Subseasonal-to-Seasonal Extreme Precipitation Events in the Contiguous United States: Generation of a Database and Climatology

TY A. DICKINSON,[a] MICHAEL B. RICHMAN,[a] AND JASON C. FURTADO[a]

[a] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 20 July 2020, in final form 25 June 2021)

ABSTRACT: Extreme precipitation across multiple time scales is a natural hazard that creates a significant risk to life, with a commensurately large cost through property loss. We devise a method to create 14-day extreme-event windows that characterize precipitation events in the contiguous United States (CONUS) for the years 1915–2018. Our algorithm imposes thresholds for both total precipitation and the duration of the precipitation to identify events with sufficient length to accentuate the synoptic and longer time scale contribution to the precipitation event. Kernel density estimation is employed to create extreme-event polygons that are formed into a database spanning from 1915 through 2018. Using the developed database, we clustered events into regions using a *k*-means algorithm. We define the ''hybrid index,'' a weighted composite of silhouette score and number of clustered events, to show that the optimal number of clusters is 15. We also show that 14-day extreme precipitation events are increasing in the CONUS, specifically in the Dakotas and much of New England. The algorithm presented in this work is designed to be sufficiently flexible to be extended to any desired number of days on the subseasonal-to-seasonal (S2S) time scale (e.g., 30 days). Additional databases generated using this framework are available for download from our GitHub. Consequently, these S2S databases can be analyzed in future works to determine the climatology of S2S extreme precipitation events and be used for predictability studies for identified events.

KEYWORDS: Extreme events; Precipitation; Databases; Statistics; Intraseasonal variability

1. Introduction

Extreme precipitation is a natural hazard that poses significant risk to life and property globally. Between 1980 and 2019, there were 49 flooding and winter storm events affecting much of the contiguous United States (CONUS) with losses exceeding $1 billion in inflation-adjusted damages (NCEI 2020). These events totaled nearly $200 billion in consumer price index–adjusted damages and caused 1603 fatalities throughout the CONUS. Impacts from extreme precipitation affect a myriad of sectors, such as water resources (e.g., Piao et al. 2010), agriculture (e.g., Klemm and McPherson 2017), energy (e.g., Schaeffer et al. 2012), and ecosystems (e.g., Knapp et al. 2008). To date, there has been an abundance of studies examining the trends, dynamics, and thermodynamics of extreme precipitation on 24- (e.g., Bradley and Smith 1994; Schumacher and Johnson 2005; Grotjahn and Faure 2008; Kunkel et al. 2012; Moore et al. 2015) and 48-h (e.g., Keim 1996; Moore et al. 2012; Warner et al. 2012) time scales. However, there remains a gap in our knowledge about the identification and dynamical characterization of extreme precipitation events on the subseasonal-to-seasonal (S2S) time scale, defined as the period ranging from 14 days to about 2 or 3 months (Vitart et al. 2017; NOAA 2018). Considerable research has been published on short-duration extreme precipitation events, as noted in a recent review by Barlow et al. (2019). Yet, 14 of the 49 aforementioned billion-dollar disasters occurred over periods of 14 days or longer, demonstrating the ability of longer-duration events to cause significant societal impacts.

Accurately predicting extremes, and communicating the appropriate risks, at any time scale is an immensely difficult problem. Forecasting precipitation extremes are particularly challenging owing to complex interactions between planetary, synoptic, and mesoscale drivers and surface hydrologic responses that determine the spatial location, scale, magnitude, and duration of the event. Forecasts of extreme precipitation have improved over time (Sukovich et al. 2014) through increased understanding of the dynamics of extreme precipitation and improvements in numerical modeling (e.g., Iyer et al. 2016), among other factors. However, improvements in short-term precipitation forecasts have not translated to the S2S time scale. Indeed, the Climate Prediction Center's week 3–4 precipitation outlook remains experimental owing to skill remaining relatively limited (CPC 2020). Pan et al. (2019) additionally showed little skill beyond 2 weeks on the West Coast in S2S model hindcasts. There have been recent improvements in precipitation forecasts beyond week 2 by leveraging large-scale climate modes such as the Madden–Julian oscillation, the quasi-biennial oscillation, and El Niño–Southern Oscillation (DelSole et al. 2017; Vigaud et al. 2017; Nardi et al. 2020). However, forecast products remain limited to tercile-like approaches (i.e., below normal, normal, above normal), providing decision-makers little information on the potential for high-impact events on such scales. Accordingly, current capabilities in predicting extreme precipitation events at the S2S time scale are low, leading to inadequate conveyance of risk to stakeholders. The development of a method to catalog S2S extreme precipitation events is the first step in extending forecast skill and ultimately improving societal responses to

---

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-20-0580.s1.

Corresponding author: Ty A. Dickinson, ty.dickinson@ou.edu

DOI: 10.1175/JCLI-D-20-0580.1

extreme precipitation events. These S2S extreme precipitation databases can then be used to identify forecasts of opportunity, calibrate and constrain S2S dynamical prediction systems, and to develop statistical models leading to more skillful forecasts of such events on the S2S time scale.

What is considered ''extreme'' is often subject to the spatial and temporal scale being analyzed; Barlow et al. (2019) document 25 different definitions of extreme precipitation in various prior studies. Here, we introduce a method for identifying extreme events that can be consistently applied across time and space. One definition of an extreme precipitation event on the S2S time scale came from Jennrich et al. (2020) where their extreme threshold was the 95th percentile of all 14-day windows between 1981 and 2010. After dividing the CONUS into six regions, using geopolitical boundaries and latitude/longitude lines, they defined an event by considering the areal extent within the region of grid boxes that exceed the 95th percentile, ensuring the areal-averaged precipitation was greater than 10 mm day^{-1} for five or more days, and requiring the day of heaviest precipitation and surrounding days to not be greater than 50% of the total precipitation. However, the division of the CONUS into regions that are not defined by the statistical properties of precipitation causes events that span across regional borders to potentially be missed, distorting the shapes of extreme events. Additionally, the continuity of grid boxes that exceed the 95th percentile was not considered, potentially merging two groups meeting their extreme criteria into one extreme event despite several hundreds of kilometers or more of geographical separation. The dynamics driving the two groups to be extreme would likely be distinct, thereby confounding the linkages in subsequent composite analysis.

The present study builds on the work by Jennrich et al. (2020) by relaxing the assumption of political boundary regions within the CONUS and extending the temporal study period. To illustrate the details and implementation of the framework, we develop a database of 14-day extreme precipitation events and present a brief climatology of that database. Databases of other lengths (e.g., 30-day events) are also available on our GitHub. Section 2 presents the algorithm for defining an extreme event with an example. Section 3 focuses on processing and aggregating all potential events into a single database. The temporal and spatial characteristics of the database are discussed in section 4. A summary and conclusions follow.

2. Event identification

a. Data

We used daily precipitation data from the Livneh (Livneh et al. 2013) and Parameter–Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994) datasets throughout the analysis. Because the Livneh dataset ends in 2011, it was augmented with PRISM through 2018 to complete the record to near present. The PRISM precipitation data were bilinearly interpolated using the Python library xESMF (Zhuang et al. 2020) from a native 4-km grid onto Livneh's grid of about 6 km. Cubic spline interpolation was tested and found to offer essentially the same precipitation field as bilinear interpolation when visually compared (not shown), and hence the simpler method was adopted. To examine the interpolation error, we took PRISM's daily precipitation between 1981 and 2010, interpolated to Livneh's grid, interpolated back to PRISM's grid, and divided the difference between PRISM's original value and its twice-interpolated value by 2. The mean absolute error over the CONUS was $O(10^{-2})$ mm or $O(10^{-3})$ mm depending on the month. However, errors approached $O(10^{1})$ mm in the Sierra Nevada and Cascade Range mountains during boreal winter. Although both the interpolation error and difference between the Livneh and PRISM datasets may become relatively large, particularly in the western one-third of the CONUS, the Livneh dataset, at least partially, is a function of the 1981–2010 PRISM climatology. This dependence makes PRISM the best choice of precipitation datasets to append to the Livneh dataset. The mean error between the Livneh and PRISM datasets is $O(10^{-3})$ mm (not shown), with Livneh generally being larger than PRISM, albeit by small amounts. Accordingly, our database extends through 31 December 2018, giving a total of 104 years for the analysis.

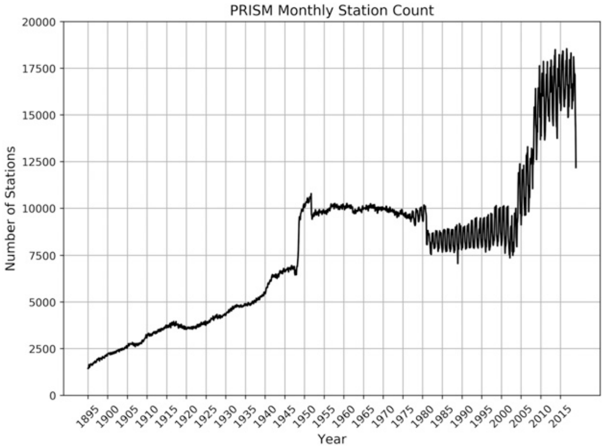

FIG. 1. Total number of stations used in processing monthly precipitation data from PRISM as a function of time.

Rainfall in the Livneh and PRISM datasets, and thus the calculated extreme-value statistics, are affected by the number of rain gauges used by the models. Figure 1 depicts a time series of the number of stations employed in calculating monthly precipitation in PRISM; the Livneh dataset does not come with a metafield showing the number of gauges used in processing a given day's precipitation field. The primary precipitation data source for both datasets is the National Weather Service Cooperative Observer Program (Daly et al. 2008; Livneh et al. 2013; Henn et al. 2018); thus it is very likely that magnitude and trend in Fig. 1 are similar for the Livneh dataset. Over time, the number of rain gauges in the network increased from under 2500 stations to about 18 000 stations. Girons Lopez et al. (2015) found that having fewer rain gauges led to decreased accuracy in estimating higher rainfall rates. Therefore, in the early twentieth century, the decreased spatial density of the rain gauges likely led to many areas having underestimated precipitation amounts relative to later in the twentieth century and beyond. Thus, a positive trend in the number of identified events, when using a stationary threshold,

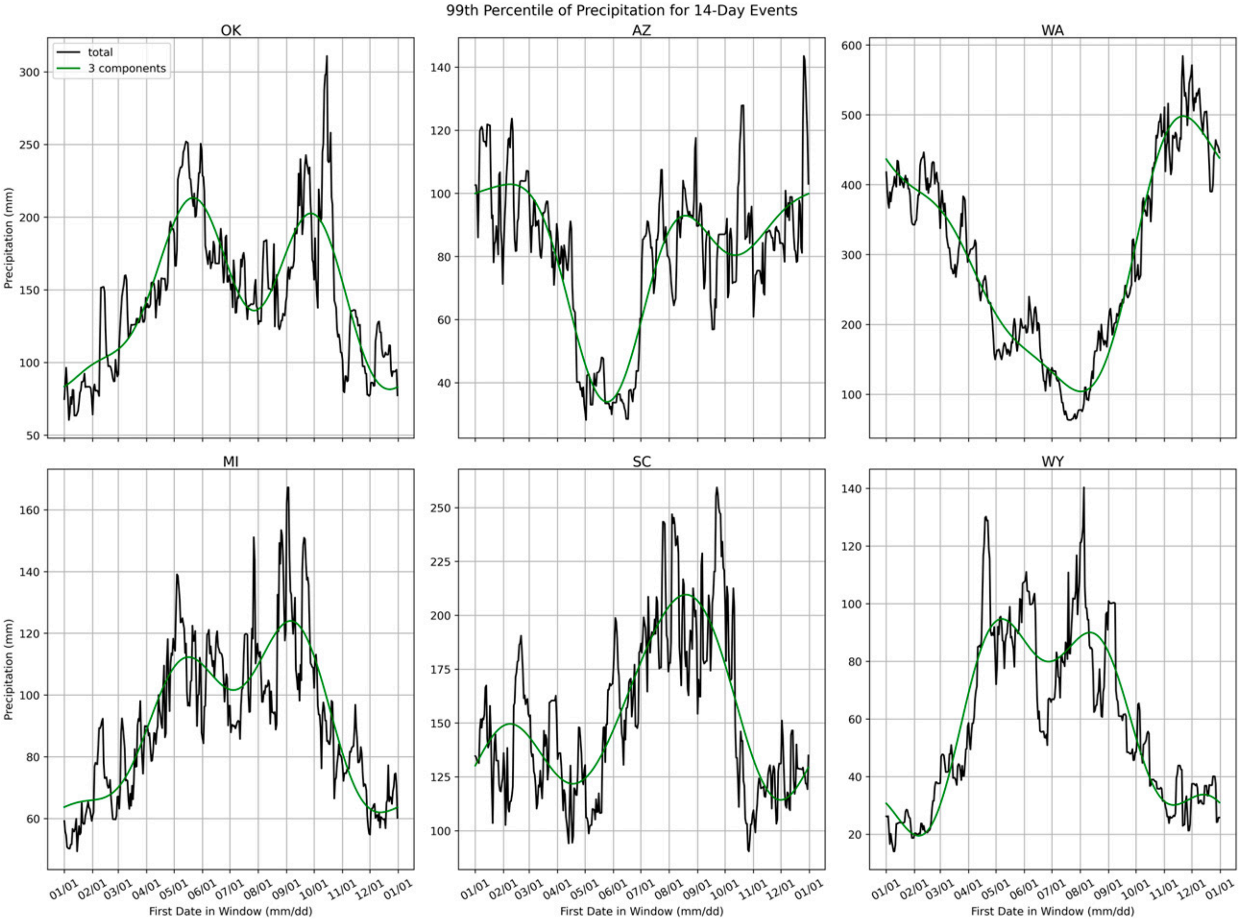

FIG. 2. Raw 99th percentile for each 14-day window and smoothed percentiles using the first three Fourier harmonics (green) for six selected grid points across the CONUS. The plotted date is the start date of each window.

in the database may be due, at least to some degree, to station density.

b. Extreme criteria

We began by finding all grid points within the CONUS that experienced extreme precipitation conditions over a given 14-day window. A point was flagged as having experienced extreme precipitation if it passed two criteria. First, total precipitation must meet or exceed the 99th percentile. We selected the 99th percentile to identify extreme events in the far-right tail of the precipitation distribution as these events are very likely to be coincident with major socioeconomic impacts. The 99th percentile was calculated using each 14-day sum over all 104 years for each grid point and each overlapping 14-day window (i.e., 1–14 January, 2–15 January, etc.). The black line in Fig. 2 depicts the derived percentiles as a function of calendar day for six points across the CONUS: Norman, Oklahoma; Sedona, Arizona; 1° longitude east of Seattle, Washington; 1° longitude west of Port Huron, Michigan; Colombia, South Carolina; and Cheyenne, Wyoming. The large degree of window-to-window noise in precipitation is apparent in the corresponding percentiles, with oscillations of ≥50 mm seen in the span of 2 weeks at every grid point. Therefore, we smoothed the raw percentiles by decomposing each time series into its Fourier coefficients. We tested reconstructed time series by retaining harmonics ranging from two to six and subjectively determined that three harmonics led to the best smoothed signal (see the green line in Fig. 2; see Figs. S1 and S2 in the online supplemental material for results of sensitivity testing). Furthermore, numerous studies have examined precipitation seasonality and shown that the first three harmonics generally explain the majority of the variance in the precipitation distribution (e.g., Scott and Shulman 1979; Kirkyla and Hameed 1989; Epstein and Barnston 1990). Figure 3 displays the total thresholds for one day in each season. Each window and each grid point in the CONUS will have its own threshold criterion. To illustrate, consider a precipitation field over the example window of 12–25 January 1937 (Fig. 4a). Points that pass the first criterion are where values in Fig. 4a exceed the extreme threshold values in Fig. 4b (e.g., parts of Arkansas, Tennessee, Kentucky, and Indiana).

The next criterion considers the duration of the precipitation during the event. Although a duration check may eliminate

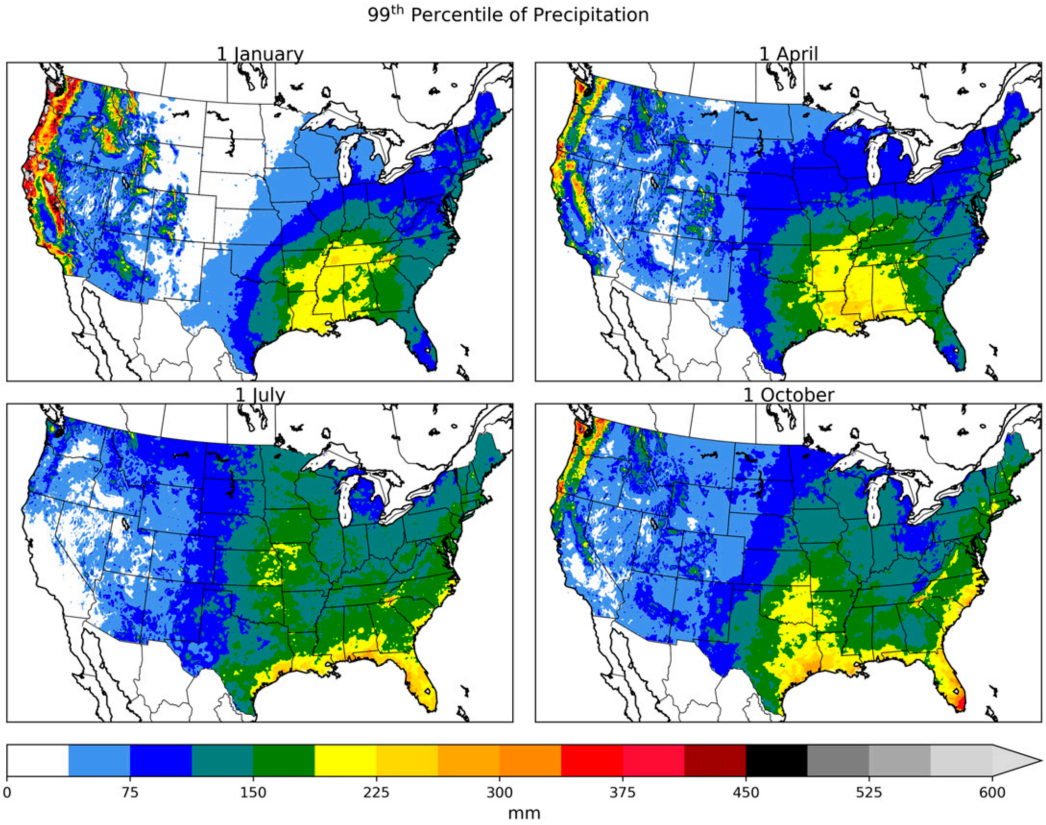

FIG. 3. The 99th percentile of precipitation in the CONUS for one 14-day window in each meteorological season.

short duration events with exceptionally heavy precipitation totals, we aim to identify events with persistent synoptic-scale forcing to maximize potential predictability at meaningful (i.e., S2S) lead times to decision-makers. Given the importance of defining persistently forced extreme precipitation events, for a grid point to be considered extreme, at least 7 of the days (i.e., half of the duration of the event) must receive greater than or equal to the long-term mean daily precipitation. We defined the climatological daily precipitation as the long-term mean of all days in a given 14-day window over all 104 years. In other words, we calculated the mean daily precipitation using all 1456 days in our period of record. The relatively large sample size used in calculating the climatological mean led to more stable intraseasonal variability. The mean and median differences between one calendar day's long-term mean and the next day's mean are $O(10^{-5})$ mm, and the maximum value is a mere 2.3 mm. Accordingly, we did not smooth further using Fourier harmonics. Grid points that meet or exceed the duration criterion are shown as medium and dark blue in Fig. 4c and are also outlined by the solid black contour.

c. Finding an extreme region

The two aforementioned criteria were applied to each grid point in the domain independently, and the points that met both criteria were flagged as extreme (e.g., Fig. 4d). See Fig. 5 for a summary of the logic. Once grid points were identified as extreme, the next step was to consider the spatial continuity of these points in defining an extreme region. To objectively identify extreme regions, kernel density estimation (KDE; Rosenblatt 1956), a nonparametric technique that fits a probability density function (PDF) to a given set of observations, was employed. Here, the geospatial coordinates of all points flagged as extreme were used to define a 3D PDF over the CONUS where the vertical coordinate (i.e., density) is greater where more extreme points are in close proximity. The density of a given point y with observations x_i, $i = 1, 2, \ldots, n$ is given by

$$\rho(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y - x_i}{h}\right), \tag{1}$$

where K is the kernel, a prescribed weighting function, and h is the bandwidth (i.e., a smoother for the resulting PDF). After considering the advantages and disadvantages of a number of kernels, we concluded the Epanechnikov kernel was best suited for both this study and future studies with globally gridded precipitation data. The Epanechnikov kernel for a random variable u is given by

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & |u| \le 1 \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

The Epanechnikov kernel was selected for its parabolic shape and the assignment of zero density outside of $|u| \le 1$ (i.e., its

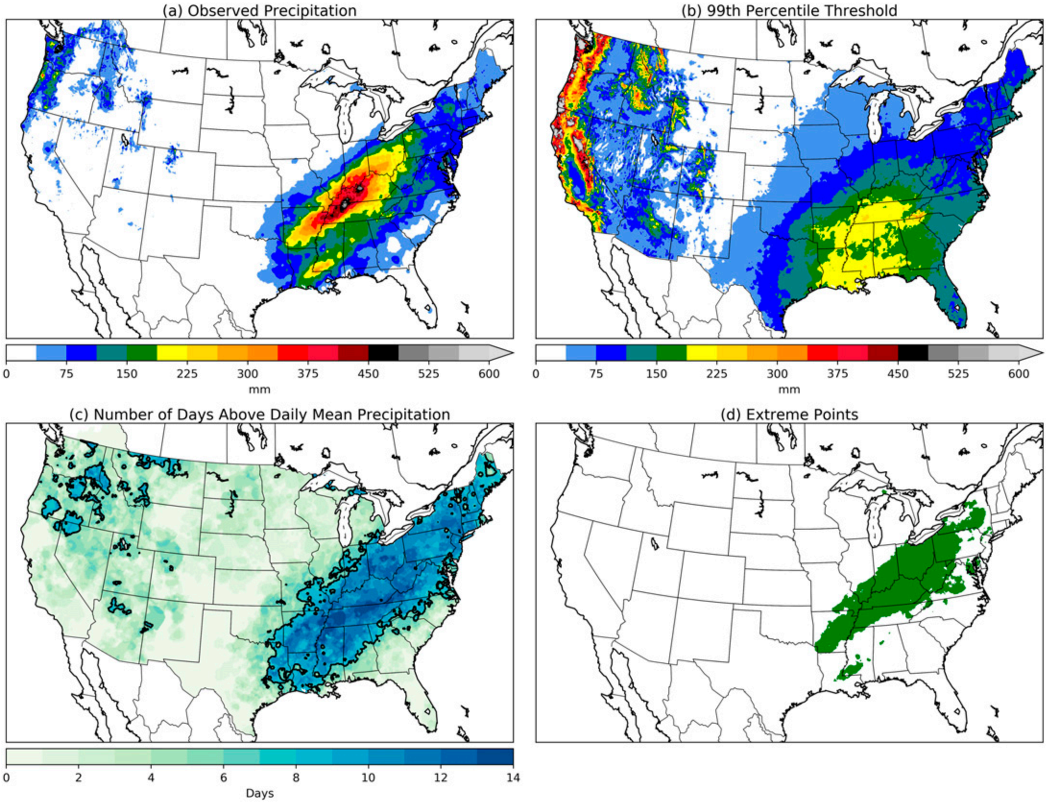

FIG. 4. (a) Total recorded precipitation (mm) between 12 and 25 Jan 1937. (b) The 99th percentile of precipitation (mm) derived for 12–25 Jan. (c) Number of days with above-normal daily precipitation during 12–25 Jan 1937. The black contour outlines 7 days, which is the threshold for our second criterion in event selection (see the text). (d) Points flagged as extreme per our algorithm (see the text for details).

support). Other kernel options, such as Gaussian and exponential, do not impose a support, which leads to a smoother density field. One benefit of Epanechnikov's more rigid kernel is its ability to effectively outline areas outside the CONUS where Livneh and PRISM lack data. Hence, polygons tend to follow the boundary and have little extent into the oceans. A Gaussian kernel, for instance, will still lack significant extent over oceans, but its smoother density field would produce a lesser gradient; however, we have no data over the oceans to support such a gradient. Cross-validation tests were conducted to determine the optimal bandwidth on a number of test events, with varying times and locations. These tests consistently gave bandwidths from 0.01 to 0.03. Therefore, we used a bandwidth of 0.02.

The resulting density was fit on a grid with 0.1° resolution in the domain 24°–50°N; 128°–66°W, where the great circle distance calculated using the Haversine formula is used as the distance metric. In the midlatitudes, a bandwidth of 0.02 equates to a point on the grid having nonzero density if it is within roughly 1.25° of an extreme point. Next, the density was normalized by dividing by the maximum value to facilitate direct comparison of potential events from different regions of the CONUS and during different times of the year (Fig. 6a). Subsequently, an extreme precipitation event region was developed using a discrete normalized density value. We found that a density value of 0.2710, the 99th percentile of all density fields between 1915 and 2018, produced representative and consistent polygons across both time and space. Figure 6b displays the polygon drawn using the single contour, outlining the area defined as extreme for the 14-day window. A 14-day extreme precipitation event was then defined to have occurred if the area inside the polygon, calculated using Green's theorem, exceeds 200 000 km^2. We calculated the areas of every KDE-derived polygon between 1915 and 2018 and examined the distribution of areal extents. An areal extent of 200 000 km^2 records the largest 5% of all areas captured and represents a "critical" point where the derivative of areas as a function of percentiles changes rapidly (see Fig. S3 in the online supplemental material).

3. Cataloging extreme events

Section 2 outlined how to identify an extreme precipitation event given a single window. The database was generated by considering all windows in the period of record and aggregating all identified events. We used a Python program (https://github.com/tydickinson29/PRES2iPpy/blob/master/pres2ippy/analysis/rainpy.py) that constructs a unique object based on

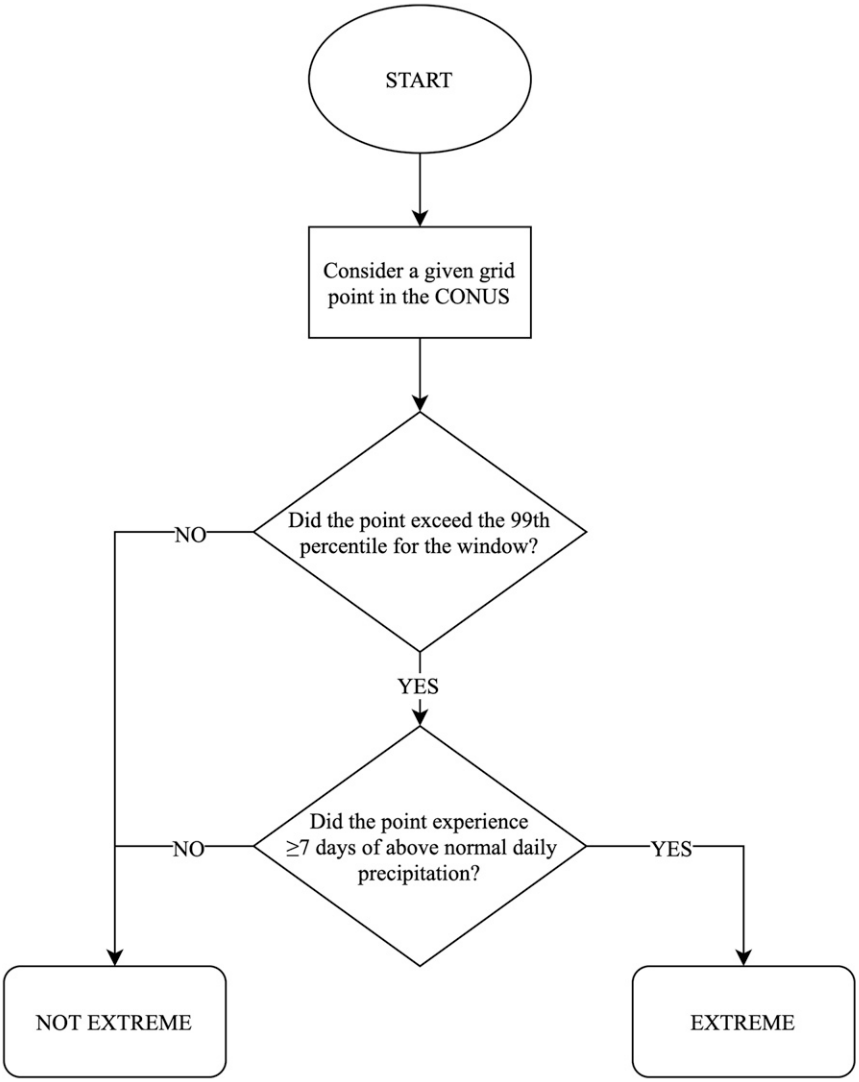

FIG. 5. Flowchart describing the logic in determining if any given point is labeled extreme. See the text for details.

input day, month, and year. The program is available for users to explore any date; for example, other researchers may want to define their own regional database of events using the same method but with a smaller areal extent. The database was generated by inputting every possible start date from 1 January 1915 to 31 December 2018 (excluding leap days) and archiving identified extreme events to a text file. However, a consequence of simply moving a window in increments of a single day is the database contained extreme events with polygons over virtually the same area for several consecutive days. Thus, we developed an objective postprocessing algorithm to group ''repeat'' events, label one of the events in the group as the extreme event, and discard all others.

The postprocessing algorithm that discards all repeats began by considering the event in the first row of the preprocessed text file (denoted as E_0; see Fig. 7). A smaller, subset file was constructed in which only the events with starting dates on or before the ending date of E_0 are considered. For example, if E_0 is 1–14 January, then we found all events that have a start date on or before 14 January. Then, events were grouped by the location and shape of their polygons. A binary matrix with the same domain and resolution as the KDE density (see section 2c) was constructed for all events in the subset file, where ''1'' denotes a grid point inside the event polygon and ''0'' denotes a grid point outside the polygon. The correlation coefficient r was calculated between all events in the subset file with respect to E_0; extreme precipitation events are considered to be representative of the same event if $r \geq 0.5$. The events are then placed into a group of similar events.

Within this subsetted group, we chose the ''most extreme'' precipitation event to label as the extreme event for the database. To make this choice, we define the total over extreme (TOE) as

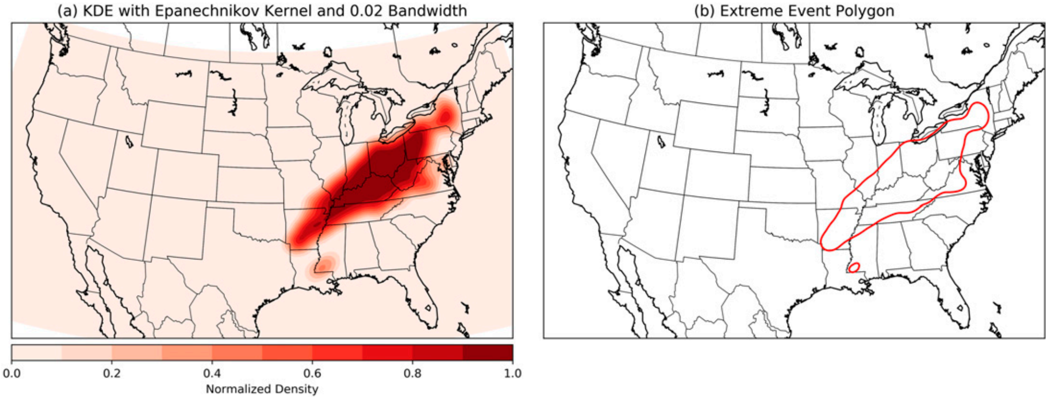

FIG. 6. (a) Full KDE normalized density field using the Epanechnikov kernel and 0.02 bandwidth for the 12–25 Jan 1937 window. (b) Extreme-event polygon drawn using the 0.2710 contour.

$$\mathrm{TOE} = \sum_{i=1}^{n} (P^{i}_{\mathrm{total}} - P^{i}_{q99}), \tag{3}$$

where i denotes a particular grid point that was flagged as extreme, n is the total number of flagged grid points within the event's polygon, P_{total} is the grid point's total 14-day precipitation, and P_{q99} is the grid point's 14-day 99th-percentile threshold. TOE considers the extremity of the event and considers the areal extent of the event indirectly since larger polygons tend to have more points inside flagged as extreme. The event with the largest TOE within the group was determined to be "extreme" and is retained for the database, whereas all other events in the group are discarded. This process of forming groups of similar events, based on the $r < 0.5$ criterion, was recursive until all events in the file either do not overlap temporally or do not have a correlation coefficient above 0.5. After postprocessing, the 14-day database contains 851 unique events.

Table 1 lists a few events identified in the database with various statistics describing each event. A full table can be accessed online (http://pres2ip.com/extreme-event-tables) where users can apply filters to find specific events based on time and/or location. The database in .csv and .shp formats and the scripts used to generate and postprocess the database are also available online (https://github.com/tydickinson29/PRES2iPpy/tree/master/pres2ippy/databases).

4. Statistical characteristics of 14-day extreme events

a. Event regionalization

Another research goal is to objectively define regions of extreme 14-day precipitation. Classifying the events in the database into geographically similar clusters allows for the identification of typical extreme-event shapes across the CONUS, as well as labeling specific extreme events as being shaped exceptionally anomalous. In addition, characteristic patterns associated with the events can be diagnosed for each region, which is the focus of a future paper.

Two commonly applied types of clustering algorithms that are implemented on gridded data are hierarchical and nonhierarchical. Hierarchical clustering tends to use an agglomerative approach where each observation begins as its own cluster and observations recursively are merged together, with the end result being visualized as a tree. Nonhierarchical methods utilize other recursive algorithms that do not form a hierarchy. In a rigorous examination of clustering techniques applied on precipitation data in the eastern two-thirds of the United States, Gong and Richman (1995) found nonhierarchical methods to outperform hierarchical methods. Therefore, we apply the k-means clustering (Forgy 1965; Lloyd 1982) implementation in the Python library scikit-learn (Pedregosa et al. 2011). The k-means algorithm clusters data into k groups of equal variance, attempting to minimize the intracluster sum of squares (i.e., inertia) using Euclidean distance. After events are clustered, the silhouette score (Rousseeuw 1987) is calculated as a metric for how well an event fits its assigned cluster. The silhouette score for a given sample is given by

$$s = \frac{b - a}{\max(a, b)}, \tag{4}$$

where a is the mean distance between a sample and all other samples in the cluster, b is the mean distance between a sample and all other samples in the next nearest cluster, and $\max(a, b)$ is a function that returns the higher value of a and b. From Eq. (4), the silhouette score can range from -1 to $+1$, where higher positive numbers indicate a good match to the assigned cluster. Events with negative silhouette scores are removed because they indicate that a sample was likely misclassified and/or is highly anomalous relative to all other events in the cluster (Rousseeuw 1987). The silhouette score has been applied to meteorological applications to remove misclassified events (e.g., Richman and Adrianto 2010; Mercer et al. 2012).

For each event, a vector consisting of 0s and 1s is constructed, where a grid point inside the polygon is assigned a 1, with the same resolution as the KDE density field. K-means clustering is then applied iteratively in the following manner:

1) Cluster n events into k clusters.
2) Calculate the silhouette score s for all n events.

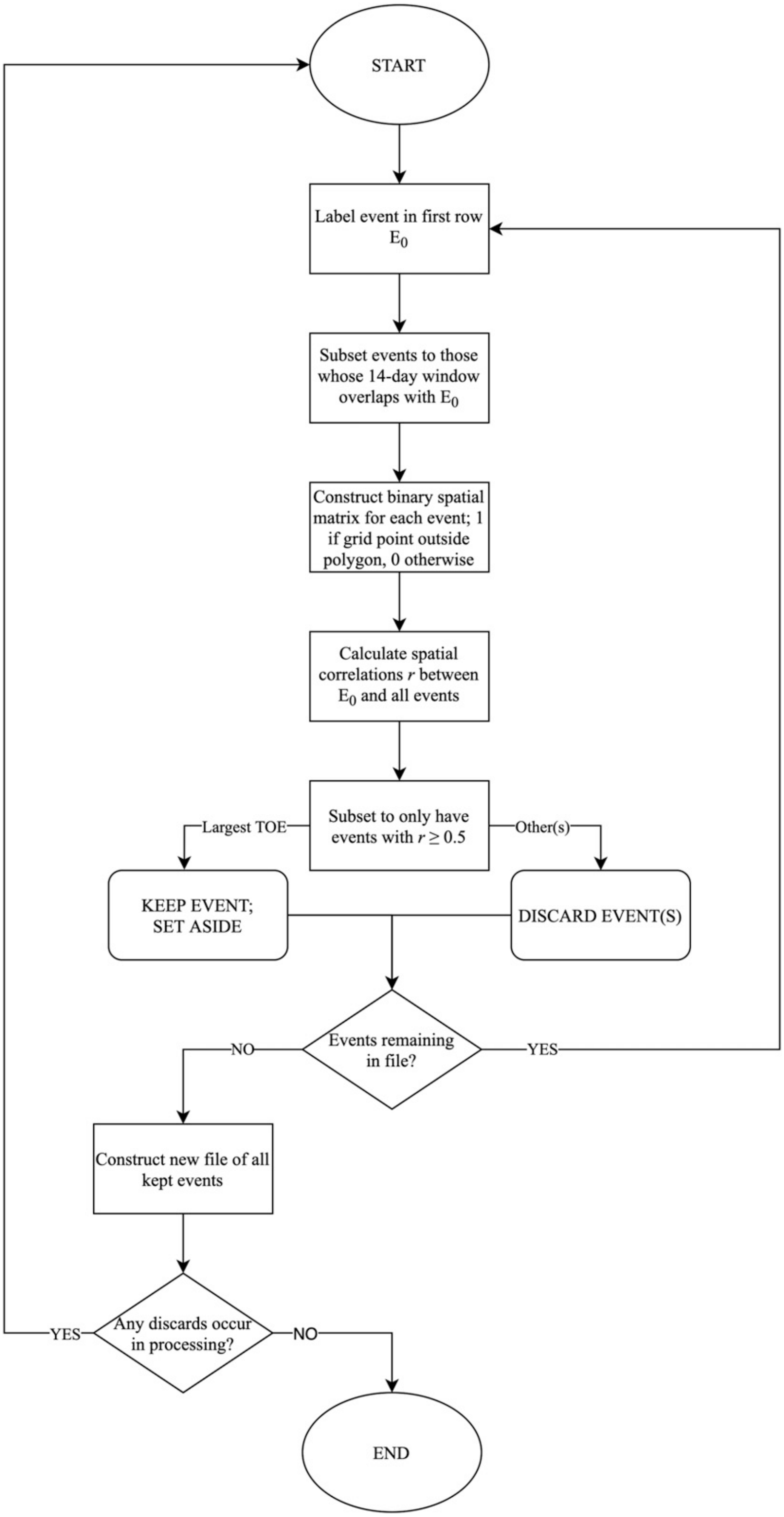

FIG. 7. Flowchart describing the logic during postprocessing of extreme precipitation events. See the text for details.

TABLE 1. A sample of events from version 1.0 of the 14-day extreme precipitations events database. $\text{Precip}_{\text{AA}}$ denotes areal-averaged precipitation in the region; $\text{Max}_{\text{total}}$ is the greatest 14-day precipitation total from any point inside the region; Max_{1day} is the greatest single-day precipitation total from any point inside the region from any of the 14 days. Area and TOE are rounded to the nearest square kilometer and millimeter, respectively, and all other entries are rounded to two decimal places.

Begin date	End date	Area (km^2)	TOE (mm)	$\text{Precip}_{\text{AA}}$ (mm)	$\text{Max}_{\text{total}}$ (mm)	Max_{1day} (mm)	Lon_{C} (°E)	Lat_{C} (°N)	Cluster
12 Jan 1937	25 Jan 1937	881 375	1 558 392	188.45	541.00	138.07	275.90	38.51	12
19 Nov 1998	2 Dec 1998	244 172	298 550	312.32	1366.15	286.08	237.52	44.28	3
7 May 2015	20 May 2015	234 242	277 190	233.28	458.09	228.37	263.36	34.58	13

3) Find the number of events with negative silhouette scores (n^*).
4) If $n^* \neq 0$, remove the events with $s < 0$ and repeat from step 1; otherwise, end.

Choosing the optimal value for k is the most difficult and subjective decision when implementing any clustering algorithm. The elbow method (Thorndike 1953) and the silhouette method (Rousseeuw 1987) are two popular tools that are used as guidance to determine the best value for k. In both methods, a range of values for k is tested; in this study, we test $k = 5, 6, \ldots, 30$. However, neither method depicts a clear signal as a tool in deciding which k to choose (Figs. 8a,b). Although the inertia continues to decrease with increasing k, there is no cusp present. Similarly, the average silhouette scores generally increase with increasing k, though a major reason for the positive correlation is from the lower number of events being clustered. In comparing Figs. 8b and 8c, it is seen that there is an inverse relationship between the average silhouette and the number of clustered events. Consequently, we define an index, which we name the hybrid index (HI), that is composed of the product of the average silhouette and number of clustered events. The best value for k is determined to be the one for which the maximum HI value occurs; for our 14-day database, the optimal number of clusters is $k = 15$ (Fig. 8d).

Figure 9 depicts the ''average'' polygons for each cluster. The average polygon is created using all events in each cluster; the average number of events per cluster is about 50. The polygons in the cluster are rasterized to a grid with 0.01° spacing where the grid bounds are the minimum and maximum longitudes and latitudes from all polygons in the cluster. The

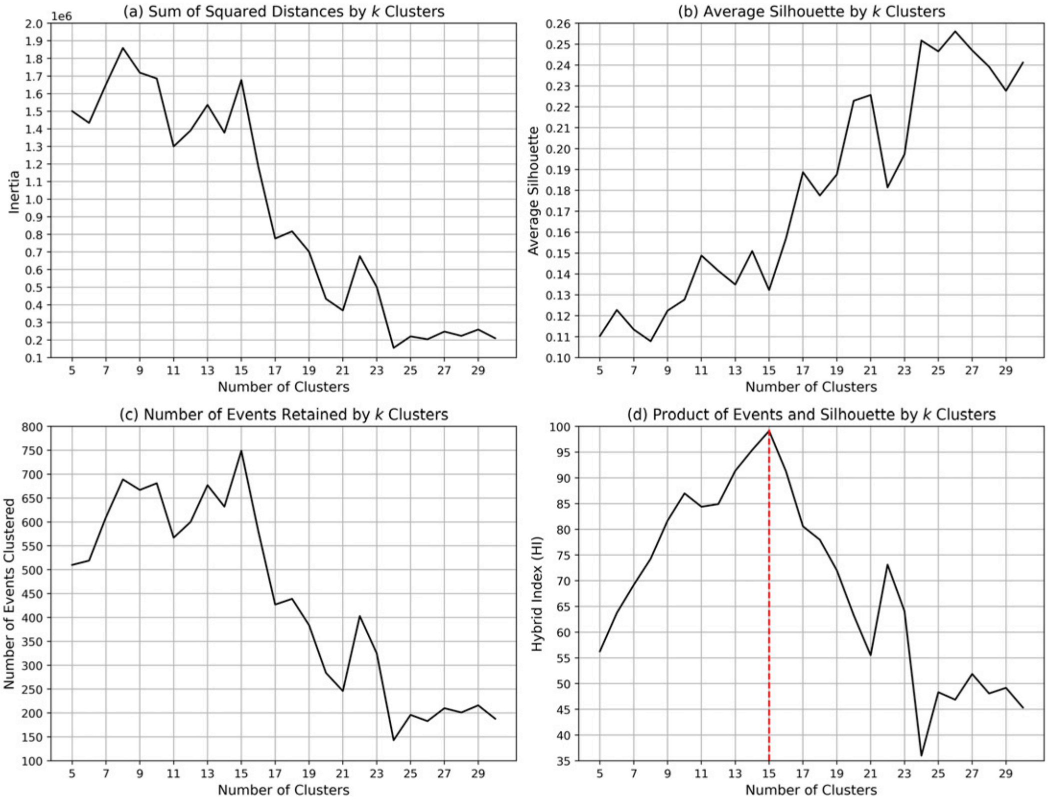

FIG. 8. Various statistics as a function of number of clusters for 14-day extreme precipitation events: (a) intracluster sum of squared distances (i.e., inertia), (b) average silhouette score, (c) total number of clustered events, and (d) product of (b) and (c) (i.e., the HI). The dashed vertical red line in (d) marks the HI maximum.

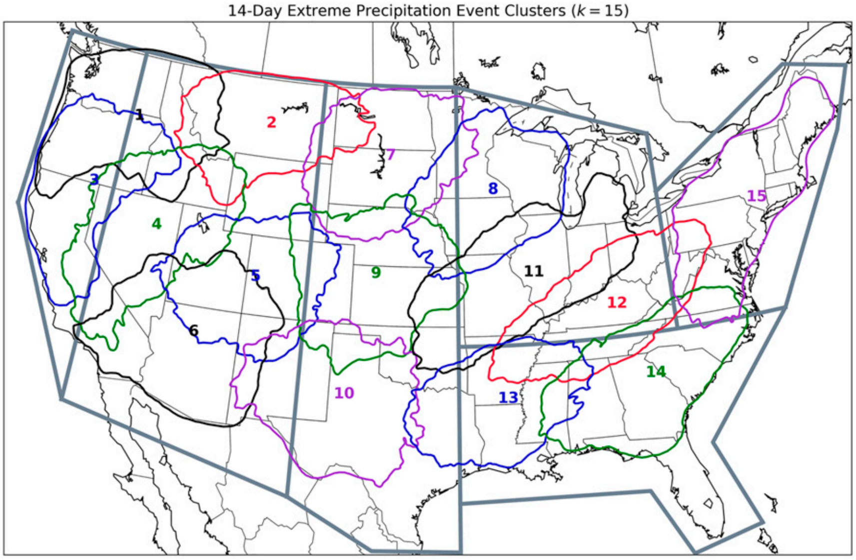

FIG. 9. Average polygons (colored contours) for $k = 15$ clusters across the CONUS for 14-day extreme precipitation events. Also shown are the regional boundaries defined by Jennrich et al. (2020) (gray outlines).

rasterized grid is normalized to represent a probability that any given grid point is within a polygon, ranging from 0 (no coverage) to 1 (total coverage). This grid is then convolved with a Gaussian kernel with a radius of 100 grid points to smooth the normalized grid. The final smoothed polygon is the quantile threshold of 0.2, thus depicting the area that is covered by ≥20% of the events assigned to the cluster. Although there are some gaps in areas of cluster coverage, the majority of the CONUS is described with the chosen number of clusters, with the gaps likely arising from geographical limitations of employing KDE on a dataset only defined over land. Maps with a smaller value for k than chosen herein leave large parts of the CONUS outside a cluster or are too general in their shapes, whereas larger values for k yield clusters that have extensive spatial overlap (i.e., large cluster intercorrelations). Overall, 102 of 851 events (~12.0%) are not assigned to a cluster.

Gong and Richman (1995) found the biggest limitation to the physical validity of k-means results is its sensitivity to the initial cluster center locations. To minimize such a problem, cluster centers are initialized using the "k-means++" method (Arthur and Vassilvitskii 2006). This method begins by placing a random cluster center, but then uses a probabilistic approach to effectively disperse the other centers among attributes to yield improved cluster assignments. Nevertheless, after choosing k, we test 10 different seed initializations to examine the spatial stability and/or reliability in each of the clusters where the final assignments are from the seed with the largest product between events clustered and average silhouette. In general, each of these seeds produced maps that are similar to those in Fig. 9 (see Fig. S4 in the online supplemental material). Although three seeds moved a cluster from the Mountain West to the Southeast (Figs. S4c,e,g), the remaining seeds yielded consistent results.

To examine cluster stability further, we also test hierarchical clustering methods using Ward's linkage and average linkage. Ward's linkage produces clusters that are mostly similar to k-means but with a few differences (see Fig. S5 in the online supplemental material). Cluster 3 is centered farther north and extends farther east than the k-means solution, and cluster 10 stretches close to the southernmost point in Texas. Cluster 5 is the most different (see the orange region in Fig. S5) between Ward's linkage and k-means with the Ward's cluster shifted northwest with a nearly zonal major axis as compared with a nearly circular region in the k-means case. The remaining clusters are very similar between the two algorithms. Conversely, average linkage produced chained results; 92% of clustered events are assigned to 1 cluster, a consequence of its minimization criterion. Such behavior was also noted in Gong and Richman (1995) for certain average linkage implementations. Seasonality is also not expected to significantly affect cluster assignments since accumulation thresholds are different for each window throughout the year and events are recorded in all clusters' warm and cool seasons.

Figure 9 displays potential weaknesses in defining regions before identifying events. Several clusters have significant portions of the average polygon in two or more regions used by Jennrich et al. (2020) (e.g., clusters 1, 11, 12, and 13). Events in these clusters may have been completely missed, owing to the event's areal extent being distributed among several regions, or counted twice but in different regions. Although the event depicted in Fig. 6b occurred before the analysis of Jennrich et al. (2020), the event polygon lies in the Great Plains, Southeast, Great Lakes, and Northeast. Additionally, some of the regions used in Jennrich et al. (2020) seem overly extensive, depicted by several of our clusters inside of a single region. For instance, clusters 2, 4, 5, 6, and a portion of 10 are inside the Mountain

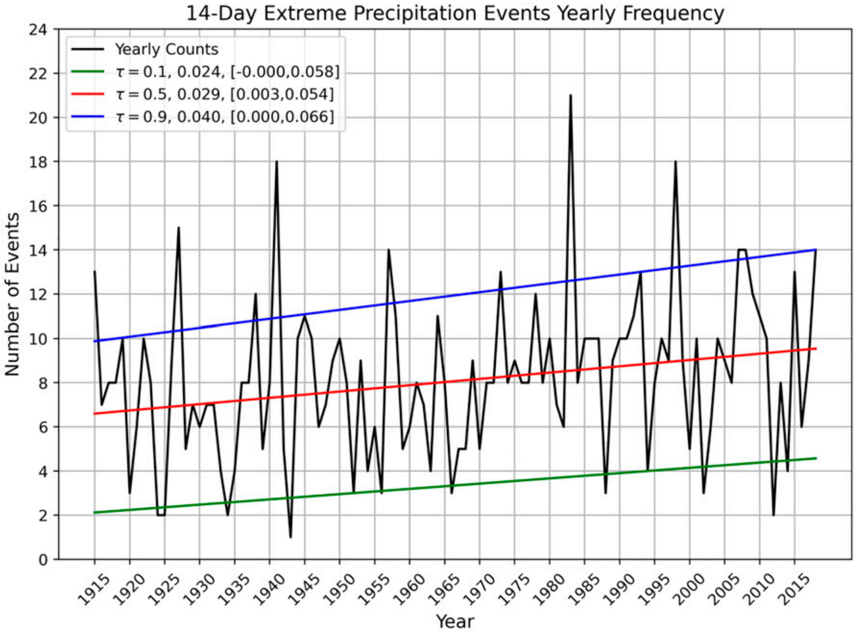

FIG. 10. Total yearly counts (black) of recorded 14-day extreme precipitation events throughout the CONUS as a function of time. QR models are fit on $\tau = 0.1$ (green), $\tau = 0.5$ (i.e., median; red), and $\tau = 0.9$ (blue) quantiles of the time series. Slopes are listed as the second number in the legend, and the 95% confidence intervals on the slope are shown in square brackets thereafter. Confidence intervals were calculated using BCa bootstrapping. Slopes and confidence interval bounds were rounded to three decimal places.

West, and several clusters cross into the Great Plains region. Our clusters suggest that the Great Plains should be divided into subregions, likely owing to contrasting atmospheric drivers for the northern and southern Great Plains (Flanagan et al. 2018). Conversely, cluster 15 is nearly identical to the Jennrich et al. (2020) Northeast region.

Although the framework here was able to eliminate most of the geopolitical boundaries within the CONUS, we are still bounded by the physical (i.e., oceans) and international borders of the CONUS. Nevertheless, the geographical locations and orientations of the clusters suggest favorable general synoptic drivers. Clusters 1, 2, and 7 likely have similar synoptic characteristics, such as the region being centered near an anomalous trough–ridge inflection point, with slightly differing phase and amplitude of the upper-level wave; Konrad (2001) showed 500-hPa cyclone centers were commonly observed upstream from extreme precipitation regions when considering 2-day extreme events. Additionally, Jennrich et al. (2020) and Zhao et al. (2017) showed similar trough–ridge dipoles across nearby areas, such as the Great Plains, Great Lakes, and northeastern United States. Clusters 8, 11, 12, 13, and 15 have a southwest–northeast orientation and have less polygon variability in the clusters, signaling the synoptic influence of a longwave trough and strong surface baroclinicity associated with extratropical cyclones. Conversely, cluster 9 has larger variability in event shape, suggesting a larger mesoscale influence within the synoptic signal (e.g., upslope flow, mesoscale convective systems; Schumacher and Johnson 2006). The broad shapes seen in clusters 6 and 10 likely have monsoonal-type flow with strong vapor transport from the south and southwest (Zhao et al. 2017; Jennrich et al. 2020). Similarly, clusters in the southeast CONUS presumably rely on strong moisture transport from the Gulf of Mexico and Atlantic Ocean and are influenced by the subtropical jet (Jennrich et al. 2020). Work is ongoing to elaborate further on these characteristics and build on the results from Jennrich et al. (2020) using these cluster regions.

b. *Event frequency*

The temporal characteristics of the 14-day extreme precipitation events are next examined. The median number of events per year within the CONUS is 8, and the mean number of events is about the same (8.2). However, these statistics are not time invariant. Figure 10 depicts the number of extreme events over time in our database and the trends in the 10th, 50th, and 90th percentiles (i.e., $\tau = 0.1$, $\tau = 0.5$, and $\tau = 0.9$ quantiles) calculated using quantile regression (QR; Koenker and Bassett 1978); QR is an extension of ordinary linear regression that incorporates a loss function such that the regression line can be fit to any part of the probability distribution. Furthermore, QR makes no distributional assumption of the predictand and inherits the same assumptions on the error terms as ordinary linear regression. All three quantiles have a positive trend with the median number of observed events increasing at a rate of about one event per 50 years. Statistical significance of the slopes was tested using bias-corrected and accelerated (BCa) bootstrapping (Efron 1987) with 10 000 replications at $p < 0.05$ and a null

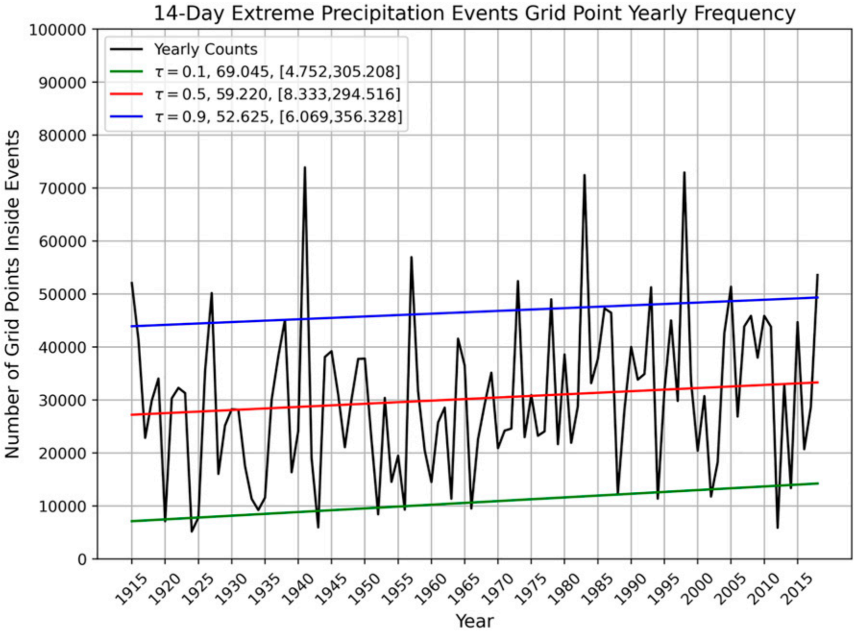

FIG. 11. As in Fig. 10, but for the total number of grid points inside extreme events as a function of time.

hypothesis of no trend. The 95% confidence intervals on the slopes are given in the caption of Fig. 10. Although we cannot reject the null hypothesis for the 10th percentile, the null hypothesis can be rejected for the 50th and 90th regression models as the associated lower bounds are positive. Furthermore, a permutation test on the means of the first and second halves of the time series yielded $p \approx 0.02$, suggesting there was a significant increase in the mean number of events CONUS-wide between 1967 and 2018 relative to 1915 and 1966. The median number of grid points within extreme-event polygons has also increased over time and is significant at $p < 0.05$ (Fig. 11). Although the 10th percentile is also increasing at a statistically significant rate, a permutation test on the areas of events in the first and second halves of our period of record yielded $p \approx 0.78$. Thus, the increasing number of grid points in extreme events cannot be attributed to increasing event size and corroborates the finding that the frequency of extreme precipitation events, defined by the 99th percentile, in the CONUS is increasing. The increased station density noted in Fig. 1 may also be a factor in the increasing frequency of extreme precipitation events, a possibility left for future research.

Note also that two of the three years with the highest number of events occurred in 1983 and 1998, corresponding to two of the strongest El Niños on record. The strongest El Niño on record occurred in 2015/16, and, although less apparent than the aforementioned years, 2015 was another active year for extreme precipitation events as it lies nearly on the $\tau = 0.9$ QR model. Influences from large scale modes of variability may be a factor in driving more, or less, extreme precipitation events in the CONUS, and exploring such links is the topic of ongoing work.

Spatially, there are relative maxima in event frequency in Idaho, Oregon, and northern Nevada and across Arkansas and western Tennessee and Kentucky (Fig. 12). There is a minimum in event frequency in central Wyoming that is collocated with the largest gap in the interior CONUS for our developed clusters (Fig. 9). There are no discontinuities in the event frequency throughout the CONUS, which would not be the case if geopolitical or latitudinal/longitudinal delineations were chosen before identifying events. The sharp gradient near the coasts and international borders arises due to both Livneh and PRISM being masked outside of the CONUS and partially from the choice of the Epanechnikov kernel (see section 2c). The use of a globally gridded dataset, such as a reanalysis product, would improve this aspect of the frequency analysis along the borders of the CONUS. However, reanalysis precipitation rates are largely model derived and are thus considered to have more uncertainty (e.g., Kalnay et al. 1996). Furthermore, the much larger horizontal grid spacing means these products will be unable to resolve small-scale features as accurately as Livneh and PRISM (Slinskey et al. 2019).

The increased frequency of grid points CONUS-wide being in extreme-event polygons is fueled largely by increasing trends in the high plains, northern Nevada, southern Michigan, and the majority of New England (Fig. 13). There are two main regions of decreasing trends: Southern California and Nevada, and Wyoming. Trends here were calculated using ordinary linear regression as bootstrapping QR trends to assess statistical significance was too computationally expensive when considering a spatial field. Bootstrapping was performed by randomly generating 10 000 new time series, sampling with replacement, forming a distribution of slopes using ordinary linear regression, and comparing the point estimate contoured

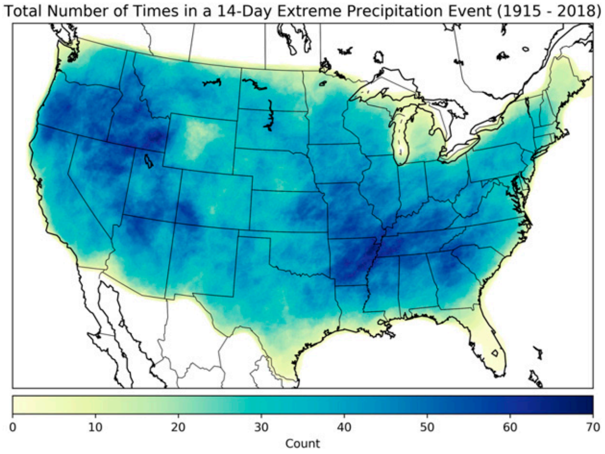

FIG. 12. Total number of times that a grid point is within an extreme-event polygon over the period 1915–2018.

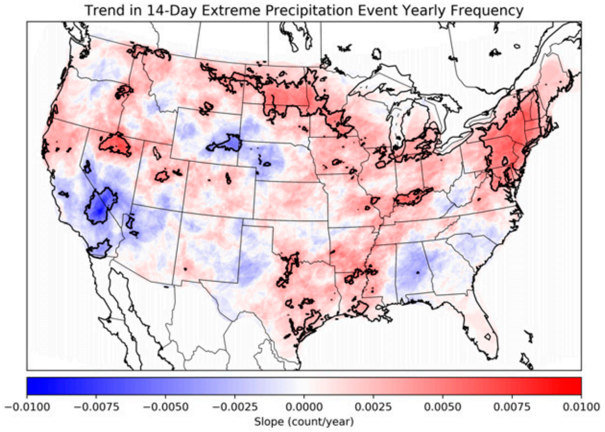

FIG. 13. Trend (number per year; shading) in yearly frequency of grid points falling inside 14-day extreme-event polygons using ordinary linear regression. The black contour delineates statistical significance of the slope at $p < 0.05$ using bootstrapping.

in Fig. 13 with the distribution. Roughly 9.8% of the grid points over the CONUS have significant trends (8.8% positive; 1.0% negative).

5. Summary and conclusions

This research is motivated by a dearth of published research examining the characteristics of extreme precipitation with time scales beyond about 7 days. Furthermore, a large number of different definitions of ''extreme'' exist in the literature. For these reasons, we developed an algorithm for identifying 14-day extreme precipitation events. The algorithm is designed to be easily adaptable within the S2S time scale (e.g., 30-day, 45-day, and 60-day, all of which are currently being computed and archived). We applied this framework to generate multiple databases available for public use. These databases are unique catalogs of extreme precipitation events in that they are identifying S2S extremes as opposed to short-term extremes within an S2S window. Grid points are designated as being extreme if they satisfy two criteria. The first criterion involves accumulation: a grid point must exceed the 99th percentile over the window being considered (e.g., 14 days). The second criterion involves duration: a grid point must record greater than normal daily precipitation on at least half of the number of days in a specified time window (e.g., 7 days). Average daily precipitation is calculated by aggregating all days in a window between 1915 and 2018 in the Livneh and PRISM datasets and calculating a mean. The combination of these two criteria ensures the database identifies periods throughout the CONUS during which anomalously heavy precipitation occurs for more than a few days. Thus, events in our databases likely have more significant, persistent global and synoptic forcings, which may be leveraged to provide forecasts of opportunity for extreme precipitation on the S2S time scale.

Grid points labeled as extreme are grouped into an extreme region using KDE. A three-dimensional KDE PDF is fit over the CONUS, with 0.1° resolution, where a larger density (normalized by the maximum value) denotes more points labeled as extreme are in close proximity. An extreme-event polygon is developed from the density field using the 0.2710 contour level. If such a polygon is greater than 200 000 km^2 in areal extent, it is identified as an extreme event.

We apply the above criteria to build a database of 14-day extreme precipitation events between 1915 and 2018. In any given year, the typical number of events CONUS-wide is eight. However, the typical number of events is increasing throughout our period of record (Figs. 10 and 11), particularly in the Dakotas and throughout much of the Northeast (Fig. 13). This increasing trend is consistent with other literature that has found increases in extreme precipitation in the CONUS (e.g., Kunkel et al. 1999; Groisman et al. 2004; Hoerling et al. 2016; Stegall and Kunkel 2019; Kirchmeier-Young and Zhang 2020). However, caution should be used since many observational-based precipitation datasets do not homogenize station records, which may affect the trends in the datasets. Nevertheless, since there are well-documented changes in extreme precipitation, future works may explore the possibility of detrending before identifying extreme precipitation events.

We also applied k-means clustering to objectively divide the CONUS into regions. Although the masking of data in Livneh and PRISM outside the CONUS presents a challenge when identifying and clustering events, the majority of the CONUS is described by the clusters, and these clusters have fewer boundary effects relative to defining regions a priori. Every cluster has events in at least nine months of the year, and many clusters qualitatively show little to no seasonality, although more work is being done to quantify the climatologies of each individual cluster. Developing the regions is key not only to identify groups of events that are similarly shaped and positioned geographically but also to isolate events that are exceptionally anomalous in that they are shaped very differently. Removing these anomalously shaped events improves regional definitions and will yield more stable climatological statistics for each cluster, which is work currently being conducted. The next step is to diagnose the large-scale characteristics that force extreme events within each cluster and compare and contrast patterns among regions in a manner similar to Jennrich et al. (2020). Our approach to objectively define regions may better isolate the dynamical signal within the extreme events. Defining regions before identifying events ignores the processes that cause the event, whereas our method is likely grouping events that have similar large-scale characteristics. Our clusters are also more representative of the atmosphere as we remove many latitude/longitude bounds used to delineate regions, though we are constrained by the CONUS geopolitical boundary. Overall, our approach provides a strong foundation to diagnose the atmospheric drivers across the CONUS, both in the wet and dry seasons, while also allowing us to isolate the physical reasons why unclustered events were atypical. Despite the continued constraint of the CONUS boundary, the orientation of the clusters follows physical intuition in many cases. For instance, many of the clusters in the southern plains into the Ohio River valley and the Northeast have their major axis in a southwest–northeast direction, consistent with upstream longwave troughing and aligned with the North American storm track.

Examination of such large-scale characteristics should be undertaken in both reanalysis and model datasets. Work is underway to outline any potential changes to the grid that the KDE-derived PDF is fit to and the KDE bandwidth needed to seamlessly apply our framework across multiple datasets. With these adaptations, additional databases are being developed in reanalysis datasets, such as ERA5 (Hersbach et al. 2020), operational seasonal forecasting models, such as the North American Multi-Model Ensemble (Kirtman et al. 2014), and climate models, such as those in phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016). Because the algorithm can be applied consistently between observations and models, the characteristics that drive S2S extreme precipitation can be compared for observed and simulated events. Diagnosing the three-dimensional dynamics and thermodynamics, and analyzing the capabilities of models to reproduce these drivers, are important to gaining a more comprehensive understanding of S2S extreme precipitation. In addition, quantifying trends in the frequency of these drivers will yield insights into regional changes in extreme precipitation. Defining databases and diagnosing large-scale characteristics, and how they may change in the future, is important as it will lend insights into predictability potential of the events to extend the current boundary in prediction of S2S extreme precipitation and perhaps other hazards.

Acknowledgments. We thank Elinor Martin, Renee McPherson, and the rest of the PRES^{2}iP team for their help with the work. We also thank the anonymous reviewers for their helpful comments and insights that provided the impetus for improving the paper. This research has been supported by the National Science Foundation (NSF) Prediction of and Resilience against Extreme Events (PREEVENTS) program, Grant ICER-1663840. Some of the computing for this project was performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU).

Data availability statement. Livneh data are provided by the NOAA/OAR/ESRL/Physical Sciences Laboratory (https://psl.noaa.gov/data/gridded/data.livneh.html). PRISM data are provided by the PRISM Climate Group (http://prism.oregonstate.edu). The code utilized in the study and the developed databases in .csv and .shp formats are available at https://github.com/tydickinson29/PRES2iPpy. Any use of the databases or programs in the GitHub repository should cite this paper.

REFERENCES

Arthur, D., and S. Vassilvitskii, 2006: k-means++: The advantages of careful seeding. Stanford InfoLab Tech. Rep. 2006-13, 11 pp., http://ilpubs.stanford.edu:8090/778/1/2006-13.pdf.

Barlow, W., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **53**, 6835–6875, https://doi.org/10.1007/s00382-019-04958-z.

Bradley, A. A., and J. A. Smith, 1994: The hydrometeorological environment of extreme rainstorms in the southern plains of the United States. *J. Appl. Meteor. Climatol.*, **33**, 1418–1431, https://doi.org/10.1175/1520-0450(1994)033<1418:THEOER>2.0.CO;2.

CPC, 2020: Week 3-4 outlooks. NOAA/NWS, https://www.cpc.ncep.noaa.gov/products/predictions/WK34/.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

——, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, https://doi.org/10.1002/joc.1688.

DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of week-3–4 average temperature and precipitation over the contiguous United States. *J. Climate*, **30**, 3499–3512, https://doi.org/10.1175/JCLI-D-16-0567.1.

Efron, B., 1987: Better bootstrap confidence intervals. *J. Amer. Stat. Assoc.*, **82**, 171–185, https://doi.org/10.1080/01621459.1987.10478410.

Epstein, E. S., and A. G. Barnston, 1990: A precipitation climatology of 5-day periods. *J. Climate*, **3**, 218–236, https://doi.org/10.1175/1520-0442(1990)003<0218:APCODP>2.0.CO;2.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Flanagan, P. X., J. B. Basara, J. C. Furtado, and X. Xiao, 2018: Primary atmospheric drivers of pluvial years in the United States Great Plains. *J. Hydrometeor.*, **19**, 643–658, https://doi.org/10.1175/JHM-D-17-0148.1.

Forgy, E., 1965: Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, **12**, 768–769.

Girons Lopez, M., H. Wennerström, L. Nordén, and J. Seibert, 2015: Location and density of rain gauges for the estimation of spatial varying precipitation. *Geogr. Ann.*, **97A**, 167–179, https://doi.org/10.1111/geoa.12094.

Gong, X., and M. B. Richman, 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Climate*, **8**, 897–931, https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2.

Groisman, P. Ya., R. W. Knight, T. R. Karl, D. R. Easterling, B. Sun, and J. H. Lawrimore, 2004: Contemporary changes of the hydrological cycle over the contiguous United States: Trends derived from in situ observations. *J. Hydrometeor.*, **5**, 64–85, https://doi.org/10.1175/1525-7541(2004)005<0064:CCOTHC>2.0.CO;2.

Grotjahn, R., and G. Faure, 2008: Composite predictor maps of extraordinary weather events in the Sacramento, California, region. *Wea. Forecasting*, **23**, 313–335, https://doi.org/10.1175/2007WAF2006055.1.

Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, 2018: An assessment of differences in gridded precipitation datasets in complex terrain. *J. Hydrol.*, **556**, 1205–1219, https://doi.org/10.1016/j.jhydrol.2017.03.008.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoerling, M., J. Eischeid, J. Perlwitz, X.-W. Quan, K. Wolter, and L. Cheng, 2016: Characterizing recent trends in U.S. heavy precipitation. *J. Climate*, **29**, 2313–2332, https://doi.org/10.1175/JCLI-D-15-0441.1.

Iyer, E. R., A. J. Clark, M. Xue, and F. Kong, 2016: A comparison of 36–60-h precipitation forecasts from convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **31**, 647–661, https://doi.org/10.1175/WAF-D-15-0143.1.

Jennrich, G. C., J. C. Furtado, J. B. Basara, and E. R. Martin, 2020: Synoptic characteristics of 14-day extreme precipitation events across the United States. *J. Climate*, **33**, 6423–6440, https://doi.org/10.1175/JCLI-D-19-0563.1.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Keim, B. D., 1996: Spatial, synoptic, and seasonal patterns of heavy rainfall in the southeastern United States. *Phys. Geogr.*, **17**, 313–328, https://doi.org/10.1080/02723646.1996.10642588.

Kirchmeier-Young, M. C., and X. Zhang, 2020: Human influence has intensified extreme precipitation in North America. *Proc. Natl. Acad. Sci. USA*, **117**, 13 308–13 313, https://doi.org/10.1073/pnas.1921628117.

Kirkyla, K. I., and S. Hameed, 1989: Harmonic analysis of the seasonal cycle in precipitation over the United States: A comparison between observations and a general circulation model. *J. Climate*, **2**, 1463–1475, https://doi.org/10.1175/1520-0442(1989)002<1463:HAOTSC>2.0.CO;2.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Klemm, T., and R. A. McPherson, 2017: The development of seasonal climate forecasting for agricultural producers. *Agric. For. Meteor.*, **232**, 384–399, https://doi.org/10.1016/j.agrformet.2016.09.005.

Knapp, A. K., and Coauthors, 2008: Consequences of more extreme precipitation regimes for terrestrial ecosystems. *BioScience*, **58**, 811–821, https://doi.org/10.1641/B580908.

Koenker, R., and G. Bassett Jr., 1978: Regression quantiles. *Econometrica*, **46**, 33–50, https://doi.org/10.2307/1913643.

Konrad, C. E., 2001: The most extreme precipitation events over the eastern United States from 1950 to 1996: Considerations of scale. *J. Hydrometeor.*, **2**, 309–325, https://doi.org/10.1175/1525-7541(2001)002<0309:TMEPEO>2.0.CO;2.

Kunkel, K. E., K. Andsager, and D. R. Easterling, 1999: Long-term trends in extreme precipitation events over the conterminous United States and Canada. *J. Climate*, **12**, 2515–2527, https://doi.org/10.1175/1520-0442(1999)012<2515:LTTIEP>2.0.CO;2.

——, D. R. Easterling, D. A. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *J. Hydrometeor.*, **13**, 1131–1141, https://doi.org/10.1175/JHM-D-11-0108.1.

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

Lloyd, S. P., 1982: Least squares quantization in CPM. *IEEE Trans. Inf. Theory*, **28**, 129–137, https://doi.org/10.1109/TIT.1982.1056489.

Mercer, A. E., C. M. Shafer, C. A. Doswell, L. M. Leslie, and M. B. Richman, 2012: Synoptic composites of tornadic and nontornadic outbreaks. *Mon. Wea. Rev.*, **140**, 2590–2608, https://doi.org/10.1175/MWR-D-12-00029.1.

Moore, B. J., P. J. Neiman, F. M. Ralph, and F. E. Barthold, 2012: Physical processes associated with heavy flooding rainfall in Nashville, Tennessee, and vicinity during 1–2 May 2010: The role of an atmospheric river and mesoscale convective systems. *Mon. Wea. Rev.*, **140**, 358–378, https://doi.org/10.1175/MWR-D-11-00126.1.

——, K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. M. Hamill, 2015: Climatology and environmental characteristics of extreme precipitation events in the southeastern United States. *Mon. Wea. Rev.*, **143**, 718–741, https://doi.org/10.1175/MWR-D-14-00065.1.

Nardi, K. M., C. F. Baggett, E. A. Barnes, E. D. Maloney, D. S. Harnos, and L. M. Ciasto, 2020: Skillful all-season S2S prediction of U.S. precipitation using the MJO and QBO. *Wea. Forecasting*, **35**, 2179–2198, https://doi.org/10.1175/WAF-D-19-0232.1.

NCEI, 2020: U.S. billion-dollar weather and climate disasters. NOAA, https://www.ncdc.noaa.gov/billions/.

NOAA, 2018: Subseasonal and seasonal forecasting innovation: Plans for the twenty-first century. National Weather Service Tech. Rep., 15 pp., https://www.weather.gov/media/sti/S2S/Annotated%20Outline%20for%20Public%20Release_7_12_18.pdf.

Pan, B., K. Hsu, A. AghaKouchak, S. Sorooshian, and W. Higgins, 2019: Precipitation prediction skill for the West Coast United States: From short to extended range. *J. Climate*, **32**, 161–182, https://doi.org/10.1175/JCLI-D-18-0355.1.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Piao, S., and Coauthors, 2010: The impacts of climate change on water resources and agriculture in China. *Nature*, **467**, 43–51, https://doi.org/10.1038/nature09364.

Richman, M. B., and I. Adrianto, 2010: Classification and regionalization through kernel principal component analysis. *Phys. Chem. Earth*, **35**, 316–328, https://doi.org/10.1016/j.pce.2010.02.001.

Rosenblatt, M., 1956: Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, **27**, 832–837, https://doi.org/10.1214/aoms/1177728190.

Rousseeuw, P. J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7.

Schaeffer, R., and Coauthors, 2012: Energy sector vulnerability to climate change: A review. *Energy*, **38**, 1–12, https://doi.org/10.1016/j.energy.2011.11.056.

Schumacher, R. S., and R. H. Johnson, 2005: Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Mon. Wea. Rev.*, **133**, 961–976, https://doi.org/10.1175/MWR2899.1.

——, and ——, 2006: Characteristics of U.S. extreme rain events during 1999–2000. *Wea. Forecasting*, **21**, 69–85, https://doi.org/10.1175/WAF900.1.

Scott, C. M., and M. D. Shulman, 1979: An areal and temporal analysis of precipitation in the northeastern United States. *J. Appl. Meteor. Climatol.*, **18**, 627–633, https://doi.org/10.1175/1520-0450(1979)018<0627:AAATAO>2.0.CO;2.

Slinskey, E. A., P. C. Loikith, D. E. Waliser, and A. Goodman, 2019: An extreme precipitation categorization scheme and its observational uncertainty over the continental United States. *J. Climate*, **20**, 1029–1052, https://doi.org/10.1175/JHM-D-18-0148.1.

Stegall, S. T., and K. E. Kunkel, 2019: Simulation of daily extreme precipitation over the United States in the CMIP5 30-Yr Decadal Prediction Experiment. *J. Appl. Meteor. Climatol.*, **58**, 875–886, https://doi.org/10.1175/JAMC-D-18-0057.1.

Sukovich, E. M., F. Ralph, F. E. Barthold, D. W. Reynolds, and D. R. Novak, 2014: Extreme quantitative precipitation forecast performance at the Weather Prediction Center from 2001 to 2011. *Wea. Forecasting*, **29**, 894–911, https://doi.org/10.1175/WAF-D-13-00061.1.

Thorndike, R. L., 1953: Who belongs in the family? *Psychometrika*, **18**, 267–276, https://doi.org/10.1007/BF02289263.

Vigaud, N., A. Robertson, and M. Tippett, 2017: Multimodel ensembling of subseasonal precipitation forecasts over North America. *Mon. Wea. Rev.*, **145**, 3913–3928, https://doi.org/10.1175/MWR-D-17-0092.1.

Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Warner, M. D., C. F. Mass, and E. P. Salathé Jr., 2012: Wintertime extreme precipitation events along the Pacific northwest coast: Climatology and synoptic evolution. *Mon. Wea. Rev.*, **140**, 2021–2043, https://doi.org/10.1175/MWR-D-11-00197.1.

Zhao, S., Y. Deng, and R. X. Black, 2017: A dynamical and statistical characterization of U.S. extreme precipitation events and their associated large-scale meteorological patterns. *J. Climate*, **30**, 1307–1326, https://doi.org/10.1175/JCLI-D-15-0910.1.

Zhuang, J., R. Dussin, A. Jüling, and S. Rasp, 2020: JiaweiZhuang/xESMF: v0.3.0 Adding ESMF.LocStream capabilities, https://doi.org/10.5281/zenodo.3700105.