.

DE-AC02-05CH11231 and used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231.

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

We thank the staff of the California-Nevada, Colorado Basin, and Northwest NOAA River Forecast Centers, particularly John Lhotak, Dan Kozlowski, and Kyle Lerman, for their help accessing data and implementing the Mountain Mapper algorithm.

We thank Bill Boos and Travis O'Brien for their detailed and insightful comments, which substantially improved this publication.

Data Availability Statement

Datasets for this research are available at the following locations:

- The NEXRAD Stage IV data product is described in Lin (2011) and can be accessed online at `https://data.eol.ucar.edu/dataset/21.093`
- The GHCN-Daily data is described in Menne et al. (2012) and can be accessed online at `https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00861`.
- The CNRFC gridded QPE product can be accessed online at `https://www.cnrfc.noaa.gov/arc_search.php`

- The PRISM NORM81 climatology is described in Daly et al. (2015) and can be accessed online at `http://www.prism.oregonstate.edu/recent/`

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131. doi: 10.1002/joc.3413

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, *116*(D2), D02115. doi: 10.1029/2010JD014741

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A., ... others (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, *111*(D5), D05109. doi: 10.1029/2005JD006290

Behnke, R., Vavrus, S., Allstadt, A., Albright, T., Thogmartin, W. E., & Radeloff, V. C. (2016). Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecological applications*, *26*(5), 1338–1351. doi: 10.1002/15-1061

Bies, A. J., Boydston, C. R., Taylor, R. P., & Sereno, M. E. (2016). Relationship between fractal dimension and spectral scaling decay rate in computer-generated fractals. *Symmetry*, *8*(7), 66. doi: 10.3390/sym8070066

Chen, C.-T., & Knutson, T. (2008, April). On the verification and comparison of extreme rainfall indices from climate models. *Journal of Climate*, *21*(7), 1605–1621. Retrieved from `https://doi.org/10.1175/2007jcli1494.1` doi: 10.1175/2007jcli1494.1

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer. doi: 10.1007/978-1-4471-3675-0

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, *33*(2), 140-158. doi: 10.1175/1520-0450(1994)033⟨0140:ASTMFM⟩2.0.CO;2

Daly, C., Smith, J. I., & Olson, K. V. (2015). Mapping atmospheric moisture climatologies across the conterminous United States. *PLOS ONE*, *10*(10), 1-33. doi: 10.1371/journal.pone.0141140

Feng, Z., Houze, R. A., Leung, L. R., Song, F., Hardin, J. C., Wang, J., ... Homeyer, C. R. (2019, September). Spatiotemporal characteristics and large-scale environments of mesoscale convective systems east of the Rocky Mountains. *Journal of Climate*, *32*(21), 7303–7328. doi: 10.1175/jcli-d-19-0137.1

Fulton, R. A., Breidenbach, J. P., Seo, D.-J., Miller, D. A., & O'Bannon, T. (1998). The WSR-88D rainfall algorithm. *Weather and forecasting*, *13*(2), 377–395. doi: 10.1175/1520-0434(1998)013⟨0377:TWRA⟩2.0.CO;2

Gervais, M., Tremblay, L. B., Gyakum, J. R., & Atallah, E. (2014). Representing extremes in a daily gridded precipitation analysis over the United States: Impacts of station density, resolution, and gridding methods. *Journal of Climate*, *27*(14), 5201–5218. doi: 10.1175/JCLI-D-13-00319.1

Groisman, P. Y., Karl, T. R., Easterling, D. R., Knight, R. W., Jamason, P. F., Hennessy, K. J., ... Zhai, P.-M. (1999). Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climatic Change*, *42*(1), 243–283. doi: 10.1023/A:1005432803188

Houze, R. A. (2004). Mesoscale convective systems. *Reviews of Geophysics*, *42*(4), RG4003. doi: 10.1029/2004rg000150

Iguchi, T., Kozu, T., Kwiatkowski, J., Meneghini, R., Awaka, J., & Okamoto, K. (2009). Uncertainties in the rain profiling algorithm for the TRMM precipitation radar. *Journal of the Meteorological Society of Japan. Ser. II*, *87*, 1–30. doi: 10.2151/jmsj.87A.1

Jones, P. G., Gladkow, A., & Jones, A. L. (2001). *Floramap: A computer tool for predicting the distribution of plants and other organisms in the wild [CD-ROM].* Centro Internacional de Agricultura Tropical (CIAT).

Kirstetter, P.-E., Delrieu, G., Boudevillain, B., & Obled, C. (2010). Toward an error model for radar quantitative precipitation estimation in the Cévennes–Vivarais region, France. *Journal of Hydrology*, *394*(1-2), 28–41. doi: 10.1016/j.jhydrol.2010.01.009

Knight, D. B., & Davis, R. E. (2009). Contribution of tropical cyclones to extreme rainfall events in the southeastern United States. *Journal of Geophysical Research*, *114*(D23), D23102. doi: 10.1029/2009jd012511

Kursinski, A. L., & Mullen, S. L. (2008). Spatiotemporal variability of hourly precipitation over the eastern contiguous United States from stage IV multisensor analyses. *Journal of Hydrometeorology*, *9*(1), 3–21. doi: 10.1175/2007JHM856.1

Lebel, T., Bastin, G., Obled, C., & Creutin, J. D. (1987). On the accuracy of areal rainfall estimation: A case study. *Water Resources Research*, *23*(11), 2123–2134. doi: 10.1029/wr023i011p02123

Lin, Y. (2011). *GCIP/EOP surface: Precipitation NCEP/EMC 4km gridded data (GRIB) stage IV data. Version 1.0.* UCAR/NCAR - Earth Observing Laboratory. Retrieved from `https://data.eol.ucar.edu/dataset/21.093` doi: 10.5065/D6PG1QDD

Lin, Y., & Mitchell, K. E. (2005, 1). The NCEP stage II/IV hourly precipitation analyses: development and applications. In *Proceedings of the 19th Conference on Hydrology* (Vol. 1.2). Retrieved from `https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm`

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., ... Lettenmaier, D. P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, *26*(23), 9384–9392. doi: 10.1175/JCLI-D-12-00508.1

Lovejoy, S. (1982). Area-perimeter relation for rain and cloud areas. *Science*, *216*(4542), 185–187. doi: 10.1126/science.216.4542.185

Lovejoy, S., & Mandelbrot, B. B. (1985). Fractal properties of rain, and a fractal model. *Tellus A*, *37*(3), 209–232. doi: 10.3402/tellusa.v37i3.11668

Maddox, R. A., Chappell, C. F., & Hoxit, L. R. (1979). Synoptic and meso-α scale aspects of flash flood events. *Bulletin of the American Meteorological Society*, *60*(2), 115–123. doi: 10.1175/1520-0477-60.2.115

McGraw, D., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2019). Precipitation frequency analyses based on radar estimates: An evaluation over the contiguous United States. *Journal of Hydrology*, *573*, 299–310. doi: 10.1016/j.jhydrol.2019.03.032

Mehran, A., AghaKouchak, A., & Phillips, T. J. (2014). Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *Journal of Geophysical Research: Atmospheres*, *119*(4), 1695–1707. doi: 10.1002/2013JD021152

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910. doi: 10.1175/JTECH-D-11-00103.1

Miniussi, A., Villarini, G., & Marani, M. (2020). Analyses through the metastatistical extreme value distribution identify contributions of tropical cyclones to rainfall extremes in the eastern united states. *Geophysical Research Letters*, *47*(7), e2020GL087238. doi: 10.1029/2020gl087238

Nelson, B. R., Prat, O. P., Seo, D.-J., & Habib, E. (2016). Assessment and implications of NCEP stage IV quantitative precipitation estimates for product intercomparisons. *Weather and Forecasting*, *31*(2), 371-394. doi: 10.1175/WAF-D-14-00112.1

NOAA. (2006). Federal meteorological handbook No. 11, Doppler radar meteorological observations, part C, WSR-88D products and algorithms (FCM-H11C-2006 ed.) [Computer software manual]. Washington, DC. (https://www.ofcm.gov/publications/fmh/FMH11/FMH-11-PartC-April2006.pdf)

Parker, M. D., & Johnson, R. H. (2000, October). Organizational modes of midlatitude mesoscale convective systems. *Monthly Weather Review*, *128*(10), 3413–3436. doi: 10.1175/1520-0493(2001)129⟨3413:omommc⟩2.0.co;2

Parra, J. L., Graham, C. C., & Freile, J. F. (2004). Evaluating alternative data sets for ecological niche models of birds in the andes. *Ecography*, *27*(3), 350–360.

Prat, O. P., & Nelson, B. R. (2015). Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002-2012). *Hydrology and Earth System Sciences*, *19*(4), 2037-2056. doi: 10.5194/hess-19-2037-2015

Risser, M. D., Paciorek, C. J., Wehner, M. F., O'Brien, T. A., & Collins, W. D. (2019). A probabilistic gridded product for daily precipitation extremes over the United States. *Climate Dynamics*, *53*(5-6), 2517-2538. doi: 10.1007/s00382-019-04636-0

Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., & Bloomfield, J. (2002). Increased crop damage in the US from excess precipitation under climate change. *Global Environmental Change*, *12*(3), 197–202. doi: 10.1016/

S0959-3780(02)00008-0

Schaake, J., Henkel, A., & Cong, S. (2004). Application of prism climatologies for hydrologic modeling and forecasting in the western US. In *Preprints, 18th Conf. on Hydrology, Seattle, WA* (Vol. 5). Amer. Meteor. Soc. Retrieved from `https://ams.confex.com/ams/84Annual/techprogram/paper_72159.htm`

Schär, C., Ban, N., Fischer, E. M., Rajczak, J., Schmidli, J., Frei, C., ... others (2016). Percentile indices for assessing changes in heavy precipitation events. *Climatic Change*, *137*(1-2), 201–216. doi: 10.1007/s10584-016-1669-2

Schumacher, R. S., & Johnson, R. H. (2006, February). Characteristics of U.S. extreme rain events during 1999–2003. *Weather and Forecasting*, *21*(1), 69–85. doi: 10.1175/waf900.1

Schumann, A. H. (2011). *Flood risk assessment and management: How to specify hydrological loads, their consequences and uncertainties*. Springer Science & Business Media. doi: 10.1007/978-90-481-9917-4

Seo, B.-C., Krajewski, W. F., Quintero, F., ElSaadani, M., Goska, R., Cunha, L. K., ... others (2018). Comprehensive evaluation of the IFloods radar rainfall products for hydrologic applications. *Journal of Hydrometeorology*, *19*(11), 1793–1813. doi: 10.1175/JHM-D-18-0080.1

Seo, D.-J., & Breidenbach, J. (2002). Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology*, *3*(2), 93–111.

Shepherd, J. M., Grundstein, A., & Mote, T. L. (2007). Quantifying the contribution of tropical cyclones to extreme rainfall along the coastal southeastern United States. *Geophysical Research Letters*, *34*(23), L23810. doi: 10.1029/2007gl031694

Sheridan, P., Smith, S., Brown, A., & Vosper, S. (2010). A simple height-based correction for temperature downscaling in complex terrain. *Meteorological Applications*, *17*(3), 329–339. doi: 10.1002/met.177

Spies, R. R., Over, T. M., & Ortel, T. W. (2018). *Comparison of NEXRAD multisensor precipitation estimates to rain gauge observations in and near Dupage County, Illinois, 2002–12* (Open-File Report 2018–1061). Reston, Virginia 2018: U.S. Geological Survey. (30 p., doi:10.3133/ofr20181061)

Stevenson, S. N., & Schumacher, R. S. (2014). A 10-year survey of extreme rainfall events in the central and eastern United States using gridded multisensor precipitation analyses. *Monthly Weather Review*, *142*(9), 3147–3162. doi: 10.1175/mwr-d-13-00345.1

Sun, X., & Barros, A. P. (2010). An evaluation of the statistics of rainfall extremes in rain gauge observations, and satellite-based and reanalysis products using universal multifractals. *Journal of Hydrometeorology*, *11*(2), 388–404. doi: 10.1175/2009JHM1142.1

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A., … others (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104*, 70–97. doi: 10.1016/j.atmosres.2011.10.021

Timmermans, B., Wehner, M., Cooley, D., O'Brien, T., & Krishnan, H. (2019). An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate dynamics*, *52*(11), 6651–6670. doi: 10.1007/s00382-018-4537-0

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, *79*(1), 61–78. doi: 10.1175/1520-0477(1998)079⟨0061:APGTWA⟩2.0.CO;2

Touma, D., Michalak, A. M., Swain, D. L., & Diffenbaugh, N. S. (2018). Characterizing the spatial scales of extreme daily precipitation in the United States. *Journal of Climate*, *31*(19), 8023–8037. doi: 10.1175/JCLI-D-18-0019.1

Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., … Frieler, K. (2019). The effects of climate extremes on global agricultural yields. *Environmental Research Letters*, *14*(5), 054010. doi: 10.1088/1748-9326/ab154b

Willie, D., Chen, H., Chandrasekar, V., Cifelli, R., Campbell, C., Reynolds, D., … Zhang, Y. (2017). Evaluation of multisensor quantitative precipitation estimation in Russian River basin. *Journal of Hydrologic Engineering*, *22*(5), E5016002. doi: 10.1061/(ASCE)HE.1943-5584.0001422

Young, C. B., Bradley, A. A., Krajewski, W. F., Kruger, A., & Morrissey, M. L. (2000). Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. *Journal of Hydrometeorology*, *1*(3), 241–254. doi: 10.1175/1525-7541(2000)001⟨0241:ENMPEF⟩2.0.CO;2