Characterizing the Spatial Scales of Extreme Daily Precipitation in the United States

DANIELLE TOUMA

Department of Earth System Science, Stanford University, Stanford, California

ANNA M. MICHALAK

Department of Global Ecology, Carnegie Institution for Science, Stanford, California

DANIEL L. SWAIN

Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, California, and The Nature Conservancy, Arlington, Virginia

NOAH S. DIFFENBAUGH

Department of Earth System Science, and Woods Institute for the Environment, Stanford University, Stanford, California

(Manuscript received 13 January 2018, in final form 25 May 2018)

ABSTRACT

The spatial extent of an extreme precipitation event can be important for a basin's hydrologic response and subsequent flood risk, and may yield insights into underlying atmospheric processes. Using a relaxed movingneighborhood approach, we develop indicator semivariograms based on precipitation records from the Global Historical Climatology Network-Daily (GHCN-D) station network to directly quantify the climatological length scales of extreme daily precipitation over the United States during 1965-2014. We find that the length scales of extreme (90th percentile) daily precipitation events vary both regionally and seasonally. Over the eastern half of the United States, daily extreme precipitation length scales reach 400 km during the winter months, but are approximately half as large during the summer months. The Northwest region, on the other hand, exhibits little seasonal variation, with extreme precipitation length scales of approximately 150 km throughout the year. By leveraging in situ station measurements, our study avoids some of the uncertainties associated with satellite or interpolated precipitation data, and provides the longest climatological assessment of length scales of extreme daily precipitation over the United States to date. Although the length scales that we calculate can be sensitive to station density, neighborhood size, and neighborhood relaxation, we find that the interregional and interseasonal differences in length scales are relatively robust. Our method could be extended to quantify changes in the spatial extent of extreme daily precipitation in the recent past, and to investigate the underlying causes of any changes that are detected.

1. Introduction

Extreme precipitation events tend to cause high surface runoff, and can subsequently lead to flooding that is capable of inflicting extensive economic, societal, and ecological damages. In 2017 alone, the United States experienced four major flooding events from extreme precipitation, resulting in 211 deaths and over \$180 billion (U.S. dollars) in damages (NOAA/NCEI 2017). Typically, extreme precipitation events are of relatively short duration (lasting from a few hours to a few days) but produce damages comparable to those resulting from long-duration droughts (which often last a year or more). Additionally, recovery can be slow and costly, as evidenced by recent examples such as the 2013 Boulder floods (Gochis et al. 2015) and the 2016 Louisiana floods (van der Wiel et al. 2017). Understanding the spatial characteristics of these extreme precipitation events is important not only because of their considerable societal impacts, but also because these characteristics have

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-18-0019.s1.

Corresponding author: Danielle Touma, detouma@stanford.edu

DOI: 10.1175/JCLI-D-18-0019.1

^{© 2018} American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

been changing in recent decades as atmospheric temperatures increase (O'Gorman 2015).

The frequency and intensity of extreme wet events have been increasing across much of the globe (IPCC 2012), including over much of the United States during the past half century (DeGaetano 2009; Pryor et al. 2009; Gleason et al. 2008). In some cases, these changes in the frequency and/or intensity have been attributed to increasing greenhouse gas concentrations (Min et al. 2013; Diffenbaugh et al. 2017; Cohen et al. 2014). The observed increase in the intensity of extreme wet events has been partly attributed to the thermodynamic contribution of increasing temperatures, and in some cases partly attributed to changes in atmospheric circulation (Trenberth et al. 2003; Trenberth 2011; O'Gorman 2015; Diffenbaugh et al. 2017). Accordingly, as atmospheric temperatures continue to rise, the frequency and intensity of extreme wet events are expected to change further in the future, although such changes are likely to manifest nonuniformly in space and time (Diffenbaugh et al. 2005; Gleason et al. 2008; Singh et al. 2013; Trapp et al. 2007).

Studies have also started to investigate the effect of rising temperatures on the spatial characteristics of precipitation. For example, Wasko et al. (2016) found that increasing temperatures reduce the spatial extents of extreme rain storms in Australia. Similarly, Chang et al. (2016) and Guinard et al. (2015) found that higher temperatures reduce the spatial extent of rainstorms over the United States. However, other studies, such as that of Dwyer and O'Gorman (2017), found inconclusive results when assessing the change in zonal length scale of precipitation events in the tropics. Although these studies (and others) have already begun to quantify changes in the spatial extent of precipitation in response to increasing temperatures, a robust assessment of the longterm historical climatology of spatial extent of extreme precipitation has not yet been developed.

A key challenge in developing such a climatological assessment is the brevity and uncertainty of pixelated or gridded datasets, including those derived from ground-based weather radar, space-based satellite precipitation records, and statistically (e.g., Daymet; Thornton et al. 1997) or dynamically (e.g., reanalysis) interpolated datasets. Decadal variability of the climate system can induce apparent (but spurious) long-term trends (Deser et al. 2014, 2012; Thompson et al. 2015; Li et al. 2017; Endo et al. 2017; Hawkins et al. 2016)—and/or mask trends arising from changes in radiative forcing—in short (~30 yr) radar and/or satellite-derived observations (Schneider et al. 2013). These assessments can be further affected by the reported overestimation of the frequency of extreme precipitation in satellite data when compared to ground and radar measurements (Mehran and Aghakouchak 2014; Aghakouchak et al. 2011). Although station-based interpolated datasets can provide longer periods of record, they consistently underestimate extreme precipitation, especially over regions with sparser ground measurements (Sun and Barros 2010; Behnke et al. 2016).

Although measurement uncertainties in precipitation gauge data have been reported in multiple studies (Rasmussen et al. 2012; Sieck et al. 2007; Tokay et al. 2010; Liu et al. 2013), assessing the characteristics of extreme precipitation using in situ data directly overcomes the need for interpolation methods to estimate extreme precipitation. Additionally, station datasets provide a much longer period of record than is available from satellites or radar, allowing robust assessments of long-term trends in the characteristics of extreme precipitation.

Our objective in the present study is to infer the spatial extent of extreme daily precipitation using station data, thereby overcoming the limitations of gridded datasets. Given that we cannot assume that precipitation varies smoothly between adjacent stations, we quantify the length scale of extreme daily precipitation (rather than the size of continuous objects). Our framework, which we describe below, employs indicator semivariograms to infer the length scale of extreme daily precipitation using the Global Historical Climatology Network-Daily (GHCN-D) dataset of U.S. stations (Menne et al. 2012). We use our approach to identify regional and seasonal variations in the climatological length scales for the 1965-2014 period. Additionally, we assess the sensitivity of those variations to our methodological choices.

2. Data

We use the GHCN-D station dataset, which contains daily precipitation records since 1861. To balance the need for a set of stations that has continuous data availability with the need for a period of record that is sufficiently long for climatological assessment, we focus our analysis on the period from 1965 to 2014. Over North America, 4512 stations in the GHCN-D dataset have (partially intermittent) records that start by 1965 and continue until or past 2014. These 4512 stations are shown in Fig. 1a. Across all stations, more than half of the observations used are based on 24-h accumulation periods recorded between 0700 and 0800 local time (LT), while 15% are recorded at 1800 LT, and approximately a tenth are recorded at midnight LT.

We first find all daily precipitation values from 1965 to 2014 that are greater than or equal to 1 mm day^{-1} for each station and each calendar month. We then define



FIG. 1. (a) Locations and NCEI regions for GHCN-D stations used in the analysis. Stations shown have been recording data between 1965 and 2014. The number of stations in each region is in parentheses. (b) The number of stations used in each 500-km radius neighborhood for p90 semivariogram analysis. (c) Frequency distribution of the number of stations used in 500-km radius neighborhoods found in each region.

the 90th percentile value of these precipitation values as the respective extreme precipitation threshold for each station and each calendar month. We summarize these 90th percentile values in 3-month (or seasonal) averages in Fig. 2, using December–February (DJF) for winter, March–May (MAM) for spring, June–August (JJA) for summer, and September–November (SON) for autumn.

Using this percentile threshold value for each month and each station, we create a binary, or indicator, extreme event dataset (hereafter "p90") for the whole period of 1965–2014, where stations that have a daily precipitation value equal to or above the corresponding monthly 90th percentile threshold are set equal to 1, and stations that have a daily precipitation value less than that threshold are set equal to 0. Stations that have no recorded data on a given day are eliminated from the binary dataset for that day.

3. Methods

a. Overview of methodology

We use indicator semivariograms to quantify the length scale of extreme precipitation using our binary

p90 dataset. Indicator semivariograms have been used in previous studies to spatially interpolate the probability of the presence of a threshold variable using indicator kriging (e.g., Berezowski et al. 2016; Goovaerts et al. 2016; Haberlandt 2007). However, in our study, we use indicator semivariogram methods to directly compute the length scales of extreme precipitation, without kriging the binary variable. The indicator semivariogram quantifies the squared difference between the values of two data points as a function of the separation distance between those points. By calculating the indicator variogram for multiple pairs of stations (as a function of their separation distance) in a given region and time, we can find the distance at which stations with p90 values are no longer substantially correlated in that region and time period. We use this distance to quantify the length scale of p90 precipitation for that given region and time.

As shown in previous studies using radar, satellite, and climate model data, length scales of precipitation can exhibit substantial regional variation over the United States (Guinard et al. 2015; Dwyer and O'Gorman 2017; Chang et al. 2016; Kursinski and Mullen 2008). To ensure that we capture the nonstationarity of spatial



Seasonal average of monthly 90th percentile precipitation 1965-2014

FIG. 2. Monthly 90th percentile of daily precipitation from 1965 to 2014 for each station averaged over each season. For each month, the 90th percentile of daily precipitation greater or equal to 1 mm over 1965–2014 is found for each station. The monthly 90th percentile value is used to create the p90 dataset (e.g., Fig. 3), but for this figure we show the seasonal average of the 90th percentile value of each month at each station.

characteristics, we directly address regional differences in length scales by using a moving neighborhood approach similar to that developed by Haas (1990). Haas (1990) estimated local semivariogram parameters for sulfate deposition data by using neighborhoods of stations centered around locations of interest in order to provide kriging estimates, and found that this method provides a more accurate representation of the semivariogram structure than when using global semivariogram parameters. The moving neighborhood approach has also been used extensively to estimate local semivariogram parameter sets for spatial prediction (e.g., Lloyd 2005, 2010; Tadić et al. 2015), resulting in decreased prediction error relative to the use of a global semivariogram parameter set.

However, when using a moving neighborhood approach, the number of observations within a neighborhood or the size of the neighborhood must first be established. While most studies use a single neighborhood size (or density) to estimate local semivariogram parameters based on well-known physical traits of the variable of interest (e.g., Alkhaled et al. 2008; Hammerling et al. 2012; Tadić et al. 2015), other studies quantify the effect of the number of observations within a neighborhood on the errors of estimation. More specifically, Lloyd (2005) found that a larger number of observations within a neighborhood produced smaller errors for monthly precipitation estimation over the United Kingdom. Likewise, Oyler et al. (2015) found that minimum and maximum temperatures can be locally estimated with little error over the United States using ~80 stations per neighborhood. In our study, we test the sensitivity of our results to the number of observations in each neighborhood, and also to the size of the neighborhood.

Similar to Alkhaled et al. (2008) and Hammerling et al. (2012), we also "relax" the boundaries of a neighborhood in order to account for larger scales of variability, and to prevent unnatural cutoffs of extreme precipitation areas when calculating the spatial extent of extreme precipitation over a certain location. We relax the neighborhood by allowing stations outside of the boundary to be included as one member of a station pair in our semivariogram calculation. We also quantify the effect of this neighborhood relaxation on our calculation of the spatial extent, and assess whether a relaxed neighborhood reduces the sensitivity to different neighborhood sizes.

The theory and calculation of indicator semivariograms is described in detail by Goovaerts (1997) and applied in many studies to understand the spatial structure of environmental variables (e.g., Berezowski et al. 2016; Goovaerts et al. 2016; Haberlandt 2007). In the following subsections, we first describe the methods for selecting the data used in the calculation of indicator semivariograms (including the moving neighborhood method) so that we capture any seasonal and regional variations in p90 length scales. We then detail the use of raw, experimental, and theoretical variograms to quantify p90 precipitation length scales.

b. Seasonal and regional considerations

To define the length scale of extreme precipitation for a station, we first delineate a 500-km-radius neighborhood around that station. Using a 500-km neighborhood allows us to capture the spatial heterogeneity of extreme daily precipitation over different regions. We then ensure that there are at least 20 stations within that 500-km neighborhood. (A total of eight stations are eliminated because their 500-km neighborhoods encompass fewer than 20 stations.)

Because we are interested in the climatological length scale of p90 precipitation for each season, we calculate the raw semivariogram for the days in which at least 10% of the stations in an individual neighborhood show a p90 event in a season (hereafter, "p90 days"). Given that our threshold for an extreme event is the 90th percentile, this 10% restriction ensures that our analysis only encompasses days that exceed the number of events expected to occur at random within each neighborhood. Because of spatial correlations in our p90 dataset, approximately 2%-12% of days in each season over the whole period (1965-2014) are selected as p90 days (see Fig. S1 in the online supplemental material). JJA shows the largest percentage of days eliminated by this criterion, especially in the eastern half of the United States. Interestingly, for neighborhoods on the Pacific coast, fewer than 50% of days exhibit at least one station with p90 precipitation (Fig. S1).

For raw semivariogram calculations, we only use pairs of stations that include at least one station with a binary value of 1 on a given day. By using only pairs of stations that both meet the threshold (hereafter, "1–1 pairs"; red–red pairs in Fig. 3a) or for which one meets the threshold and one does not (hereafter, "1–0 pairs"; red-black in Fig. 3a), we ensure that pairs of two nonthreshold stations (hereafter, "0–0 pairs"; black-black in Fig. 3a) are not included in our characterization of the length scale of p90 daily precipitation. To implement our relaxation technique, we also use 1–1 and 1–0 pairs that are partly in the neighborhood (i.e., with one station inside the neighborhood and the other station outside the neighborhood; red–pink, red–gray, and black–pink pairs in Fig. 3a) to capture length scales that are representative of the neighborhood but that may be longer than the neighborhood dimensions. Pairs of stations that fall completely outside the neighborhood are excluded (pink–pink, gray–gray, and pink–gray pairs in Fig. 3a). All raw semivariogram calculations from all pairs on all p90 days are weighted equally.

To assess regional variations in the length scales, we summarize our results using nine U.S. regions, as defined by the National Centers for Environmental Information (NCEI) (Karl and Koss 1984). The NCEI regions are shown in Fig. 1a. (Note that there are several stations in Mexico and Canada that are included in our single-station analysis but that are not included in any of the regional summaries.) To assess the statistical significance of regional differences in length scales within each season we use the Mann-Whitney U test. To assess the statistical significance of the seasonal variations in each region, we use the paired Wilcoxon signed-rank test, which is a nonparametric paired difference test (McDonald 2009). We report the p values of the tests, where the p values have been adjusted using the Bonferroni method, which penalizes p values when conducting multiple tests using the same sample (Wright 1992)-that is, 6 intraseasonal comparisons per region and 36 intraregional comparisons per season.

c. Raw and experimental semivariograms

On any given day, the raw semivariogram value can be calculated for a pair of selected stations using the 500-km relaxed moving neighborhood around a station for distances up to 500 km for p90. The raw semivariogram is calculated as shown in Eq. (1) and in Fig. 3b:

$$\gamma(h) = 0.5 \times [Z(x) - Z(x+h)]^2, \tag{1}$$

where *h* is the distance between two stations, Z(x) is the p90 event value at a station at location *x*, and Z(x + h) is the p90 event value at a station that is at distance *h* away from *x*. Because the p90 event value, Z(x), is equal to 1 when a station's precipitation exceeds or equals the p90 threshold and is equal to 0 when it falls below the threshold, the semivariogram $\gamma(h)$ has only two possible outcomes: 0 when two stations have the same Z(x) value on a given day, and 0.5 when two stations have different



FIG. 3. Illustration of method for SON climatology range calculation for a station in Colorado using the 133 p90 days selected in SON over 1965–2014 for the station (as described in section 3). (a) The p90 station data for days 1, 2, 3, and 133. The stations within the neighborhood are shown in red and black and stations outside the neighborhood are shown in gray and pink. Stations that exceed the p90 threshold for these particular days are in red and pink, and stations that are below the p90 threshold are shown in black and gray. (b) Raw indicator semivariograms for days 1, 2, 3, and 133. The number of pairs that have a semivariogram value of 0 (1–1 pairs in purple) or 0.5 (1–0 pairs in blue) in each separation distance interval (*x* axis). (c) The sum of all raw indicator semivariograms over 133 selected days in SON at 0 or 0.5 (purple bars for 1–1 pairs and blue bars for 1–0 pairs). To calculate the experimental semivariogram (black circles), all semivariogram values in each separation distance interval (delineated by gray dashed vertical lines) are averaged to obtain one semivariogram value for the center of the distance interval. The exponential semivariogram (gray line) is fit to the experimental semivariogram as described in section 3, and the practical range, or length scale (vertical black line and arrow), is retrieved from the fitted exponential semivariogram model.

Z(x) values on the same day. Because we are only using pairs of stations that have at least one Z(x) value of 1 (i.e., 1–1 or 1–0 pairs, or red–red, red–black, red–pink, red–gray, and black–pink pairs in Fig. 3a), all raw

semivariogram values of 0 are calculated from 1-1 pairs and all values of 0.5 are calculated from 1-0 pairs.

Raw semivariogram values are calculated for p90 days in each season and for each station, and then grouped into 20 equally spaced intervals *i* of separation distances *h* (e.g., SON Station 599; Fig. 3b). We then calculate the experimental semivariogram $\overline{\gamma}(h_i)$ by taking the mean of raw semivariogram values in each interval *i* (black circles in Fig. 3c). In the case of raw indicator semivariogram values, we calculate a simple ratio between the number of 1–1 pairs $(N_{1-1,i})$ and the number of 1–0 pairs $(N_{1-0,i})$ of stations in each interval:

$$\overline{\gamma}(h_i) = 0.5 \left[\frac{N_{1-0,i}}{N_{1-0,i} + N_{1-1,i}} \right].$$
(2)

The experimental semivariogram (Fig. 3c; black circles) increases with greater separation distance and eventually asymptotes at a value close to 0.5. The shape of the relationship shows that stations farther apart are less likely to have p90 precipitation on the same day, meaning that there are more 1–1 pairs at smaller separation distances, while the number of 1–0 pairs mostly increases with larger distances, as expected.

d. Length scale of extreme precipitation calculated from fitted semivariograms

To characterize the relationship between the experimental semivariogram and the distance h between station pairs we fit a theoretical semivariogram to the experimental semivariogram. Various models of semivariograms are commonly used in the geostatistical literature (Chiles and Delfiner 2012). Following visual inspection of the experimental semivariograms and recommendations from previous studies (e.g., Berndt et al. 2014; Western et al. 1998), we fit the exponential model for p90 experimental semivariograms as defined in Eq. (3):

$$\gamma_{\exp}(h) = \begin{cases} c + b \left[1 - \exp\left(-\frac{3h}{\alpha}\right) \right], & h > 0\\ 0, & h = 0 \end{cases}$$
(3)

where *c* is the nugget, *b* is the partial sill, and α is the practical range of the semivariogram. The nugget *c* reflects measurement errors or microscale variability. The partial sill *b* represents the asymptotic value of the exponential semivariogram at a large separation distance. In this study, we focus on the practical range α , with which we represent the length scale of p90 daily precipitation (Goovaerts 1997; Haberlandt 2007; Fig. 3c). To fit exponential semivariograms to the experimental semivariograms calculated for each neighborhood during each season, we use the *variofit* function from the *geoR*

package (Ribeiro and Diggle 2016) in R (R Core Team 2015). By using this function, we can automatically fit exponential semivariograms to the experimental semi-variograms from all seasons and all neighborhoods.

Before presenting the results for the climatological length scale of extreme precipitation, we assess the results for two selected days for one station in Colorado, with the aim of assessing the implications of the relaxed moving neighborhood. For 10 September 2002, the calculated length scale was 242 km. For 10 September 2013, the calculated length scale was 73 km. The 2013 date indeed had much smaller clusters of stations with extreme precipitation, covering a smaller area, while the examined date in 2002 had a larger cluster of stations with extreme precipitation (see Fig. S2 for more details). The large extreme rainfall cluster centered over Illinois, Missouri, and Iowa in 2013 shows the importance of constraining the raw semivariogram calculations to pairs of stations with at least one station within a neighborhood, in order to eliminate any influences from distinct precipitation areas. In contrast, 2002 shows the importance of relaxing the neighborhood to include stations exhibiting p90 precipitation outside of the edge of the neighborhood, and to prevent an arbitrary, unnatural cutoff from the p90 stations located inside the neighborhood (e.g., see the northern edge of the neighborhood in Fig. S2a).

e. Sensitivity tests

We test the sensitivity of our results to a number of our methodological choices. First, to test the impact of the "relaxed" method, we recalculate the semivariograms using a restricted neighborhood, i.e., only using 1-1 and 1-0 pairs of stations within the neighborhood (red-red and red-black pairs in Fig. 3a). Second, to understand the effect of the neighborhood size, we recalculate the semivariograms using a 300-km neighborhood radius and a 700-km neighborhood radius, for both the relaxed and restricted neighborhood method. Last, we test the sensitivity of the number of stations per 500-km neighborhood by recalculating the semivariogram using a uniform number of 100, 200, and 300 stations within each neighborhood. For stations with neighborhoods that encompass greater than 100, 200, or 300 stations, we randomly subselect 100, 200, or 300 stations without replacement. We do not assess the sensitivity of the length scales for stations that originally had fewer than 100, 200, or 300 stations in their neighborhood. When selecting pairs of stations for calculating the raw semivariogram, stations outside the neighborhood remain intact. As a reference, 1%, 7%, and 30% of stations have fewer than 100, 200, and 300 stations, respectively, in their 500-km neighborhood.



Climatological length scale of daily p90 precipitation events p90 days | 1965-2014

FIG. 4. (a) The climatological length scale (km) for each season over p90 days 1965–2014 for daily p90 extreme rainfall events for all neighborhoods centered at all stations. (b) The spread of the p90 precipitation climatological length scale for each NCEI region (see Fig. 1 for region delineations) and for each season on p90 days, where the center line in the box is the 50th percentile of the length scales in the region, the bottom and top boundaries of the box represent the 25th and 75th percentiles, respectively, and the bottom and top lines extend to the 1st and 99th percentiles, respectively (EN = East North, WN = West North).

4. Results

a. Seasonal length scale of p90 precipitation

We calculate the seasonal climatology of the length scale extracted from the fitted semivariograms for extreme precipitation across the United States using a 500-km relaxed neighborhood (Fig. 4a). We find that the winter months (DJF) generally have the longest length scales of extreme precipitation and the summer months (JJA) generally have the shortest (Fig. 4a). Additionally, the longest length scales generally occur in the Central, East North Central, Northeast, South, and

West regions, while the smallest occur across the Northwest, Southwest, and West North Central regions. We find that the Central region has the longest median length scale in DJF, MAM, and SON, while the West region has the longest median length scale in JJA (Figs. 4 and 5a).

Variation in length scales between regions is highest during winter, and subtler during summer (Fig. 5a), although the differences in length scales between regions are statistically significant in most cases (p < 0.01). In DJF, the largest regional differences are between the Northwest region and the South, Central, and East

Variations in climatological length scales 500km relaxed neighborhood



FIG. 5. Magnitude and significance of (a) the intraregional differences in median p90 climatological length scales for each season and (b) the median intraseasonal differences in p90 climatological length scales for each region. The shading of each grid represents the difference of the region and season in (a) and (b), respectively, on the y axis from the region or season on the x axis. [For example, the blue bottom-right corner of the DJF panel in (a) shows that the difference of the median p90 climatological length scale of the Central region from the median p90 climatological length scale of the Central region from the median p90 climatological length scale of the WN Central region is between -45 and -90 km.] The symbol inside each box represents the adjusted p value of the Mann–Whitney U test when testing for intraregional differences in (a) and of the Wilcoxon signed-rank sum test when testing for intraseasonal differences in (b) (EN = East North, WN = West North).

North Central regions (differences of >135 km in each case) (Figs. 4b and 5a). In contrast, the largest JJA variations occur between the West region (median length scale \sim 220 km) and the Southwest region (median length scale of \sim 130 km) (Figs. 4b and 5a).

In addition to these regional differences in median length scale, we find that within-region spatial heterogeneity of length scales also varies across regions and seasons. For example, in the Northeast, the DJF interquartile range (IQR; \sim 90 km) is triple the JJA IQR (\sim 30 km). In contrast, the West exhibits an IQR of \sim 50–60 km in all seasons (Fig. 4b).

Although the magnitude of seasonal variations in climatological length scales differ across regions, the differences between seasonal length scales over each region are generally statistically significant (p < 0.001; Fig. 5b) and are plausibly related to underlying differences in the seasonally and regionally varying atmospheric processes that generate precipitation extremes. In most regions, JJA median length scales are significantly smaller than the median length scales in other seasons (p < 0.001), and DJF median length scales are significantly larger (p <0.001). (In the Northwest, MAM median length scales are smaller than the median length scales in other seasons, and SON median length scales are larger.) The difference between JJA and DJF median length scales is largest in the South (difference of >135 km), and the differences between JJA and SON length scales are largest (differences of >90 km in each case) over the East North Central, South, and West North Central regions (Figs. 4b and 5b). Although there is still a clear seasonal cycle in the Northwest and Southwest (Fig. 4b), the magnitude of the variations among median seasonal length scales remain below 45 km (Fig. 5b).

b. Sensitivity of p90 length scales to neighborhood size, relaxation, and station density

While the magnitude of length scales is sensitive to neighborhood size, relaxation, and station density, the overall patterns of regional and seasonal variations are relatively robust across all methodological choices assessed in our sensitivity analysis. Compared to length scales when using a 500-km relaxed neighborhood, using a smaller 300-km neighborhood yields a shorter length scale in 90%-95% of stations across seasons, and using a larger 700-km relaxed neighborhood yields a longer length scale in 87%-93% of stations across seasons (see Fig. S4). Compared to length scales calculated using relaxed neighborhoods, using a restricted neighborhood decreases extreme precipitation length scales in DJF, MAM, and SON by less than 25% for the majority of stations (67%, 57%, and 57%, respectively); also, 70% of the length scales in JJA increase by less than 10% (see Fig. S4). Likewise, using uniform station densities decreases length scales by up to 40% (see Fig. S5).

Nonetheless, the regional variations in length scales remain robust in the sensitivity tests. For example, the Central region consistently has significantly longer length scales than any other region in DJF, MAM, and SON, while the Southwest region has the longest length scales in JJA (see Fig. S6). The highest dependence of regional variations on different methodological choices is found in DJF. For instance, when using a 700-km relaxed neighborhood, the difference between the Northwest and West North Central length scale becomes positive and statistically significant. However, large differences in interregional variations are rare.

The seasonal variations are also robust to using different neighborhood sizes and station densities, as well as when using a restricted neighborhood method. For example, DJF length scales are largest in the Central, East North Central, South, Southeast, and West regions, while JJA length scales are smallest in all regions except for the Northwest over all methodological choices. The most prominent exceptions are the Northeast and West North Central regions, where methodological choices have a strong impact on the seasonal variations in length scales (see Fig. S7 for more details).

We note that these sensitivity analyses do not test all of the methodological choices in our framework. Though we test the sensitivity to the neighborhood size and the number of stations, we do not account for gradients in station density within the neighborhood. Given that the semivariogram calculation relies on the separation distance between stations, such gradients could ultimately bias our length scale calculation. For example, denser station availability in one part of a neighborhood could lead to a greater number of station pairs included in the experimental semivariogram at smaller separation distances from that part of the neighborhood. These station pairs would weight the experimental semivariogram at smaller separation distances, and consequently impact our length scale calculation.

Additional untested factors include the size of separation distance intervals for the calculation of the experimental semivariogram, the maximum separation distance to which the semivariograms are calculated, the years over which p90 thresholds and length scales are calculated, and the 10% threshold for choosing p90 days. However, our analysis shows that while the absolute magnitude of length scales can be sensitive to methodological choices (especially neighborhood size), the seasonal and regional differences in length scales are relatively robust.

5. Discussion and conclusions

This study presents a method for quantifying regional and seasonal variations in the spatial extent of extreme precipitation using station data, which is foundational to investigating changes in extreme precipitation in a warming world. Our method is advantageous in that the relaxed moving neighborhood allows for an assessment over a large region, and the geostatistical techniques we use allow us to employ longer station datasets without relying on interpolated precipitation data. Until the representations of extreme precipitation in satellite, radar, interpolated, and modeled datasets improve-and/or have been continuously observing the climate system for a longer period of time-our method can provide a foundation to understand the characteristics of extreme precipitation using direct, in situ measurements.

1 October 2018

a. Linking length scales to seasonal and regional atmospheric phenomena

The robust seasonal and regional variations in length scales appear to correspond to well-known atmospheric phenomena. The significantly shorter length scales of extreme precipitation in the summer months over most of the continental United States (with the exception of the Pacific Northwest and northern Great Plains) is characteristic of the intense but relatively localized downpours associated with convective storms—which tend to dominate warm-season precipitation in these areas. The longest length scales in each region, on the other hand, occur during winter (as do the highest median length scales in many areas)—coinciding with the passage of larger "synoptic scale" cyclones as the jet stream and associated storm track shift southward during the cool season (Schneider et al. 2011).

However, our findings show that smaller-scale processes could also be important in modifying the spatial extent of extreme precipitation during the cooler seasons. Short autumn, winter, and spring length scales in the northwest (Fig. 4) are consistent with Rutz et al. (2014), who show that heavy rainfall from atmospheric rivers (ARs) rarely penetrate past the Cascade Range in Washington and Oregon. In the rain-shadowed region immediately to the east of the Cascades, regional topography allows both westerly and southwesterly flows to generate heavy precipitation regimes (e.g., Rutz et al. 2014)-resulting in a more heterogeneous distribution of extremes, and shorter length scales. Farther to the south (across California and western Nevada), however, longer p90 length scales do extend farther inland. As in the Pacific Northwest, this may be largely a product of the coastal topography: California's coastal mountains are not as tall as the Cascades, and form a less perfect barrier to east-moving Pacific moisture, allowing longer length scales to extend inland across most of the state (Rutz et al. 2014). The much taller Sierra Nevada range, located along the eastern margin of California, has a similar effect as the Cascades, resulting in a strongly rain-shadowed region across the Great Basin in Nevada, and an eastward decrease in p90 length scales.

We note that p90 length scales across California are generally larger than across the Pacific Northwest, despite the predominance of AR-driven autumn–winter–spring extremes in both regions. We hypothesize that latitudinal asymmetry in the spatial orientation of ARs may be partially responsible for this difference. ARs may more frequently be zonally (east–west) oriented in the north as opposed to meridionally (north–south) oriented in the south (Rutz et al. 2015), meaning that these extreme events may affect a broader section of the north–south-oriented coastline in Washington/Oregon than in California. An assessment of the anisotropic length scales of extreme precipitation along zonal and meridional axes could yield further insights into topographical and storm track controls along the Pacific coast.

Conversely, topography in the northeastern United States plays a relatively smaller role in producing heterogeneity in winter length scales. Here, a significant fraction of winter extreme precipitation occurs during strong coastal cyclones (known regionally as "nor'easters"). These nor'easters generate large but coastally confined swaths of extreme precipitation from the upper mid-Atlantic (as far south as Delaware) to northern New England (as far north as coastal Maine) (Agel et al. 2015), which are shown by longer extreme precipitation length scales along the coast. Shorter p90 length scales are found in the vicinity of the Great Lakes, where a significant fraction of extreme wintertime precipitation results from intense but highly localized convective "lake-effect snow" bands (Niziol et al. 1995). During seasons other than winter, Northeastern p90 length scales are more spatially uniform-likely owing to the importance of eastward-moving, large-scale cyclonic storms in triggering precipitation extremes (Murray and Colle 2011).

Although summer length scales are more spatially uniform across the United States (relative to the cooler seasons), we still observe differences in length scales that coincide with regionalized summertime precipitation regimes. For example, the band of longer length scales across the Great Plains and portions of the Midwest (from southeast Texas to the edge of Lake Michigan) coincides with the region where mesoscale convective complexes (MCCs) most commonly form during the warm season. MCCs are organized storm systems, but they tend to have intermediate length scales (i.e., between larger-scale cyclones and individual thunderstorms), and frequently result in intense rainfall and severe weather across relatively broad regions (Laing and Fritsch 1997). On the other hand, parts of the Southeast, especially the Florida Peninsula and other areas along the Gulf Coast, feature intense but highly localized pulse or "popcorn" thunderstorms in the summer months, yielding much shorter characteristic length scales (Miller and Mote 2017). We note, however, that subdaily precipitation in the summer follows a strong diurnal cycle (Higgins et al. 1997), meaning that calculated length scales may be sensitive to the time of observation of daily precipitation accumulation among nearby stations.

In fact, given the daily temporal resolution of the station dataset, interpretations of the length scales of extreme precipitation in our results are not necessarily absolute. Longer length scales of daily extreme precipitation can be a result of a slow and large precipitation system, or a smaller but fast-moving system. An assessment of the length scales of subdaily or multiday precipitation using our method could yield more insight into the temporal controls on the length scales of extreme precipitation. Similar to Zhou and Matyas (2017), who assess the spatial characteristics of precipitation along hurricane tracks, our method could also be used to quantify the length scale of extreme precipitation along the trajectory of a given precipitation system by calculating the length scale of extreme precipitation around the centroid of the system. By following different storm tracks, we could disentangle the physical processes that produce different length scales of extreme precipitation for different types of precipitating systems, such as tropical storms or atmospheric rivers.

b. Strengths, limitations, and applications of the method

In contrast to the distinct regional and seasonal variations in climatological length scales found in our results, recent analyses of the spatial extents of extreme precipitation using radar and satellite data show more limited variations in climatological length scales within and between regions, and between seasons. For example, Guinard et al. (2015), who use 1-hourly radar data to assess precipitation objects over the United States from 1992 to 2001, present subtler seasonal variations in the size of precipitation objects over the Northeast, Central, and Southeast regions. Similarly, we show much more pronounced regional variations in p90 precipitation length scales in the eastern half of the United States (ranging from \sim 120 to 350 km; Fig. 5) than Dwyer and O'Gorman (2017), who used satellite data from 1998-2015 to assess the zonal length scale of p99 precipitation.

The magnified regional and seasonal variations found in our climatological quantification of p90 precipitation length scales may be a result of using a longer dataset, and/or of using in situ measurements of extreme precipitation directly. Applying our method to the same gridded datasets used in previous studies would help to identify the source of discrepancies in the regional and seasonal climatologies of extreme precipitation length scales. On the other hand, if the gridded dataset in question was originally oversmoothed, then applying our method could still result in misleading magnitudes of length scales due to that oversmoothing.

Discrepancies between our results and those of previous analyses could also be due to different thresholds of precipitation and different periods over which thresholds are calculated. In our study, long length scales of p90 precipitation over regions with large heterogeneities in p90 thresholds, such as east and west of the Sierra Nevada, may not necessarily represent extents of uniformly intense precipitation. In such regions with small-scale variations in p90 precipitation intensities resulting from topographic influences on storm tracks, quantifying length scales of extreme precipitation using a single, uniform, absolute precipitation threshold (rather than a percentile) may produce a more meaningful assessment of high-impact precipitation events, and may show more correspondence with previous uniform-threshold analyses of gridded datasets.

We also note that gauge measurements of rain and snow possess their own uncertainties due to varying gauge types, wind effects, and other factors (Rasmussen et al. 2012; Sieck et al. 2007), which could impact our extreme precipitation dataset and, consequently, the calculated length scales. Although previous experimental efforts have assessed measurement uncertainty in rain and snow gauges in different locations (e.g., Tokay et al. 2010; Liu et al. 2013), we do not quantify these uncertainties for all ~4500 stations in this study. However, we directly use the station measurements to infer length scales without smoothing or interpolating the dataset, eliminating further uncertainties that our extreme precipitation dataset could incur. Further, our method provides a clear framework to assess the sensitivity of our quantification of the length scale of extreme precipitation to methodological choices.

Given that our methodological choices result in differences in the magnitude of length scales, but not in differences in relative regional and seasonal variations, our framework could be used to assess relative changes in the length scales over time, and to diagnose the changes in atmospheric variables that shape any trends that are detected. Although anthropogenic warming has generally increased the intensity of extreme precipitation (e.g., O'Gorman 2015), Guinard et al. (2015) show that the spatial extents of precipitation decreased in 2002-11 compared to 1992-2001. Similarly, Wasko et al. (2016) show that the spatial extents of rain storms decrease with increased local temperatures (although intensities of the storms increase). However, these studies have been limited in their temporal and/or spatial extent. Our method introduces the potential to provide more robust assessment of trends in the length scales over large continental areas, using longer station records. For example, using our method, we could quantify the length scale of extreme daily precipitation at each location for each year, and then evaluate trends in the length over the 50-yr period, while accounting for spatial and temporal dependencies.

Moreover, the substantial heterogeneity in climatological length scales within some of the "climatically consistent" NCEI regions (e.g., the Northeast, South, and West North Central regions) points to potentially complex relationships between changes in the spatial extent of extreme precipitation and changes in atmospheric processes. For example, after calculating trends in length scale of extreme precipitation, we could subsequently assess the thermodynamic and dynamic contribution to those trends by quantifying the corresponding trends in local temperature and precipitable water (Trenberth et al. 2003; Trenberth 2011; O'Gorman 2015), along with changes in the magnitude and spatial patterns of geopotential heights and winds (Horton et al. 2015). In this way, we can begin to identify the causes of changes in the spatial extent of extreme precipitation in the recent past.

Our method and results can also be used to test relationships between length scales of extreme precipitation and associated hydrologic responses—which may enable a more comprehensive understanding of regional flood risk. While the p90 precipitation threshold encompasses events that are likely to have substantial hydrological impacts, a higher threshold (i.e., p99 or p99.9) may be better suited to examining events that are most likely to result in severe flood events. Future work will evaluate changes of precipitation length scales using a range of precipitation thresholds, with a focus on thresholds that are most likely to cause adverse impacts to human and environmental systems.

Acknowledgments. We thank three anonymous reviewers for insightful and constructive feedback. We thank NOAA for providing the GHCN station data, and the Stanford Research Computing Center and Stanford's Center for Computational Earth and Environmental Science for providing computational resources and support. We acknowledge funding support from the U.S. Department of Energy and Stanford University. D.L.S. was supported by the UCLA Sustainable LA Grand Challenge and the NatureNet Science Fellows Program through a collaboration between The Nature Conservancy and the University of California, Los Angeles.

REFERENCES

- Agel, L., M. Barlow, J.-H. Qian, F. Colby, E. Douglas, and T. Eichler, 2015: Climatology of daily precipitation and extreme precipitation events in the northeast United States. *J. Hydrometeor.*, 16, 2537–2557, https://doi.org/10.1175/ JHM-D-14-0147.1.
- Aghakouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, **116**, D02115, https://doi.org/10.1029/ 2010JD014741.

- Alkhaled, A. A., A. M. Michalak, S. R. Kawa, S. C. Olsen, and J. W. Wang, 2008: A global evaluation of the regional spatial variability of column integrated CO₂ distributions. *J. Geophys. Res.*, **113**, D20303, https://doi.org/10.1029/2007JD009693.
- Behnke, R., S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, 2016: Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecol. Appl.*, 26, 1338–1351, https://doi.org/10.1002/15-1061.
- Berezowski, T., M. Szczesniak, I. Kardel, R. Michalowski, T. Okruszko, A. Mezghani, and M. Piniewski, 2016: CPLFD-GDPT5: High-resolution gridded daily precipitation and temperature data set for two largest Polish river basins. *Earth Syst. Sci. Data*, 8, 127–139, https://doi.org/10.5194/ essd-8-127-2016.
- Berndt, C., E. Rabiei, and U. Haberlandt, 2014: Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. J. Hydrol., 508, 88–101, https://doi.org/10.1016/j.jhydrol.2013.10.028.
- Chang, W., M. L. Stein, J. Wang, V. R. Kotamarthi, and E. J. Moyer, 2016: Changes in spatiotemporal precipitation patterns in changing climate conditions. *J. Climate*, **29**, 8355–8376, https://doi.org/10.1175/JCLI-D-15-0844.1.
- Chiles, J.-P., and P. Delfiner, 2012: Geostatistics: *Modeling Spatial Uncertainty*. John Wiley & Sons, 726 pp.
- Cohen, J., and Coauthors, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, 7, 627–637, https://doi.org/10.1038/ngeo2234.
- DeGaetano, A. T., 2009: Time-dependent changes in extremeprecipitation return-period amounts in the continental united states. J. Appl. Meteor. Climatol., 48, 2086–2099, https:// doi.org/10.1175/2009JAMC2179.1.
- Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, 2, 775–779, https:// doi.org/10.1038/nclimate1562.
- —, A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, 27, 2271– 2296, https://doi.org/10.1175/JCLI-D-13-00451.1.
- Diffenbaugh, N. S., J. S. Pal, R. J. Trapp, and F. Giorgi, 2005: Finescale processes regulate the response of extreme events to global climate change. *Proc. Natl. Acad. Sci. USA*, **102**, 15 774– 15 778, https://doi.org/10.1073/pnas.0506042102.
- —, and Coauthors, 2017: Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci. USA*, **114**, 4881–4886, https://doi.org/10.1073/ pnas.1618082114.
- Dwyer, J. G., and P. A. O'Gorman, 2017: Changing duration and spatial extent of midlatitude precipitation extremes across different climates. *Geophys. Res. Lett.*, 44, 5863–5871, https:// doi.org/10.1002/2017GL072855.
- Endo, H., A. Kitoh, R. Mizuta, and M. Ishii, 2017: Future changes in precipitation extremes in East Asia and their uncertainty based on large ensemble simulations with a high-resolution AGCM. *Sci. Online Lett. Atmos.*, **13**, 7–12, https://doi.org/ 10.2151/sola.2017-002.
- Gleason, K. L., J. H. Lawrimore, D. H. Levinson, T. R. Karl, and D. J. Karoly, 2008: A revised U.S. Climate Extremes Index. J. Climate, 21, 2124–2137, https://doi.org/10.1175/ 2007JCLI1883.1.
- Gochis, D., and Coauthors, 2015: The great Colorado flood of September 2013. Bull. Amer. Meteor. Soc., 96, 1461–1487, https://doi.org/10.1175/BAMS-D-13-00241.1.

- Goovaerts, P., 1997: Geostatistics for Natural Resources Evaluation. Oxford University Press, 483 pp.
- —, M. T. D. Albuquerque, and I. M. H. R. Antunes, 2016: A multivariate geostatistical methodology to delineate areas of potential interest for future sedimentary gold exploration. *Math. Geosci.*, **48**, 921–939, https://doi.org/10.1007/ s11004-015-9632-8.
- Guinard, K., A. Mailhot, and D. Caya, 2015: Projected changes in characteristics of precipitation spatial structures over North America. *Int. J. Climatol.*, 35, 596–612, https://doi.org/10.1002/ joc.4006.
- Haas, T. C., 1990: Kriging and automated variogram modeling within a moving window. *Atmos. Environ.*, 24A, 1759–1769, https://doi.org/10.1016/0960-1686(90)90508-K.
- Haberlandt, U., 2007: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J. Hydrol., 332, 144–157, https://doi.org/10.1016/ j.jhydrol.2006.06.028.
- Hammerling, D. M., A. M. Michalak, and S. R. Kawa, 2012: Mapping of CO₂ at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2. *J. Geophys. Res.*, **117**, D06306, https://doi.org/10.1029/ 2011JD017015.
- Hawkins, E., R. S. Smith, J. M. Gregory, and D. A. Stainforth, 2016: Irreducible uncertainty in near-term climate projections. *Climate Dyn.*, 46, 3807–3819, https://doi.org/ 10.1007/s00382-015-2806-8.
- Higgins, R. W., Y. Yao, E. S. Yarosh, J. E. Janowiak, and K. C. Mo, 1997: Influence of the Great Plains low-level jet on summertime precipitation and moisture transport over the central United States. J. Climate, 10, 481–507, https://doi.org/10.1175/ 1520-0442(1997)010<0481:IOTGPL>2.0.CO;2.
- Horton, D. E., N. C. Johnson, D. Singh, D. L. Swain, B. Rajaratnam, and N. S. Diffenbaugh, 2015: Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature*, **522**, 465–469, https://doi.org/10.1038/ nature14550.
- IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. C. B. Field et al., Eds., Cambridge University Press, 582 pp., https://www.ipcc.ch/pdf/ special-reports/srex/SREX_Full_Report.pdf.
- Karl, T. R., and W. J. Koss, 1984: Regional and national monthly, seasonal, and annual temperature weighted by area, 1895–1983. National Climatic Data Center, 38 pp., https://repository.library. noaa.gov/view/noaa/10238.
- Kursinski, A. L., and S. L. Mullen, 2008: Spatiotemporal variability of hourly precipitation over the eastern contiguous United States from Stage IV multisensor analyses. J. Hydrometeor., 9, 3–21, https://doi.org/10.1175/2007JHM856.1.
- Laing, A. G., and J. M. Fritsch, 1997: The global population of mesoscale convective complexes. *Quart. J. Roy. Meteor. Soc.*, 123, 389–405, https://doi.org/10.1002/qj.49712353807.
- Li, J., Z. Zhu, and W. Dong, 2017: Assessing the uncertainty of CESM-LE in simulating the trends of mean and extreme temperature and precipitation over China. *Int. J. Climatol.*, 37, 2101–2110, https://doi.org/10.1002/joc.4837.
- Liu, X. C., T. C. Gao, and L. Liu, 2013: A comparison of rainfall measurements from multiple instruments. *Atmos. Meas. Tech.*, 6, 1585–1595, https://doi.org/10.5194/amt-6-1585-2013.
- Lloyd, C. D., 2005: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. J. Hydrol., 308, 128–150, https://doi.org/10.1016/ j.jhydrol.2004.10.026.

- , 2010: Nonstationary models for exploring and mapping monthly precipitation in the United Kingdom. *Int. J. Climatol.*, 30, 390–405, https://doi.org/10.1002/joc.1892.
- McDonald, J. H., 2009: *Handbook of Biological Statistics*. 2nd ed. Sparky House Publishing, 313 pp.
- Mehran, A., and A. Aghakouchak, 2014: Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations. *Hydrol. Processes*, 28, 2262–2270, https://doi.org/10.1002/hyp.9779.
- Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network–Daily database. J. Atmos. Oceanic Technol., 29, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1.
- Miller, P. W., and T. L. Mote, 2017: A climatology of weakly forced and pulse thunderstorms in the southeast United States. *J. Appl. Meteor. Climatol.*, 56, 3017–3033, https://doi.org/ 10.1175/JAMC-D-17-0005.1.
- Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2013: Corrigendum: Human contribution to more-intense precipitation extremes. *Nature*, **498**, 526, https://doi.org/10.1038/ nature12197.
- Murray, J. C., and B. A. Colle, 2011: The spatial and temporal variability of convective storms over the northeast United States during the warm season. *Mon. Wea. Rev.*, **139**, 992– 1012, https://doi.org/10.1175/2010MWR3316.1.
- Niziol, T. A., W. R. Snyder, and J. S. Waldstreicher, 1995: Winter weather forecasting throughout the eastern United States. Part IV: Lake effect snow. *Wea. Forecasting*, **10**, 61–77, https:// doi.org/10.1175/1520-0434(1995)010<0061:WWFTTE>2.0.CO;2.
- NOAA/NCEI, 2017: U.S. billion-dollar weather and climate disasters. National Centers for Environmental Information, accessed 25 April 2018, https://www.ncdc.noaa.gov/billions/.
- O'Gorman, P. A., 2015: Precipitation extremes under climate change. Curr. Climate Change Rep., 1, 49–59, https://doi.org/ 10.1007/s40641-015-0009-3.
- Oyler, J. W., A. Ballantyne, K. Jencso, M. Sweet, and S. W. Running, 2015: Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *Int. J. Climatol.*, **35**, 2258–2279, https://doi.org/10.1002/ joc.4127.
- Pryor, S. C., J. A. Howe, and K. E. Kunkel, 2009: How spatially coherent and statistically robust are temporal changes in extreme precipitation in the contiguous USA? *Int. J. Climatol.*, 29, 31–45, https://doi.org/10.1002/joc.1696.
- Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, 93, 811–829, https://doi.org/10.1175/ BAMS-D-11-00052.1.
- R Core Team, 2015: R: A language and environment for statistical computing. https://www.r-project.org/.
- Ribeiro, P. J., Jr., and P. J. Diggle, 2016: geoR: Analysis of geostatistical data. http://cran.r-project.org/package=geoR.
- Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Mon. Wea. Rev.*, 142, 905–921, https://doi.org/10.1175/MWR-D-13-00168.1.
- —, —, and —, 2015: The inland penetration of atmospheric rivers over western North America: A Lagrangian analysis. *Mon. Wea. Rev.*, **143**, 1924–1944, https://doi.org/10.1175/ MWR-D-14-00288.1.
- Schneider, D. P., C. Deser, J. Fasullo, and K. E. Trenberth, 2013: Climate data guide spurs discovery and understanding. *Eos*,

Trans. Amer. Geophys. Union, 94, 121–122, https://doi.org/ 10.1002/2013EO130001.

- Schneider, S. H., T. L. Root, and M. D. Mastrandrea, Eds., 2011: Encyclopedia of Climate and Weather. 2nd ed. Oxford University Press, 1488 pp., http://www.oxfordreference.com/view/ 10.1093/acref/9780199765324.001.0001/acref-9780199765324.
- Sieck, L. C., S. J. Burges, and M. Steiner, 2007: Challenges in obtaining reliable measurements of point rainfall. *Water Resour. Res.*, 43, W01420, https://doi.org/10.1029/2005WR004519.
- Singh, D., M. Tsiang, B. Rajaratnam, and N. S. Diffenbaugh, 2013: Precipitation extremes over the continental United States in a transient, high-resolution, ensemble climate model experiment. J. Geophys. Res. Atmos., 118, 7063–7086, https://doi.org/ 10.1002/jgrd.50543.
- Sun, X., and A. P. Barros, 2010: An evaluation of the statistics of rainfall extremes in rain gauge observations, and satellite-based and reanalysis products using universal multifractals. *J. Hydrometeor.*, **11**, 388–404, https://doi.org/10.1175/2009JHM1142.1.
- Tadić, J. M., X. Qiu, V. Yadav, and A. M. Michalak, 2015: Mapping of satellite Earth observations using moving window block kriging. *Geosci. Model Dev.*, 8, 3311–3319, https://doi.org/ 10.5194/gmd-8-3311-2015.
- Thompson, D. W. J., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. J. Climate, 28, 6443–6456, https://doi.org/10.1175/JCLI-D-14-00830.1.
- Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol., 190, 214–251, https://doi.org/ 10.1016/S0022-1694(96)03128-9.
- Tokay, A., P. G. Bashor, and V. L. McDowell, 2010: Comparison of rain gauge measurements in the mid-Atlantic region.

J. Hydrometeor., **11**, 553–565, https://doi.org/10.1175/2009JHM1137.1.

- Trapp, R. J., N. S. Diffenbaugh, H. E. Brooks, M. E. Baldwin, E. D. Robinson, and J. S. Pal, 2007: Changes in severe thunderstorm environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing. *Proc. Natl. Acad. Sci. USA*, **104**, 19719–19723, https://doi.org/ 10.1073/pnas.0705494104.
- Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Res.*, 47, 123–138, https://doi.org/10.3354/ cr00953.
- —, A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, 84, 1205–1217, https://doi.org/10.1175/BAMS-84-9-1205.
- van der Wiel, K., and Coauthors, 2017: Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrol. Earth Syst. Sci.*, 21, 897– 921, https://doi.org/10.5194/hess-21-897-2017.
- Wasko, C., A. Sharma, and S. Westra, 2016: Reduced spatial extent of extreme storms at higher temperatures. *Geophys. Res. Lett.*, 43, 4026–4032, https://doi.org/10.1002/2016GL068509.
- Western, A. W., G. Blöschl, and R. B. Grayson, 1998: How well do indicator variograms capture the spatial connectivity of soil moisture? *Hydrol. Processes*, **12**, 1851–1868, https:// doi.org/10.1002/(SICI)1099-1085(19981015)12:12<1851::AID-HYP670>3.0.CO;2-P.
- Wright, S. P., 1992: Adjusted P-values for simultaneous inference. Biometrics, 48, 1005–1013, https://doi.org/10.2307/2532694.
- Zhou, Y., and C. J. Matyas, 2017: Spatial characteristics of stormtotal rainfall swaths associated with tropical cyclones over the eastern United States. *Int. J. Climatol.*, **37**, 557–569, https:// doi.org/10.1002/joc.5021.